Reprinted from the Journal of Geophysical Research
The MSATT logo incorporates Viking images of the southern hemisphere limb of Mars, depicting the surface and atmosphere in early and late winter (NASA Photos 53A65 and 79B06). Design by Ronna Hurd, Lunar and Planetary Institute.
Martian Surfaces and Atmosphere Through Time

Introduction to the MSATI special section (Paper 95JE00223)  Robert M. Haberle and James F. Bell

Viking Lander image analysis of Martian atmospheric dust  (Paper 94JE002648)  James B. Pollack, Maureen E. Eckert-Bell, and Michael K. Shepard


Temperature measurements of a Martian local dust storm  (Paper 94JE02766)  Ralph Kahn

Atmospheric effects on the mapping of Martian thermal inertia and thermally derived albedo  (Paper 94JE02449)  Joan N. Hayashi, Bruce M. Jakosky, and Robert M. Haberle

Thermal, inertias in the upper millimeters of the Martian surface derived using Phobos' shadow  (Paper 94JE00226)  Bruce H. Betts, Bruce C. Murray, and Tomáš Světek

Mid-infrared transmission spectra of crystalline and nanophase iron oxides/oxhydroxides and implications for remote sensing of Mars  (Paper 94JE01389)  James F. Bell, Ted L. Roush, and Richard V. Morris

Thermal emission measurements 2000-400 cm⁻¹ (5-25 µm) of Hawaiian palagonitic soils and their implications for Mars  (Paper 94JE02448)  Ted L. Roush and James F. Bell

Hematite, pyroxene, and phyllosilicates on Mars: Implications from oxidized impact melt rocks from Manicouagan Crater, Quebec, Canada  (Paper 94JE01500)  Richard V. Morris, D. C. Golden, James F. Bell III, and H. V. Lauer Jr.

SNC: Multiple source areas for Martian meteorites  (Paper 94JE02184)  Allan H. Treiman

Simultaneous adsorption of CO₂ and H₂O under Mars-like conditions and application to the evolution of the Martian climate  (Paper 94JE01899)  Aaron P. Zent and Richard C. Quinn


Low-temperature and low atmospheric pressure infrared reflectance spectroscopy of Mars soil analog materials  (Paper 94JE03333)  Janice L. Bishop and Carl H. Pieters

Longitudinal dunes on Mars: Relation to current wind regimes  (Paper 95JE00225)  Pascal Lee and Peter C. Thomas

Polar margin dunes and winds on Mars  (Paper 95JE002659)  Peter C. Thomas and Peter J. Gierasch

Geology and landscape evolution of the Hellas region of Mars  (Paper 94JE02804)  Kenneth L. Tanaka and Gregory J. Leonard

The circum-Chryse region as a possible example of a hydrologic cycle on Mars: Geologic observations and theoretical evaluation  (Paper 94JE02805)  Jeffrey M. Moore, Gary D. Bow, Wanda L. Davis, Virginia C. Galic, David R. Janke, Christopher P. McKay, Carol R. Stoker, and Aaron P. Zent

A search for nitrates in Martian meteorites  (Paper 94JE02803)  Monica M. Grady, I. P. Wright, and C. T. Pillinger

A slightly more massive young Sun as an explanation for warm temperatures on early Mars  (Paper 94JE03008)  Daniel P. Whitmire, Laurence R. Doyle, Ray T. Reynolds, and John J. Matese

Diagnostic calculations of the circulation in the Martian atmosphere  (Paper 94JE03207)  Michelle L. Santee and David Crisp

The sensitivity of the Martian surface pressure and atmospheric mass budget to various parameters: A comparison between numerical simulations and Viking observations (Paper 94JE03079)  
Frédéric Hourdin, François Forget, and Olivier Talagrand 5501

Post-Phobos model for the altitude and size distribution of dust in the low Martian atmosphere (Paper 94JE03363)  
Eric Chassefière, P. Drossart, and O. Korablev 5525

A reexamination of the relationship between eddy mixing and O$_2$ in the Martian middle atmosphere (Paper 95JE00324)  
Jan Rosenqvist and E. Chassefière 5541

A case for ancient evaporite basins on Mars (Paper 95JE00325)  
Randall D. Forsythe and James R. Zimbelman 5553
Introduction to the MSATT special section

Robert M. Haberle and James F. Bell III
Space Sciences Division, NASA Ames Research Center, Moffett Field, California

The papers published here are based on a workshop entitled "Mars: Past, Present, and Future: Results from the MSATT Program." MSATT (Mars Surface and Atmosphere Through Time) was the last of the Mars data analysis programs and functioned mainly through a series of focused workshops, the final one being held at the Lunar and Planetary Institute in Houston, Texas on November 15-17, 1993. The program began and ended with workshops that brought the entire MSATT community together. At the kickoff workshop (held in Boulder, Colorado, September 23-25, 1991), research areas were identified and collaborations were formed. At the final workshop, progress was assessed and results were presented.

This special MSATT section of JGR-Planets is dedicated to the memories of Roger Burns and Jim Pollack. Both men were MSATT investigators, and Roger served on the steering committee. Their dedication to the program, and to planetary science in general, has served the community well, and they will be sorely missed.

Though ill and undergoing chemotherapy, Roger Burns presented a paper at the final workshop and chaired the session on Mineralogy and Future Observations. In his usual gentle, patient style, Roger led a spirited discussion of ongoing controversies in Mars surface science. His presence was inspirational and illustrative of his devotion to the field. Roger's main planetary science interests were on the mineralogy and weathering history of the Martian surface. Along with Amos Banin, Roger organized a lively and successful workshop on Mars chemical weathering, held at Cocoa Beach, Florida, in September 1992. At that meeting, Roger presented some of his first results on the presence of dehydroxylated clay silicates on Mars, as well as more advanced work on the oxidation of Martian basalts.

Roger was a champion of the search for exotic ferric- and sulfate-bearing phases on Mars because he, more than anyone else, was keenly aware of the potential of such materials for revealing substantial information on the style and rate of Martian weathering and climatic variations.

Jim Pollack was unable to attend the final workshop, as he was then convalescing from his second surgery. But, like Roger, Jim did not let his illness prevent him from participating in the program. He attended three workshops prior to the final one, and continued his research at home. Jim's main MSATT interests were characterizing the optical properties of airborne dust and quantifying the various factors that control the seasonal CO₂ cycle. He realized the importance of suspended dust on the thermal structure and heating of the Martian atmosphere. So, with Maureen Ockert-Bell, he used Viking images to better define the size distribution and single scattering properties of this ubiquitous aerosol. With François Forget, he carried out a reanalysis of the Mariner 9 IRIS and Viking IRIS data sets with a particular focus on determining the presence and radiative effects of CO₂ ice clouds in the Martian polar regions. The results of the Ockert-Bell collaboration are reported in this issue; the CO₂ ice cloud studies continue and will be reported on later.

This paper is not subject to U.S. copyright. Published in 1995 by the American Geophysical Union.

Paper number 95JE00223.
Of course, both Roger and Jim were interested in all MSATT research, and the papers published here, as well as those from the kickoff workshop (see the special sections of *JGR-Planets*, February and June 1993), illustrate how broad the program really was. Here you will find papers that address the geology, mineralogy, and meteorology of Mars in an effort to assess how the surface and atmosphere of this fascinating planet have evolved over time. Could early Mars have been warmed by a brighter young sun instead of a massive greenhouse effect? Were glaciers and hydrological cycles part of Mars' relatively recent past, or was aeolian activity responsible for the putative glacial features? Do the SNCs come from a single source region, or is more than one site involved? And what really are the properties of Martian soils and what do they tell us about the weathering environment? Clearly, these are difficult questions, but progress toward answers can be found in this issue. Also contained in this issue are a mix of theoretical and observational papers that deal with the general circulation of the current atmosphere, the factors that drive it (dust properties), and the role it plays in controlling the current climate system.

Looking back on the MSATT program, which sponsored nine workshops in two years, it is safe to say that it was an incredibly productive program. Those of us who participated will miss the frequent focused workshops that stimulated our thinking and guided our research. We can only hope that soon we will obtain new data pertaining to the Martian surface and atmosphere system, and that soon we will have another data analysis program as successful as MSATT.

J.F. Bell III and R.M. Haberle, Space Sciences Division, NASA Ames Research Center, Moffett Field, CA 94035.

(Received January 13, 1995; accepted January 13, 1995.)
1. Introduction

Solar heating of the Martian atmosphere and surface plays a fundamental role in determining the temperature structure and dynamics. As with Earth, the heating of the atmosphere provides the energy to drive the winds and general circulation pattern. Unlike Earth, a majority of the absorbed solar and infrared energy comes from absorption by suspended dust particles (Gierasch and Goody, 1972; Pollack et al., 1979; Haberle et al., 1982). Understanding the interactions between the atmospheric dust and solar energy enhances our ability to understand and predict the seasonal and interannual climate cycle.

Over the past 20 years, scientists have analyzed data obtained by various spacecraft to define the key radiative properties of airborne dust, summarized in Table 1 of this paper and by Kahn et al. (1992) and Murphy et al. (1993). Mariner 9 and the Viking orbiters and landers have provided a wealth of information. The investigation of atmospheric dust particles has been centered around several key descriptive parameters: the particle size distribution (characterized by the geometric cross-section weighted mean radius and variance $\nu_{p}$), the shape of the particles, the single-scattering albedo $\omega_{s}$, and the imaginary index of refraction $\nu$. Given this information, one can calculate the single-scattering phase function $p(O)$, and the extinction coefficient $\kappa_{\alpha}$. Each of these parameters is important in determining the extent of energy absorption and scattering by the dust particles.

The single-scattering properties of the airborne dust particles can be derived, in part, from their size distribution, such as the modified-gamma or lognormal size distributions. Toon et al. (1977) fit brightness temperature models to the Mariner 9 infrared interferometer spectrometer (IRIS) observations in the thermal infrared to obtain the dust composition and size. A good fit to the spectra was found in montmorillonite. Using the optical properties of montmorillonite and the modified-gamma distribution, they extracted the mean infrared properties of the dust. Their best fit size distribution for $\gamma=2$ and $\nu=0.5$ was $r_{m}=0.4 \mu m$. Using these values, it is possible to calculate the first two moments of the size distribution: $r_{m}=2.70 \mu m$ and $\sigma_{r}=0.38 \mu m$.

Pollack et al. (1979) also assumed a modified-gamma size distribution to analyze Viking Lander sky images. They found the Toon et al. size parameters to be consistent with Viking Lander image results. Pollack (1982) presented a number of single-scattering parameters that were determined from these
observations. The single-scattering albedo ($\omega_0 = 0.67 - 0.99$) and asymmetry parameter ($g = 0.74 - 0.90$) presented in Table 1 of Pollack [1982] are significantly different from the many values at the visible wavelengths contained in the review by Murphy et al. [1993].

Clancy and Lee [1991] used Viking orbiter infrared thermal mapper (IRTM) emission-phase-function (EPF) measurements of several regions to better constrain the particle single-scattering behavior. To first order, their phase function results correspond well with those of Pollack et al. [1977, 1979]. However, the asymmetry factor was lower than that of Pollack et al. [1979] ($g = 0.55$ versus $g = 0.79$), the single-scattering albedo was higher ($n = 0.92$ versus $n = 0.85$), and the cross-section-weighted mean particle size differed by nearly a factor of 10 ($r_m = 0.4 \mu m$ versus $r_m = 2.5 \mu m$).

The Phobos mission provided another chance to collect data that could supply estimates of the first two moments of the dust particle size distribution. Chance [1992] and Korenbrev et al. [1993] analyzed solar occultations measured with the IR spectrometer to obtain the volume extinction coefficient at 3.7 and 1.9 $\mu m$ in the altitude range of 12-35 km. The modified-gamma size distribution (with $\alpha = 2$ and $\gamma$ varied) and the optical indices of a basalt-type mineral were used to extract the dust particle size distribution information. Nine observation sequences were analyzed. Korenbrev et al. concluded that $r_m = 1.26 \pm 0.2 \mu m$ for $\omega_0 = 0.2$ to $0.1 \mu m$ was the best fit to the data averaged over all altitudes. The discrepancies between these results and earlier results were attributed to the atmospheric conditions during the observations (clear (Phobos) versus dusty (Viking)) and to the difference in altitude (limited altitude region (Phobos) versus column integrated (Viking)). Yet in trying to rectify the size values to the Toon et al. sizes, Korenbrev et al. found that the number density profiles were nonphysically constant and would require an enormous flux to maintain.

Clancy et al. [this issue] combined analysis of the Mariner 9 IRIS data and Viking IRTM EPF, and assumed a "palagonitelike" composition for their analysis. They found two good fits to the data: (1) $r_m = 1.2 \mu m$ for $\omega_0 = 0.4 \mu m$ and (2) $r_m = 1.8 \mu m$ for $\omega_0 = 0.2 \mu m$. Clancy et al. favor the second size distribution because this distribution fits the 20 to 30 $\mu m$ dust opacity as well as the visible to 9 $\mu m$ dust opacity ratio found by Martin [1986].

We have carried out a new analysis of Viking Lander 1 and 2 (VL1 and VL2) images of the Martian atmosphere to evaluate the radiative properties of the atmospheric dust particles. The properties of interest were the first two moments of the particle size distribution, the imaginary index of refraction, and other single-scattering properties, such as the single-scattering albedo and phase function. The method of analysis was improved by taking advantage of vignetting (i.e., partial shadowing of the camera) to allow us to obtain data closer to the sun and thus refine previous estimates of the particle size parameters, especially $r_m$. We also included a better representation of light scattered by randomly oriented nonspherical particles. The refined estimate of particle size and shape has improved the estimates of the other single-scattering properties.

In section 2 we present the data and the methods used in this analysis. Section 3 presents the results, including some

<table>
<thead>
<tr>
<th>Year</th>
<th>$r_m$</th>
<th>$\omega_0$</th>
<th>$g$</th>
<th>$\alpha$</th>
<th>$\gamma$</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>0.2</td>
<td>0.265</td>
<td>0.35</td>
<td>1.8</td>
<td>0.02-0.01</td>
<td>Mariner 9 [Pollack et al., 1976]</td>
</tr>
<tr>
<td>1971</td>
<td>0.2</td>
<td>0.268</td>
<td>0.35</td>
<td>1.8</td>
<td>0.004-0.015</td>
<td>Mariner 9 [Chayke and Groom, 1978]</td>
</tr>
<tr>
<td>1977</td>
<td>0.38</td>
<td>143.5</td>
<td>0.5-0.9</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Viking lander [Pollack et al., 1979]</td>
</tr>
<tr>
<td>1977</td>
<td>0.5</td>
<td>143.5</td>
<td>0.5-0.9</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Viking orbiter IRTM EPF [Clancy and Groom, 1981]</td>
</tr>
<tr>
<td>1977</td>
<td>0.5</td>
<td>208.1</td>
<td>0.5-0.9</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>this work</td>
</tr>
<tr>
<td>1977</td>
<td>0.42</td>
<td>143.5</td>
<td>0.5-0.9</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>this work</td>
</tr>
<tr>
<td>1977</td>
<td>0.023</td>
<td>343-348</td>
<td>12-35 km altitude</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Mariner 9 IRIS [Toon et al., 1977]</td>
</tr>
<tr>
<td>1977</td>
<td>0.25</td>
<td>2-18, 1989</td>
<td>0.76-3.14</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Mariner 9 IRIS [Sato and Crisp, 1993]</td>
</tr>
<tr>
<td>1977</td>
<td>0.25</td>
<td>2-18, 1989</td>
<td>0.76-3.14</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Phobos ISM [Drossart et al., 1991]</td>
</tr>
<tr>
<td>1977</td>
<td>0.2</td>
<td>2-18, 1989</td>
<td>0.76-3.14</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Phobos [Koroblev et al., 1993]</td>
</tr>
<tr>
<td>1977</td>
<td>0.2</td>
<td>2-18, 1989</td>
<td>0.76-3.14</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Phobos [Koroblev et al., 1993, Figure 7]</td>
</tr>
<tr>
<td>1977</td>
<td>0.8</td>
<td>295-350</td>
<td>12-35 km altitude</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Phobos [Koroblev et al., 1993, Figure 7]</td>
</tr>
<tr>
<td>1977</td>
<td>0.8</td>
<td>295-350</td>
<td>12-35 km altitude</td>
<td>1.8</td>
<td>0.9-1.5</td>
<td>Phobos [Koroblev et al., 1993, Figure 7]</td>
</tr>
</tbody>
</table>

Here, $r_m$ refers to the dust storm of November 1971. Unless otherwise noted, all work assumes nonspherical particles.
Table 2. Viking Lander Image Information

<table>
<thead>
<tr>
<th>Image ID</th>
<th>Sol</th>
<th>$\lambda$</th>
<th>$\lambda_e$</th>
<th>Solar Distance</th>
<th>Azimuth Angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>12A202R</td>
<td>0</td>
<td>97.0</td>
<td>0.50</td>
<td>37.0</td>
<td>0-220</td>
</tr>
<tr>
<td>21B220/BGR</td>
<td>97</td>
<td>143.5</td>
<td>0.36</td>
<td>27.7</td>
<td>10-85</td>
</tr>
<tr>
<td>21B220/R1</td>
<td>97</td>
<td>143.5</td>
<td>0.36</td>
<td>26.4</td>
<td>10-85</td>
</tr>
<tr>
<td>12C118/BOR</td>
<td>211</td>
<td>208.1</td>
<td>1.93</td>
<td>35.5</td>
<td>5-60</td>
</tr>
<tr>
<td>12C119/R1</td>
<td>211</td>
<td>208.1</td>
<td>1.93</td>
<td>34.4</td>
<td>5-60</td>
</tr>
</tbody>
</table>

Image ID is given, where the first digit represents lander number, the second is camera number, and the subsequent digits are the picture and filter identification. The effective wavelength of the filters is as follows; blue=0.49 pm, green=0.55 pm, red=0.67 pm, IR=0.86 pm, and survey=0.89 pm.

Sol is given as the number of days since V11 touchdown. All images were taken in the afternoon.

The optical depth of the dust as given by a sunshade measurement taken close in time.

2. Data and Methods

The data that were used for this study include several images taken by VL1 and VL2 (Table 2). The images were converted to brightness values using the ground-based calibration files of Hock et al. [1975]. Brightness was expressed as the ratio of the observed brightness to that of a perfectly reflecting Lambert surface, normally illuminated by sunlight at Mars distance from the sun (UIF) or the radian factor, as described by Happke [1981]). The images were corrected for vignetting (described below) and were translated from the camera coordinate system (CCS) to the local horizon coordinate system (LHCS) [Tucker, 1978]. The data were scanned at a constant observer elevation, so that azimuthal distance from the sun would closely match the scattering angle of the observation.

Above 20° in elevation CCS, vignetting due to the camera contamination cover was important. When the contamination cover was closed, it partially shadowed the camera lens, with the shadow fraction monotonically increasing with increasing elevation angle. Thus vignetting partially offset the tendency of the sky to brighten when observations were made close to the Sun. This helped to reduce the saturation at angles close to the Sun. Here, we carefully selected scans that were unsaturated at the smallest angular distance from the Sun to get as far into the particles' diffraction pattern as possible. Note that brightness values at an azimuthal distance of less than 10° from the sun are suspect [Pollack et al., 1977] and thus were not used. We used calibration files for the fractional reduction in brightness due to vignetting to make a good correction for it in scans of interest [Wolf, 1976].

Figure 1 illustrates the importance of vignetting to the work that is presented above. This comparison shows the difference between a real scan taken from an image where there was vignetting (solid line) and what the scan would have looked like had the image not had vignetting (dotted line). The scans are in pseudo-DN number with saturation equal to a value of 63. Had the contamination cover not shadowed the camera, saturation would have occurred at 15° scattering angle. With vignetting, we are able to extract meaningful data down to 10° scattering angle. Accounting for vignetting also allows for a greater dynamic range in the data and thus portrays a better representation of the particles' diffraction pattern.

Equally important in our data selection was the choice of scanning elevations and azimuths. Although the phase function relies on all of the particles-scattering properties, certain areas can be used to optimize some of the parameters. Figure 3 illustrates this idea.

Figure 2, zone 1. The shape of the curve at small scattering angles (0°) was used to find the particle size distribution information ($r_p$ and $\lambda_{so}$). An elevation as close as possible to that of the Sun was used to minimize the angular distance of the last unsaturated data point.

At small scattering angles the shape of the observed intensity curve is dominated by single scattering. This area of the single-scattering phase function (zone 1) is highly diagnostic of particle size, but it is not sensitive to particle shape [Hansen and Travis, 1974; Pollack et al., 1977; Pollack and Cuzzi, 1980]. To illustrate the sensitivity of the shape of the intensity curve versus azimuth in zone 1, consider a limiting case (Figure 2b, $r_p = 0.1 \mu$m). The single-scattering phase curve of very small particles closely resembles that of Rayleigh scatterers [Hansen and Travis, 1974]. For the geometry of our images, this means that we expect at most a 5% increase in $I / I_0$ between 20° and 10° azimuth angle. The observed $I / I_0$ curve (cf. Figures 3-7) actually increases by a factor of 3-5 over this same range of angles. Thus the mean particle size of Martian dust must have a size parameter, $x = 2\pi r_p / \lambda_{so}$, $>>0.1$ or $r_p > 0.01 \mu$m. For large particles (e.g., Figure 2b, $r_p = 2.5 \mu$m), the shape of the phase function is much steeper in zone 1 with as much or a factor of 10 or more increase. For these particles, $x > 1$, the phase function is controlled by light diffracted around the particle with the amplitude of the phase function increasing and its width decreasing as a function $x$ [Pollack and Cuzzi, 1980]. Both criteria imply that $x < 500$ or $r_p < 10 \mu$m.
Figure 2. Comparison of five single-scattering phase functions, demonstrating the effect of particle (a) shape and (b) size on the scattering behavior. The numbers denote the specific areas that were chosen for optimizing the phase function parameters: zone 1, particle size distribution; zone 2, imaginary index of refraction and single scattering albedo; and zone 3, shape.

Figure 2, zone 2. The absolute value of the brightness at scattering angles of 30°-50° was used to determine the value of the single-scattering albedo, which in conjunction with the size parameters found from zone 1, yielded estimates of the particles' imaginary index of refraction (for an assumed value of their real index). The scattering angles of zone 2 were chosen because there is significant multiple scattering at these angles [e.g., Hansen and Travis, 1974], which enhances particle absorption, and because the deduced single-scattering albedo depends least on knowledge of the particles' single-scattering phase function, a corollary to Russell's rule [see Veverka, 1971]. An elevation below the area of vignetting was chosen.

Figure 2, zone 3. The data at intermediate and larger scattering angles determined several of the particle shape-dependent parameters. This point is illustrated by the divergent behavior of the phase function curves of Figure 2a for scattering angles greater than 50° (also see Pollack and Cuzzi [1980]). Fits in this zone yield the shape parameters related to the slope in this zone (PTB or SLP). They are fully defined later in this section in the discussion of the procedure used to evaluate single-scattering properties. An elevation below the area of vignetting was chosen for these scans as well.

Figure 3. Scan of image 12A002 (sol 156) taken at the 15° elevation angle (solid line) plotted with the best fit model (dotted line), where SLP = 0.013, ThMin = 118.3.
A number of constraints were placed on the selection of the Viking Lander images. (1) They needed to be afternoon images to minimize the interfering effect of diurnal water ice clouds (Colburn et al., 1989). (2) They needed to be close in time to a multidiode image to give a value for the atmospheric dust optical depth (see below). (3) The solar elevation angle and azimuth angles close to the Sun had to be in the image to provide the zone 1 information, or azimuth angles opposite the Sun had to be in the image to provide zone 3 information.

Multispectral imaging on the Viking landers was accomplished by alternately selecting three diodes (blue, green, red, or IR1, IR2, IR3) for three vertical scans at each azimuth position (Tucker, 1978). Images 21B220 (color triplet) and 21B221 (IR1 from the IR triplet) supplied information on zone 1 unknowns (\( v_{\text{eff}} \) and \( v_{\text{eff}} \)), and zone 2 unknowns (\( \sigma_{\text{in}} \) and \( \sigma_{\text{in}} \)). These images provided multiwavelength coverage at a time well before major dust storms. Additionally, these were the same images analyzed in our earlier studies (Pollack et al., 1977, 1979) and thus provided ready comparisons with our earlier work. Images 12C118 and 12C139 provided comparable information to 21B220 and 21B221, but near the peak of the first of two global dust storms that occurred during the time that Viking was operational. We selected image 12A002 for the analysis of zone 3 unknowns because it covered a wide range of azimuth angles at intermediate and large scattering angles.

The process of simulating the IRF observed by the Viking Lander employed a number of theoretical models. We had to determine the scattering properties of the airborne species, the scattering behavior of the surface, and combine the results in a realistic manner. Below we describe the models used to determine the light-scattering properties of the airborne dust, the surface scattering, the radiative transfer model used to combine these scattering models, and the method used to determine a best fit to the free parameters.

The shape of a particle is particularly important when considering the light-scattering behavior of airborne dust particles at intermediate and large scattering angles. We used the semi-empirical theory for scattering and absorption by randomly oriented nonspherical particles of Pollack and Cuzzi (1980) to define the single-scattering properties of Martian airborne dust particles. Here, we briefly summarize key aspects of this theory. Mie theory was used for particles where the ratio of the circumference \( 2\pi r \) to the wavelength \( \lambda \) is small.
The scattering characteristics of the large, nonspherical particles are divided into three components: a diffracted component, an externally reflected component, and internally refracted and transmitted components. Nonspherical theory was most important for the internal components of the scattered light, the other two were equivalent to their counterparts for spheres. Pollack and Cuzzi [1980] empirically fit laboratory measurements and found that the data had a linear behavior, $I(r) = F - r$, which can be described by the slope ($F_T$) of the empirical function in the high scattering angle region. Showalter et al. [1992] added a quadratic term to this theory, based on the results of Liu et al. [1983], to include a point (TMin) at which the slope (SLP, which is evaluated at 30° scattering angle) may change, in order to accommodate a modest backscattering peak. Finally, the scattering portion, but not the absorption portion, of the interaction cross section was enhanced by the ratio of the irregular particle's surface area to that of an equal-volume sphere (SAR).

The remaining free parameters in the nonspherical theory are $X_1$ and $X_2$. Mie theory was used to define single-scattering properties for the smallest particles, i.e., those with a size parameter $x < X_1$. The nonspherical particle theory was used for the larger particles, i.e., those with a size parameter $x > X_2$. A weighted average of the two methods was employed at intermediate values of $x$ to provide a smooth transition from one regime to the next, as opposed to the abrupt transition between the two regimes ($X_0$) originally suggested by Pollack and Cuzzi [1980]. Thus in the current version of this nonspherical particle theory there are a total of at least four free parameters: SAR, $X_1$, $X_2$, and $F_T$. In the quadratic theory, $F_T$ is substituted with SLP and TMin, making a total of five free parameters. Realistically, we can determine only two of these parameters, while assuming values for the remainder (reasonable values were found by Pollack and Cuzzi [1980] for a number of particle shapes). Here, we let SLP and TMin be the two free parameters because of their direct relationship to the single-scattering phase function at intermediate and large angles of scatter.

In our analysis we assumed a lognormal size distribution for the dust. This distribution function was chosen because it provides a useful representation for many aerosol species on Earth and because there are only two free parameters, $r_8$ and $\sigma$:

$$n(r) = \frac{5}{r^2} \exp \left( -0.5 \frac{(\ln r - \ln r_8)^2}{(\ln \sigma)^2} \right)$$

$$\ln r_8 = \frac{1}{\int \ln r n(r) \ dr}$$

Figure 5. (continued)
where $c$ is a constant, $r_e$ is the modal radius, and $\sigma$ is a parameter giving the width of the distribution. Note that our parameter $\sigma$ is equivalent to $\exp(\alpha)$ of Hansen and Travis [1974]. The two $\sigma$ are not the same because of slightly different definitions of the lognormal size distribution. The two free parameters ($r_e$ and $\sigma$) are associated with the first two moments of the size distribution, $r_{ef}$ and $\sigma_{ef}$:

$$r_{ef} = \exp \left( -2.5 (\ln \sigma)^2 \right)$$

$$\sigma_{ef} = \exp \left( (\ln \sigma)^2 \right) \cdot 1$$

It is important to note that $r_{ef}$ can be found from the value of $\sigma$, but $\sigma_{ef}$ is dependent on both the values of $r_e$ and $\sigma$. Hansen and Travis [1974] have shown that in many applications, many different particle size distribution functions having the same first two moments have nearly identical single-scattering properties. Conversely, it is difficult to derive much more size distribution information than the first two moments using brightness data alone.

Table 3 summarizes the parameters used to determine the light-scattering properties of the dust particles. Some parameters were fixed to the "nominal" values shown in this table. The range of values used for the free parameters are also listed in Table 3. In section 3 we present the results of a series of sensitivity calculations, in which the impact of varying each of these parameters is tested.

Multiple scattering from the surface can account for less than 1% or as much as 10% of the total intensity. The photometric properties of the surface were modeled using the Hapke [1981, 1986] photometric function. The Hapke function has at least four independent variables related to the surface particle characteristics: the single-scattering albedo, the average macroscopic surface roughness, and the width and magnitude of the opposition effect. Arvidson et al. [1989] analyzed several Viking Lander images of the surface using the color filters to better quantify the surface-scattering parameters. The values presented in their Table 1 were used in this study.

A single-scattering description of the atmosphere was inadequate for our purposes [Pollack et al., 1977]. Hence we used a model that employs multiple scattering in a plane-parallel atmosphere based on the doubling method [Hansen, 1969]. We assumed that the single-scattering properties of...
the particles were invariant with altitude. The surface was included via the adding method described by Hansen [1969]. The specification of atmospheric optical depth was the only input to this program aside from the single-scattering particle properties. For each of the sky images the optical depth was obtained directly from a CENDI image (λ = 0.67 μm) taken as close in time as possible to the image exposure, usually a few hours or at most a few days [Collaborative et al., 1989]. At other wavelengths, the optical depth scaled linearly with extinction coefficient.

We used an iterative method to fit the modeled I/F to the observed data of a given zone. We solved for the best fit parameter values and their uncertainties by finding the values that minimized \( \chi^2 \), the weighted sum of squares of the residuals:

\[
\chi^2 = \frac{1}{N} \sum_{j} \sum_{i} w_{ij} (I/F_{\text{model}} - I/F_{\text{observed}})^2
\]

where \( N \) is the total number of observations modeled. The weighting factor \( w \) estimates variations among the different filters in the total measurement analysis error. A weighting scheme was devised to account for leaks in the detectors, assumed to be the greatest error in the observations. The scaling factor \( h \) is chosen to minimize \( \chi^2 \) for each model parameter and thus fits the modeled data to the shape of the curve. This protocol was followed whenever feasible to remove the influence of errors in absolute calibration, almost always much larger than the relative errors of the derived model parameters. When the absolute value of the data was modeled, \( h \) was given a value of 1.

We sequentially analyzed zones 1, 2, and 3 to derive the optimum values of the free parameters of one regime, while using the best current estimates of the free parameters of the other two regimes. Each parameter for a zone was systematically varied, one at a time, and \( \chi^2 \) was minimized. This procedure was iterated until the free parameters of the zone of current interest changed minimally. We then proceeded to the next zone using these newly determined values from the previous zone. In general, this process had to be repeated several times for the best fit for all of the parameters to be found.

Figure 6. Scans of image 21B220, 21B221 (sol 97) (solid line) taken at the solar elevation angle for relative azimuth angle <5° (with vignetting) and at 15° elevation for angles between 30° and 50° (without vignetting). The scans are plotted with the best fit model (dotted line) \( \mu = 0.5 \text{ μm, } r_{\text{m}} = 1.85 \text{ μm} \). The discontinuity at 30° most likely is due to the uncertainty in the absolute calibration of the vignetting correction.
3. Results

In this section we apply the protocol of section 2 to determine the size, shape, and radiative properties of Martian airborne dust during two very different dust-loading conditions. The first refers to the summer season in the northern hemisphere far in time from the season of global dust storms, when dust loading (or equivalently, optical depth) is close to its minimum value of a few tenths. The second period occurs during midfall in the northern hemisphere at a time close to the peak of the first global dust storm of Viking year 1 (the 1977a storm). In each case, a four-color image sequence (blue, green, red, IR1) was used to determine \( r_{\text{eff}}, \sigma_{\text{eff}}, \) and \( \sigma_0 \), and hence \( n_0 \), from analyses of zones 1 and 2. The optical depth of these afternoon images at an effective wavelength of 0.67 \( \mu \text{m} \) was obtained from sundial images taken in the afternoon of the same Martian day (Colburn et al., 1989). We simulated zone 3 data from the panorama image obtained shortly after touchdown of VL1. The dust optical depth was obtained from the closest relevant data points of Colburn et al. (1989). This image was obtained during early northern summer at a time of low optical depth (cf. Table 2) as were the northern midsummer images taken at VL2. We analyzed the two sets of images concurrently to provide the best overall fits for zones 1, 2, and 3. Few images met the necessary constraints for images used in analyzing zone 3 parameters. Thus the shape parameters found for image 12A002 were applied directly and without change to the analyses of the dust storm images. The best fit values of \( \sigma_{\text{eff}} \) and \( \Theta_{\text{Min}} \) are summarized in Table 5. The simulation of the panorama data is shown in Figure 3. A very good, if not perfect, fit is obtained over a large range of scattering angles. The value of \( \Theta_{\text{Min}} \) (\( <180^\circ \)) implies the presence of a backscattering peak in the dust particles' single-scattering phase function.

3.1. Particle Properties

As alluded to earlier, it is possible to describe a size distribution well using the first two moments, \( r_{\text{eff}} \) and \( \sigma_{\text{eff}} \), when one's interest centers on such radiative transfer quantities as the single-scattering phase function. With the lognormal size distribution, we attempted to find the first two moments that could best account for the scattering behavior of the observed data between 10° and 50° relative azimuth (zone 1). For the purposes of this work, relative azimuth (sun azimuth minus observed azimuth) is roughly equal to
It is clear from Figure 5 that only a crude estimate of $x_{\text{eff}}$ can be made when the primary goal is to get a fine estimate of $r_{\text{eff}}$. The largest values of $x_{\text{eff}}$, which have size distributions skewed toward numerous large particles, provide noticeably poorer fits to the data than do calculations made for smaller values of $x_{\text{eff}}$ (compare Figure 5c with Figures 5a and b). However, airborne dust in the Earth’s troposphere lofted by dust storms tends to have broad size distributions [Weitzhak et al., 1987]. For these reasons, we selected a median value as the best fit, $x_{\text{eff}} = 0.5 \mu m \pm 0.25 \mu m$ (see Table 5).

Estimates of the best fit values of the single-scattering albedo $\omega_0$ were obtained from simulations of zone 2. In modeling zone 2 data, the imaginary index of refraction was systematically varied. This required a fit to the absolute value of the sky brightness. Our best fit parameters, such as $\omega_0$, fit the observed data very well (Figure 6). Values of $\omega_0$ presented in Table 5 vary strongly with wavelength in the visible and near IR due to a strong wavelength dependence of the absorption coefficient of ferric iron contained in the dust [Soderblom, 1992]. The efficiency of single scattering by the dust particles in the red wavelengths could account for the reddish color of the Martian sky [Pollack et al., 1977] and the similar in color between optically thick dust storms and bright surface features [McCord et al., 1977]. The results given in Table 5 are in qualitative agreement with these expectations with one notable exception: $\omega_0$ for the IR filter in the non-dust storm time period was found to be slightly smaller than that of the red filter. We expected that $\omega_0$ for the IR filter would have been at least as large as $\omega_0$ for the red filter. We advocate that caution be used when examining the value of $\omega_0$ for the IR filter because this filter was harder to calibrate. It was partially transparent, to multiple scattering of photons off of the back surface was an important source of photons. Also, this filter was susceptible to degradation by neutrons originating from the RTG power sources. The blue, green, and red filters were not. Although calibration measurements of this effect were taken on the ground prior to launch, they were limited in scope. We therefore believe that the IR filter has a $\omega_0$ that is at least as large as that of the red filter, as is indicated in ground-based measurements of bright surface features [Mustard and Bell, 1994].

We also obtained estimates of the free parameters of zones 1 and 2 for the dust storm images of Table 2 (sol 211). The

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Nominal Value</th>
<th>Range of Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{eff}}$</td>
<td>free</td>
<td>0.01 - 10</td>
<td></td>
</tr>
<tr>
<td>$x_{\text{eff}}$</td>
<td>free</td>
<td>0.25 - 1.0</td>
<td></td>
</tr>
<tr>
<td>$n_1$</td>
<td>free</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>fixed</td>
<td>0.0 - 0.1</td>
<td></td>
</tr>
<tr>
<td>S&amp;W</td>
<td>fixed</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>SLP</td>
<td>free</td>
<td>0.0 - 0.5</td>
<td></td>
</tr>
<tr>
<td>ThMin</td>
<td>fixed</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>fixed</td>
<td>5.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Model Parameters and Their Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Nominal Value</th>
<th>Range of Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{eff}}$</td>
<td>free</td>
<td>0.01 - 10</td>
<td></td>
</tr>
<tr>
<td>$x_{\text{eff}}$</td>
<td>free</td>
<td>0.25 - 1.0</td>
<td></td>
</tr>
<tr>
<td>$n_1$</td>
<td>free</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>fixed</td>
<td>0.0 - 0.1</td>
<td></td>
</tr>
<tr>
<td>S&amp;W</td>
<td>fixed</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>SLP</td>
<td>free</td>
<td>0.0 - 0.5</td>
<td></td>
</tr>
<tr>
<td>ThMin</td>
<td>fixed</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>fixed</td>
<td>5.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Size Parameters for a Lognormal Size Distribution Composed of Nonspherical Particles

<table>
<thead>
<tr>
<th>Sol 97</th>
<th>Sol 211</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{eff}}$, $\mu m$</td>
<td>$r_{\text{eff}}$, $\pm \mu m$</td>
</tr>
<tr>
<td>$0.25$</td>
<td>$1.6$</td>
</tr>
<tr>
<td>$0.51$</td>
<td>$1.9$</td>
</tr>
<tr>
<td>$0.86$</td>
<td>$2.2$</td>
</tr>
</tbody>
</table>

Table 5. Radiative Parameters for a Lognormal Size Distribution Composed of Nonspherical Particles

<table>
<thead>
<tr>
<th>Sol</th>
<th>Effective Wavelength</th>
<th>$\omega_0$</th>
<th>$\sigma$</th>
<th>$\sigma_{\text{eff}}$</th>
<th>$n_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>0.49</td>
<td>0.79</td>
<td>0.68</td>
<td>2.82</td>
<td>0.010</td>
</tr>
<tr>
<td>55</td>
<td>0.66</td>
<td>0.84</td>
<td>0.66</td>
<td>2.89</td>
<td>0.007</td>
</tr>
<tr>
<td>0.66</td>
<td>0.04</td>
<td>0.63</td>
<td>3.04</td>
<td>0.0025</td>
<td></td>
</tr>
<tr>
<td>0.86</td>
<td>0.89</td>
<td>0.65</td>
<td>3.11</td>
<td>0.0065</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>0.49</td>
<td>0.73</td>
<td>0.71</td>
<td>2.83</td>
<td>0.018</td>
</tr>
<tr>
<td>0.55</td>
<td>0.81</td>
<td>0.67</td>
<td>2.94</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>0.66</td>
<td>0.93</td>
<td>0.63</td>
<td>3.13</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>0.86</td>
<td>0.92</td>
<td>0.65</td>
<td>3.23</td>
<td>0.0085</td>
<td></td>
</tr>
</tbody>
</table>

For sol 97, $r_{\text{eff}}=1.85 \mu m$ and $x_{\text{eff}} = 0.51 \mu m$.

For sol 211, $r_{\text{eff}}=1.53 \mu m$ and $x_{\text{eff}} = 0.51 \mu m$.

Here $\omega_0$, $\sigma$, and $\sigma_{\text{eff}}$ are good to within 5%; $n_1$ is good to within 10%.
ability of the derived best fit values to fit the observed sky brightness is shown in Figure 7. Note that the model fits to the data of each filter are not nearly as good as for the low dust optical depth case (Figure 6). In this case it was difficult to find an average value of $r_m$, that fits each filter well. It is likely that this is due to the increased importance of multiple scattering in the high optical depth case, which would make the results much more sensitive to slight differences in $n_i$ and $r_m$.

Table 4 gives the numerical values and associated error bars of $r_m$ for each value of $n_i$. Table 5 summarizes the best fit values for all free parameters and the single-scattering properties. A most surprising finding is that the best fit values of $r_m$ at times of low and large dust optical depth are very similar. This interesting result will be discussed in the next section of the paper.

3.2. Sensitivity Testing

In carrying out the simulations described above, we fixed the values of a number of parameters (cf. Table 3). To determine whether the inferred best fit values of the "free parameters" were robust, we systematically varied, at a time, the values of the fixed and free parameters. The results of this sensitivity study are summarized in Table 6. We find that the nonspherical parameters are mostly robust (changes of 10% in one parameter effect the others by less than 5%) and they do not have a large effect on the size distribution and imaginary index. SAR was the exception; changes in SAR increased the values of the phase function at the large scattering angles and required compensation by changes in either the value of the imaginary index or the mean size of the particle. This is because SAR provides a weight to the relative contribution of small (or $g > X$) particles to the total scattering properties [see Pollack and Cuzzi, 1980]. The free parameters (size distribution parameters and the imaginary index) are connected by the transition from zone 1 data to zone 2. That is, the shape of the curve in zone 1 (size parameters) and the absolute value of the data in zone 2 (optical constants) are bound. Decreases in one optical constant (e.g., $n_i$) can be compensated by increases of the other (e.g., $n_x$) or by decreases in the mean particle size. The size parameters can also compensate for changes within their own zone (an increase in $v_n$ leads to an increase in $r_m$) or by a change in one of the optical constants. We believe that we have learned enough to distinguish the connections of these parameters and to determine the best fit given the information content of the data.

Table 6. Sensitivity Testing of the Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_m$</td>
<td>$n_i$ 10% or $v_n$ 10%</td>
</tr>
<tr>
<td>$v_n$</td>
<td>$n_i$ 10% or $r_m$ 10%</td>
</tr>
<tr>
<td>$n_i$</td>
<td>$n_i$ inversely correlated 10%</td>
</tr>
<tr>
<td>$r_m$</td>
<td>other 10% change in $n_i$ or $r_m$</td>
</tr>
<tr>
<td>SAR</td>
<td>5-10% change in $n_i$ or $r_m$</td>
</tr>
<tr>
<td>SLP</td>
<td>robust</td>
</tr>
<tr>
<td>ThMin</td>
<td>robust</td>
</tr>
<tr>
<td>X1, X2</td>
<td>2% change in $r_m$ or 10% in SLP</td>
</tr>
</tbody>
</table>

Each parameter was varied by 10% from their best fit values, and the impact on the other variables was assessed.

The shape of the single-scattering phase function at the effective wavelengths of the four filters is illustrated in Figure 8. We used the best fit parameters found from the northern summer images to calculate these phase functions. Very similar curves pertain to the best fit parameters of the northern spring dust storm. At scattering angles less than about 30°, with the exact value of this boundary depending on $r_m$ and $X$, the phase function increased progressively more sharply with decreasing scattering angle due to diffracted light [Hansen and Travis, 1974; Pollack and Cuzzi, 1980]. As expected, the phase function $p(0)$ has its largest amplitude and narrowest width at the smallest scattering angles for the blue filter due to the size parameter $x$ increasing as $X$ decreases.

Double refraction events into and out of the particles, and to a lesser degree external reflection, lead to a more gradual fall of $p(0)$ at scattering angles lying between about 50° and 500°. Both these types of interactions depend only weakly on the particles' shape. Note that the value of $p(0)$ is close to 1 in zone 2, a result anticipated in an earlier section of this paper. It results from the phase function normalization condition that demands that the solid angle average of $p(0)$ equal 1. Finally, shape-dependent behavior occurs at scattering angles greater than about 50° due to more complicated paths that are followed by some photons.

A modest backscatter peak at $X = 180°$ (Figure 8) gives additional clues about particle shape, suggesting one or more internal reflections per scattering event [Pollack and Cuzzi, 1980]. The amount of light scattered into the large scattering angles is enhanced by partial or total internal reflection and by particle internal geometry, especially equidimensionality. Internal reflections are abated by sharp corners, such as occur in a cube or parallelepiped. They are also strongly affected by the real index of refraction through the Brewster angle condition for total internal reflection. Zerull et al. [1977] note that a backscattering enhancement can be seen in "fluffy" aggregate particles as well.

4. Discussion

4.1. Comparisons With Other Work

Using a number of data sets from past spacecraft missions, as well as Earth-based platforms, several investigators have
derived mean values for the dust particles’ single-scattering properties and their size and shape (see Table 1 of this paper, Rhode et al., [1992], and Murphy et al., [1993, Table 1]). Most of the work that has been performed to date has concentrated on the particle single-scattering albedo \(\omega(\lambda)\) is quite high (greater than 0.7) and the asymmetry factor \(g\) is approximately 0.6. This research agrees with those results. However, the many different studies do not agree well on the mean particle radius. The use of the different size distributions among the various research projects has very little effect on the deduced value of the mean size, as long as it is equated to the cross-section weighted mean radius \(\langle r_m \rangle\) and the cross-section weighted mean variance \(\langle \sigma_m \rangle^2\). Here we make further comparisons with other investigators’ values of \(r_m\) and \(\sigma_m\) to see how well our values correspond. We attempt to understand the reasons for some major differences in the results and give our opinions and reasoning for an optimum choice.

Even though the present work examined some of the same images, the mean particle size values \(r_m\) are smaller than those presented by Pollack et al. [1979]. This can be attributed to the addition of the vignetting correction to the data processing algorithm, to the improved method of data selection at small azimuth angles, to the use of four concurrent narrow filter images, and to the improved computer codes: the \(\chi^2\) fitting, the particle properties, and the radiative transfer programs. While Pollack et al. [1979, 1979] were aware that vignetting occurred, they could not correctly correct for it and so used data below the area of vignetting (15° elevation angle). Also Pollack et al. [1979] only used data from the blue and IR1 filters to obtain their results. Both of these filters have calibration problems, as shown by Hack et al. [1975].

The single-scattering properties do not agree well with those presented by Pollack [1982], particularly the asymmetry factor \(g\) and \(\omega(\lambda)\). The smaller value of \(g\) and higher value of \(\omega(\lambda)\) in this work are consistent with the smaller particle size and smaller values of imaginary index found in this work.

The new values for \(g\) obtained in the present work agree well with other estimates, such as those given by Clancy and Lee [1991]. The latter work was found by fitting a variety of Heney-Greenstein phase functions to intensity measurements made at intermediate and large scattering angles with the Viking Lander camera (1977). The IRTM solar channel \(r_m(\lambda = 0.67 \text{ micron})\), which is close to the \(r_m\) of the visible channels of the Viking Lander cameras, has a much broader bandwidth (0.3-3.0 microns). Since \(g\) varies slowly with wavelength for our deduced value of \(r_m\), this difference in the width of the bands for the two data sets should not produce major differences in \(r_m\) values. Thus the value for \(g\) obtained by Clancy and Lee [1991] \((g = 0.35)\) is viable; it was found with a more empirical approach than ours and was anchored to data other than Lander images at intermediate and large angles of scatter. We prefer the \(g\) values given in this paper \((g = 0.63 \text{ at } \lambda = 0.65 \text{ micron})\) because they are based on fits to one data set over almost a complete suite of scattering angles.

As discussed in the introduction to this paper, a wide range of values has been given for \(r_m\). Here we compare our values with those earlier investigations to see the degree to which these differences stem from time variations, differences in protocol, and the adequacy of the data set for estimating \(r_m\).

Our new values are smaller than the value \(-2.5 \text{ micron}\) given by Pollack et al. [1979], who also analyzed our northern summer data set of images. We attribute this difference to the factors enumerated earlier in this subsection, namely, our more thorough data analysis and more careful attempt to get useful \(g\) values at the smallest possible scattering angles. Significant differences are found when our results are compared with the \(r_m\) value \((2.5 \text{ micron})\) found by Toon et al. [1979], who simulated the Mariner 9 IRIS spectra during the decay of the great global dust storm of 1971. We believe these differences, if real, may reflect differences in location of the two data sets. The IRIS spectra were all situated near the subsolar point in the southern hemisphere of the dust region of the dust storm source material and were taken at a local time close to the expected diurnal maximum for the ground temperature. Thus the dust from these regions may be the largest value for \(r_m\) since some of the biggest particles lofted into the atmosphere from their initial source region may not have had enough time to settle back to the ground. Conversely, all of the data used in this paper were taken at the Viking Lander sites in the northern hemisphere. Assuming the source region is in the southern hemisphere, all of the dust seen in the Viking Lander images would have been advected from the southern hemisphere by the storm, allowing for considerable sedimentation time. Indeed, it is surprising that these two estimates do not diverge more than they do. Interannual variations of dust particles due to local dust storms and ice hazes are other possible sources of the particle size difference.

The values of \(r_m\) in this work and in Clancy and Lee’s [1991] analysis of the IRTM emission phase function (EPF) sequences do not agree. When investigating the Viking EPF measurements, Clancy and Lee [1991] found \(r_m < 0.4 \text{ micron}\) (Table 1), assuming a montmorillonite composition and modeling the visible/9-\(\mu\text{m}\) dust opacity ratio rather than the single-scattering phase function. The Viking IRTM data are of special value, since they come from specially designed sequences that view the same area for a variety of emission and phase angles. Since they encompass only scattering angles that are greater than 60°, they have a much greater sensitivity to particle shape than to particle size. In their analysis they used only measurements made at scattering angles larger than 60°, and they note that this could cause problems in determining particle size. We believe and R. T. Clancy agrees that this is likely (personal communication, 1993).

In their more recent work, Clancy et al. [this issue] use the Mariner 9 IRIS spectra of the 1971 global dust storm in the southern subtropics (as do Toon et al. [1977]) to model the particle size distribution assuming a "palagonitellite" composition. The cross-section weighted mean particle radius and variance presented in that study \((r_m = 0.68 \text{ micron}, \sigma_m = 0.8 \text{ micron})\) correspond very well to the work presented here, possibly too well, considering the factors described above for the Toon et al. work.

The \(r_m\) values presented by Korahliev et al. [1993] also agree well with our values (cf. Tables 1 and 4). Korahliev et al. present (their Figure 7) altitude profiles of particle effective radius for one of the sequences. It is interesting to note that for the lowest altitudes shown in their plot (12 km) and for a value of \(X_m\) close to the value that we quote (0.4), the particle effective radius is approximately 1.8 \(\mu\text{m}\). The Viking data refer to the vertically averaged column of dust (effective altitude 11 km). It would appear from their altitude profiles, then, that we are close to perfect agreement on the mean particle size. As mentioned in the above, we would expect to see large size differences between these two data sets due to the
time and space differences of the observations. Additionally, the method of the Phobos analysis was constrained by the limited wavelength baseline and by the use of Mie theory to extract extinction efficiencies. By contrast, we use data at four wavelengths and model the shape of the diffraction peak, which is more dependent. However, there is a remarkable correlation between the two data sets.

4.2. Dust Particle Properties

Table 5 summarizes the best fit values of the single-scattering properties and size, and shape parameters of Martian airborne dust particles above the Viking landers at a time of low optical depth (northern summer) and near the peak of the first global dust storm (northern fall), as deduced from the analyses of this paper. They may also represent optimum values of these key properties, as argued in section 4.1. Figure 8 shows the dust particles' single-scattering phase function at the effective wavelengths of the blue, green, red, and IR filters of the Viking Lander cameras. Here, we describe and try to understand several striking aspects of these results.

First, to within the error bars of the current analyses, the $r_{	ext{dep}}$ values derived at times of two very different dust optical depths and activity are close to the same value. This similarity is much greater than might be expected if one postulates greatly reduced dust raising between global dust storms and less dust to the surface due to a size-dependent sedimentation rate, with the biggest particles being lost the quickest. One possible reason for the similarity in $r_{	ext{dep}}$ might be the presence of eddy mixing, a mechanism that is size independent, as a result of the removal process than sedimentation during most seasons of the year. The key region where eddy mixing must dominate over sedimentation in transporting dust downwind to the planet's boundary layer. This mechanism has been invoked to explain the surprising lack of size evolution during global dust storms (Conrath, 1975; Toon et al., 1977). Eddy mixing could be enhanced by local dust storms continually supplying fresh material to the atmosphere. Many local dust storms occur between times of global ones (Murphy et al., 1993). We suspect that both processes play roles in explaining the similarity in $r_{	ext{dep}}$. The combination of these two mechanisms does not allow the dust sufficient time to undergo much size evolution.

Local dust storms as a mechanism makes particular sense if there is a "universal dust," i.e., one having the same composition and size distribution, at source regions on the ground. Some support for there being a "universal dust" on Mars is provided by the similarity in values of the imaginary index of refraction ($n_r$) and $r_{	ext{dep}}$ for the two cases of dust loading analyzed here (cf. Tables 4 and 5). Additional support is given by the similarity in geometric albedos and visible and near IR spectra of optically thick airborne dust (McCord et al., 1977) and various bright surface regions (McCord and Westphal, 1971; Bell et al., 1990). The assumption is that the dust originates in bright surface source regions is limited by the spatial resolution of the two ground-based data sets. Mustard and Bell (1994) note in their comparison of Phobos images and ground-based telescopic images of bright areas that there are many more features in the high-resolution spacecraft data than previously seen in the low-resolution telescopic data.

Finally, we can make a crude estimate of the amount of visible/near IR light absorbing mineral(s) in the airborne dust. Analysis of high resolution spectra of the wavelength region from 0.4 to 0.9 μm strongly suggest that the light absorbing minerals are chiefly iron oxides (Soderblom, 1992). Huffman and Young (1978) measured the optical constants of magnetite, one of the more promising candidate iron oxides. Assuming that this mineral is one of the light absorbing minerals in Martian dust, we obtain a very rough estimate of its fractional volume by dividing the $r_{	ext{dep}}$ of airborne dust that was found here by the imaginary index of magnetite at the various wavelengths of interest. In this way we obtain a fractional volume of 1.2% magnetite in the Martian airborne dust. It would be interesting to compare the imaginary index found in our work with those of a number of other rocks and minerals. As yet, such a library of optical constants for the iron oxide minerals and rocks that are likely to be abundant on Mars does not exist.

4.3. Solar Heating

The revised values of some of the radiative properties of the airborne dust particles may engender significant changes in the amount of solar radiation that is absorbed within the atmosphere and at the surface, as compared with that found by Pollack (1982). Here, we make crude estimates of the magnitude of these changes to obtain a preliminary estimate of their significance. A more complete and quantitative analysis of this issue will be made in paper 2 of this series on airborne dust properties, at which point we will have extended the definition of these monochromatic values across essentially the entire solar spectrum.

In the present paper, we have obtained modest to major changes in the values of three key properties of airborne dust in the Martian atmosphere that could alter the solar energy deposition profile, as deduced from the properties given by Pollack (1982). These are the effective particle radius $r_{	ext{dep}}$, the single-scattering asymmetry parameter $g$, and the single-scattering albedo $a_{	ext{eff}}$ for the IR filters.

Our preferred values for $r_{	ext{dep}}$ of 1.5-1.9 μm are somewhat smaller than values of 2.5-2.7 μm obtained in earlier work. The particle size affects the solar energy deposition profile chiefly by playing a major role in determining the wavelength dependence of the dust optical depth (Figure 9). The phase function of light scattering is described by the forward scattering factor $Q_{	ext{scat}}$.
Therefore the altered value of magnitude because the revised value of $g$ leads to an enhancement in the amount of sunlight deposited within the atmosphere. Alteration of $\omega_0$ in the near IR has the opposite effect. These two changes, even the change in particle size, should not greatly disturb current estimates of solar energy deposition in the Martian atmosphere.

5. Summary

In this paper we have presented results of radiative transfer modeling of Viking Lander images. The data analysis and modeling have resulted in significant improvements over past Viking Lander image analyses (Pollack et al., 1977, 1979) to values of a number of key properties of the Martian airborne dust. We also explored some of the implications of these new results.

1. We have analyzed three sets of Viking Lander images to determine the size, shape, and radiative properties of Martian airborne dust during two very different dust-loading conditions as summarized in Table 5 and Figures 3-6. One image was taken by VL2 during the summer season in the northern hemisphere, when dust optical depth was close to its minimum value of a few tenths. The second set of images was taken at VL1 during midfall in the northern hemisphere at a time close to the peak of the first global dust storm of Viking year 1. The third image was a panorama taken by VL1 just after touchdown, when the dust optical depth was quite low. The first two sets of images were used to find the particle cross-section weighted mean radius $(r_{mp})$ and variance $(\delta r_{mp})$ and their wavelength dependent single-scattering albedos $(a_\lambda)$. The best fit $r_{mp}$ for both images was found to be $0.5 \pm 0.2 \, \mu m$. The variance of $r_{mp}$ varied from 1.85 to 0.3 for the low dust optical depth VL2 image to 1.52 ± 0.3 $\mu m$ for the high dust optical depth VL1 image. The panorama image was used to determine the degree of irregularity of the dust particles. The particles are nonspherical. They scatter light fairly efficiently into the backward hemisphere, which implies equidimensionality and possibly very sharp corners.

2. These values of $r_{mp}$ and $\omega_0$ are surprisingly consistent between the two dust loadings and with the latest research done with Phobos and Mariner 9 IRIS data (see Table 1). One would expect a much greater discrepancy due to the uplifting of dust during the global dust storm. Eddy mixing can be invoked as a more important dust removal mechanism than sedimentation for the lifting of dust from the surface, since eddy mixing is size independent. This is in combination with frequent local dust storms which would account for the lack of change of mean particle size with time and location. A second explanation may be the presence of dust source regions where the dust is the same composition and size, a "universal" dust.

3. The single-scattering properties summarized in Table 5 are consistent with other investigators' results. Over the Viking wavelength range (0.49 to 0.86 $\mu m$), the single-scattering albedo ranges from 0.6 to 0.9, the asymmetry factor from 0.6 to 0.7, and the extinction coefficient from 2.8 to 3.2. Also shown in Table 5 are the values of the imaginary index of refraction for each image, assuming that the real index of refraction is 1.5. The imaginary index ranges from 0.01 to 0.0025 for the low dust-loading example (sol 97) and from 0.018 to 0.004 for the dust storm case (sol 211). Due to ground-based calibration and possible RTG damage, absolute
IIF values for the IRI filter have large errors and thus do so at eq and $\eta_b$. However, the trends shown in Table 5 (such as the increasing value of imaginary index) appear to be consistent with ground-based results [Pollack et al., 1993].

4. Though the revised values of $r_e$ are smaller than the values found by Toom and Pollack [1977] and Pollack et al. [1979], the extinction values are not largely different. Therefore the solar energy deposition within the atmosphere and at the surface of the planet due to the different values of the mean particle size would change little. Major changes of several tens of percent may result from the revised values of $g$. The larger value of $g$ leads to an enhancement in the amount of sunlight absorbed by the atmospheric dust. A reduction of absorbed light of corresponding size and opposite magnitude may exist due to a higher $\eta_b$ in the near IR than quoted in Table 5.

5. It is possible to determine the fractional composition of the dust particles from the imaginary index of refraction, given the $\eta_b$ of other minerals or rocks. An example of 1.2% composition of maghemite is found. It would also be possible to compare the shape of the imaginary index curve to those of iron oxide rocks and minerals and find one or more good candidates for the dust composition. Unfortunately, an extensive library of the complex index of refraction for iron oxide minerals and rocks in the visible wavelengths does not exist yet.

6. Notation

c.o.v. size distribution adjustable constants.
CCS Viking Lander camera coordinate system.
$\chi^2$ sum of the squares of the residuals.
FTL ratio of integral forward light transmission to backward.
$g$ scattering asymmetry factor.
$I/F$ ratio of observed brightness to that of a Lambert surface.
$N$ wavelength of light in micrometers, $\mu m$.
$\Delta$ areonometric longitude of Mars.
$\eta_b$ differential number of particles with radius $r$.
$\eta_i$ imaginary index of refraction.
$\eta_r$ real index of refraction.
$\rho(0)$ normalized phase function.
$\sigma_e$ particle extinction efficiency.
r particle radius.
$\theta$ geometric cross-section weighted mean particle radius.
r $r$ modal radius.
$\sigma$ size distribution width parameter.
SAR ratio of irregular particle surface area to equal volume sphere.
SLP slope of phase function at $0^\circ$.
SOL number of Martian days since VL1 touchdown.
$\psi$ scattering angle.
TMMin location of phase function minimum.
$\omega$ geometric cross-section weighted variance.
$\omega_s$ single-scattering albedo.
$z$ particle size parameter, equal to $2\pi r / \lambda$.
$X_1$ maximum value of $x$ for Mie particles.
$X_2$ minimum value of $x$ for nonspherical particles.

Acknowledgments. The authors wish to thank R. E. Arvidson, J. Bell, A. R. C. Bridge, R. Haberle, J. Murphy, and T. Roush for helpful discussions and insights. The authors are indebted to M. Showalter for versions of his Mie and nonspherical computer code, radiative transfer code, and for his help in implementing changes to these codes. The time and effort put into the review of this manuscript by T. Z. Martin and S. W. Lee are greatly appreciated. Funds for the support of this study have been allocated to the San Jose State University Foundation by NASA Ames Research Center under interagency NASA/90-4. This work was supported in part by a grant from NASA's Planetary Atmospheres program, RTOF 154-160-90-01.

References


Pollack et al.: Properties of Martian Atmospheric Dust


M. E. Ockert-Bell, NASA Ames Research Center 245-3, Moffett Field, CA, 94035. (e-mail: mbell@nasch.arc.nasa.gov)

M. K. Shepard, CEPS MRC 515, National Air and Space Museum, Smithsonian Institution, Washington, D. C., 20560. (e-mail: mshepard@ceps.nasa.gov)

(Received March 28, 1994; revised October 3, 1994; accepted October 5, 1994.)
A new model for Mars atmospheric dust based upon analysis of ultraviolet through infrared observations from Mariner 9, Viking, and Phobos


Abstract: We propose key modifications to the Toon et al. (1977) model of the particle size distribution and composition of Mars atmospheric dust, based on a variety of spacecraft and wavelength observations of the dust. A much broader (r_mode=0.02 μm), smaller particle size distribution coupled with a "palagonite-like" composition is argued to fit the composite ultraviolet-to-30-μm absorption properties of the dust better than the montmorillonite-basalt, r_mode=0.4 μm, r_mode=0.40 dust model of Toon et al. Mariner 9 (infrared interferometer spectrometer) IRIS spectra of high atmospheric dust opacities during the 1971-1972 Mars global dust storms are analyzed in terms of the Toon et al. dust model, and a Hawaiian palagonite sample (Roush et al., 1991) with two different size distribution models incorporating smaller dust particle sizes. Viking Infrared Thermal Mapper (IRTM) emission-phase-function (EPF) observations at 9 μm are analyzed to retrieve 9-μm dust opacities coincident with solar band dust opacities obtained from the same EPF sequences (Clancy and Lee, 1991). These EPF dust opacities provide an independent measurement of the visible/9-μm extinction opacity ratio (ε2) for Mars atmospheric dust, which is consistent with a previous measurement by Martin (1986). Model values for the visible/9-μm opacity ratio and the ultraviolet and visible single-scattering albedos are calculated for the palagonite model with the smaller particle size distributions and compared to the same properties for the Toon et al. model of dust. The montmorillonite model of the dust is found to fit the detailed shape of the dust 9-μm absorption well. However, it predicts structured, deep absorptions at 20 μm which are not observed and requires a separate ultraviolet-visible absorbing component to match the observed behavior of the dust in this wavelength region. The modeled palagonite does not match the 8- to 9-μm absorption presented by the dust in the IRIS spectra, probably due to its low SiO2 content (31%). However, it does provide consistent levels of ultraviolet/visible absorption, 9- to 12-μm absorption, and a lack of structured absorption at 20 μm. The ratio of dust extinction opacities at visible, 9 μm, and 30 μm are strongly affected by the dust particle size distribution. The Toon et al. dust size distribution (r_mode=0.40, r_mode=0.4 μm, r_mode=2.7 μm) predicts the correct ratio of the 9- to 30-μm opacity, but underpredicts the visible/9-μm opacity ratio considerably (1 versus 2). A similar particle distribution width with smaller particle sizes (r_mode=0.17, r_mode=0.4 μm, r_mode=1.2 μm) will fit the observed visible/9-μm opacity ratio, but overpredicts the observed 9-μm/30-μm opacity ratio. A smaller and much broader particle size distribution (r_mode=0.02, r_mode=0.08 μm, r_mode=1.8 μm) can fit both dust opacity ratios. Overall, the nanocrystalline structure of palagonite coupled with a smaller, broader distribution of dust particle sizes provides a more consistent fit than the Toon et al. model of the dust to the IRIS spectra, the observed visible/9-μm dust opacity ratio, the Phobos occultation measurements of dust particle sizes (Chassefiere et al., 1992), and the weakness of surface near IR absorptions expected for clay minerals (Clark, 1992; Bell and Crisp, 1993).

Introduction

The long-standing model of Mars atmospheric dust is a predominantly montmorillonite composition with a particle size distribution corresponding to a cross-section weighted mean radius (r_mode) of 2.7 μm (r_mode=0.4 μm). The basis for this model was an analysis of Mariner 9 infrared interferometer spectrometer (IRIS) observations (5-50 μm) of the 1971 global dust storm, performed by Toon et al. (1977). The actual conclusions from the Toon et al. analysis regarding the dust composition were not as specific as the montmorillonite model generally adopted by the Mars science community. Toon et al. concluded that the dust was a mixture of weathering products from silicic rocks (SiO2>60%) and a secondary component such
relative point out that the ratio of Toon dust, but relied upon Mariner 9 UVS determinations of dust inaccurate for the determination of this ratio. More recent studies predicted by the standard dust model of IRTM measurements near 0.92 versus a value of 0.86 from Pollack based on the wavelength center of the dust at band absorption in the IRIS spectra and was considered by that smaller particle sizes for the dust would be consistent with a simple montmorillonite composition. a conclusion which measurements during the 1977 dust storms were not consistent with a simple montmorillonite composition, with two important modifications. A non spherical shape for the dust particles was modeled to explain the observed single-scattering phase function of the dust, and a separate visible/ultraviolet absorbing component was postulated to match these Viking Lander observations and earlier Mariner 9 ultraviolet spectrometer (UVS) observations (Pang and Ajello, 1977). Pollack et al. (1979) suggested that a severe percent component of magnetite in the Mars atmospheric dust could provide the observed visible and ultraviolet absorption by the dust, since neither montmorillonite nor basalt absorbs visible or ultraviolet light efficiently.

Since these key studies, a number of important measurements ranging from analysis of Viking 9 micron dust opacities (Martin, 1980) to Phobos near-IR extinction observations of Mars dust (Chaseyferre et al., 1992) have been obtained. Hunt (1979) pointed out that the observed ratios of 9- to 20-μm dust absorption in Viking infrared thermal mapper (IRTM) measurements during the 1977 dust storms were not consistent with a simple montmorillonite composition, a conclusion which Toon et al. (1977) also determined from their analysis of the Mariner 9 IRIS spectra. Both Toon et al. and Hunt suggested that an additional component, such as basalt, might account for the relative weakness of the 20-μm dust absorption. Zurek (1982) pointed out that the ratio of visible-to-IR extinction opacities for the dust, obtained from comparison of the Viking Lander and IRTM measurements (-2), was not consistent with the ratio predicted by the standard dust model of Toon et al. (+1). Toon et al. (1977) also examined the visible-to-IR opacity ratio of the dust, but relied upon Mariner 9 UVS determinations of dust opacity (Pang and Hord, 1973), which were apparently inaccurate for the determination of this ratio. More recent studies include an analysis of Viking solar-bond IRTM emission-phase-function (EPF) sequences by Clancy and Lee (1991), which suggested smaller absorption (visible single-scattering albedo near 0.92 versus a value of 0.86 from Pollack et al. (1979)) and larger backscattering (single-scattering asymmetry parameter of 0.55 versus 0.79 from Pollack et al. (1979)) for Mars atmospheric dust at visible/IR wavelengths. Clancy and Lee also noted that smaller particle sizes for the dust would be consistent with the observed visible/IR opacity ratio. Photos solar occultation measurements of dust extinction at wavelengths of 0.75-3.15 μm indicated a value for rν 71 μm of 0.2-1.5 μm (rν 71 μm = 0.2 μm at 15-25 km altitude (Chaseyferre et al., 1992). With regard to the dust composition, Clark (1992) interpreted ground-based near-IR spectra of Mars surface reflectance to place a very low limit (<1%) on the amount of crystalline montmorillonite-like clay materials. Bell and Crisp (1993) identified weak absorption features in the 2.20- to 2.25- μm region from near-IR reflectance observations, which suggested a small but measurable clay component to the Mars bright soil. In either case, the dominant character of the Mars bright regions appears more consistent with amorphous or poorly crystalline materials such as palagonite (e.g., Singer, 1982; Morris et al., 1990; Oosenberg and Handy, 1992; Banin et al., 1992; Soderblom, 1992; Roush et al., 1993; Bell et al., 1993; Marchie et al., 1993), even though the small amount of extinction observed by IRTM is consistent with the presence of small particle sizes for the dust.

In the following analysis, we consider a "palagonite-like" dust composition and a smaller and considerably broader particle size distribution than the Toon et al. size distribution as an alternative model for Mars atmospheric dust. We do not attempt to model a wide range of materials and particle size distributions, and hence, cannot claim a unique solution for the composition and particle sizes of Mars atmospheric dust. In fact, the term "palagonite" accommodates a wide range of compositions and component crystallinity (usually amorphous to noncrystalline), since it refers to the weathering product of basaltic glass. Rather, we attempt to demonstrate that a single, plausible component and particle size distribution can fit the full range of dust observations to a much better degree than does the currently accepted model of Mars dust. Our demonstration is made with a particular sample of palagonite from the upper slopes of Mauna Kea (Singer, 1982; Clark et al., 1990; Roush et al., 1991), which has been found to be a good spectral analog for Mars dust (e.g., Singer, 1982; Morris et al., 1990). One of the key elements in conducting this analysis is the ability to compute 7- to 30-μm dust absorption spectra for comparison with the 1971 Mariner 9 IRIS spectra, as well as the Viking IRTM 9-μm band observations in 1977.

**Infrared Spectroscopic Modeling of Mars Atmospheric Dust**

In order to reanalyze the key IRIS spectral observations as well as Viking IRTM EPF 9-μm sequences, one of us (R. Gladstone) has modified a doubling-and-adding radiative transfer code developed for terrestrial atmosphere studies (Gladstone et al., 1984). The adapted code is capable of accurate dust emission-absorption-scattering radiative transfer calculations over the 5- to 30-μm wavelength region for variable dust composition and particle size inputs, and incorporates both the Viking IRTM channel weights and the Mariner 9 IRIS wavelength resolution for direct comparisons to these data sets. Two to five kilometer vertical resolution between the surface and 50-km altitude and 4 quadrature zenith angles characterize the radiative transfer modeling. We adopt atmospheric temperature profiles according to the algorithm of Marvin (1986) in the case of the Viking IRTM comparisons, and estimated Mariner 9 IRIS temperature profiles from the 15-μm CO₂ band from IRIS spectra. For comparison with the IRIS ratios, atmospheric opacity from CO₂, which is the primary molecular absorber in the 7- to 30-μm wavelength region, is not modeled in the dust radiative transfer. Hence, the 12- to 18-μm region, where the CO₂ opacity is important, is excluded in our analysis, as it was in the Toon et al. analysis. We model the dust absorption/scattering properties with Mie single scattering theory and the optical constants for a particular specimen of Hawaiian palagonite as retrieved by Roush (Roush et al., 1991) for the 5- to 25-μm wavelength region, and over the 0.3- to 5.0-μm region as described below. For comparison, we also model the infrared absorption spectrum of montmorillonite, employing the infrared optical constants of montmorillonite 2197 from the analysis of Toon et al. (1977).

We consider palagonite as the primary alternative to the montmorillonite model of Mars atmospheric dust, based on several considerations. As a primary weathering product of
basalt, palagonite is a very plausible constituent of the bright and dark regions on Mars. Palagonite absorbs in the ultraviolet and visible wavelength region due to its Fe$^{3+}$ content. Montmorillinite does not absorb at UV/visible wavelengths, although it can present such absorption if coated by a ferric oxide stain (Hawt et al., 1975; Toon et al., 1977). Palagonite may be consistent with the observed lack of weakness of 2.2- to 2.25-$\mu$m clay absorption for the bright regions of the Mars surface, due to the amorphous to nanocrystalline structure of many palagonites such as the modeled Mauna Kea sample. Furthermore, as apparent in the palagonite sample we model below, palagonite does not necessarily display strong, structured absorption near 20 microns as does montmorillinite. The observed lack of this 20-$\mu$m absorption feature in Mariner 9 IRIS spectra of the 1971 dust storm was emphasized by Toon et al. (1977) as an indication that pure montmorillinite was an incomplete model of Mars atmospheric dust. In fact, all of the additional compositions considered by Toon et al. (quartz, basalt, granite, andesite, basaltic glass, and obsidian) exhibited distinctive 20-$\mu$m absorptions which were not apparent in the IRIS spectra of Mars atmospheric dust. Montmorillinite was notable primarily for its ability to match the observed 9-$\mu$m absorption of Mars atmospheric dust. Montmorillinite with 25-50% basalt by weight led to a better fit of the observed 20-$\mu$m absorption, but also seriously degraded the model fit to the observed 9-$\mu$m absorption (Toon et al., 1977).

Mauna Kea Palagonite

The palagonite sample was originally collected at several locations on the upper slopes of Mauna Kea, Hawaii (Roush, 1987). Subsequent to collection, the sample whose optical constants are used here was separated from the bulk sample by liquid suspension in methanol. Initial characterization of this sample included X-ray diffraction and determination of major elements by X-ray fluorescence analysis. The sample was X-ray amorphous and the major elements are provided by Roush et al. (1991). Subsequent to the initial analyses, the sample was studied using transmission electron microscopy (TEM) [Roush and Blake, 1991]. The point-to-point resolution of the microscope in conventional imaging mode was about 0.5 nm, the lateral spatial resolution for elemental analysis was 30-50 nm, and selected area electron diffraction patterns could be recorded from particles as small as 100 nm laterally. Qualitative elemental abundances of 101 grains showed that Si, Al, O, Fe, and Ti were most abundant and lesser amounts of Ca, Cr, Mg, and P were present in some samples (Roush and Blake, 1991). Grain-to-grain compositions were highly variable, but three major classes of grains having Si, Al, and Fe as the most abundant element were identified (Roush and Blake, 1991). A limited number of grains from each major class were selected for characterization by electron diffraction. Most grains did not exhibit electron diffraction patterns indicating that they were amorphous. The diffraction patterns observed were consistent with layer lattice silicates (Roush and Blake, 1991). Five of six grains that had Al as the most abundant element did not yield a diffraction pattern. A particle size distribution was determined by measuring the major (a) and minor (b) axes of 610 grains from micrographs calculating the projected elliptical area (SA), and the equivalent spherical particle diameter of this sample was determined to be 1.0220.66 $\mu$m (Roush and Blake, 1991).

Over the infrared wavelength region (5-25 $\mu$m) we adopt the palagonite optical constants presented by Roush et al. (1991). The imaginary index versus wavelength is presented in Figure 1a. In the near infrared (0.7 to 4.6 $\mu$m), the indices of refraction were derived using the techniques described by Clark and Roush (1984) and Clark et al. (1990). This technique uses the measured reflectance of the sample, knowledge of the grain size, assumes the real index does not vary significantly, and iteratively applies Hopke's (1993) radiative transfer model to vary the imaginary index until the computed reflectance matches the measured reflectance. The initially derived values of the imaginary index using this technique are plotted, along with the infrared values in Figure 1b. Dust particle size distributions, based on the modified gamma distribution. All of the model calculations presented employ one of these size distributions, for which the parameters are provided in Table 2. Size distribution 1 corresponds to the Toon et al. model; size distribution 2 exhibits a size distribution width similar to the Toon et al. model, but with roughly one half the average particle sizes. Size distribution 3 exhibits a very broad size distribution width, and is found to obtain a dust opacity variation between visible and 30-$\mu$m wavelengths most appropriate to the observed variation.
[Roush et al., 1991], as the dotted line in Figure 1a. Clearly, there is a discrepancy in the imaginary index near 5 μm, between the long wavelength region of the near-infrared and the short wavelength region of the infrared values. To resolve this discrepancy, we note that there are fewer assumptions used in deriving the infrared values than the near-infrared values. As a result, we believe that the infrared imaginary indices are more accurate than the near-infrared values. Additionally, Clark et al. [1990] derived 0.3- to 5-μm absorption coefficients for a similar material, and when the absorption coefficients are converted to imaginary indices, they compare quite well with the values derived in the infrared for the sample studied here. Hence, we have arbitrarily scaled the near-infrared imaginary indices derived for the sample studied here to agree with the derived infrared values. These rescaled values are also shown in Figure 1a, as the solid line curve, which now merges with the mid-infrared values from Roush et al. [1991], and with the ultraviolet and visible values adopted from Clark et al. [1990].

In the visible and near-IR, the spectral properties of palagonites exhibit some variability (e.g., Soderblom et al., 1978; Evans et al., 1981; Singer, 1982; Morris et al., 1990; Bell et al., 1993; Goddard et al., 1993; Bishop et al., 1993), so one would expect some variability in their infrared spectral behavior as well. Roush [1992] investigated the transmission properties of several palagonites collected from different rainfall regimes on the island of Hawaii. The transmission behavior of these samples generally fell into two distinct categories based on observed spectral features. One category exhibited a Si-O stretching fundamental that occurred at wavelengths 1.5 μm, a weak feature near 10.53 μm that can be associated with Fe-OH stretching mode, and a doublet in the region Si-O bending fundamental, with maximum absorptions near 18.2 and 23.9 μm. The other exhibited a Si-O stretching fundamental that occurred at wavelengths 210 μm and a single feature in the region Si-O bending fundamental, at approximately 16.1 μm. Roush et al. [1993] show the transmission spectra of two other palagonites that also illustrate this behavior.

Comparisons to Mariner 9 IRIS Spectra

We present four models of palagonite and montmorillonite 219S to four individual IRIS spectra, obtained early during the Mariner 9 mission, when the atmospheric dust optical depths were large. The four Mars regions corresponding to the modeled IRIS observations are in Hellas Planitia (40°S, 306°W), Solis Planum (34°S, 10°W), Argyre Planitia (55°S, 51°W), and southern Chryse Planitia (4°N, 25°W). The pole longitudes of these IRIS spectral observations are Lp = 29° (orbit 17), 208° (orbit 18), 301° (orbit 28), and 305° (orbit 40), respectively. Table 1 presents a list of observational parameters for each of these IRIS spectral measurements. Because we present individual IRIS spectra for all four cases (i.e., no spatial or temporal averaging), the spectral noise is quite high in the 7- to 11-micron wavelength region. Toon et al. [1977] presented model comparisons to average sets of 9 spectra within the Hawaii Planum region (20°S-40°S, 190°W-285°W) for orbits 8, 56, 80, 138, and 154. Toon et al. found the largest dust IR optical depths for orbit 8 (τ9μm = 1.5) and 56 (τ9μm = 0.8), and no clear indication of any change in the dust IR optical properties over the range of time studied (τp=295-350). As demonstrated below, we derive comparable dust opacities (τ9μm = 1.1-0.7) and spectral lines fits (with the montmorillonite 219S, τ9μm = 2.7-μm model) for the three more southern regions (56, 50, and 34°S) we analyze here. Our more northern region (4°N) presents a smaller dust optical depth (τ9μm = 0.5), suggesting a possible hemispheric gradient in the atmospheric dust loading at this time (τp = 300°). Perhaps more interesting, we find the highest dust opacity for the Solis Planum region, which has often been cited as a region of ongoing dust storm activity. This observation corresponds to a high-altitude region (τ9μm = 1.1, altitude = 7 km), and yet it indicates a significantly higher dust opacity than observed over the lower-elevation Hellas region (τ9μm = 0.7, altitude = 4 km) at the same time (i.e., orbit 17 versus orbit 18).

To resolve the discrepancy, we believe that the infrared imaginary indices are more accurate, as they compare quite well with the derived infrared values. These rescaled values are also shown in Figure 1a, as the solid line curve, which now merges with the mid-infrared values from Roush et al. [1991], and with the ultraviolet and visible values adopted from Clark et al. [1990].

Table 1. Mariner 9 IRIS Spectra

<table>
<thead>
<tr>
<th>Region</th>
<th>Latitude, Longitude</th>
<th>Local Time, hours</th>
<th>Emission Angle, deg</th>
<th>DAS</th>
<th>Orbit, Lp, deg</th>
<th>Surface Pressure, mbar</th>
<th>Surface Temperature, K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argyre</td>
<td>55.0°S, 13.95°W</td>
<td>52.8</td>
<td>2605528</td>
<td>28</td>
<td>301.1</td>
<td>5.5</td>
<td>233</td>
</tr>
<tr>
<td>Planitia</td>
<td>50.7°W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hellas</td>
<td>39.8°S, 12.2°W</td>
<td>46.7</td>
<td>2209744</td>
<td>17</td>
<td>297.8</td>
<td>8.7</td>
<td>246</td>
</tr>
<tr>
<td>Planitia</td>
<td>306.4°W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solis</td>
<td>24.2°W, 13.3°S</td>
<td>61.3</td>
<td>2245531</td>
<td>18</td>
<td>298.1</td>
<td>2.9</td>
<td>273</td>
</tr>
<tr>
<td>Planum</td>
<td>109.1°W, 10.3°S</td>
<td>73.7</td>
<td>3032992</td>
<td>40</td>
<td>304.7</td>
<td>6.7</td>
<td>238</td>
</tr>
<tr>
<td>Southern</td>
<td>4.3°N, 25.2°W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chryse</td>
<td>Planitia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DAS, Data Acquisition Subsystem time (incremented 1 count per 1.2 seconds time)
CLANCY ET AL.: NEW MODEL FOR MARS ATMOSPHERIC DUST

insensitivity is due to the fact that absorption dominates over scattering within the dust 9-um extinction over a wide range of dust parameters. As a result, the Viking IRITM 9-um dust opacities derived by Marvin (1986) on the basis of the Toon et al. model of Mars atmospheric dust are not overly dependent on this model, although their interpretation in terms of dust mass would be. More important, the detailed fit of the wavelength dependence of the 9-um dust absorption and the relative depths of the visible, 9-um, and 20-um dust absorptions are very sensitive to the dust composition and particle size distribution.

Besides definition of the atmospheric temperature profile, it is also necessary to constrain the IR brightness temperatures of the surface regions on Mars. It is possible to use the 5- to 7-um brightness temperatures from the IRIS spectra to constrain the 5- to 7-um surface brightness temperature, since the dust absorption is smaller (but not negligible) at these wavelengths. At longer IR wavelengths the atmospheric dust absorption is more significant. Toon et al. (1977) assumed a wavelength-independent surface emissivity/brightness temperature, based on comparison to an IRIS spectrum (orbit 200) taken well into the decay of the 1971-1972 dust storm. A recent analysis of IRIS emission angle sequences for orbits 174 and 176 also indicated no gradient in surface brightness temperatures between 5 and 30 um (McElroy et al., 1995).

In the following analysis, we consider the effects of a wavelength-dependent surface emissivity on the surface brightness temperatures, based on Viking IRITM observations of 7- to 11-um surface brightness variations (Christensen, 1982, 1986) and IRIS observational limits on the 5- to 30-um variation in surface emissivity. Christensen (1982) found a significant variation in the IR emissivity of lower albedo (albedo=0.28) regions on Mars versus a wavelength-independent IR surface emissivity for the brightest (albedo=0.80) regions on Mars. The maximum 7- to 20-um variation in the IR emissivity of the Mars surface (Syria Major, albedo=0.10) is 0.98 to 0.62, corresponding to a decrease in surface brightness temperature of roughly 8 K between wavelengths of 7 and 20 um. In fact, Christensen found that much of the decrease in the surface emissivity occurred at wavelengths between 7 and 11 um, which would lead to considerable ambiguity in separating the 9-um atmospheric dust absorption from the surface emissivity effects. Since the IRITM analysis of Christensen (1982) is based in nature and indicates considerable variability and/or uncertainty in the detailed wavelength dependence of the surface IR emissivity, we model only the effects of a simple linear decrease in the surface brightness temperature between 5 and 30 um. The implications of 8- to 11-um spectral structure in the surface emissivity are considered in the discussion of our modeling results. A separate wavelength dependence in the surface brightness temperatures of Mars results from differences in the thermal properties of the dust and rock components of the Mars surface. Christensen (1986) analyzed the nighttime differences in the 7- and 20-um IRITM brightness temperatures to map rock abundance on the Mars surface. However, this effect is nearly zero during the daytimes, when the above IRIS spectra were taken, and does not influence the current analysis.

We model three separate particle size distributions for the Mars atmospheric dust, all based on the modified gamma distribution (e.g., Hansen and Travis, 1974) for direct comparison to the Toon et al. model. The modeled size distributions, presented in Figure 1b and Table 2, include (size 1) the Toon et al. model ($r_{mod}=0.40$ microns), (size 2) a size distribution with roughly two times smaller particles ($r_{mod}=0.17$ microns) but the same size distribution width as the Toon et al. model (varr=0.4 microns), and (size 3) a size distribution with much smaller particles ($r_{mod}=0.02$ microns) combined with a much broader size distribution ($r_{mod}$ variance=0.8 microns) than the Toon et al. model. At visible wavelengths, the effective (cross-section weighted, r_eff) radii for these three size distributions are 2.7, 1.2, and 3.8 microns, respectively. As demonstrated below, the particle size distribution is the primary influence on the relative variation in the dust optical depth at visible, 9-um, and 30-um wavelengths. The Toon et al. size distribution was derived primarily on the basis of the 7- to 10-um absorption properties of the atmospheric dust. The second particle size distribution allows a better fit to the relative visibility-9-um variation in dust opacity and the Phobos near-IR observations of the dust extinction opacities, but requires a substantial 5- to 30-um variation in surface emissivity to fit the IRIS brightness temperatures at 18-30 um. The third particle size distribution is chosen to provide agreement with the optical depth variation of the dust from visible, near-IR, 9-um, and 30-um observations, for the case of minimal wavelength dependence of surface emissivity between 5 and 30 um. As seen in Figure 1b, this particle size distribution exhibits a great deal of very fine dust ($r < 0.1$ microns), yet provides a significant amount of 1- to 10-um sized particles due to the broadness of the size distribution.

Argyre Planitia

The bottom half of Figure 2a presents the brightness temperature calculations for the standard model of Toon et al. (monteirenillo 219b, size 1, r_eff = 2.7 microns) as compared to Table 2, Modified Gamma Size Distribution Parameters

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>r_{mod}</th>
<th>r_{eff}</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size 1</td>
<td>2</td>
<td>3.32</td>
<td>0.5</td>
<td>0.40</td>
<td>2.7</td>
</tr>
<tr>
<td>Size 2</td>
<td>2</td>
<td>9.6</td>
<td>0.5</td>
<td>0.17</td>
<td>1.2</td>
</tr>
<tr>
<td>Size 3</td>
<td>1</td>
<td>12.0</td>
<td>0.3</td>
<td>0.014</td>
<td>1.8</td>
</tr>
</tbody>
</table>
the observed IRIS brightness temperature spectrum (data points) of the Argyre Planitia region (see Table 1). The surface emissivity was assumed independent of wavelength for this modeled spectrum. The top half of Figure 2a indicates the wavelength-dependent single-scattering albedo (SSA, dotted line) and extinction opacity (solid line) for the modeled dust properties. In this case, the peak 9-μm dust opacity is 0.8. The fit of the Toon et al. dust model to the IRIS spectrum appears very good in the 9-μm region but much less satisfactory in the 20-μm region, consistent with the original analysis of Toon et al. [1977]. In Figure 2b, we present the fit of the montmorillonite 219b optical constants for a smaller particle size distribution (size 2, \( r_{wp} = 1.2 \) μm), indicating the sensitivity of the detailed shape of 9-μm absorption band and the 18- to 30-μm dust absorption to the dust particle sizes. Notice that the larger dust particle sizes (size 1, \( r_{wp} = 2.7 \) μm) better fit the IRIS spectra beyond 15 μm with a wavelength-independent surface emissivity, whereas the smaller dust particle sizes (size 2, \( r_{wp} = 1.2 \) μm) would require a significant (~5 K) decrease in the modeled surface brightness temperature between 5 and 30 μm to fit the IRIS spectra in the same region. This conclusion holds true for the palagonite as well as the montmorillonite composition. The palagonite model, presented in Figure 2c for the smaller dust particle sizes (palagonite, size 2, \( r_{wp} = 1.2 \) μm) and wavelength-dependent (dashed line) and wavelength-independent (solid line) models of the surface brightness temperature, provides a reasonable fit to the observed 9- to 12-μm dust absorption, but does not present sufficient absorption in the 8- to 9-μm wavelength region. On the other hand, this model leads to an absence of structured absorption in the 20-μm wavelength region, which is much more consistent with the IRIS spectral observations in this region. We present results for the third, broad particle size distribution with a palagonite composition in Figure 2d (palagonite, size 3, \( r_{wp} = 1.3 \) μm), which incorporates a wavelength-independent surface brightness temperature. The model fit to the IRIS observations is very good for the 9- to 12-, 20-, and 30-μm regions, but still does not provide sufficient absorption in the 8- to 9-μm region. For an additional compositional comparison, we present a model calculation for the basalt optical constants provided by Toon et al. [1977]. Figure 2e indicates a best fit for these basalt optical constants and the Toon et al. size distribution (size 1). This model does not fit either side of the 9-μm absorption very well, but does lead to much subdued absorption structure at 20 μm. All five of the model cases presented in Figure 2a-2e indicate 9-μm extinction opacities of 0.7-0.8.

### Helias Planitia

Both the Argyre and Helias Planitia spectra were observed at moderate emission angles (53° and 47°, respectively). The Helias spectrum (Figure 3) was taken over the deeper western portion of the Helias basin (Table 1), but still exhibits fairly moderate 9-μm dust absorption relative to the two other southern regions presented (i.e., \( \sim 0.6 \) versus 0.8 and 1.1). The three basic models of dust are compared to the Helias observation in Figures 3a-3e. These include the standard Toon et al. model (Figure 3a, montmorillonite 219b, size 1), the palagonite model with a factor-of-two reduction in particle sizes (Figure 3b, size 2), and the palagonite model with a very broad particle size distribution (Figure 3c, size 3). These comparisons are very similar to those presented for the Argyre Planitia region. The standard Toon et al. model provides the best fit to the observed 9-μm absorption, but

![Figure 2b](image-url)

**Figure 2b.** Same as Figure 2a, but for montmorillonite 219b with the smaller particle size distribution 2.

![Figure 2c](image-url)

**Figure 2c.** Same as Figure 2a, but for a Hawaiian palagonite composition with particle size distribution 2, including a case in which the surface brightness temperature decreases by 6 K between 5 and 30 μm (dashed line).

![Figure 2d](image-url)

**Figure 2d.** Same as Figure 2a, but for a Hawaiian palagonite composition with the very broad particle size distribution 3.

![Figure 2e](image-url)

**Figure 2e.** Same as Figure 2a, but for a basaltic composition with the Toon et al. particle size distribution 1.
CLANCY ET AL.: NEW MODEL FOR MARS ATMOSPHERIC DUST

Figure 3a. Same as Figure 2 for an IRIS spectrum over Hellas Planitia (see Table 1). The modeled dust case is the Toon et al. model of montmorillonite 219b with particle size distribution 2.

Figure 3b. Same as Figure 3a, but for a Hawaiian palagonite composition with the very broad particle size distribution 3.

Figure 3c. Same as Figure 3a, but for a Hawaiian palagonite composition with the very broad particle size distribution 3.

Figure 3d. Same as Figure 2 for an IRIS spectrum over Solis Planum (see Table 1). The model of atmospheric dust presented is the Toon et al. model of montmorillonite 219b with particle size distribution 1.

overestimates the 20-μm feature with distinctive structure that is not present in the observations. It also leads to an incorrect ratio of the visible-9-μm dust extinction opacity, as described below. The palagonite model with size distribution 2 (Figure 3b) does not match the detailed shape of the observed 9-μm absorption. It also does not provide sufficient absorption in the 20- to 30-μm region, such that it would require a significant decrease (~6 K) in the surface emissivity in this wavelength region to fit the observed 20- to 30-μm brightness temperature. The palagonite model with the very broad size distribution 3 (Figure 3c) provides the best overall match to the IRIS spectrum, although it still does not exhibit sufficient absorption in the 8- to 9-μm region.

Solis Planum

We present the three basic models of Mars atmospheric dust for comparison to the Solis Planum spectrum in Figures 4 a-4c. These models include the Toon et al. case (Figure 4a), montmorillonite 219b, size 1, f_MCM = 2.7 μm, wavelength-independent surface emissivity, the palagonite model with a narrow size distribution (Figure 4b, palagonite, size 2, f_MCM = 1.2 μm, wavelength-dependent and independent surface emissivity), and the palagonite model with a very broad size distribution (Figure 4c, palagonite, size 3, f_MCM = 1.8 μm, wavelength-independent surface emissivity). The comparisons of Figures 4a-4c suggest similar conclusions to those indicated for the Argyre Planitia and Hellas Planitia model-data comparisons, although all of the model fits for the Solis Planum spectrum are somewhat poorer in the 7-9 and 18-30 μm regions. The 20 μm absorption of the montmorillonite model is clearly out of line with the observed spectrum. The palagonite models do not produce sufficient absorption in the 8- to 9-μm region, but do provide a better match to the observed 20 μm region than the montmorillonite composition. Palagonite with the small, narrow particle size distribution (Figure 4b, size 2, f_MCM = 1.2 μm) fits the observed spectrum in the 18- to 30-μm region if the surface brightness temperature is decreased by ~15 K between 5 and 20 μm. This is a substantially larger wavelength gradient in surface emissivity than required for the Argyre Planitia and Hellas Planitia observations. Although it is possible that the higher emission angle of observation for the Solis Planum spectrum (61°) may lead to a more extreme wavelength dependence for surface emissivity, the requirement of a 15 K gradient in surface brightness temperature between 5 and 20 μm appears unlikely in light of the analysis of IRIS emission pair sequences by McMillan et al. [1997]. Palagonite with the broad particle size distribution 3 (Figure 4c) provides a reasonable approximation to the observed spectrum in the 18- to 30-μm region without requiring a substantial wavelength dependence for the surface emissivity. We also present a baseline model calculation for the Solis Planum observation, since this spectrum indicates the largest 9-μm dust absorption of the spectra we analyzed. Figure 4d displays the resulting best fit model spectrum with the Toon et al. size distribution (size 1) and basalt optical constants, which does not match the observed absorption structure of the 9- to 20-μm regions well, but does exhibit appropriate levels of absorption at 30 μm due to the larger particle sizes employed.

Another distinction in the Solis Planum IRIS spectrum is a relative maximum at 7- to 8-μm wavelength, which is not common in the spectra for the other regions studied. A similar feature is observed in several of the IRIS spectra analyzed by Toon et al. [1977]. Model calculations with quartz showed
somewhat similar structure, and there are weak CO₂ bands in the 7.4-μm region which may contribute to this feature (Toon et al., 1977). Notice also that the palagonite models lead to weak absorptions just beyond 6 μm which are apparently related to the presence of water in the palagonite sample (Roush et al., 1991).

Southern Chryse Planitia

The southern Chryse observation, taken near the crater Galliata, represents the highest emission angle (74°) and lowest dust opacity (τ_0 = 0.5) of the four spectra presented. The same three basic models of dust are presented against the observed spectrum in Figures 5a-5c. The results of these comparisons are very similar to those for the Solis Planum region in particular.

Toon et al. model (Figure 5a, montmorillonite 219b, size 1, r_mw = 2.7 μm) leads to deep, structured 20-μm absorption that is not observed and a very low ratio for the visible-9-μm extinction opacity ratio. The palagonite model for the smaller, narrow size distribution (Figure 5b, size 2, r_mw = 1.2 μm) does not fit the observed 20-μm absorption, unless an unreasonably large increase in the surface emissivity is assumed for this wavelength region. The palagonite model with the broad particle size distribution (Figure 5c, size 3) provides the best relative fit among the visible, 9-μm, and 30-μm absorption spectrum of the dust, but neither palagonite model provides sufficient absorption in the 8- to 9-μm region.

All of the dust models we consider fail to match the observations in some way. A smaller particle size distribution than the Toon et al. model is required to match the observed visible-9-μm opacity ratio and the Phobos near-IR observations.

Figure 4a. Same as Figure 4c, but for a Hawaiian palagonite composition with particle size distribution 2, including a case in which the surface brightness temperature decreases by 15 K between 5 and 30 μm (dashed line).

Figure 4b. Same as Figure 4a, but for a Hawaiian palagonite composition with particle size distribution 2, including a case in which the surface brightness temperature decreases by 15 K between 5 and 30 μm (dashed line).

Figure 4d. Same as Figure 4a, but for a basalt composition with the Toon et al. particle size distribution 1.

as discussed below. Simply decreasing the mode radius of the Toon et al. particle size distribution by a factor of 2 (size 2) helps to match these observations, but appears to under predict the 20- to 30-μm extinction opacity of the dust. A very broad particle size distribution with a very small modal radius for the dust particle size (size 3) may provide the best fit to the observed variation in the dust extinction opacity between visible and 30-μm wavelengths. A dust composition of montmorillonite provides the best fit to the shape of the observed 9-μm absorption, but is not consistent with the shape of the observed 20-μm absorption of the dust. The modeled palagonite composition can provide a good fit to the observed 20-μm and 9- to 13-μm absorption of the dust, but lacks sufficient absorption in the 8- to 9-μm region. Palagonite also exhibits UV and visible absorption roughly consistent with observations of Mars atmospheric dust (see below), while montmorillonite does not. Several effects may contribute to the poor match of the modeled palagonite to the observed 8- to 9-μm absorption of the dust. The composition of the modeled palagonite is unlikely to be the same as for the Mars dust. Terrestrial palagonites exhibit a wide range of compositions, depending on the composition of the basalt which is weathered to produce the palagonite. The frequency of peak 8- to 9-μm absorption in igneous silicates is known to be sensitive to the SiO₂ content in particular (Hanel et al., 1972). The SiO₂ content of the modeled palagonite (31%) is much lower than the SiO₂ content of montmorillonite 219b (63%), which is quite a bit higher than the Mars soil SiO₂ content (42-43%) determined from the Viking Lander X-ray fluorescence experiment (Clark et al., 1982). Singer (1982) analyzed a terrestrial palagonite sample containing ~48% SiO₂.

Figure 5a. Same as Figure 2 for an IRIS spectrum over southern Chryse Planitia (see Table 1). Modeled dust case is the Toon et al. model of montmorillonite 219b with particle size distribution 1.
heterogeneous material such as palagonite involves a certain modelling of the have on infrared cloud optical depth, which varies as a function of wavelength. Finally, we do not know the detailed nature of the Mars surface emissivity within the sufficiently clear atmosphere observations to determine potential temperature variations of the Mars surface within the 9-µm dust opacity than the single emission angle 9-µm dust opacity inversion algorithm described by Martin [1986]. The surface brightness temperature over the IRTM 9-µm bandpass is assumed constant (no wavelength or emission angle dependence) and is based on the 7-µm brightness temperature from the IRTM 7-µm channel brightness. The atmospheric temperature profile is constructed from the IRTM 7- and 15-µm channel brightness temperatures, with the assumption of a 10-20 K offset between the surface and lower atmosphere temperatures. This temperature offset is chosen according to the local time of the individual EPF sequences modeled [e.g., see Martin, 1986]. The radiative scattering/extinction calculations employ the IRTM bandpass weighting and model dust parameters for both montmorillonite and montmorillonite and palagonite with size distributions 1 and 2, respectively. In Figures 6a and 6b we present model fits to 9-µm EPF brightness temperature from the IRTM channel over Valles Marineris during the 1977 global dust storm. The smooth model curves represent predicted emission angle dependences for the 9-µm brightness temperatures for 3 atmospheric 9-µm dust opacities, based on the Toon et al. dust model.
model.

Figure 6b. An EPF 9-μm sequence observed over Chryse Planitia is compared to three 9-μm opacities, based on the Toon et al. dust model.

sequences over Valles Marineris (9.0 ± 5.5, 71.7 ± 2.0°W) and Chryse Planitia (23.2 ± 23.2°, 68.6 ± 6.9°W). The observations are indicated by individual points; model results are presented as lines for three values of the zenith 9-μm optical depth, for the montmorillonite 219b, single model with 1 dust properties. In Figure 6b, we show the fit of the palagonite, single model with 2 dust properties to the same Chryse Planitia observations of Figure 6b. Both of these models indicate best fit values of -0.3 for the 9-μm dust optical depth for the Chryse Planitia EPF sequence. Similarly, either dust model leads to a 9-μm dust opacity of -0.8 for the Chryse Marineris EPF sequence.

Figure 7 presents the visible 9-μm opacity ratio versus the visible opacity, based upon the 9-μm dust extinction opacities derived from 13 such EPF sequences and the solar band extinction opacities derived for the same 13 sequences (from Toon and Lee, 1991). For visible opacities less than 1, the visible/9-μm opacity ratio is consistent with a value of 2. For visible opacities greater than 1, there appears to be an increase in the visible/9-μm opacity ratio to values as high as 3-4. However, the uncertainties in these ratios are very large for the higher dust opacities such that this apparent trend of the visible/9-μm opacity ratio with dust opacity is not reliably defined. The largest source of uncertainty in the opacity ratios of Figure 7 result from

Figure 7. The ratio of visible and 9-μm dust opacities obtained from solar band (Clancy and Lee, 1991) and 9-μm Viking IRTM EPF sequences (such as in Figure 6) are plotted versus the visible (solar band) dust opacity for 13 analyzed EPF sequences. The predicted ratio of visible/9-μm dust opacity is presented for the Toon et al. dust model (dashed line, montmorillonite 219b with size distribution 1) and the Hawaiian palagonite model with smaller dust particle sizes (dotted line, size distributions 2, 3, 4). The model dependent uncertainties in the visible dust opacities, which are particularly dependent on the dust single-scattering albedo for the higher dust opacities. Uncertainties in the 9-μm dust opacities are contributed primarily by the uncertain surface brightness and atmospheric temperatures employed in the modelling. Martin et al. (1986) estimated uncertainties in 9-μm dust opacities, using this technique, which are less than 15% for the range of 7-9-μm IRTM brightness temperatures employed in this current analysis. The difference in the 9-μm opacity between the two models of the dust properties is small, for the reasons described above. Also indicated on Figure 7 are the predicted visible/9-μm extinction opacity ratios for the three basic models of Mars dust considered in this analysis. A ratio of unity results from the Toon et al. model (montmorillonite 219b, size 1, rcwp = 2.7 μm). The two palagonite models (size 2, rcwp = 1.2 μm; and size 3, rcwp = 1.8 μm) both lead to visible/9-μm opacity ratios of 2. Higher ratios result directly from smaller average particle sizes for the dust. Although dust composition plays a lesser role in determining the visible/9-μm opacity ratio, montmorillonite actually requires substantially smaller particle sizes than the palagonite to obtain a visible/9-μm opacity ratio of ~2.

The Phobos Infrared Spectroscopic Modeling Dust Particle Size Determination

Another important measurement regarding the dust particle sizes are the Phobos near-IR (1.9 and 3.7 μm) occultation measurements of dust extinction profiles (e.g., Krasnopolsky et al., 1989). A more recent analysis of these data in terms of the dust particle sizes is given by Chassefiere et al. (1992), who find that the effective radius (rcwp) of the dust increased from 0.9 to 1.3 μm between the altitudes of 25 and 15 km. These are
significant measurements because all of the particle sizes considered for Mars dust place the effective radii of the dust at sizes comparable to near IR wavelengths. Hence, an optimum measurement of the wavelength dependence in the dust scattering extinction efficiency would be at wavelengths of 2-4 μm. The results of Chassefiere et al. [1992] are generally consistent with the smaller particle sizes we argue for on the basis of the visible/9-μm dust opacity ratios. The width of the particle size distribution from their analysis is quite similar to the size distribution 2 in Figure 1b, which we argue is not consistent with the 20- to 30-μm absorption of the dust indicated in the IRIS spectra. However, Chassefiere et al. indicate very large uncertainties (50%) in their determination of the particle size distribution width.

**Dust UV/Visible Absorption**

Although the detailed variation of the atmospheric dust absorption at visible and ultraviolet wavelengths is not well measured, it is known that the dust exhibits single-scattering albedos of order 0.4-0.6 at 3000 Å [Pang and Ajello, 1977] and 0.86 [Pollack et al., 1977] to 0.92 [Clancy and Lee, 1991] in the solar band average (-6000-10000 Å). A reanalysis of the multifilter photometry of Mars dust from the Viking landers is provided in this special issue by Pollack et al. (this issue). The ultraviolet/visible absorption of the dust has led to arguments for anatase (TiO2) [Pang and Ajello, 1977] and magnetite [Pollack et al., 1977, 1979] components in the Mars atmospheric dust. In fact, the elemental composition of the modeled palagonite is consistent with significant, nanocrystalline components of titanium and iron oxides (4.1 % TiO2, 14.6 % Fe2O3, 1 % FeO [Roark et al., 1991]). We compute ultraviolet (3000 Å) and visible (6000 Å) single-scattering albedos for palagonite dust, based on 3000-10000 Å optical constants of another Maa Kaa palagonite taken from Clark et al. [1990]. The sharp rise in the imaginary index of this palagonite over the 0.3- to 0.7-μm wavelength region is shown in Figure 1a. Ultraviolet single-scattering albedos of 0.56 and 0.57 are derived for particle size distributions 2 and 3 (Figure 1b), respectively. Visible single-scattering albedos of 0.94, and 0.93 are computed for palagonite with the same particle size distributions.

These values are reasonably consistent with the range of observed single-scattering albedos for Mars atmospheric dust. Of course, no conclusions regarding the detailed composition of Mars dust can be drawn for these very coarse components of single-scattering albedos. They merely point out the plausibility of a "palagonite-like" composition for Mars atmospheric dust in terms of their broad UV-visible absorption properties. It is very unlikely that the composition of the modeled palagonite applies to Mars dust. But it appears possible that a "single component" model for Mars atmospheric dust, based on a heterogeneous, nanocrystalline or amorphous weathering product of basalts, may account for the complete set of ultraviolet-to-30 μm observations of Mars atmospheric dust.

**Conclusions**

1. A combined analysis of the Viking EPF visible and 9-μm micron sequences indicates a visible-to-IR dust extinction ratio of 2-3, consistent with the earlier comparison by Martin [1986] of 9-μm IRRTM dust opacity determinations and the Viking lander extinction opacity observations [Pollack et al., 1977]. The observed ratio of 2-3 is significantly different from the predicted ratio of 1.0 from the current model of Mars atmospheric dust.

2. The ratios of visible and 9-μm dust extinction efficiencies for a palagonite or montmorillonite composition are much closer to the observed ratio for smaller particle sizes of the dust (ρmode=0.17 versus 0.49 μm). Palagonite dust for two size distribution models presented (size 2 and 3, Figure 2) yields a visible-to-IR extinction ratio of 2, consistent with the measurements of this ratio, as well as the Phobos near-IR extinction measurements of the dust particle sizes [Clancy et al., 1992]. A montmorillonite composition would require smaller particle sizes (ρmode=0.57 μm, rmode=0.10 μm) than palagonite to obtain the same visible-to-IR opacity ratio of 2, such that it might not be consistent with the Phobos particle size determination for the dust, and would no longer fit the IRIS 9-μm or 18- to 30-μm absorptions well.

3. Palagonite dust can also lead to a much improved fit to the IRIS observations near 20-μm wavelengths, since the sample of palagonite modeled displays no distinct 20-μm absorption. However, the same sample of palagonite does not exhibit sufficiently strong absorption in the 8.5-μm wavelength region relative to the IRIS measurements of this absorption. This difference is significant within the IRIS measurement uncertainties, but may be accommodated by a different Mars palagonite composition and/or uncertainties in detailed modeling of the 9-μm dust absorption.

4. Palagonite dust with a ρmode of 1.2-1.8 μm (ρmode <0.17 microns) leads to single-scattering albedos (ω0), which are roughly consistent with their observed values at ultraviolet (ωuv-0.6 - 0.5) and (ωuv-0.9 -0.6) visible wavelengths. Hence, a single component model for Mars atmospheric dust may explain the existing observations of the dust absorption from 0.3 to 20 μm within the modeling and measurement uncertainties.

5. Comparison of palagonite and montmorillonite dust with the same size distribution gives similar 9- to 30-μm opacities, except from the deep doublet structure at 20-μm and positive frequency shift of the 9-μm band for montmorillonite versus the modeled palagonite. However, the average 18- to 30-μm opacity is very different for ρmode=0.27 versus 1.2 μm (ρmode=0.40 versus 0.17 μm) particle sizes of either composition. The key uncertainty in differentiating size distributions 1 and 2 (which possess roughly the same size distribution width, a rvariance=0.4 μm) on the basis of the 18- to 30-μm dust opacity is the degree to which the surface emissivity may vary from 5 to 30 μm. The large decreases (6-13 K) in surface brightness temperatures between 5 and 30 μm required by size distribution 2 seem unlikely, and indicate the possibility of a third model for the Mars atmospheric dust distribution. In this case, a very broad particle size distribution (ρmode variance-0.8 μm) is modeled to provide the observed, relative variation of the dust opacity between visible, 9, and 30-μm wavelengths without the requirement of a substantial wavelength dependence for the surface emissivity. This particle size distribution leads to large numbers of very fine dust particles (ρmode <0.02 microns) to provide the observed visible/9-μm opacity ratio and significant numbers of μm-sized particles (ρmode=1.8 μm) to fit the 18- to 30-μm region of the IRIS spectral observations.

In summary, we submit that palagonite with a smaller and broader particle size distribution than that of the Toon et al. determination provides a much improved model to Mars atmospheric dust. Since palagonite is a common and fairly primitive weathering product of terrestrial basalts, palagonite could sensibly be considered as a major surface component on Mars (e.g., Soderblom et al., 1978; Singer, 1982; Roark et al., 1993). The weakness or absence of reflection signatures from clay minerals on the surface of Mars [Bell and Crisp, 1993; Clark, 1992] could also be consistent with a palagonite
composition for Mars dust if the conditions for basal weathering on Mars produced sufficiently enobrous forms of palagonite. On the other hand, such clay bands may be spectrally masked by another component [Doremus and Haux, 1992] or the clay minerals, themselves, may be dehydroxylated [Barus, 1993]. Variations in palagonite composition and textures are large for terrestrial samples [Dove et al., 1981; Golden et al., 1993], which can lead to significant variations in their 7- to 20-μm absorption properties [Ainsworth et al., 1993]. It is noteworthy that the palagonite composition modeled in this work is very low in SiO₂ abundance (31%, Roach et al. [1991]) as compared to the SiO₂ content of the montmorillonite 2:1:6 sample (68%, Toon et al. [1977]). As noted by Hanel et al. [1972] and Toon et al. [1977], the 8- to 9-μm absorption properties of silicates are particularly sensitive to the SiO₂ abundance. In addition, the Mie scattering theory employed in the radiative transfer modelling may not accurately determine the 9-μm band position and shape (e.g., West et al., 1999), and uncertainties in the wavelength-dependent emissivity of the surface may prevent an accurate fit. In any case, the key distinction in our new model versus the old model of Mars dust may follow less from differences in composition as from the low degree of crystallinity of palagonite versus montmorillonite, and the smaller particle sizes derived for the atmospheric dust. Both conditions tend to suppress the 20-μm absorption relative to the 9-μm absorption and increase the visible/IR dust opacity ratio. Such a change in the model of Mars atmospheric dust is clearly relevant to a similar controversy regarding the degree of crystallinity and clay mineralization in the soil covering the Mars bright regions (e.g., Soderblom, 1992; Bains et al., 1992).

It is also worth emphasizing that the particle sizes of Mars dust are remarkably small for all three models. The modal radii of the atmospheric dust-particle sizes range from <0.02 μm for our preferred, broad size distribution model to 0.60 μm for the Toon et al. [1977] model. The method of lifting Mars dust into the atmosphere is not entirely certain, but may involve radiation of larger particles (e.g., r ≈ 100 μm Greeley et al., 1980) from a much finer dust component (e.g., r < 2 μm). The longer settling times of the fine dust particles would lead to their availability at the surface for longer periods of time, subject to many other processes such as surface bonding and condensation. Our model of smaller particle sizes for Mars dust (mod = 0.17 μm < r < 0.02 μm) would increase the settling lifetime of Mars atmospheric dust. The very fine particles (r < 0.02 μm) might require condensation processes for their removal from the atmosphere over annual timescales, or they may in fact remain as a constant background aerosol in the Mars atmosphere. Current dynamical models of Mars dust transport point to the importance of the dust size distribution on the transport and temporal evolution of atmospheric dust [Morphy et al., 1993]. Smaller average dust particle sizes also lead to significant changes in the radiative forcing of the Mars atmospheric dust on atmospheric temperatures. The factor of 2 or more decrease in particle sizes for our models versus the Toon et al. [1977] model of the dust are directly related to the factor of 2 increase in the visible/IR opacity ratio between the same models. Changing this opacity ratio can dramatically influence dust heating/cooling rates for the Mars atmosphere. In fact, the thermal modelling analysis of Hapke and Jakosky [1991] requires a visible/IR opacity ratio of 2.25 to reproduce the sign of the temperature change within the lower Mars atmosphere during the initiation of 1977 global dust storm. In this case, the difference in the visible/IR opacity ratio between the dust particle size models leads to conditions of atmospheric cooling versus heating.

Acknowledgements. The authors are indebted to S. Murchie, J. Bishop, B. Clark, and P. Barnes for helpful discussions regarding our current understanding of the Mars surface soil. We also benefited from very constructive reviews by J. Bell and J. Crisp. Assistance in preparation of the manuscript was provided by A. Allen, J. Lucey, and B. Droege. Funding for this work was provided through a NASA Mars Surface and Atmospheric Through Time (MAST) grant.

References


Hansen, G. E., Thermal infrared properties of the Martian atmosphere, 4,


Rosh, T. L., Characterization of the spectral reflectance of mafic silicates, hydrous silicates, and hydrous silicon-water ice mixtures in the 0.6 to 4.5 μm wavelength region and applications to planetary science, Ph.D. dissertation, 130 pp., Univ. of Hawaii, Honolulu, 1987.


Page intentionally left blank
Temperature measurements of a Martian local dust storm

Ralph Kahn
Jet Propulsion Laboratory, California Institute of Technology, Pasadena

Abstract. A technique for estimating the ground and near-ground atmospheric temperatures within a Martian local dust storm is presented. It is applied to soundings taken by the Viking orbiter infrared thermal mapper (IRTM) instrument at four times-of-day for one storm. Essentially, a comparison is made between infrared radiances emerging from the storm interior and those from the region surrounding the storm. Particle extinction properties are assumed to be independent of position in the storm region, and scattering properties must be selected arbitrarily. For the storm studied here, the ground temperature in the interior is at least 6 K cooler, whereas the near-ground atmospheric temperature may be less than or comparable to, those of the surroundings. The thermal structure of the storm interior did not change measurably between 11.5 and 16.6 hours local time. These observations favor theories of dust storm development in which regional winds rather than local, dust-driven convection initiate the mobilization of dust from the surface. It is also concluded that the optical properties of dust particles in this local storm differ from those observed by Mariner 9 during the 1971-1972 global dust storm.

1. Introduction

Although the occurrence of local and regional dust storms on Mars is well documented (Bragg et al., 1979; Peterfreund and Kiefer, 1979; Zarek, 1982), there is still some question as to the means by which dust is initially mobilized at the Martian surface (e.g., Groswy et al., 1992). The idea that regional winds superimposed on the general circulation provide the shear stress necessary to lift dust at the surface is supported both by theory and observation (Leovy et al., 1975; Zarek et al., 1992). Similarly, there is evidence that direct heating of the surface creates free convective vortices that mobilize dust locally as dust devils (Gierasch and Goody, 1973; Ryan and Lucich, 1983; Thomas and Gierasch, 1985). Constraints on the evolution of temperature within Martian local dust storms would provide important clues as to the mechanisms that operate under different environmental conditions on Mars.

This paper demonstrates a technique that places some very crude but relevant constraints on the temperatures of Martian local dust storms. The method relies upon broadband infrared soundings from the infrared thermal mapper (IRTM) experiment flown on the Viking mission to Mars. The IRTM made observations over an extensive part of the planet, covering a wide range of seasons and local conditions. Although four of the five detector channels were designed to examine surface emissions (Kiefer et al., 1972), the substantial infrared opacity of the dusty Martian atmosphere, even during relatively clear conditions, produced a measurable atmospheric contribution to the observed infrared radiances (Kiefer et al., 1977). These data represent the most detailed and extensive spatial and temporal record of the Martian atmosphere yet available. In particular, IRTM data include observations made during the development of a few local dust storms (Peterfreund and Kiefer, 1979). Thermal signatures at these sites show distinct spectral characteristics that contain information about the thermal behavior of the storms.

Current knowledge of infrared absorption and scattering properties of particles in the Martian atmosphere, the surface emission properties, and our understanding of the vertical distribution of optical depth in the Martian sky, are limited. Interpretation of broadband radiances, which sample the surface together with the lower atmosphere, is therefore only partly constrained by existing data. This paper explores a number of the most reasonable hypotheses about the atmospheric optical and thermal structure for consistency with the IRTM observations over a dust storm. The technique essentially makes a comparison between infrared radiances emerging from the storm interior and those coming from the area surrounding the storm. Data taken over the Viking Lander 1 (VL1) site are also examined.

The paper begins by presenting the IRTM data used in this study, and the model used to simulate IRTM spectra. Then the thermal signatures of an area adjacent to a local dust storm are examined. In the dust storm surroundings, the vertical temperature structure is easier to constrain than in the interior. IRTM soundings in the surroundings are used to obtain ratios of particle extinction optical properties from constraints on the temperature structure. Particle scattering properties must be selected arbitrarily for this calculation, but we can place sufficient constraints on the atmospheric optical depth and vertical temperature structure to obtain meaningful extinction coefficients. Next, an attempt is made to calculate the mean atmospheric infrared spectral extinction over the VL1 site, as was done in the dust storm surroundings. However, the low atmospheric optical depth at the VL1 site makes the inversion particularly sensitive to the unknown ground temperature and emissivity. Finally, particle properties are assumed to be independent of position in the storm region. With the help of the constraints on particle properties from the storm surroundings, IRTM soundings of the storm interior are used to obtain an...
effective dust cloud temperature and the ground temperature at the storm center.

The results of these studies represent a step toward understanding the mechanisms of dust cloud generation and growth. The data examined in this paper are of low surface spatial resolution, so it is not possible to perform critical tests on length scales small enough to rule out either of the proposed dust storm mechanisms. However, this work establishes the viability of a technique that can be applied to data with higher spatial resolution.

2. IRTM Dust Storm Data

The data used in this study were taken by the Viking Orbiter IRTM experiment in four broadband channels centered near 7.9, 11, and 20 µm. The IRTM channel passbands are given in Figure 1. The 7.9-, 9-, and 11-µm bands are each about 2 µm wide. The 20-µm band is nearly 6 µm wide. Following conventions established in earlier presentations of IRTM data (e.g., Kieffer et al., 1977), the observed radiances are given in terms of brightness temperatures T1, T9, T11, and T20, measured in the bands centered near 7.9, 11, and 20 µm, respectively.

The IRTM instrument made 28 simultaneous soundings: four measurements directed at each of seven spots on the surface. The instrument contained four telescopes. Each telescope sampled light from all seven spots, and focused this radiation onto seven detectors. The telescopes were aligned so that the same seven surface locations, labeled spots one to seven, were observed by each telescope. One telescope served seven 20-µm detectors, another provided light for seven 11-µm detectors, so there is an 11-µm and a 20-µm measurement for every spot in the IRTM data. A third telescope made seven all-red measurements, which are not used in this study. The fourth telescope had three 7-µm detectors, three 9-µm detectors, and one 15-µm channel to measure atmospheric emissions. Therefore, simultaneous 7- and 9-µm measurements cannot occur at any spot. For this reason all the spectral measurements used here are compared with the 11-µm data, which was obtained for every location. Further details of the IRTM are given by Kieffer et al. (1977).

A local dust storm that occurred in the Solis Planum region of Mars during summer in the southern hemisphere (Ls = 237°) was selected for this study. The storm appeared in four separate sequences of IRTM observations on the same day, between 11.5 and 16.6 hours local time (Peterfed et al., 1979). The orbiter cameras recorded a dust cloud in this region on the subsequent day. Figure 2 illustrates the dust storm data. Each graph contains the spectral brightness temperature differences along an east to west traverse, transecting the region of the storm with the most pronounced temperature differences relative to the surroundings. Spectral temperature differences are calculated from pairs of observations with the same telescope spot number to minimize sampling errors in regions of large thermal contrast.

Although brightness temperature values exhibit a 3 K to 5 K scatter for adjacent measurements, temperature differences taken in this way are consistent to better than 1 K in most instances. Some sampling error does occur, since for fixed temperature, the slope of the Planck function depends on wavelength. This causes the spectral detectors to have varying sensitivity to thermal inhomogeneity in the field of view (Martin et al., 1979; Jakosky, 1979). The spatial resolution on the Martian surface of the four sequences shown lies between 220 km and 230 km. For this study, data were selected within a 3° by 4° strip centered about the latitude shown, so the centers of the sampling footprints are within 180 to 240 km of the latitude of the traverse. The physical size of the dust storm, as judged from the perturbation of the temperature in IRTM contour maps (Peterfed and Kieffer, 1979) is about 450 km by 850 km. This size is comparable to that of the cloud photographed in the same region on the following day by the orbiter cameras. In summary, the available data average out any thermal features with length scales less than 200 km, and there are three to four resolution elements across the storm itself.

The accuracy of the brightness temperatures varies among the channels. The absolute errors are less than 0.5 K in the T1, T9, T11, and T20 bands for the temperature ranges considered here, according to Kieffer et al. (1977).

The storm center is taken to be at the size of minimum T11. (Maps of T11, showing the locations of the minima for several storms, are given by Peterfed and Kieffer [1979]). In Figure 2, note that T9 - T11 reaches a maximum near the storm center. In three of the traverses, T9 - T11 has a maximum near the storm center, whereas T20 - T11 shows a relative minimum. The temperature differences are asymmetric about the extremes, with a considerably steeper gradient in T9 - T11 (where λ = 7, 9, or 20) to the east of the storm center, and a progressively shallower slope to the west for later measurements, except for T20 - T11 in the fourth sequence. From these data, representative models of the dust storm center and surroundings were constructed, to be used in the numerical simulations. These are presented in Table 1. For this storm, the brightness temperatures decrease about 35 K, the T8 - T11 and T9 - T11 contrasts increase, whereas T20 - T11 decreases in the dust storm interior relative to the surroundings.

3. Numerical Simulation of Spectra

To explore the dependence of infrared signatures on physical conditions, synthetic spectra were produced to compare with the IRTM data. A two-stream radiative transfer model similar to that of Toon et al. (1977) was used. Toon et al. analyzed data from the infrared interferometer spectrometer (IRIS) that flew on the Mariner 9 mission to Mars. The instrument obtained high-resolution spectra for wavelengths between 5 and 50 µm. Vertical temperature profiles were calculated from an analysis of CO2 bands in the IRIS data. The dust model required input values of the ground temperature, the optical depth at one wavelength, and the particle extinction and scattering properties. Toon et al. assumed spherical particles, and let the real and imaginary parts of the index of refraction at each wavelength, as well as the particle size distribution, be free parameters giving the particle scattering properties in their model. They obtained particle indices of refraction, size distribution, and atmospheric optical depth at each wavelength by comparing the model results with the IRIS data. The data analyzed by Toon et al. were taken during the Martian global dust storm of 1971-1972, when the atmospheric dust loading was near its maximum.
Figure 2: IRTM dust storm data. Each longitude traverse contains data from a latitude strip 3' to 4' wide. Temperature differences (K) and the 11-km brightness temperature are plotted as functions of longitude for the four IRTM sequences. Sequence 210180 was taken at 11.5 hours local time, with a surface resolution of 220 km, 210181 is at 12.8 hours and 230-km resolution, 210182 at 14.6 hours and 230-km resolution, and 210183 at 16.6 hours and 225-km resolution.

For the present study, source functions were obtained at all levels using upward and downward intensities calculated from the two-stream equations. This method was selected because it is the simplest one that contains two parameters describing Martian dust properties; this is as much information about dust properties as is available in the data. The method also allows direct comparison with the results of Toon et al. The technique approximates the angular dependence of the radiation field due to scattering anisotropy by taking hemispherical average values in two directions (appendix). To obtain the emergent intensity, the source function in the upward direction is integrated along the path selected by the viewing angle, thereby preserving the geometrical angular dependence of the result. The equation for the upward two-stream intensity is written as a second-order differential equation with two-point boundary values, solved using a standard numerical method [Lindzen and Kuo, 1969].

The downward two-stream intensity is subsequently obtained as an initial value problem using a Runge-Kutta method of order 3.

Table 1. Representative Dust Storm Thermal Signature

<table>
<thead>
<tr>
<th>Storm Center</th>
<th>Storm Surrounding</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0 - T11</td>
<td>16.</td>
</tr>
<tr>
<td>T0 - T11</td>
<td>7.</td>
</tr>
<tr>
<td>T00 - T11</td>
<td>0. to -2.</td>
</tr>
<tr>
<td>T11</td>
<td>240.</td>
</tr>
<tr>
<td>T11</td>
<td>275.</td>
</tr>
</tbody>
</table>

Details of the calculation are given in the appendix. The procedure was tested by calculating the examples given by Toon et al., and obtaining the same results to two significant figures or better.

Unlike the high-resolution IRIS data, this study relies upon broadband observations, with a direct measurement of the atmospheric temperature at only one level, near 24 km, obtained with the IRTM 15-µm channel. The particle scattering properties at each wavelength are given by two parameters, the single scattering albedo (ω0) and the integral of the normalized single scattering phase function weighted by the cosine of the scattering angle, which is called the asymmetry factor (β). These parameters cannot be determined from the IRTM data available in the region of interest; initially, they are assumed to be equal to those derived by Toon et al. The sensitivity of the results to small changes in these values is examined later. Vertical temperature structure, spectral optical depth, and ground temperature are free parameters, as discussed below.

To produce a model spectrum, the ground temperature, the atmospheric thermal profile, the vertical distribution of optical depth, the spectral extinction coefficients Q(λ), the scattering parameters ω0 (λ) and β (λ), and the surface emissivity must be specified. The model allows arbitrary selection of vertical temperature structure. It is parameterized here by the ground temperature T0, the atmospheric temperature at the ground T00, the boundary layer height ZBL, and boundary layer lapse rate Tα, the atmospheric lapse rate Tα, and the height of the tropopause ZTOP, above which the temperature is taken to be independent of
height. These quantities are illustrated in Figure 3a. The thermal structure used for dust storm interior models, which are discussed in a later section, are shown in Figure 3b. The emission angle for most of the dust storm observations used in this study lies between 30° and 45°. In most of the models, this parameter is set to 35°, which happens to be the angular region where the two-stream method most closely approximates detailed radiative transfer calculations for expected particle size distributions (Foss et al., 1977). The calculated spectral brightness temperature differences vary less than 1 K when the input emission angle is changed by 5° around the nominal value selected.

4. Models of the Dust Storm Surroundings

We seek information about atmospheric and ground temperatures in the interior of a dust storm. The general radiative transfer problem requires a knowledge of both the particle properties and the vertical temperature structure. Reasonable estimates of some of the particle properties can be obtained using an assumed vertical temperature in the dust storm surroundings. Some parameters in the temperature profile of the storm interior can then be retrieved using the deduced particle properties.

In this section, we calculate the ratios of the spectral extinction coefficients required to reproduce the observed spectra for the dust storm surroundings. These results will depend upon the unknown vertical temperature structure in the dust storm surroundings, so a range of possible values is obtained for the spectral extinction coefficients, based upon different choices of $T_{BL}$, $T_{TROP}$, and $T_{g}$. However, with the available data, reasonable limits are placed on the values of these parameters in the storm surroundings. Extinction coefficients are represented as ratios $R_{\lambda} = \frac{Q_{\lambda}}{Q_{11}}$, where $Q_{\lambda}$ is the extinction coefficient in the wavelength band $\lambda$, which may be 7, 9, or 20 $\mu$m, and $Q_{11}$ is the extinction coefficient in the 11-$\mu$m band. For each assumed temperature profile, these ratios of extinction coefficients are calculated, together with a value of the optical depth at 11 $\mu$m, so that the four observed brightness temperatures match the four observed brightness temperatures of the storm interior.

Since the ratios of extinction coefficients are less dependent on the assumed temperature profile than are the optical depths at each wavelength, these ratios are utilized for later use in the dust storm interior models. For the interior models, the 11-$\mu$m dust cloud optical depth ($\tau_{11}$) is treated as a free parameter. With these inputs, and the observed spectrum at the storm center, values of $T_{g}$ and the effective dust cloud temperature $T_{e}$ are constrained using a least squares scheme.

Figure 4 shows the results of numerical modeling of the dust storm surroundings. We now discuss the choice of parameter space and sensitivity of the results to each of the parameters. The uncertainty in $R_{\lambda}$ is primarily due to the indeterminacy of $T_{g}$. $T_{g}$ varied from 288 K to 292 K. Near midday, and where the atmospheric visible optical depth is less than 1, $T_{g}$ is expected to be greater than the atmospheric temperature, based on theoretical modeling (Gierasch and Goody, 1972; Pollack et al., 1979). The largest observed brightness temperature in the dust storm surrounding is 286 K, which is therefore a lower bound on the choice of $T_{g}$. The upper bound on $T_{g}$ in the parameter space is arbitrary. A reasonable guess at an upper bound on $T_{g}$ is made by noting that $T_{g}$ is a sensitive function of the 11-$\mu$m atmospheric optical depth. There is no direct measurement of the 11-$\mu$m optical depth, but the visible optical depth in the dust storm surroundings is typically less than 1, as evidenced by the appearance of ground features in the orbiter images of local dust storms [Briggs et al., 1979], including one taken in the same region as the storm used in this study, on the subsequent day.

Next, the 11-$\mu$m optical depth needs to be related to the estimated visible optical depth. The relative particle extinction cross sections for Martian dust in the visible and infrared regions is not yet well established, but probably has a value around 2 [Zurek, 1982; Martin, 1986]. Pollack et al. [1979] use visible properties obtained from the Viking Lander camera experiments and infrared properties derived from the Mariner 9 IRIS measurements, thereby assuming that the atmospheric particles are the same in these two situations. Using these data, the 11-$\mu$m optical depth would be less than the visible optical depth.

![Figure 3: Parameters for model thermal profiles. (a) Model parameters for atmospheric temperature structure. Here input ratios $\Gamma_{T} = \Gamma_{BL} / \Gamma_{TROP}$ and $\Gamma_{g} = \Gamma_{BL} / \Gamma_{TROP}$. (b) Model parameters for dust storm interior temperature structure.](image-url)
Unfortunately, we later conclude that the particle properties observed over the dust storm must differ from those of the Martian 9, so we have no way to make a direct comparison between visual and 11-μm opacities, which would set a firm upper bound on the value of $T_g$. However, the infrared extinction cross sections for micrometer-sized dust particles are typically smaller than the visible extinction cross sections. Therefore, values of $T_g$ were selected for which the 11-μm optical depth is less than 1. The chosen range of $T_g$ values results in a 30% uncertainty in $R_{11}$, and smaller uncertainties in $R_{09}$ and $R_{07}$.

The surface emissivities were all chosen to be equal to 1. For the most probable surface materials, the 7-μm emissivity will be near 1 [Martin et al., 1979]. Emissivities in the 9-, 11-, and 20-μm bands will be more dependent upon surface composition and structure, and may be as low as 0.85 for some likely situations [Kiefer et al., 1969], but probably fall between 0.92 and 1 [Christensen, 1982]. The ground temperature affects the radiances in the 7-μm band more than the other bands. The 10% to 15% uncertainty in surface emissivity can contribute as much as a 10% error to the derived extinction coefficient ratios, particularly $R_{11}$. It is more likely that the 9- and 11-μm emissivities, if less than 1, are within a few percent of each other, so that $R_{09}$ will exhibit smaller errors.

Values of $T_a$ (see Figure 3a) are taken to fall between 220 K and 260 K, which covers the range of possible values consistent with radiative-convective models of the Martian atmosphere in the afternoon at midlatitudes [Gierasch and Goody, 1972; Pollack et al., 1979]. These models were confirmed to a limited degree by values of $T_a$ measured at the Viking Lander sites by the Viking meteorology experiment [Hess et al., 1977]. The upper limit to this range is also constrained by limiting the value of $T_1$ to unity. The boundary layer height was varied between 3 and 8 km, and adiabatic and isothermal lapse rates in the boundary layer were chosen as limiting cases. The range of boundary layer height includes the theoretical boundary layer thicknesses in the afternoon [Gierasch and Goody, 1968; Pollack et al., 1979] and covers the dust cloud height estimates made from Viking imaging and infrared data [Peterfreund and Kieffer, 1979]. The atmospheric lapse rate above the boundary layer was determined by fixing the atmospheric temperature at 24 km to be 200 K, as measured by the 15-μm channel of the IRTM instrument in the region of the storm observation [Martin and Kieffer, 1979]. The atmospheric lapse rate above the boundary layer was determined by fixing the atmospheric temperature at 24 km to be 200 K, as measured by the 15-μm channel of the IRTM instrument in the region of the storm observation [Martin and Kieffer, 1979]. An exponential vertical distribution of optical depth, with a scale height of 10 km, was used for all models of the dust storm surroundings, to match the results obtained from Viking imaging experiments [Kahn et al., 1981]. The atmospheric lapse rate above the boundary layer was determined by fixing the atmospheric temperature at 24 km to be 200 K, as measured by the 15-μm channel of the IRTM instrument in the region of the storm observation [Martin and Kieffer, 1979]. An exponential vertical distribution of optical depth, with a scale height of 10 km, was used for all models of the dust storm surroundings, to match the results obtained from Viking imaging experiments [Kahn et al., 1981].
moved from 50 to 64 km, for a model with other inputs set in the middle of the parameter space. Figure 4 presents a series of model results covering the parameter space discussed above. For all cases, mean values of $R_b$ and $\beta$ weighted by the Planck function are used, as obtained for the particles observed by Mariner 9. This assumption is discussed subsequently. The weighted average of property $A$ at temperature $T$ is defined as:

$$A(T) = \frac{\int d(\lambda) \sigma(\lambda,T) f(\lambda) d\lambda}{\int d(\lambda) f(\lambda) d\lambda}$$

(1)

where $\sigma(T)$ is the Planck function at temperature $T$ and $f(\lambda)$ is the relative IRTM response for band $i$ at wavelength $\lambda$, given by Kiefer et al. [1977]. Values of the dust particle parameters calculated in this manner are shown in Table 2. From Figure 4, $T_1$ increases with $R_b$, but the ratios $R_b$ are virtually independent of $T_2$ and $T_3$. $R_b$ depends heavily on the choice of $T_2$, whereas $R_b$ and $R_a$ are much less affected. This confirms that the 7-µm channel samples the ground much more than the other channels, as was suggested by the higher observed brightness temperatures at 7 µm by the IRTM instrument. Table 3 presents the range of values of the extinction coefficient ratios deduced from Figure 4. For comparison, the ratios of extinction coefficients for the particles observed by Mariner 9 [Toon et al., 1977] are also given in Table 3, and are plotted as circles in Figure 4. The derived extinction coefficient ratios are not consistent with those of the Mariner 9 dust. Note that by increasing $R_b$ the ratios $R_a$ are not consistent with those of the values of the extinction coefficient ratios deduced from Figure 4. There is insufficient information to make a direct determination of these values from the Viking data. The particle scattering properties are treated as free parameters in this study, avoiding the questions of particle composition and size distribution. The sensitivity of the conclusions to these choices is tested below. A number of perturbation studies were also performed on the values of $\alpha_0$, $\beta$, and $\gamma$, although there is no formal basis on which to limit the choices. For 10% changes in $\alpha_0$, $\beta$, and $\gamma$, we obtain less than a 3% change in the derived value of $Q_b$, a 10% change in $Q_b$ and $Q_1$, and less than a 3% change in $Q_2$. One alternative to this approach is to vary the particle composition and possibly the size distribution to obtain particle optical parameters which yield agreement between the model and the observed spectra. This approach has been employed by Haur [1979], who concludes that the thermal signature of a dust cloud is reproduced using dust composed of two common terrestrial minerals, with the same size distribution as the Mariner 9 particles. However, this result relies on specific assumptions about the temperature structure, and especially upon the particle size distribution and allowed composition of the particles. (For example, water ice was not included as a possible contributor to the column optical depth.) These assumptions are of necessity arbitrary, since there are only four directly measured quantities in the problem. The introduction of an additional free parameter increases the difficulty in extracting valid information from the model.

The derived ratios of extinction coefficients are model dependent, particularly with regard to the assumed scattering properties of the particles. The most likely values of $R_b$, based on a wide range of possible thermal conditions, are given in Table 3. In summary, for all reasonable thermal profiles, the observed particle properties differ from those of the Mariner 9 dust. Given the arbitrary assumption that the particle scattering properties are similar to the Mariner 9 dust, this approach sets useful limits on the ratios of extinction coefficients from the IRTM data.

5. Viking Lander 1 Sol 0 Models

The best available constraints on Martian atmospheric structure were acquired over the VL1 site at the time of the VL1 entry (sol 0). The vertical temperature profile was measured directly by probes on the entry vehicle [Gillett and Kirk, 1977]. We also have constraints on the vertical distribution of optical depth locally from lander camera experiments [Kahn et al., 1981]. This appears to be an ideal place to test the procedure for retrieving particle extinction properties from IRTM data that was presented in the previous section. With the additional information, the particle properties would seem to be better determined than for the dust storms surroundings case. However, the unknown surface emissivities, ground temperature, and particle scattering properties are still required for the model. The atmospheric vertical optical depth is more than a factor of 2 lower at the VL1 site than in the dust storm surroundings, based on imaging data. The differences between brightness temperature values in the IRTM channels are much smaller at VL1 than in the dust storm surroundings. This is a manifestation of the lower optical depth at the VL1 site, where most of the radiation observed at the IRTM detector comes directly from IRTM.

In Table 4 the IRTM observations are listed which most closely correspond to the VL1 site at sol 0. The landing occurred at 16.13 hours local time. Table 4 shows the orbiter revolution and IRTM sequence identification number, mean viewing angle of the observation, approximate local time of day, and surface resolution of the observations. It also contains the mean and standard deviation of the IRTM 11-µm brightness temperature, and of the IRTM channel differences used in this study. For each IRTM sequence, all the data in a 1° box around the landing site have been accumulated, and the brightness temperature differences were calculated using pairs of measurements made simultaneously on the same telescope spot number. From these data, a representative thermal signature was abstracted (Table 5), to be compared with synthetic spectra. These results are modeled using a temperature profile with $T_m = 241.5$ K, the top of the boundary layer at 7.5 km with temperature 215.95 K, and a 140 K

<table>
<thead>
<tr>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.29</td>
<td>0.403</td>
<td>1.24</td>
</tr>
</tbody>
</table>

Table 3. Particle Extinction Coefficient Ratios
Md is composed of basalt, the average surface emissivity in each of the vertical columns over the properties. These include differences in particle composition and the IRTM properties deduced for the dust storm area (Figure 4). There are many possible reasons for the difference in calculated particle properties. These include differences in particle composition and size distribution, differences in surface emissivity, and the size distribution, differences in surface emissivity, and the possible presence of additional constituents in the vertical column, as for example, water ice haze. The VLI site results are also smaller than in the dust storm area so the ground emissivity for the surface is set at 1. The total atmospheric optical depth was set to 1. The extinction ratios for Mariner 9 observations are given as symbols for optical depth is set at 10⁶. The scale height of the cloud, since there is only one dust layer phenomenon, with a characteristic temperature T₀ and an optical depth at 11 μm T₁₁. Above the cloud, the vertical structure of the atmosphere is assumed to be similar to the surroundings, and unaffected by the presence of the cloud. This structure is illustrated in Figure 2. The temperature measurements, which sample at about 24 km above the surface, are unaltered by the presence of the dust cloud (Pollack et al., 1979). Visible images of local dust clouds also exhibit sharp upper boundaries (Briggs et al., 1979). In addition, observed radiances are heavily weighted to the more optically dense lower atmosphere even in the dust storm surroundings, as discussed in section 4. Since the dust cloud has a larger optical depth than the boundary layer of the dust storm surroundings, the contribution of the atmosphere above the cloud top to the total radiance will not have an important effect upon the results.

6. The Dust Storm Interior

Using the range of particle properties deduced in the dust storm surroundings, we now construct a model of the atmosphere at the dust storm interior. The dust cloud is treated as a boundary layer phenomenon, with a characteristic temperature T₀ and an optical depth at 11 μm T₁₁. Above the cloud, the vertical structure of the atmosphere is assumed to be similar to the surroundings, and unaffected by the presence of the cloud. This structure is illustrated in Figure 2. The temperature measurements, which sample at about 24 km above the surface, are unaltered by the presence of the dust cloud (Pollack et al., 1979). Visible images of local dust clouds also exhibit sharp upper boundaries (Briggs et al., 1979). In addition, observed radiances are heavily weighted to the more optically dense lower atmosphere even in the dust storm surroundings, as discussed in section 4. Since the dust cloud has a larger optical depth than the boundary layer of the dust storm surroundings, the contribution of the atmosphere above the cloud top to the total radiance will not have an important effect upon the results.

The temperature structure for the dust storm interior model contains two parameters, T₀ and T₁₀, that are fit in a least squares sense by the four spectral radiances obtained over the dust storm interior (see Figures 7b). For this procedure to be meaningful, the four atmospheric weighting functions must overlap in the region of the cloud, since there is only one free parameter in the vertical temperature profile above the ground. Knowledge of the vertical distribution of optical depth within the cloud is insufficient to justify a more elaborate temperature structure model. The
dust

Increasing the optical depth in the boundary layer, to simulate a by calculating the overlap integral of the weighting functions.

surroundings,
do not overlap. For 

band

0.20 and 0.25. These three channels sample the atmosphere at nearly the same level. Overlap with the atmospheric contribution to the measurements can be evaluated approximately equal to the fractional overlap of the two channel atmosphere, and approaches 1 if the channel sampling functions between about 0.16 and 0.3. given the models in

cases. In Figure 6, 

cloud may be as

small as 50% for some models. This channel samples closer to the ground than the others, and also receives a larger proportion of net radiation from the ground itself. By calculating one representative atmospheric temperature in the inversion scheme, in addition to the ground temperature, the value of will be aliased if the cloud is not isothermal. However, this effect is small, as is shown in the discussion of perturbation studies below.

Figure 6 illustrates the least squares solution to for the dust storm interior as a function of 11-pm. The particle properties obtained for the dust storm surroundings, as given in Table 3, have been used. The vertical dust distribution is discontinuous, with the cloud top set at 4 km, and the scale height of dust set to 10 km in both the cloud and the atmosphere above. The 11-pm optical depth of the atmosphere above the cloud is fixed at 0.4, which is equal to the 11-pm atmospheric optical depth above 4 km in the middle of the parameter space for the dust storm surrounding models. The atmospheric temperature at the cloud top is taken as 220 K, and the atmospheric lapse rate is determined so that at 24 km the temperature is 200 K, to agree with the IRTM measurements at 15 pm (Martin et al., 1979), as was done in the cloud-free area. Convergence is achieved to better than 1 K per channel, which is comparable to the experimental accuracy of the IRTM, for 11-pm less than 3.

Solutions are less accurate for large cloud optical depth, and for 11-pm = 5 the least squares solution for 11-pm deviates by 5 K from the observed value.

In Figure 6, below the dust cloud is less than 280 K for all choices of 11-pm. For the dust storm surroundings model, the absolute lower bound on 11-pm came to 286 K. The value of 11-pm is less certain, and depends strongly on the value of . In particular, 11-pm varies most rapidly when . This is explained as follows: from Figure 4, 11-pm is most sensitive to the value of and in Table 3, the ratio of extinction coefficients for the 7- and 11-pm bands is about one third. So when 11-pm is

Figure 5: Viking Lander 1 site models. Ratios of extinction coefficients and the 11-pm optical depth are plotted as functions of ground temperature for a model in which the atmospheric weighting functions must also be insensitive to the rest of the atmospheric structure, and must sample the ground relative to the cloud to varying degrees. The last two properties are strongly suggested by the results shown in Figure 4, and can be checked post hoc by the convergence and stability of the least squares procedure. From Figure 4, for the cloud is the most sensitive to the value of whereas of all the channels, the response of is greatest to changes in . The relative independence of the atmospheric contribution to the measurements can be evaluated by calculating the overlap integral of the weighting functions, which is a measure of the coincidence of the regions sampled by any pair of channels:

\[
W_{ij} = \left| 1 - \int_0^\infty K_i K_j dt \int_0^\infty K_i^2 dt \right| \quad (2)
\]

where is the derivative of the transmission function for spectral band i (e.g., Rogers, 1976). This quantity is equal to zero if the two channels sample exactly the same vertical region of the atmosphere, and approaches 1 if the channel sampling functions do not overlap. For between about 0.0 and 0.3, is approximately equal to the fractional overlap of the two channel sampling functions in these models. In the dust storm surroundings, the distribution of optical depth leads to values of which range between 0.16 and 0.3, given the models in section 4. Thus the overlap of sampling functions is very high. Increasing the optical depth in the boundary layer, to simulate a dust cloud, produces values of 11-pm, 22-pm, and 28-pm between 0.30 and 0.25. These three channels sample the atmosphere at nearly the same level. Overlap with the 7-pm channel may be as

Figure 6: Dust storm interior model. Ground temperature (upper curve) and cloud temperature (lower curve) are plotted as functions of the 11-pm dust cloud optical depth. The dust storm top is fixed at 4 km, and the 11-pm optical depth above 4 km is set to 0.4. is set to 220 K and .
increased to 3, the 7-μm optical depth reaches unity, and the ratio of ground to atmospheric radiation becomes less dependent upon $\tau_{7\mu m}$. No specific conclusions can be drawn about the effective temperature of the dust cloud as related to the boundary layer temperature in the surroundings because of the large uncertainty in both these quantities. Theoretical modeling of the boundary layer temperature in the midlatitude at a similar latitude in the northern hemisphere, for early summer, yields temperatures near 230 K (Pollack et al., 1979). The increased visual optical depth (which is still less than 1) and the greater solar input in the dust storm surroundings over those of the VLI site models can raise this temperature of the order of 10 K (Figures 7 and 8 in Pollack et al. 1979). This places the current estimate of the near-ground temperature at the upper end of the range of possible values of $T_{g0}$. From these results, note that $T_{g0} - T_{g}$ is less than 40 K in the dust storm interior. These observations are discussed in the next section.

A number of perturbation studies have been performed on the results in Figure 6. In general, the least squares character of the method makes the solution less sensitive to changes in a single input quantity than the single-point inversions of the previous sections. Varying $\tau_{10}$ and 3 by 10% causes $T_{g0}$ to change by less than 1 K, whereas $T_{g0}$ will change less than 6 K. Altering the model brightness temperatures $T_{b}, T_{s1},$ and $T_{s2}$ independently by up to 3 K, effect $T_{g0}$ less than 1 K. and $T_{g}$ less than 2 K. If $T_{s}$ is perturbed while $\tau_{10}$ is less than 3, the change in $T_{g0}$ will be comparable to the size of the perturbation. For $\tau_{10} > 3$, altering $T_{s}$ or the 7-μm extinction coefficient will have an increasingly large effect on $T_{g}$. For $\tau_{10} > 5$, the input parameters do not produce a satisfactory fit to the observed radiance.

In summary, an attempt has been made to set meaningful limits on $T_{g}$ and $T_{s}$ based on the assumption that particle properties do not vary between the dust storm surrounding and the interior. $T_{g}$ is not constrained by the available data, but it is probably in the lower range of the likely ground temperature in the surroundings. Below the storm, $T_{s}$ must be lower than the ground temperature in the surroundings.

7. Discussion

One of the key questions that may be addressed by studying the thermal structure of a Martian dust storm is whether local thermodynamic effects play an important role in generating the surface heat stress necessary to lift dust. Small, intense convection cells occur in the form of dust devils on Earth, in places where the planetary boundary layer is stably unstable over regions tens to hundreds of meters wide (Ryan and Carroll, 1970). Similar features were found on Mars in the Viking orbiter camera images (Thomas and Gierasch, 1985). A surface resolution about 3 orders of magnitude greater than that available in the data of this study would be required to observe these features with infrared imaging. However, the results of the present work suggest that at least over surfaces on the scale of 200 km, vigorous convection at ground level is unlikely in the interior of the dust storm. For the storm studied here, the ground temperature in the storm interior is at least 6 K cooler than the ground temperature of the surroundings. Also, thermal contrast between the ground and the atmospheric boundary layer ($T_{g} - T_{s}$) in the storm surroundings at midlatitude is 30 to 60 K over most of the parameter space. From Figure 4, the dust storm interior shows this difference to be less than 40 K for $\tau_{10}$ between 1 and 3, suggesting that the boundary layer in the cloud region may be more stable against convection than in the surroundings. Both these observations are contrary to expectation for a situation in which direct heating at ground level predominates in the storm interior.

The evolution of the storm as shown in Figure 2 makes the possibility that the convective process operates even at the storm edges appear unlikely. The thermal structure of the storm over regions 200 km in size does not change measurably between 11.5 and 16.6 hours local time. If the storm were generated by direct thermal forcing at the ground during part of this period, the static stability would be lowest at the ground, and vigorous convection would transport heat and carry dust from the ground (Gierasch and Goody, 1973). In this scenario, the static stability would first decrease in the region of cloud formation as the atmosphere was heated by direct absorption of solar radiation by the dust near the ground. Later, as the dust filled the boundary layer, increased atmospheric heating would cause the static stability to increase. The calculations of this study do not rule out the presence of a region of vigorous convection at the cloud tops, nor the possibility of dust plumes or dust devils on a smaller spatial scale. The observations presented here are more favorable to arguments that the storm is maintained either by the flow of cold air in the form of density currents, or the combined effects of winds forced on scales larger than the size of the storm (Leovy et al., 1973; Pollack et al., 1979). These two hypotheses may be tested when the direction of storm motion can be compared with regional slopes and prevailing winds. Both mechanisms would allow the ground temperatures to be relatively cool in the region where dust is being lifted, and neither mechanism is excluded by the apparently unchanging vertical thermal gradient in the boundary layer within the dust storm during the late morning and afternoon.

This study of the Viking IRMT data in the area of one local dust storm led to the conclusion that the dust in the cloud region differs from that observed during the 1971-1972 global dust storm observed by Mariner 9. This may be due to differences in the type of dust distributed in regions of the planet, or to selectivity in the mechanisms which lift dust in the two instances. Given the assumptions that the particle scattering (but not extinction) properties are similar to those observed by Mariner 9, and dust dust properties do not vary horizontally in the storm region, it is possible to obtain an estimate of the ground temperature and to achieve a crude determination of the atmospheric temperature near the ground within a local dust storm. This technique may be useful in further exploring the thermodynamics of Martian dust storms when more extensive data set is available.

Appendix: Two-Stream Radiative Transfer Scheme

Following Toon et al. (1977), the equation of radiative transfer in a plane parallel atmosphere is written for monochromatic radiation, given the positive direction as upward and optical depth $\tau$ as zero at the top:

$$\frac{d\mu}{d\phi} = -J(\tau) - J(\tau)$$  \hspace{1cm} (A1)

where $\mu$ is the cosine of the emission angle, $\phi$ is the scattering angle, and $J$ is the source function. In the two-stream approximation,

$$\frac{d\mu}{d\phi} = f - J^+$$  \hspace{1cm} (A2)

$$\frac{d\mu}{d\phi} = g - J^-$$  \hspace{1cm} (A3)
Here \( f \) and \( g \) are the upward and downward two-stream intensities, respectively.

\[
J^o = \alpha_0^2 \left( 1+h \right) f + \alpha_0^2 \left( 1-h \right) g + \left[ 1 - \alpha_0^2 \right] \beta^o(T)
\]  
(A4)

With \( B_v(T) \) as the Planck function at frequency \( v \) for the temperature at atmospheric level \( T \), and

\[
J^o = \alpha_0^2 \left( 1+h \right) f + \alpha_0^2 \left( 1-h \right) g + \left[ 1 - \alpha_0^2 \right] \beta^o(T)
\]  
(A5)

Rewriting equations (A2) and (A3) using (A4) and (A5),

\[
f = \varphi' \beta^o - \nu c
\]  
(A6)

\[
g' = \varphi q + \nu^2 c + \nu c
\]  
(A7)

where

\[
a = \sqrt{1 - \alpha_0^2 \left( 1+h \right)} \]

\[
b = \sqrt{1 - \alpha_0^2 \left( 1-h \right)} \]

\[
c = \sqrt{1 - \alpha_0^2 \beta^o(T)}
\]  
(A8)

\( \beta_0 \) and \( \beta \) are the single-scattering properties discussed in the text.

By cross-differentiating equations (A6) and (A7), a second-order differential equation for \( f \) is obtained:

\[
f'' + \left( \nu^2 - \nu' \right) f' + \left( \nu^2 - \nu' \right) f = \left( \nu^2 \right) \beta^o - \alpha_0^2 \beta^o - \alpha_0^2 \beta^o - \nu^2 c - \nu c
\]  
(A9)

The lower boundary condition is

\[
f(\tau^*) = R_{B_0} \left( T_B \right)
\]  
(A10)

where \( \tau^* \) is the total atmospheric optical depth and \( \tau \) is the ground emissivity. The boundary condition at the top of the atmosphere is chosen as

\[
g(0) = 0
\]  
(A11)

Using (A2), the upper boundary condition becomes

\[
f(0) = \varphi' \beta^o c + \nu c = 0
\]  
(A12)

The solution to (A9), with (A10) and (A12), is obtained using the method of Lindzen and Kao [1969]. Then \( g \) is obtained as an initial value problem using a Runge-Kutta method with (A11) as the initial value. The solution to the equation of radiative transfer for the upward intensity is

\[
f(0) = R_{B_0} \left( T_B \right) - \frac{\nu}{\nu' \beta^o} \int_0^{\tau^*} \frac{\varphi' \beta^o \delta(\tau)}{\rho(\tau) \rho(T)} d\tau
\]  
(A13)

where \( f(0) \) is the source function. \( f(0) \) is replaced with the source function in the upward direction \( J^o \) from equation (A4), and the integration is performed numerically to obtain the emergent intensity. Contributions from atmospheric gases are neglected.

The CO2 opacity averaged over the 7-9, 11, or 20- \( \mu \)m IRM passbands is at least 3 orders of magnitude less than the observed optical depth in these bands for a 7-mbar atmosphere.

Acknowledgments. I thank R. Goody and R. Zurek for their encouragement of this work, F. Paluch for providing initial access to the IRIS data used in this study, and T.Z. Martin and F. Paluch for careful reviews of the manuscript. The early part of this work is supported in part by NASA grant NGR 73599 to Harvard University. Later portions of this project were performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration, through the NASA Planetary Atmospheres Program.

References


KAHN: TEMPERATURES OF MARTIAN LOCAL DUST STORM


(Received March 31, 1994; revised October 23, 1994, accepted October 25, 1994.)

R. Kahn, Jet Propulsion Laboratory, MS 169-237, 4800 Oak Grove Drive, Pasadena, CA 91109.
Page intentionally left blank
Atmospheric effects on the mapping of Martian thermal inertia and thermally derived albedo

Joan N. Hayashi
Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder

Bruce M. Jakosky
Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder

Robert M. Haberle
Space Sciences Division, NASA Ames Research Center, Moffett Field, California

Abstract. We examine the effects of a dusty CO₂ atmosphere on the thermal inertia and thermally derived albedo of Mars and present a new map of thermal inertias. This new map was produced using a coupled surface atmosphere (CSA) model, dust opacities from Viking infrared thermal mapper (IRTM) data, and CO₂ columns based on topography. The CSA model thermal inertias are smaller than the 2% model thermal inertias, with the difference largest at large thermal inertias. Although the difference between the thermal inertias obtained with the two models is moderate for much of the region studied, it is largest in regions of either high dust opacity or of topographic lows, including the Viking Lander 1 site and some geologically interesting regions. The CSA model thermally derived albedos do not accurately predict the IRTM measured albedos and are very similar to the thermally derived albedos obtained with models making the 2% assumption.

Introduction

The thermal inertia of a material is a measure of its temperature responses to energy input or loss; the temperature of a low thermal inertia material responds quickly, while the temperature of a high thermal inertia material responds slowly. Understanding thermal inertia has implications for understanding surface properties and surface temperatures. For silicate materials, thermal inertia is principally determined by the thermal conductivity and can be used to infer particle sizes at the surface. Also, a knowledge of surface temperatures is necessary to understand the formation of water and carbon dioxide frost and ice which, in turn, influences both the geology and the climate.

The most widely used thermal inertia data for Mars assumes an atmosphere whose only effect is a contribution to the downwelling radiation which is constant and equal to 2% of the maximum solar insolation (the "2% assumption") [Kieffer et al., 1977; Palluconi and Kieffer, 1981]. Haberle and Jakosky [1991] investigated the differences in the thermal inertias derived by making the 2% assumption versus that derived by including the effects of a dusty CO₂ atmosphere and sensible heat exchange with the surface. Specifically, they explored this effect for the location of the Viking Lander 1 (VL1). Bridges [1994] did first-order corrections for the effects of elevation on determining the thermal inertia and the effect of variations in pressure on translating thermal inertias into particle sizes. Bridges used both the 2% model and elements of Haberle and Jakosky's [1991] analysis to obtain two particle size maps. To correct the 2% model thermal inertias, Bridges assumed that at 0 km the back radiation was 2% of the maximum solar insolation, but that this decayed with elevation with a scale height of 10 km. He then assumed that both this 4% back radiation and the dust opacity (fixed at 0.4 at -2 km) decayed with elevation with a scale height of 10 km. He then incorporated the resulting thermal inertias in terms of particle sizes, accounting for the variation of conductivity, and thus thermal inertia, with pressure.

In this paper we utilize some aspects of both the Haberle and Jakosky [1991] and the Bridges [1994] analyses. We use the coupled surface-atmosphere (CSA) model presented by Haberle and Jakosky [1991] to investigate the effects of a dusty CO₂ atmosphere on the thermally derived albedo. The thermally derived albedo is the surface albedo which, together with the thermal inertia, provides model surface temperatures which best match the observed temperatures; it may differ from a measured albedo. We also present a new map of thermal inertia obtained utilizing the CSA model. This map differs from Bridges' [1994] thermal inertias, as
the effect of the atmosphere is calculated with a one-dimensional radiative-convective model rather than scaled from the maximum solar insolation. However, in interpreting the thermal inertias for certain regions of Mars in terms of particle sizes, we utilize results from Bridges [1994] to account for elevation dependent pressure variations.

Method

The CSA model is a one-dimensional diurnal model; it is described in detail by Haberle and Jakosky [1991]. The atmospheric portion of the model is a radiative-convective model which allows for absorption and emittance of radiation by CO₂ and dust, convective adjustment of the lower atmosphere, and sensible heat exchange between the surface and atmosphere. The surface and subsurface portion of the model allows for thermal diffusion and exchange of energy with the atmosphere. The CSA model parameters which were varied in this study were thermal inertia, surface albedo, latitude, atmospheric pressure at the surface, and atmospheric dust opacity. In contrast, the parameters which were varied in a diurnal model making the 2% assumption (the 2% model) were thermal inertia, surface albedo, and latitude; the 2% model temperatures do not depend on atmospheric surface pressure or dust opacity.

Two avenues of investigation were used to examine the differences between the 2% model and the CSA model. First, the differences in both the thermal inertias and thermally derived albedos due to varying dust opacities were explored at a single latitude and surface pressure. Second, the differences in the models’ mapped thermal inertias and thermally derived albedos due to spatial variations in latitude, dust opacities, and surface pressure were explored.

First, the differences between the CSA model and the 2% model at a single latitude and surface pressure were investigated for the latitude and pressure of the VLI site. Haberle and Jakosky [1991] examined the difference in thermal inertia between the two models; we explored the difference in thermally derived albedo and also included the differences in thermal inertia for completeness. The 2% model was run at a series of albedos and thermal inertias to obtain minimum and maximum surface temperatures, which were then fit using the CSA model to derive inertia and albedo. To examine the effects on thermally derived albedo, the CSA model was run at a given thermal inertia and various albedos and dust opacities. Minimum and maximum surface temperatures were obtained for each CSA model combination of albedo and dust opacity. The 2% model minimum and maximum temperatures were then interpolated to fit these CSA model temperatures and determine the 2% model albedo and dust opacity. To examine the effects on thermal inertia a similar procedure was used, except the CSA model temperatures were obtained for a given albedo and various thermal inertias and dust opacities.

Second, to examine the spatial differences between the 2% model and the CSA model thermal inertias and thermally derived albedos, it was necessary to first produce maps of the CSA model thermal inertias and thermal inertias. As the Palluconi and Kieffer [1981] 2%
Table 1. Thermal Inertias and Thermally Derived Albedos Calculated Using the Temperature at the Surface ($I_0$, $A_0$) and at the Top of the Atmosphere ($I_0^*$, $A_0^*$) for Two Different Pairs of $f$ and $A$

<table>
<thead>
<tr>
<th>$f$</th>
<th>$I_0$</th>
<th>$A_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0$</td>
<td>0.10</td>
<td>0.18</td>
</tr>
<tr>
<td>$A_0^*$</td>
<td>0.18</td>
<td>0.18</td>
</tr>
</tbody>
</table>

From Paige et al. (1994).

The pressure for each bin was determined from the digital topographic maps, assuming 6.1-mbar pressure at 0 km elevation and a constant scale height of 10 km. The pressure of 6.1 mbar is the nominal average pressure at 0 km, and is the constant pressure surface to which the elevations are referenced in the topographic map (U.S. Geological Survey [USGS], 1989). Seasonal pressure variations are ignored, as they are of the same order as the pressure variations introduced by the uncertainties in the topography (61.0 to 61.5 km, depending on the latitude [USGS, 1989]).

The dust opacity was approximated by averaging Martin and Richardson's (1993) 9-µm dust opacities for the period between $L_\text{p}=345\text{°}-125\text{°}$ (Figure 1). The $L_\text{p}=345\text{°}-125\text{°}$ was the same period which Pallucconi and Kieffer (1981) thermal inertias and thermally derived albedos used to create their map of the thermally derived albedo; we chose to map the CSA model thermal inertias and thermally derived albedos at the same resolution.

The latitude, surface pressure, dust opacity, and minimum and maximum temperatures were needed for each map point to determine the CSA model thermal inertia and thermally derived albedo. Surface temperatures were obtained by running the 2% model for each map point using the Pallucconi and Kieffer (1981) thermal inertias and thermally derived albedos at $L_\text{p}=0\text{°}$. As the Pallucconi and Kieffer did not publish their map of the thermally derived albedo, we obtained this map directly from them.

Errors arising from uncertainties in the surface pressure due to seasonal pressure variations and uncertainties in the topography are small in both the thermally derived albedo and the thermal inertia. Model comparisons using different pressures suggest the errors are less than approximately 0.01 in albedo and less than approximately 10 W m$^{-2}$ K$^{-1}$ in thermal inertia. We use SI units for thermal inertia; however, for brevity, we hereafter omit the units. To convert to units of cal cm$^{-2}$ s$^{-1/2}$ K$^{-1/2}$, divide by 4186. Allowing the inferred to visible dust opacity ratio to vary from 0.4 to 0.55 gives rise to small uncertainties in albedo (< 0.03) and thermal inertia (< 5). Errors are also introduced into the CSA thermal inertia and thermally derived albedo due to fitting temperatures as would be observed at the top of the atmosphere (20-µm infrared thermal mapper (IRM) brightness temperatures) with CSA model surface temperatures. The CSA model included the absorption and emission of radiation by dust and CO$_2$ in the atmosphere and their effects on the surface temperature; however, the CSA model did not calculate the effects of this absorption and emission on the observed brightness temperature. Paige et al. (1994) examined this problem. At latitudes of 60° - 85° and $L_\text{p} = 110°$, they calculated both surface temperatures and temperatures at the top of the atmosphere for several different combinations of atmosphere and thermally derived albedos was chosen to best reproduce the observed 20-µm brightness temperatures, using them in the interpolation region. The number of bins which are thus excluded is less than 2% of the total bins. As such a small percentage of bins is excluded, we consider that this has a negligible contribution to the error.

Errors arising from uncertainties in the surface pressure due to seasonal pressure variations and uncertainties in the topography are small in both the thermally derived albedo and the thermal inertia. Model comparisons using different pressures suggest the errors are less than approximately 0.01 in albedo and less than approximately 10 W m$^{-2}$ K$^{-1}$ in thermal inertia. We use SI units for thermal inertia; however, for brevity, we hereafter omit the units. To convert to units of cal cm$^{-2}$ s$^{-1/2}$ K$^{-1/2}$, divide by 4186. Allowing the inferred to visible dust opacity ratio to vary from 0.4 to 0.55 gives rise to small uncertainties in albedo (< 0.03) and thermal inertia (< 5). Errors are also introduced into the CSA thermal inertia and thermally derived albedo due to fitting temperatures as would be observed at the top of the atmosphere (20-µm infrared thermal mapper (IRM) brightness temperatures) with CSA model surface temperatures. The CSA model included the absorption and emission of radiation by dust and CO$_2$ in the atmosphere and their effects on the surface temperature; however, the CSA model did not calculate the effects of this absorption and emission on the observed brightness temperature. Paige et al. (1994) examined this problem. At latitudes of 60° - 85° and $L_\text{p} = 110°$, they calculated both surface temperatures and temperatures at the top of the atmosphere for several different combinations of atmosphere and thermally derived albedos was chosen to best reproduce the observed 20-µm brightness temperatures, using them in the interpolation region. The number of bins which are thus excluded is less than 2% of the total bins. As such a small percentage of bins is excluded, we consider that this has a negligible contribution to the error.

Errors arising from uncertainties in the surface pressure due to seasonal pressure variations and uncertainties in the topography are small in both the thermally derived albedo and the thermal inertia. Model comparisons using different pressures suggest the errors are less than approximately 0.01 in albedo and less than approximately 10 W m$^{-2}$ K$^{-1}$ in thermal inertia. We use SI units for thermal inertia; however, for brevity, we hereafter omit the units. To convert to units of cal cm$^{-2}$ s$^{-1/2}$ K$^{-1/2}$, divide by 4186. Allowing the inferred to visible dust opacity ratio to vary from 0.4 to 0.55 gives rise to small uncertainties in albedo (< 0.03) and thermal inertia (< 5). Errors are also introduced into the CSA thermal inertia and thermally derived albedo due to fitting temperatures as would be observed at the top of the atmosphere (20-µm infrared thermal mapper (IRM) brightness temperatures) with CSA model surface temperatures. The CSA model included the absorption and emission of radiation by dust and CO$_2$ in the atmosphere and their effects on the surface temperature; however, the CSA model did not calculate the effects of this absorption and emission on the observed brightness temperature. Paige et al. (1994) examined this problem. At latitudes of 60° - 85° and $L_\text{p} = 110°$, they calculated both surface temperatures and temperatures at the top of the atmosphere for several different combinations of atmosphere and thermally derived albedos was chosen to best reproduce the observed 20-µm brightness temperatures, using them in the interpolation region. The number of bins which are thus excluded is less than 2% of the total bins. As such a small percentage of bins is excluded, we consider that this has a negligible contribution to the error.

Errors arising from uncertainties in the surface pressure due to seasonal pressure variations and uncertainties in the topography are small in both the thermally derived albedo and the thermal inertia. Model comparisons using different pressures suggest the errors are less than approximately 0.01 in albedo and less than approximately 10 W m$^{-2}$ K$^{-1}$ in thermal inertia. We use SI units for thermal inertia; however, for brevity, we hereafter omit the units. To convert to units of cal cm$^{-2}$ s$^{-1/2}$ K$^{-1/2}$, divide by 4186. Allowing the inferred to visible dust opacity ratio to vary from 0.4 to 0.55 gives rise to small uncertainties in albedo (< 0.03) and thermal inertia (< 5). Errors are also introduced into the CSA thermal inertia and thermally derived albedo due to fitting temperatures as would be observed at the top of the atmosphere (20-µm infrared thermal mapper (IRM) brightness temperatures) with CSA model surface temperatures. The CSA model included the absorption and emission of radiation by dust and CO$_2$ in the atmosphere and their effects on the surface temperature; however, the CSA model did not calculate the effects of this absorption and emission on the observed brightness temperature. Paige et al. (1994) examined this problem. At latitudes of 60° - 85° and $L_\text{p} = 110°$, they calculated both surface temperatures and temperatures at the top of the atmosphere for several different combinations of atmosphere and thermally derived albedos was chosen to best reproduce the observed 20-µm brightness temperatures, using them in the interpolation region. The number of bins which are thus excluded is less than 2% of the total bins. As such a small percentage of bins is excluded, we consider that this has a negligible contribution to the error.

Table 2. Relevant Portion of the Wentworth Grain Size Classification

<table>
<thead>
<tr>
<th>Classification</th>
<th>Grain Size, µm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse silt</td>
<td>31.25-62.5</td>
</tr>
<tr>
<td>Very fine sand</td>
<td>62.5-125</td>
</tr>
<tr>
<td>Fine sand</td>
<td>125-250</td>
</tr>
<tr>
<td>Medium sand</td>
<td>250-500</td>
</tr>
<tr>
<td>Coarse sand</td>
<td>500-1000</td>
</tr>
<tr>
<td>Very coarse sand</td>
<td>1000-2000</td>
</tr>
<tr>
<td>Granules</td>
<td>2000-4000</td>
</tr>
</tbody>
</table>

After Wentworth (1922).
spheric conditions, surface thermal inertia (I), and surface albedo (As). They then fit the temperatures at the top of the atmosphere (20-μm brightness temperatures) with their basic thermal model, which did not include any atmospheric contribution, to obtain \( I' \) and \( As' \). They also fit the surface temperatures with their basic thermal model to obtain \( I'' \) and \( As'' \). Thus the difference between \( I \) and \( As \) and \( I'' \) and \( As'' \) is the difference between applying both included atmospheric effects and calculating the brightness temperature correctly or having ignored atmospheric effects. The difference between \( I \) and \( As \) and \( I'' \) and \( As'' \) is the difference between having included atmospheric effects to calculate the surface temperature, but not calculating the brightness temperature correctly, or having ignored atmospheric effects. To estimate the error introduced between including atmospheric effects and calculating the brightness temperature correctly, or having ignored atmospheric effects, we compare the differences between Paige et al.'s \( I' \) and \( As' \) and \( I'' \) and \( As'' \). Table 1 summarizes their results for a latitude of 60° and their atmosphere 5. Atmosphere 5 was chosen, as it most closely resembled the CSA model atmosphere; atmosphere 5 had a visible dust opacity of 0.2 as well as an infrared to visible dust opacity ratio of 0.5. The difference between \( I' \) and \( I'' \) is 10 to 20 and the difference between \( As' \) and \( As'' \) is 0.0 to 0.01. Therefore, the error introduced by utilizing surface temperature rather than the brightness temperature at the top of the atmosphere to derive thermal inertia and albedo is small when the visible dust opacity is 0.2 or less. Based on the distribution of dust opacities shown in Figure 1b, we conclude that the error in the thermal inertia introduced by ignoring the effects of the dust on the upgoing radiation is usually less than 20 and is almost always less than 40.

Another source of error is that Palluconi and Kieffer [1981] fit an annual thermal model to the IRITM measurements from \( L_p=344°-125° \), while we used their thermal inertia and thermally derived albedo in a diurnal 2% model to calculate temperatures which were matched by the diurnal CSA model temperatures. To estimate the errors due to running the models at only one \( L_p \) (\( L_p=0° \)), we also ran the 2% and CSA models at another \( L_p \) (\( L_p=55° \)). At southern high latitudes, \( CO_2 \) frost remained on the surface, and we therefore ignored the bias between 39° and 39° latitude (approximately 20% of our total bins) in this analysis. The difference between the CSA model thermally derived albedo at \( L_p=0° \) and \( L_p=55° \) had a mean of 0.01 and a standard deviation of 0.03. The difference between the CSA model thermal inertia at \( L_p=0° \) and \( L_p=55° \) had a mean of 0.7 and a standard deviation of 8.7. The error introduced by using a single \( L_p \) is small for the following reason. Palluconi and Kieffer fit an annual model (which included seasonal effects) to real temperatures (which obviously include seasonal effects); however, we then compared results from two diurnal models, and thus the lack of seasonal effects very nearly canceled out.

Finally, an error is introduced by assuming the surface emissivity is always unity. As discussed by Christensen [1982], the thermal emissivity is correlated with albedo. At albedos greater than 0.28 the emissivity is unity, but at lower albedos the emissivities can approach 0.9. In the case where the actual emissivity is 0.9, assuming a model emissivity of 1.0 results in albedos being underestimated by about 0.1 and thermal inertias being overestimated by up to 40.

Based on the above considerations, we estimate the errors in the CSA model thermal inertia are of the order of 10-20% for regions of high albedo (≥ 0.28) and may be as much as 50-60% for regions of low albedo (≤ 0.28).

There is also an additional uncertainty which is introduced in translating thermal inertias into particle
size. For geologic materials, the thermal inertia is principally determined by the material's conductivity; the conductivity is in turn a function of particle size and pressure. Obviously, a single particle size is not an actual representation of the surface. The particle size inferred from the thermal inertia is an effective particle size, which assumes that the surface is composed of uniform particles. Also, although some data exist on the variation of conductivity with pressure [Masamune and Smith, 1963; Wechsler and Glaser, 1965], the scatter within a particular data set as well as scatter between data sets adds additional uncertainties in converting thermal inertias into particle sizes. Due to these uncertainties, we have chosen to interpret particle sizes corresponding to the thermal inertias in terms of Wentworth's [1922] grain size classification (see Table 2) rather than stating specific particle sizes.

Results

In this section we first present the effects of varying dust opacities on the thermal inertia and albedo at a single latitude and surface pressure. Next, we explore the differences in the CSA and 2% model thermal inertias and thermally derived albedos due to spatial variations in latitude, dust opacities, and surface pressure. The effect of varying dust opacity on the CSA model thermal inertia versus the 2% model thermal inertia is shown in Figure 2, which is similar to Haberle and Jakosky's [1991] Figure 13. This figure shows the difference between the 2% model-derived thermal inertias and the CSA model-derived thermal inertias for a variety of dust opacities and a CSA model albedo of 0.30. The 2% model thermal inertia is always larger than the CSA model thermal inertia, with the difference being greatest at largest dust opacities. Therefore, particle sizes inferred from the 2% model thermal inertias are always too large.
Figure 3 is a similar figure for the thermally derived albedo. This figure shows the difference between the 2% model thermally derived albedos and the CSA model thermally derived albedos for a variety of dust opacities and a CSA model thermal inertia of 200. At low dust opacities the 2% model thermally derived albedos are always smaller than the CSA model thermally derived albedos. As there is little dust present at such low opacities, this is attributed solely to the greenhouse effect of the CO₂ in the atmosphere. At high dust opacities, the behavior is different at high and low albedos. At low albedo and increasing dust opacity, a constant CSA thermally derived albedo is matched by a larger 2% model thermally derived albedo. This is attributed to scattering by dust in the atmosphere. However, at high albedo and increasing dust opacity, a constant CSA thermally derived albedo is matched by a smaller 2% model thermally derived albedo. This is attributed to warming of the atmosphere due to the presence of dust.

Next, we look at the spatial differences between the CSA and the 2% model thermal inertias and the significance of the CSA model-derived thermal inertias. Figure 4 is a map of the CSA model-derived thermal inertias. Figure 5 shows the relationship between the CSA model and the 2% model thermal inertias as a two-dimensional histogram. The 2% model thermal inertias are larger, and the difference between the two is greatest at large thermal inertias. Figure 6 is a histogram of the difference between the two model thermal inertias. The difference has a mean of -46 and a standard deviation of 23.

A map of the differences between the CSA model and the 2% model thermal inertias is presented in Figure 7.
Areas with a high dust opacity (e.g., Acidalia, Chryse, and Hellas Planitiae) have some of the largest differences; the CSA model thermal inertias are about 100 smaller than the 2% model thermal inertias for these areas. Some regions which are topographic lows (e.g., Valles Marineris, Vastitas Borealis, and Idris Planitia) also have large differences (~40 to ~100) between the two model thermal inertias. Therefore, particle sizes for these regions based on the 2% model thermal inertias would be too large. For example, the 2% model thermal inertia in Hellas is about 340, while the CSA model thermal inertia is about 230. As Hellas has a high albedo (~0.33), the emissivity should be unity and the uncertainties are only about 20. From Bridges' [1994] Figure 2, which graphically shows the relationship between pressure, thermal inertia, and particle size, and using Wentworth's [1922] grain size classification (Table 2), the 2% model thermal inertia would correspond to medium sand, while the CSA model thermal inertia would correspond to fine sand; the CSA grain sizes are about 40% smaller than the 2% grain sizes. However, Acidalia Planitia, which has a 2% model thermal inertia of about 440 and a CSA model thermal inertia of about 350, has an albedo of about 0.16. Due to this low albedo and presumably low emissivity, the thermal inertias may be underestimated by 60. Thermal inertias of 440 and 350 would correspond to coarse sand and medium sand, respectively, but if the thermal inertias were less by 60, the 2% thermal inertia would correspond to grain sizes near the boundary between coarse sand and medium sand, while the CSA thermal inertia would correspond to medium sand. In both cases the CSA grain sizes are about 25% smaller than the 2% grain sizes. Some areas of topographic highs (e.g., Olympus Mons, the Tharsis Volcanoes, Alba Patera, and Elysium Mons) have CSA model thermal inertias which are very similar to or larger than the 2% model thermal inertias. These regions have differences of <5 to >60 and appear as bright regions on Figure 7. For example, the 2% model thermal inertias at Ascraeus Mons are about 95, while the CSA model thermal inertias are about 110. The albedo of Ascraeus Mons, so the uncertainties in thermal inertia should be less than 20. These thermal inertias correspond to fine sand; the CSA grain sizes are 25% smaller than the 2% grain sizes. The Viking Lander sites are perhaps the most studied sites on Mars due to additional data provided by the landers which is not available for other regions. Both lander sites have high albedos, presumably near-unit emissivity, and uncertainties in thermal inertia of about 20. The VL1 site has a 2% model thermal inertia of 347 and a CSA model thermal inertia of 266, with a difference of ~81. Both these thermal inertias correspond to medium sand, with the CSA grain sizes again about 25% smaller. The Viking Lander 2 (VL2) site has a difference between the model thermal inertias of ~57, which is close to the mean difference. The 2% model thermal inertia is 239, and the CSA model thermal inertia for the site is 256. The 2% thermal inertia corresponds to grain sizes close to the boundary between medium and coarse sand, while the CSA thermal inertia corresponds to medium sand; the CSA grain sizes are about 25% smaller. Although the model thermal inertias for VL2 have a difference close to the mean, it lies within a few degrees of some of the largest differences between the model thermal inertias.

Thus, the importance of utilizing a CSA model thermal inertia is greater than indicated by the mean and standard deviation of the difference between the two models, as some of the most geologically interesting regions as well as the two lander sites are located within or near the regions which show the largest difference between the two models.

Finally, the CSA model thermally derived albedos are similar to the 2% model thermally derived albedos; neither of the thermally derived albedos accurately model the measured albedos. As seen in Figures 2 and 8, two-dimensional histograms produced from maps of the 2% model thermally derived albedo and the IRTM measured albedos, the 2% model overestimates the albedo at low albedo and underestimates the albedo at high albedos. Figure 9 is a two-dimensional histogram of the CSA model thermally derived albedo and the measured albedos and shows a pattern similar to Figure 8. The difference between the 2% model and CSA model thermally derived albedo has a mean of 0 and a standard deviation of 0.02. Therefore, including the effects of a dusty CO₂ atmosphere does not account for the difference between the thermally derived albedo and the measured albedos. The causes of this difference is not currently understood.

Conclusions

Including a dusty CO₂ atmosphere when modeling the thermal inertia of Mars is an improvement over making the 2% assumption. The mean difference between the CSA model thermal inertias and the 2% model thermal inertias is of the order of 50 units. The CSA model thermal inertias are generally smaller, indicating that particle sizes have probably been overestimated for much of the planet. However, the importance of utilizing the CSA model thermal inertias may be greater than indicated by the mean difference, as the largest differences occur at or near some of the more geologically interesting areas as well as near the lander sites. The map of CSA model thermal inertias presented in Figure 4 is currently the best estimate of Martian thermal inertias between ~60° and 60° latitude.

Acknowledgments

We thank Mike Mellon for many useful discussions and for assistance in preparing the figures. We also thank Dave Paige and Phil Christensen for their thoughtful, constructive reviews. This research was supported in part by JPL through contract 957571 from the Mars Observer program and by NASA through grant NAGW-771.

References


Cheney, R. T., S. W. Lee, G. R. Gladstone, W. McMillan, and


Palluconi, F. D., and H. H. Kieffer, Thermal inertia mapping of Mars from 60°S to 60°N, Icarus, 45, 415-426, 1981.


Wennerstrom, A. K., A scale of grain and class terms for clastic sediments, J. Geol., 30, 377-392, 1922.


R. M. Haberle, Space Sciences Division, MS 245-S, NASA Ames Research Center, Moffett Field, CA 94035.

J. K. Hayashi and B. M. Jakosky, Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80309.

(Received February 24, 1994; revised August 22, 1994; accepted September 19, 1994.)
Thermal inertias in the upper millimeters of the Martian surface derived using Phobos' shadow

Bruce H. Betts, Bruce C. Murray, and Tomáš Sviták

Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena

Abstract. The first thermal images of Phobos' shadow on the surface of Mars, in addition to simultaneous visible images, were obtained by the Phobos '88 Termoskan instrument. The best observed shadow occurrence was on the flanks of Arsia Mons. For this occurrence, we combined the observed decrease in visible illumination of the surface with the observed decrease in brightness temperature to calculate thermal inertias of the Martian surface. The most realistic of our three models of eclipse cooling improves upon our preliminary model by including nonisothermal initial conditions and downward atmospheric flux. Most of our derived inertias fall within the range 38 to 59 J m\(^{-2}\) s\(^{-0.5}\) K\(^{-1}\) (0.9 to 1.4 \(10^{3}\) cal cm\(^{-2}\) s\(^{-0.5}\) K\(^{-1}\)), corresponding to dust-sized particles (for a homogeneous surface), consistent with previous theories of Tharsis as a current area of dust deposition. Viking infrared thermal mapper (IRTM) inertias are diurnally derived and are sensitive to centimeter depths, whereas the shadow-derived inertias sample the upper tenths of a millimeter of the surface. The shadow-derived inertias are lower than those derived from Viking IRTM measurements (84 to 147), however, uncertainties in both sets of derived inertias make conclusions about layering tenuous. Thus, near-surface millimeter versus centimeter layering may exist in this region, but if it does, it is likely not very significant. Both eclipse and diurnal inertias appear to increase near the eastern end of the shadow occurrence. We also analyzed a shadow occurrence near the crater Herschel that showed no observed cooling. This analysis was limited by cool morning temperatures and instrument sensitivity, but yielded a lower bound of 80 on eclipse inertias in that region. Based upon our results, we strongly recommend future spacecraft thermal observations of Phobos' shadow, and suggest that they will be most useful if they improve upon Termoskan's geographic and temporal coverage and its accuracy.

1. Introduction

Mars' moon Phobos orbits Mars in a roughly 8-hour, circular, equatorial orbit at an altitude of approximately 6000 km. During the time periods surrounding the Martian equinoxes, Phobos casts a completely penumbral shadow on the surface of Mars' equatorial regions during portions of each orbit. A passing of the shadow would be viewed by an observer on the surface as a partial eclipse lasting roughly 20 s. The Termoskan instrument on board the Soviet Phobos '88 spacecraft obtained the first thermal images of Phobos shadow on the surface of Mars. Simultaneous visible images were also obtained.

Termoskan was an optical-mechanical scanning radiometer with one visible channel (0.5-1.0 \(\mu\)m) and one thermal infrared channel (8.5-12.0 \(\mu\)m) (see Sestrom et al. [1989], Mervay et al. [1991], and Sestrom [1993] for more information). The instrument was fixed to the spacecraft, pointing in the anticlinal direction. Termoskan used a scanning mirror to build up the north-south component of image panoramas and the spacecraft's motion to build the east-west component.

Termoskan observed the shadow of Phobos on the surface of Mars during two of its four panoramas. Using these observations, we have been able to combine the observed decrease in visible illumination of the surface with the observed decrease in brightness temperature to calculate thermal inertias of the uppermost tenths of a millimeter of the Martian surface. Thermal inertia, a bulk measure of the resistance of a unit surface area to changes in temperature, is commonly used to characterize the insulating properties of planetary surfaces. It is defined as \(I = \frac{(\rho+c)p}{k}\)\(^2\), where \(k\) is the thermal conductivity, \(p\) is the bulk density, and \(c\) is the specific heat of the material. Low-inertia materials have smaller thermal skin depths and heat and cool more quickly than high inertia materials. In this paper, thermal inertia values are given in SI units (J m\(^{-2}\) s\(^{-0.5}\) K\(^{-1}\)) with inertias values in the units (10\(^3\) cal cm\(^{-2}\) s\(^{-0.5}\) K\(^{-1}\)) often used for the Martian surface (e.g., Kieffer et al., 1977) given in parentheses following the SI values.

Mervay et al. [1991] presented a summary of the preliminary results of an analysis of the best observed and least complicated occurrence of the shadow (near Arsia Mons; see Figure 1). We now present for the first time (1) a refined analysis of that shadow occurrence that incorporates two additional model factors: nonisothermal start conditions and atmospheric re-radiation, and that includes inertias as a function of longitude in the shadowed region; (2) an analysis of another shadow occurrence which was observed near the crater Herschel; (3) a detailed description of the models used; (4) a comparison with the diurnal inertias of Hsyaiishi et al. [this issue], which are based upon the combined surface atmosphere (CSA) model of Haberle
Section 2 of this paper gives an overview of the observations and the thermal models we used. Section 3 compares the Termoskan thermal data with the model results to derive thermal inertias. Section 4 discusses the results, and Section 5 discusses and the potential for future Phobos shadow research. The appendix describes in detail the three thermal models, including the inputs that were used from the Termoskan data.

2. Overview of Observations and Modeling

The elongated shape of the Phobos shadow (Figure 1) in the Termoskan panoramas was due to a fortuitous combination of the scanning nature of the instrument, unusual orbital geometry (the spacecraft and Phobos were nearly co-orbiting), the antisolar orientation of the instrument, and a slight rocking of the spacecraft. The similarity of the spacecraft's restaurant field of view rocked into and out of three different locations, and their basic differences are summarized in Table 1. All are centered roughly on 14°S latitude.

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>Scan Location</th>
<th>West Longitude Covered, deg</th>
<th>Times of Day Covered, hours</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Flanks of Arsia Mons</td>
<td>110 to 120</td>
<td>9.3 to 10.1</td>
<td>Analyzed here.</td>
</tr>
<tr>
<td>2</td>
<td>S. of western half of Valles Marineris</td>
<td>75 to 105</td>
<td>10.5 to 12.7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Eastern end is in the crater Herschel</td>
<td>228 to 255</td>
<td>7.4 to 9.5</td>
<td>No cooling apparent in thermal data. Analyzed here.</td>
</tr>
<tr>
<td>4</td>
<td>North of Malem Vallis</td>
<td>172 to 210</td>
<td>10.8 to 13.8</td>
<td>No visible channel data for eastern end of shadow. Dropped last in visible data that did exist.</td>
</tr>
</tbody>
</table>

All shadow occurrences are centered approximately upon 14°S latitude.

and Jakosky [1991]; and (5) implications for the planning and analysis of future similar observations.

Section 2 of this paper gives an overview of the observations and the thermal models we used. Section 3 compares the Termoskan thermal data with the model results to derive thermal inertias. Section 4 discusses the results, and Section 5 discusses and the potential for future Phobos shadow research. The appendix describes in detail the three thermal models, including the inputs that were used from the Termoskan data.
involved effects of the short duration of the eclipse relative to a diurnal period, and the relatively high daytime temperatures (which cause surface emission to dominate atmospheric emission). In order to test this and to ascertain the magnitude of atmospheric effects, we created model 3 by adding a downward atmospheric flux term to model 2. Model 3 is the most realistic of the three models, but due to the uncertainty in the downward atmospheric flux, it is useful to consider model 2 separately. For the Arsia occurrence, we present the results of all three models here for comparison and completeness (see the appendix for detailed descriptions of all three models.)

3. Comparison of Data With Models

3.1. Arsia Occurrence

To best compare Termoskan eclipse temperatures with model results, we used temperature drops within the eclipse rather than absolute temperatures. Variations in absolute temperature are observed outside the eclipse due to variations in albedo, inertia, and time of day. These factors also will affect the observed temperatures within the eclipse. To minimize these effects, we did two things to the data used for comparison. First, for points within and outside the shadow, we averaged 10-pixel (in the east-west direction) by 1-pixel strips. Second, the data we compare with models are estimated temperature drops within the eclipse. We find these temperature drops by averaging the temperatures of points to either side (north and south) of the shadow and then subtracting the temperature observed in the middle of the shadow. Using these temperature drops rather than absolute temperatures is particularly effective at removing time of day effects. Also, because the area happens to be reasonably bland thermally, gradual albedo and diurnal thermal inertia variations are also removed. In addition, by using temperature drops, we reduce the effects of atmospheric scattering, which should act approximately equally inside and outside the shadow.

To plot the data versus the model curves, these temperature drops were subtracted from 255 K, the pre-eclipse temperature used in the models. Figure 2 shows the data plotted versus model 1 curves. Figure 3 shows an analogous plot for model 2. Figure 4 shows curves for model 3 with a downward atmospheric flux (Al) of 20% of the absorbed pre-eclipse solar flux, and Figure 5 shows curves for model 3 with A_l = 10%. Data error bars are discussed in section 4.2.1., and although they are of interest in interpreting physical results, they do not affect comparisons of different models. Most of the data fall between inertias of 38 (0.9) and 50 (1.2) for model 1, between 41 (1.0) and 67 (1.6) for model 2, between 35 (0.8) and 50 (1.2) for model 3 with A_l = 20%, and between 38 (0.9) and 59 (1.4) for model 3 with A_l = 10%. In all cases, inertias are slightly higher later (towards the east) in the shadowed region.

3.2. Herschel Occurrence

For the Herschel occurrence the shadow appears clearly in Termoskan's visible channel, but cooling from the shadow does...
Eclipse model 2 results. Same as Figure 2 but for model 2 results. Model 2 added non-isothermal with depth pre-eclipse conditions to model 1. Most of the observational data values fall between model curves corresponding to thermal inertias of 41 (1.0) to 67 (1.6).

Figure 3. Eclipse model 2 results. Same as Figure 2 but for model 2 results. Model 2 added non-isothermal with depth pre-eclipse conditions to model 1. Most of the observational data values fall between model curves corresponding to thermal inertias of 41 (1.0) to 67 (1.6).

not show up in the thermal channel. Thus, for this occurrence, we cannot measure temperature drops. However, we can estimate the level of cooling that could have been detected by THEMIS, and compare that value to model results to derive a lower bound on eclipse inertias for the region near the center of the shadow. Modeling was carried out as before but with the appropriate changes in time of day, insolation, albedo, and starting temperatures. We only ran one of our models: Model 3

Figure 4. Eclipse model 3, \(A_\text{e} = 20\%\). Same as Figure 3 but for model 3 results. Model 3 added atmospheric downward flux to model 2. The curves plotted here are for a downward atmospheric flux of 20% of the absorbed pre-eclipse solar flux. Most of the observational data values fall between model curves corresponding to thermal inertias of 35 (0.8) to 50 (1.2).
with $A_f = 10\%$, which we determined to be our most realistic model (discussed below). We compared the maximum amount of modeled cooling for several thermal inertias with an estimate of the amount of cooling that could have been detected in the Termoskan data either by image processing or digital data analysis.

We estimate for this occurrence that we could have detected shadow cooling corresponding to a drop in the infrared channel of approximately 3 DN (data numbers) relative to the surrounding unshadowed area. This is based upon tested image processing of the data, comparisons with other occurrences of the shadow, and attempts to detect simulated shadow cooling effects. Infrared channel variations in this region due to inherent terrain variations are also on the order of 2 or 3 DN, but we are aided in looking for shadow cooling by the image nature of the data and the fact that we know where the shadow should be based upon the visible channel data.

For the temperatures in the Herschel region, 3 DN corresponds to approximately 2.5 K. Comparing with our model results, we find that a 2.5 K temperature drop corresponds to an eclipse inertia of approximately 80 (2). Thus, the data provide an approximate lower bound of 80 (2) for the thermal inertia in the upper millimeter of the region near the center of the shadow (approximately 14ºS, 24ºW). At first, it may seem counterintuitive that no observed cooling would only constrain the inertias to be above a value as low as 80 (2). After all, the very obvious cooling observed in the Arsia occurrence corresponds to inertias near 40 (1). However, the following factors make more accurate constraint impossible: (1) The Herschel occurrence was much cooler than the Arsia occurrence due primarily to the earlier morning observations as well as the inherently higher channel inertias in the Herschel region. Termoskan had less sensitivity at cooler temperatures (see Murrey et al. [1991] or Betts [1993] for the infrared response characteristics), e.g., 3 DN at 230 K correspond to about 2.5 K, but 3 DN at 255 K (characteristic of the Arsia occurrence) correspond to approximately 1.5 K. (2) Due to the strong dependence of the radiative cooling rate upon temperature, less shadow-induced cooling occurs when the initial temperatures are cooler. (3) The surface temperature drop during eclipse is very nonlinear with inertia, which causes less precision distinguishing between high inertia values than between low inertia values. For example, models of the Herschel occurrence yield temperature drops of 2.5 K, 1.3 K, 0.9 K, 0.7 K, and 0.6 K for inertias of 84 (2), 168 (4), 252 (6), 336 (8), and 420 (10), respectively.

4. Discussion

The results from the Arsia occurrence show that model 2 raises the derived inertias compared to the less realistic (isothermal) model 1. Adding atmospheric downward flux in model 3 reduces the derived inertias. First, we consider why the differences between the models cause these effects. Then, we consider the models' implications for the surface of Mars.

4.1. Model Differences and Realistic Inertias

Here we compare and evaluate our three models using the Arsia occurrence. Including the nonisothermal, initial temperature-versus-depth profile in model 2 allows the model surface to cool more quickly. For these low inertias, this dominates the simultaneous opposite (heating) effect from the additional pre-eclipse insolation caused by the inclusion of a nonzero conduction term in the surface boundary condition (see equations (3) and (5) in the appendix). Overall, the faster cooling causes higher inertias to be derived in model 2 than in model 1.
The atmospheric downward flux added to model 3 stays constant throughout the eclipse (the validity of this assumption is discussed in the appendix), and thus keeps the total radiation higher throughout the eclipse. This causes the surface to cool more slowly. This causes lower inertias to be derived.

Model 3 should be the most realistic model because it includes both nonisothermal starting conditions and downward atmospheric flux. The next question is, what value of atmospheric fluxes in model 3 is the most realistic? This is hard to answer precisely, but fortunately the most likely possibilities make little difference in our results and do not affect our eventual scientific conclusions. The atmospheric flux depends upon the atmospheric optical depth, which is not well known for the time of the observations. Analyses of data from other Phobos VII instruments has given optical depths that range from about 0.2 to 0.6, with the most favored values somewhere near the middle of that range. According to modeling by Haberle and Jakosky [1991, Figure 4], for this range of optical depths and the local time of day of the observations, the atmospheric flux will range from 6% to 11% of the solar flux. Thus, the 20% model run shown in Figure 4 is probably an extreme and should give lower bounds on inertia. Even the 10% run shown in Figure 5 probably gives inertias that are somewhat low. This is particularly true because the shadow was observed on a surface roughly 9 km (almost 3 mile) above the 6.1 mbar reference altitude used in the Haberle and Jakosky figure that the flux percentages came from. Atmospheric flux contributions will be reduced at higher altitudes.

The model 2 runs, which do not have any atmospheric flux contribution, should give inertia upper bounds. The real values should lie between the model 2 and model 3-20% values and are probably a little higher but not too far from the model 3-10% values. Thus, inertias of 38 (0.9) and 59 (1.4) probably bracket the majority of the eclipse-derived inertias in this shadow region. Ironically, because the nonisothermal effects nearly balance the atmospheric effects for these inertias, these results are nearly the same as the model 1 results (which were originally reported in Murray et al. [1991] and Bets et al. [1990]).

4.2. Implications for the Martian Surface

4.2.1. Arsia occurrence. The inertias derived from all the models are consistent with dust-sized particles. This is consistent with previous studies that proposed that the Tharsis region has a dust covering and is currently an area of dust deposition [Kieffer et al., 1977; Zimbelman and Kieffer, 1979; Palluconi and Kieffer, 1981; Christensen, 1986]. It is also reasonably consistent with the low rock abundances (about 5%) in this region [Christensen, 1982, 1983, 1986]. Assuming a homogeneous surface, inertias of 38 to 59 (0.9 to 1.4) likely imply particle sizes of approximately 5 to 10 microns (μm) [Haberle and Jakosky, 1991, Kieffer et al., 1973, Jakosky, 1986]. Note, however, that these values are theoretical extrapolations from laboratory data that only go down to particle sizes of approximately 50 μm [Jakosky, 1986]. Also, all of these particle sizes are for homogenous surfaces, thus the actual surface could be composed of all 2-μm particles and an occasional rock or any number of other combinations that yield a thermally averaged surface of 5- to 10-μm particles.

For the very low inertias involved in this study, diurnally derived inertias are representative of thermal skin depths of 1 to 2 cm. Due to the short duration of the eclipse relative to the length of a Martian day, the eclipse-based inertia determinations are sensitive to thermal skin depths of only a few tenths of a millimeter. Comparison of inertias derived by the two methods can indicate the degree of layering within the upper centimeters of the surface.

We compare our eclipse inertias for this region with Viking infrared thermal mapper (IRTM)-based diurnal inertias (*2×2′′ bins) calculated by both Hayashi et al. [this issue] and Palluconi and Kieffer [1981]. Hayashi et al. [this issue] used the combined surface-atmosphere model of Haberle and Jakosky [1991] along with the results of Palluconi and Kieffer [1981] and average dust opacities from Martin and Richardson [1993] in order to more accurately model the atmospheric contribution to the surface heat balance. Our eclipse inertias of 38-59 (0.9-1.4) compare to diurnal inertias for the same region of 86-133 (2.1-3.2) [Hayashi et al., this issue] and 50-147 (2.3-3.5) [Palluconi and Kieffer, 1981]. Note that the atmospheric corrections of Hayashi et al. [this issue] have relatively little effect on inertias when compared with Palluconi and Kieffer [1981] for this region due to relatively high altitude and low average dust opacities. Corrections are significantly larger elsewhere on Mars [Hayashi et al., this issue].

Thus, our derived eclipse inertias are significantly lower than diurnally derived inertias. However, the uncertainties in both sets of inertias make conclusions about near-surface layering tenuous. Most of the uncertainty for the Arsia occurrence eclipse inertias comes from a combination of approximating temperature drops within the eclipse (about ±0.4 K) and the precision of the instrument (about 0.5 K). Because these errors are independent and presumably random, they combine to provide an uncertainty of approximately ±0.7 K (which is the data uncertainty plotted in Figures 2 through 5). This translates to an inertia uncertainty of less than ±12 (0.3) for the majority of the inertias derived. It is worse for higher inertias where there is less temperature distortion between inertias, and for locations far from the shadow center. In addition, there are uncertainties due to possible inaccuracies in model inputs such as atmospheric downward flux (±0.1, based upon trying several values), the temperature versus depth profile (±0.1), would be worse for higher inertias, and other parameters such as visible flux, flux drop with time, and exact duration of eclipse (±0.2, based upon tests varying these parameters). Thus, the total uncertainty for the Arsia occurrence is approximately ±16 (0.4). So far, we have assumed that the uncertainty was the same, plus or minus. In reality the minus uncertainty is less than 16 (0.4) because inertia values are more easily distinguished at lower inertias. Note also that these uncertainties are for the Arsia occurrence and will vary even for the Termoskan instrument based on time of day, surface temperature, and inertia. Therefore, future observations can be better optimized that the Termoskan observations as discussed in section 5.

Hayashi et al. [this issue] state their uncertainty as 10-20 (0.2-0.5) for regions where albedo 20.28 and less than 50-60 (1.2-1.4) for albedos 50.28. The average albedo for the Arsia occurrence was 0.27 (it was 0.17 for the Herschel occurrence) [based upon Fleckor and Miner, 1981]. For the Arsia occurrence, we assume that the emissivity component of error in the Hayashi et al. [this issue] inertias is small since the average albedo of the region is nearly that of the cutoff albedo (0.28), above which emissivity errors are negligible (because the emissivity effectively equals 1 an assumed). However, since it is slightly below this cutoff, we include some emissivity error for comparison's sake. In this case, we have assumed that
the emissivity error is less than 10 (0.2) in the Hayashi et al. [this issue] diurnal inertias for the Arsia occurrence region.

Figure 6 compares eclipse inertias (with uncertainties found independently for each point) with the Hayashi et al. [this issue] inertias, which are plotted with uncertainties of -20 and -30 (emissivity error only not in the negative direction since it causes the inertias to be overestimated). Note that the eclipse inertias are, as discussed, consistently lower than the diurnal inertias, but that error bars often overlap. Also, notice that the eclipse inertias trend upwards towards the east as do the diurnal inertias. Unfortunately, the eclipse errors become large in that region because the eclipse is ending and the inertias are higher. However, the trend appears in the last few eclipse inertias points, making the simultaneous upward trend of both eclipse and diurnal inertias more believable than the error bars from any one point would indicate.

Thus, due to the uncertainties in both our eclipse inertias and in diurnally derived inertias, we are unable to say for certain that there is layering in the Arsia occurrence region. If there is, we can say that it is mostly fine dust over mostly less fine dust. The absence and most interesting locations to look for layering with future missions will be in regions that have higher diurnal inertias (indicative of average particle sizes significantly larger than dust) than those of the Arsia occurrence. The few tenths of a millimeter sampled by eclipse cooling measurements should present a fair representation of the surface that is sensed by optical and near-infrared instruments. Thus, for the region studied, it is unlikely that optical and near-infrared instruments would sample any significant amount of bare (not covered by dust) rock surfaces.

4.2.2. Herschel occurrence. The inertia lower bound of 80 (2) that we derive for the Herschel occurrence is consistent with the lower bound for the shadowed region: 198-252 (4.7-6.0) [Hayashi et al., this issue] and 239-301 (5.7-7.2) [Palluconi and Kieffer, 1981]. These diurnal inertias are consistent with a homogeneous surface of sand-sized particles. Unfortunately, because we are only able to derive a lower bound for this occurrence, and because that lower bound is so low, we are unable to derive any strong physical conclusions about the surface. If eclipse inertias are actually as low as 80 (2), then one could say there is a dust covering, but since they also could be significantly larger than 80 (2), the surface may be totally devoid of layering. Note, that if more advantageous observations (midday, greater sensitivity, etc.) are obtained by future spacecraft, the question of layering can be addressed effectively by the shadow analysis method.

4.3. Summary

Our analyses have shown that thermal and visible observations of the shadow of Phobos can be used to derive physical information about the upper millimeters of the Martian surface. Combining our preferred eclipse model with the Termoskan thermal data implies inertias for the Arsia Mons shadow occurrence that mostly fall within the range 38 to 59 (0.9 to 1.4). These inertias correspond to dust-sized particles (5 to 10 μm for a homogeneous surface based upon theoretical extrapolations of laboratory data). The presence of dust at the surface is consistent with previous theories of Tharsis as a current area of dust deposition. Thus, most of the upper couple centimeters is likely composed of unbounded, few micron dust particles. There may be a slight degree of layering, and the eclipse inertias and diurnal inertias both appear to increase towards the eastern end of the shadow. However, uncertainties
5. Future Phobos Shadow Research

Our results indicate that analyses of Martian surface cooling, due to the passage of the shadow of Phobos, can yield unique and interesting results about the current physical state of the upper millimeter of the Martian surface. However, the fortuitous Termonkan observations of the shadow are very limited in geographic coverage and, with the exception of the Armin occurrence, are very limited in accuracy. Because of these limitations, some of the key questions that could be addressed by this type of observation remain unanswered, e.g., is there a dust layer of at least millimeters in thickness covering much of Mars; or are dust layers regional in nature? Thus, we recommend that observations of the shadow be attempted with future missions. Future observations should improve upon the Termonkan data in geographic coverage, accuracy of the observations, and temporal coverage.

Ideal characteristics of future observations would include targeting midday shadow passages to maximize the pre-eclipse temperature, which will maximize the shadow temperature drops, the distinction between different inertiases during eclipse, and the signal to noise ratio, utilizing framing rather than (Termoskan-like) scanning imaging systems so the entire image is taken simultaneously, and taking several images as the shadow crosses the surface so that pre-eclipse, eclipse, and posteclipse temperatures are more accurately known for any given location, and having at least one broad band thermal channel and one broad band visible channel.

Observations will need to be taken during periods surrounding the equinoxes when the shadow will cross the Mars surface. The least shadow distortion and atmospheric interaction will occur in the equatorial region at equinox and during periods near that time. Because the Clifford model takes several hours, there should be ample opportunity to observe several shadow crossings at several latitudes. The most advantageous orbits for the observations will be equatorial (as is Phobos orbit) so that the shadow can be followed across the surface and so pre-eclipse and posteclipse observations can easily be obtained. One ideal observing situation would be to have the instruments on a lander on the Mars facing side of Phobos. Then, observations could be taken merely by pointing in an antisolar direction. They will be particularly useful if the field of view is wide enough and if observations are taken frequently enough so a number of locations are observed, with temporal coverage for each location providing pre-eclipse, eclipse, and posteclipse temperatures.

Appendix: Thermal Models of the Eclipse Cooling

In order to derive thermal inertia from the eclipse observations, we modeled the cooling of the surface for several thermal inertias and compared the results to the observed cooling. We present here descriptions and results from three different, but related, models: model 1, the isothermal model, which assumes that initially all depths are at the same temperature; model 2, the nonisothermal model, which utilizes an initial temperature with depth profile derived from the Clifford et al. [1987] diurnal thermal model, and model 3, the nonisothermal atmospheric model, which adds a downward atmospheric flux term to model 2. All three models are presented here for comparison and completeness. Model 1 was used to produce the results in Kieffer et al. [1991]. Model 2 improves on model 1. Model 3 is the most complete model because it includes atmospheric re-radiation, however, the amount of atmospheric flux is highly dependent upon poorly known atmospheric conditions at the time of the observation. Thus, it is useful to consider model 2 separately from model 3. All of the models solve the heat equation:

\[ \frac{\partial T}{\partial t} = \frac{1}{\rho c} \frac{\partial}{\partial z} \left( k \frac{\partial T}{\partial z} \right) \]  

where \( T \) is temperature, \( t \) is time, \( \rho \) is density, \( c \) is specific heat, \( k \) is thermal conductivity, and \( z \) is depth. This equation is solved using the same numerical method used by Clifford et al. [1987]. This is essentially the same diurnal model described by Kieffer et al. [1977]. Modeling the eclipse rather than diurnal and seasonal temperature variations requires several important differences between the eclipse models and the overall Clifford et al. model. Parameters such as the timescales used and inputs such as the decrease in visible flux with time in the eclipse must be changed. Where possible, we use the actual Termonkan data to make these changes, rather than relying on theoretical calculations. For example, we use the Termonkan visible data to derive the decrease in absorbed solar flux with time. Using the data where possible decreases the potential error introduced by poorly known factors such as the complex geometry including spacecraft rocking, atmospheric effects, and uncertainties in albedo.

All of the models produce model temperatures as a function of time in eclipse for a given value of thermal inertia. Actual thermal inertias are estimated by first running a model for several values of thermal inertia; then, model temperatures are compared with Termonkan data temperatures to find inertias for given values of time eclipsed. This comparison was discussed in section 3.

In the rest of this section we describe the three models by pointing out their differences from the Clifford et al. [1987] and Kieffer et al. [1977] models. In the first five points below, we present the elements that apply to all three models. Then, in the last two points, we address the differences in the three models. Hence, the following are the eclipse model features that are different from the Clifford et al. [1987] and, by analogy, the Kieffer et al. [1977] model.

1. Timescales (depth scales). We have adjusted all time-related variables to account for the 23-s eclipse timescale, rather than diurnal timescales. The iteration time, along with the modeled inertia, determines the thickness and depths of the model compartments based upon finite difference stability criteria [e.g., Clifford et al., 1987]. We changed the interval between iterations from several minutes to 0.001 s. Models with smaller iteration times were tested but showed no appreciable difference from the models run with 0.001-s iterations.

2. Use of orbital information and albedo. Neither Mars orbital information nor albedo information is used in the eclipse models. How these are avoided will become more clear in the specific discussions of elements below. We generalize this point here to emphasize the major differences with the Clifford et al. models.
3294

BEITEL ET AL.: MARTIAN INERTIAS DERIVED USING PHOEBOS SHADOW

[1987] model. Note that although the eclipse models do not use orbital information or albedo information directly, models 2 and 3 do use them indirectly because we ran the Clifford et al. [1987] model to generate the initial temperature with depth profiles (discussed in point 6 below). Errors in albedo introduce only very minor errors when used in this indirect fashion.

3. Downward atmospheric flux model. The downward atmospheric flux model term is set equal to the noontime solar insolation. Models 1 and 2 have no downward atmospheric flux and model 3 incorporates it elsewhere within the model.

4. Initial surface temperature. Rather than calculating this from orbital and albedo considerations, the initial (pre-eclipse) surface temperature is assigned a value that is representative of observed Terrestrial temperatures for regions just outside the shadow (e.g., 255 K for the Arsia Mons occurrence).

5. Relative insolation as a function of time. In order to model the cooling as a function of time, the model needs a description of the total absorbed insolation as a function of time within the eclipse. For models 1 and 2 this is equivalent to the absorbed solar insolation as a function of time. For model 3 this includes atmospheric as well as solar insolation as discussed more fully in point 7 below. We derive the shape of the insolation-versus-time function from the visible data and the magnitude from the thermal data. In point 7 below we consider the magnitude of the pre-eclipse insolation used, which differs for each of the three models.

Here we discuss the shape of the insolation versus time function. We use this same relative insolation function in all three models. To find the relative insolation, we did the following: A - calculated the duration of the eclipse for a point on the surface; B - derived a function that connects location within the shadow in the data to the amount of time that location has been in eclipse; C - used the visible channel eclipse profile to fit the decrease in absorbed solar insolation as a function of location within the shadow; and D - combined B and C to derive relative insolation as a function of time in eclipse. These steps are accomplished in the following manner.

A.

Combining the maximum width of a north-south profile, the orbital speed of Phobos, and the rotational speed of Mars gives an eclipse duration of just under 23 s.

B. We assumed the E-W rocking motion to be uniform over the brief Arsia Mons shadow observation. This assumption is justified by the short duration of the observation and by the smoothness and shape of the E-W visible shadow profile. Thus, we were able to develop a linear relation between the number of lines into eclipse and the time a location had been in eclipse using the observed E-W size of the shadow (approximately 180 lines for the Arsia occurrence) and the eclipse duration calculated above (approximately 22.6 s). So, for example, the 90th line after the beginning of eclipse had been in some form of eclipse for approximately 11.3 s.

C. The next step was to determine the relative decrease in solar insolation as a function of location within the shadow. We assumed that the visible channel signal was linearly proportional to visible flux, consistent with the visible detector's characteristics and with all approximate flux calibrations of the visible channel [Beitz, 1993]. Thus, the E-W visible channel signal profile should approximate the visible flux decrease within the shadow. We used this profile to quadratically fit the relative decrease in solar insolation as a function of location within the shadow. Note that in the darkest part of the shadow the flux decreased by approximately 30%. The advantage of using the visible channel profile instead of theoretical calculations of the flux decrease is that it already incorporates atmospheric effects as well as geometric distortion due to the spacecraft rocking. Position within the visible channel can be directly correlated with position within the thermal channel to within approximately 1 pixel [Beitz, 1993; 1995].

D. Combining B and C gives the relative absorbed solar insolation as a function of time eclipsed, f(t). This function can vary from 0 to 1, where 1 represents the pre-eclipse insolation. The only remaining step for modeling the insolation is to get an absolute pre-eclipse absorbed solar insolation that this relative function can be mated with. This is discussed separately below in point 7. First, we discuss the initial temperature with depth profile used because that will be important for point 7.

6. Initial temperature with depth profile. Model 1 is isothermal before the onset of eclipse with all depths set equal to the pre-eclipse representative surface temperature (255 K for the Arsia occurrence). Models 2 and 3 utilize more realistic nonisothermal subsurface pre-eclipse temperature profiles. As mentioned, establishing a reasonable estimate for this temperature with depth profile is the only place in the models where either albedo or Mars orbit information is used. Even here, they are used indirectly. Before each run of either models 2 or 3, we ran our adaptation of the Clifford et al. [1987] diurnal model utilizing the appropriate seasons (o = 18°), approximate latitude (14°S), and approximate albedo (0.27 for the Arsia occurrence and 0.17 for the Heresland occurrence as taken from the 1st x 1st bandit bolometric albedos of Phobos and Idunn [1981]). Using the diurnal model output corresponding to the local time of day of the eclipse, about 10 H for the Arsia occurrence (where 24 H = 1 Martian day), we estimated the pre-eclipse temperature as a function of depth with a linear approximation:

\[ T_j = T_0 - \frac{D_j (T_m - T_0)}{D} \] (2)

where \( T_j \) is the temperature of the jth compartment, \( T_0 \) is the initial surface temperature, i.e., 255 K for the Arsia occurrence, \( D_j \) is the depth of the top of the jth eclipse depth compartment, \( T_m \) is the surface temperature in the diurnal model, \( T_0 \) is the temperature of the first depth compartment in the diurnal model, and \( D \) is the depth of the first depth compartment in the thermal model. Note that this pre-eclipse temperature with depth profile depends upon the thermal inertia used in the diurnal model. Thus, for each value of inertia run in the eclipse model, we first modeled the depth profile using that inertia in the Clifford et al. [1987] diurnal model. This gave new values of \( T_{0m} \) and \( T_{1m} \) and to use in equation (2) and the eclipse model. Then, for a given inertia, all values of \( T_j \) could be found as a linear function of \( D_j \).

7. Pre-eclipse absorbed insolation. Finally, we needed the value of the pre-eclipse absorbed insolation to use with the relative function found in part 5 above. To do this, we use the surface equilibrium boundary condition discussed by Kieffer et al. [1977] and elsewhere:

\[ \frac{S(1 - A_{vis})}{R^2} + F_d = \frac{\kappa T^4}{A} = \frac{dF}{dt} \] (3)

where \( S \) is the solar constant, \( A \) is the bolometric Bond albedo, \( \kappa \) is the solar incidence angle, \( R \) is the Mars heliocentric distance, \( F_d \) is the downward atmospheric flux, \( c \) is the surface emissivity, \( T \) is the surface temperature, \( k \) is the thermal conductivity, \( z \) is the depth, \( dF/dt \) is the change
in temperature with depth evaluated at the surface, \( L \) is the CO₂ latent heat of condensation, and \( M \) is the mass per unit area of CO₂ frost.

The terms on the left in equation (3) represent the total absorbed insolation, i.e., what we need to know for pre-eclipse conditions. The first term on the left is the absorbed solar insolation. The second term on the left represents the downward infrared atmospheric flux that reaches the surface. This thermal infrared flux is assumed to be absorbed completely by the surface. The first term on the right is the surface emission. Here we assume the emissivity, \( \varepsilon \), to be 1, as was done by Kieffer et al. [1977]. The second term on the right is the surface conduction term. The third term on the right is the term representing latent heat from carbon dioxide condensation or condensation. The temperatures involved in this shadow analysis are far above the carbon dioxide condensation temperature, so this term is set to zero.

The three models differ in which terms are ignored in calculating the total absorbed insolation. Model 3 only ignores the CO₂ latent heat term. Model 2 also ignores the atmospheric term, and model 1 additionally ignores the conduction terms.

In principle, the total absorbed insolation could be calculated directly in any of these models from the terms on the left side of equation (3). Instead, we choose to solve for the total insolation by finding the terms on the right side. There are three reasons to do this. One, by more explicitly using the temperature data, we avoid uncertainties in albedo, and the orbital geometry, and we significantly reduce the effects of atmospheric scattering. Two, utilizing the same initial surface temperature, 225 K, that is already used in the pre-eclipse temperature profile will result in greater initial consistency within the model. Three, in model 3, to calculate the terms on the left side, we also would need to directly calculate the atmospheric downward flux. This is very difficult to do accurately due to the poor knowledge of the state of the Martian atmosphere at the time of the observations. By using the surface emission and conduction terms to determine the total insolation, we avoid these problems.

We summarize the differences between our three eclipse models in Table 2. Here we describe in detail the differences in the insulations used. In model 1, which ignores the atmospheric, conduction, and latent heat terms, the pre-eclipse absorbed insolation, \( I_0 \), is approximated by \( \alpha T^4 \), where \( \alpha \) is the Stefan-Boltzmann constant and \( T \) is a representative value of the surface temperature derived from the data itself (225 K for the Arsia occurrence).

Non-eclipsed surface temperatures, and the corresponding non-eclipsed absorbed insolutions, will change over the region covered by the eclipse and differ from the representative surface temperature (225 K for Arsia), however, the short duration and small area covered by the Arsia eclipse keep these variations small. Also, as discussed in section 3, we use temperature drops within the eclipse rather than absolute temperatures for comparison with the model to minimize errors caused by insolation or initial temperature errors in the model.

Ignoring the conduction term, \( \kappa(dT/dz) \), in model 1 is consistent with the initial isothermal assumption in this model which implies \( dT/dz = 0 \). The insolation as a function of time, \( I(t) \), combines \( I_0 \) with the relative flux decrease as a function of eclipsed time, \( R(t) \), that was derived from the visible channel data in step 5. Thus, for model 1,

\[
I(t) = I_0 R(t) \tag{4}
\]

Model 2 differs from model 1 by approximating the pre-eclipse absorbed insolation, \( I_0 \), by

\[
I_0 = \alpha T^4 - \frac{k T^4}{R} \tag{5}
\]

Note that including the conduction term here is consistent with the nonisothermal initial conditions of this model. After the initial temperature profile with depth is derived for this model as described in step 6 above, \( dT/dz \) is approximated by Taylor expansions of the first three depth steps and their temperatures [Clifford et al., 1987]. Note that for the mid-morning Arsia eclipse observation, \( dT/dz \) is negative. Thus, \( I_0 \) will be larger in model 2 than in model 1. The total insolation as a function of time, \( R(t) \), is found in model 2 in the same way as in model 1, using equation (4).

Model 3 determines the value of \( I_0 \) in the same manner as model 2, using equation (5). Model 3 differs by assuming \( I_0 \) to be composed of an atmospheric component (downward infrared flux) in addition to the solar component. Thus, model 3 differs from model 2 in the calculation of \( R(t) \). The relative flux drop as a function of time derived in step 5 is based upon the visible channel, which sensed only the solar flux drop. We assume that the absorbed downward atmospheric flux, \( I_a \), remains constant throughout the eclipse. This assumption is very reasonable because (1) the shadowed portion of the atmosphere represents less than approximately 0.5% of the volume of atmosphere that contributes infrared atmospheric flux to the shadowed surface, and (2) the flux from the shadowed atmosphere will only drop on the order of 10-20% even assuming very large amounts of atmospheric cooling, so the total atmospheric flux variation will be less than 0.1%. Thus, in our model, \( I_a \) remains constant throughout the eclipse, whereas the absorbed solar insolation, \( I_s \), will vary with eclipse time as described by the function \( R(t) \) that was derived in step 5. Thus, rather than using equation (4) as models 1 and 2 did, model 3 uses

\[
I(t) = I_s + I_a R(t) \tag{6}
\]

Here, \( I_s \), which is what we derived from the surface emission and conduction terms in equation (5), represents the contribution of \( I_a \) and \( I_s \), i.e.,

\[
I_s = I_s + I_a \tag{7}
\]

\( I_a \) will depend strongly upon the amount of dust in the atmosphere as well as to some extent the elevation of the surface. In order to conveniently model different values of \( I_s \), we define \( I_s \)
as a fraction, \( A_0 \), of \( I_0 \). Thus,

\[
I_0 = A_0 I_s
\]  

(8)

Combining equations (6), (7), and (8) gives an equation dependent upon only \( I_0 \) and \( A_j \):

\[
I(t) = \frac{\frac{1}{2} I_s (1 + A_j)}{1 + A_j}
\]  

(9)

which reduces to equation (4) when \( A_j = 0 \). So, using equation (9), the model can be run for several values of \( A_j \) to determine the effects of different amounts of downward atmospheric flux.

Acknowledgments. We thank Ken Heinzenhof and Ken Elgerd for their helpful and detailed reviews, Doug Nash, Denny Mulhine, and Andy Leighton for helpful comments on the manuscript, Ioan Hayashi for providing digital versions of the Hayashi et al. (this issue) kernels; and Bruce Jakosky, David Paige, and Hugh Kieffer for insightful discussions. We also thank A. Selivanov, M. Naraeva, V. Kharlamov, and V. Gektin for assistance with the data, and B. Ely, R. McLendon, K. Hudleston, and S. Welah for assistance with tasks at SSI. Funding for this research was provided by NASA grant NASG-1426 and NASG-2491. Division of Geological and Planetary Sciences, California Institute of Technology contribution 5284.

References


B. H. Betts, San Juan Capistrano Research Institute, 3182 Camino Capistrano, San Juan Capistrano, CA 92675. (e-mail: betts@pmail.sjc.r.i.c.)

B. C. Murray, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125 (e-mail: bcharles@msir2.ida.caltech.edu)

T. Svitek, Orbital Sciences Corporation, 21700 Atlantic Blvd., Dulles, VA 20166.

(Submitted March 1, 1994; revised January 13, 1995; accepted January 13, 1995.)
Mid-infrared transmission spectra of crystalline and nanophase iron oxides/oxyhydroxides and implications for remote sensing of Mars

James F. Bell III1
National Research Council/NASA Ames Research Center, Space Science Division, Moffett Field, California

Ted L. Roush2
Department of Geosciences, San Francisco State University, San Francisco, California

Richard V. Morris
Planetary Science Branch, NASA Johnson Space Center, Houston, Texas

Abstract. Ferric-iron-bearing materials play an important role in the interpretation of visible to near-IR Mars spectra, and they may play a similarly important role in the analysis of new mid-IR spacecraft spectral observations to be obtained over the next decade. We review existing data on mid-IR transmission spectra of ferric oxides/oxyhydroxides and present new transmission spectra for ferric-bearing materials spanning a wide range of mineralogy and crystallinity. These materials include 11 samples of well-crystallized ferric oxides (hematite, maghemite, and magnetite) and ferric oxyhydroxides (goethite, lepidocrocite). We also report the first transmission spectra for purely nanophase ferric oxide samples that have been shown to exhibit spectral similarities to Mars in the visible to near-IR and we compare these data to previous and new transmission spectra of terrestrial palagonites. Most of these samples show numerous, diagnostic absorption features in the mid-IR due to Fe3+-O2 symmetry vibrations, structural and/or bound OH, and/or silicates. These data indicate that high spatial resolution, moderate spectral resolution mid-IR ground-based and spacecraft observations of Mars may be able to detect and uniquely discriminate among different ferric-iron-bearing phases on the Martian surface or in the airborne dust.

Introduction and Background

Ferric iron (Fe3+) iron-bearing minerals are present on the surface of Mars. Information regarding the crystalline or amorphous nature of the iron-bearing (and other) surface materials on Mars can provide insight into the availability of liquid water at the surface and the duration, mode, and extent of chemical weathering. The presence of ferric iron has been established by a variety of remote sensing and in situ analysis measurements performed over the past four decades by Earth-based telescopes and the Mariner, Viking, and Phobos spacecraft missions [e.g., Dulac, 1955; Skwarek, 1961; Adams and McCord, 1969; McCord and Westphal, 1971; Hargreaves et al., 1977; Tanumihardjo et al., 1977; Clark et al., 1982; Pollack et al., 1977; Soderblom et al., 1978; Singer, 1982; Bell et al., 1990; Brand et al., 1991; Bell, 1992; Soderblom, 1992; Murck et al., 1993; Roush et al., 1993].

1Also at University of Washington, Geological Remote Sensing Laboratory, Seattle.
2Also at NASA Ames Research Center, Space Science Division, Moffett Field, California.

Copyright 1995 by the American Geophysical Union. Paper number 94JE01389. 0148-0227/95/502E-01389-$05.00

However, since the Viking Landers did not carry any instruments capable of determining mineralogy, the identities of the specific iron-bearing phases remain uncertain.

The existing remote sensing data indicate that Mars ferric minerals exist in two distinct but co-existing phases: a poorly crystalline or even amorphous phase similar to nanophase ferric oxides [e.g., Morris et al., 1989; Morris and Lauer, 1990; Banin et al., 1993] and much more crystallized ferric oxide phases such as hematite and magnetite [e.g., Pollack et al., 1977; Morris et al., 1989; Bell et al., 1990]. Analysis of visible to near-IR spectra and comparisons to laboratory mineral samples and spectral analogs have shown that these two phases are intimately associated: the bulk ferric oxides seem to occur within a matrix of the more fine-grained, possibly nanophase, ferric oxides, with the former accounting for specific charge-transfer absorption features and the latter being the primary pigmenting agent in the Martian soils [e.g., Morris et al., 1990, 1993; Bell et al., 1993; Golden et al., 1993].

In the near-to mid-infrared (400-4000 cm−1, 2.5-25 μm), spectral features arise from vibrational motions of atoms and molecules which comprise the materials. Fundamental modes of these vibrations yield IR absorption features that are much more intense than any associated combination and/or overtones of these modes, and hence IR remote sensing observations are extremely sensitive to minor concentrations
of these absorbing species [e.g., Farmer, 1974; Ryskin, 1974; Gaddaden, 1975; Salisbury et al., 1991]. Mid-IR ground-based and spacecraft observations of Mars have been presented and discussed by Hsu et al. [1973], Toon et al. [1977b], Pollack et al. [1990], Roush et al. [1991, 1992], and Moersch et al. [1995] among others.

This study represents one aspect of a broader project intended to provide data in the mid- and far-infrared spectral regions for both well-characterized iron oxides/oxyhydroxides and poorly crystalline or amorphous materials like palagonites and nanophase ferric oxides [Roush, 1992; Bell and Roush, 1993; Roush and Bell, this issue]. In this paper we review the currently available transmission data for many ferric oxides/oxyhydroxide minerals that could plausibly be considered to exist on Mars (e.g., Bell, 1992), present new transmission measurements from a suite of well-characterized ferric-bearing materials, and discuss the implications of these data on the interpretations of existing and potential future remote sensing observations of Mars.

Existing Data
Mid-IR transmission spectra of uncharacterized samples of hematite (α-Fe₂O₃), goethite (α-FeOOH), lepidocrocite (γ-FeOOH), akaganeite (β-FeOOH), and magnetite (Fe₃O₄) were discussed and reviewed by Nyquist and Kogel [1971], Hsu et al. [1973], Farmer [1974], Ryskin [1974], and Gaddaden [1975]. The results in the mid-IR can be summarized as follows: hematite spectra showed strong absorption features near 633, 544, and 465 cm⁻¹ (15.8, 18.4, and 21.5 μm); goethite spectra showed features near 3125, 893, 800, 667, 592, and 461 cm⁻¹ (3.2, 11.2, 12.5, 15.0, 16.9, and 21.7 μm); lepidocrocite spectra showed bands near 3333, 1163, 752 cm⁻¹ (3.0, 8.6, and 13.3 μm); akaganeite has absorption bands near 3333, 847, 676, and 450 cm⁻¹ (0.0, 11.8, 14.8, and 22.2 μm); and magnetite has absorption bands near 576 and 380 cm⁻¹ (17.3 and 26.3 μm). Little effort was made in these reviews to assign a specific origin to these features, except for the OH-bearing species. For those samples, bands near 3333 cm⁻¹ (3.0 μm) were assigned to the OH stretching fundamental; bands near 893 to 1163 cm⁻¹ (8.6 to 11.2 μm) were assigned to OH in-plane bending modes; and bands near 752 to 800 cm⁻¹ (12.5 to 13.3 μm) were assigned to OH out-of-plane bending modes [Ryskin, 1974]. The subhydrous phases exhibited many fewer absorption features in the mid-IR than the OH-bearing phases.

More recently, Salisbury et al. [1991] have presented mid-IR transmission and reflectance spectra for four petrographically and chemically analyzed samples of hematite, goethite, and magnetite. The major bands identified from our analysis of their digital data are similar to those discussed above and are included in Table 1 below.

New Samples
Because there can be considerable physical, chemical, and structural variability among samples, we obtained transmission measurements of a different set of well-characterized crystallized ferric materials for comparison with previously obtained lab data and currently available mid-IR Mars remote sensing data. We also report the first mid-IR transmission measurements of purely nanophase ferric oxides shown to be good spectral analogues to Mars at visible wavelengths. Finally, we measured the transmission spectrum of a naturally occurring palagonite sample for comparison with the nanophase and well-crystallized ferric materials and with previously reported transmission measurements of palagonites [Roush, 1992; Crisp and Bartholomew, 1992; Roush et al., 1993].

The well-crystallized ferric oxide samples include four hematites (HMS5, HMS13, HMS14, HMS15), three goethites (GTS2, GTS3, GTS5), two lepidocrocites (γ-FeOOH), MHS4, one lepidocrocite (LPS2), and one magnetite (MTS4). All of these samples are submicron powders that have been chemically and physically characterized by Morris et al. [1985] using a variety of laboratory techniques including Mössbauer and visible to near-IR reflectance spectroscopy, X-ray fluorescence, X-ray diffraction (XRD), thermogravimetric analysis (TGA), vibrating sample magnetometry (VSM), and transmission electron microscopy. All of the samples were found to be reasonably pure, with the major impurities being SiO₂ and S₉O₇ which varied in abundance between samples at the 0.5 to 2.5% level [Morris et al., 1985].

The nanophase iron oxide samples include two pure synthetic samples that have been heated to different temperatures (TNA13 at 210⁰ and 260⁰C, and TNA30 at 215⁰ and 290⁰C), one synthetic ferrihydrite sample (S11-4-DDS), and one sample of nanophase ferric oxide that was dispersed in a silicon gel matrix (S6FN28). The pure TNA13 and TNA30 samples were prepared by oxidation in air of trinucleate sodium hydroxyiron(III) nitrate as described by Morris et al. [1991]. The low-temperature samples (210⁰ and 215⁰C) have been characterized as nanocrystalline by Mössbauer and reflectivity measurements that do not show any indication of crystallized ferric oxide phases [Morris et al., 1991]. This same study showed that the higher-temperature samples (260⁰ and 290⁰C) begin to show evidence of crystalline hematite. Golden et al. [1994] have determined a mean particle size of 6.1 μm for TNA13/210 and 8.5 μm for TNA13/260 using high-resolution transmission electron microscopy techniques. These authors also used electron diffraction to determine that the composition of the TNA13 nanophase particles is hematite and/or ferrihydrite. The S11-4-DDS synthetic ferrihydrite sample is a 2-line ferrihydrite (verified by Mössbauer spectroscopy), prepared and provided to us by D.G. Schulte. The S6FN28 sample dispersed in silica gel was prepared by impregnating gel particles having 6-nm-diameter pores with a ferric nitrate solution, followed by calcination at 550⁰C [Morris et al., 1989]. The sample was characterized chemically and physically by Morris et al. [1989] using Mössbauer and visible to near-IR reflectance spectroscopy, instrumental neutron activation analysis, XRD, TGA, and VSM. The same techniques were used by Bell et al. [1993] to characterize the naturally occurring palagonite tektite sample PF-1. This sample underwent post-impact thermal alteration, which oxidized some of the preexisting basaltic tektite particles. The ferric iron mineralogy of the sample measured here, from the finest size fraction of PF-1 (< 20 μm), was found to be dominated by nanophase ferric oxide, with minor amounts of well-crystallized hematite and magnetite. Other phases identified in the sample include olivine, pyroxene, plagioclase, and very minor smectite clay [Bell et al., 1993].

Measurement Technique
A small amount (1 μg) of each sample was mixed with KBr powder while both were immersed in alcohol. After the al-
cobalt evaporated, 200 mg of the resultant mixture was pressed into a cohesive pellet using a standard 13-mm KBr pellet die. Using the facilities at the NASA Ames Astrochemical Laboratory, with the cooperation of Robert Bohn, Scott Sanford, and Louis Allamandolla, the transmission spectrum of each pellet was obtained at a resolution of 4 cm⁻¹ using a Nicolet 7199 Fourier transform spectrometer in the 4000 to 400 cm⁻¹ spectral region, and was subsequently ratioed to the transmission of a pure KBr pellet.

Results

Well-Crystallized Iron Oxides/Oxyhydroxides

Our transmission measurements of pure crystalline ferric oxide and oxyhydroxide powders are shown in Figures 1 and 2. A compilation of the major spectral features in these data is presented in Table 1. The transmission spectra of the four hematite samples (Figure 1) show strong absorptions in the 450 to 600 cm⁻¹ (16 to 22 μm) region separated by a narrow maximum near 500 cm⁻¹ (20 μm). These spectra are consistent with previously reported data for hematite [e.g., Farmer, 1974; Gadsden, 1975; Salisbury et al., 1991]. Because of the known purity of the samples and the relative weakness of transmission features due to adsorbed or structurally contaminating water, it is likely that the 450 to 600 cm⁻¹ bands are the result of Fe³⁺-O⁻ vibrational transitions in the hematite hexagonal close-packed (hcp) structure [Klein and Hurlbut, 1985]. The strong drop-off in transmission towards the highest frequencies for samples HMS13 and HMS15 most likely reflects the decrease in signal due to scattering of light out of the incident beam as the wavelength of the incident radiation begins to approach the mean particle size of the samples. Morris et al. [1985] showed that both of these samples have a much larger mean particle size than the other two hematites studied here; this observation is consistent with our scattering interpretation.

The transmission spectra of our two maghemite samples (Figure 1) are substantially more complex. There is a broad similarity to the hematite spectra in that there are strong absorption features in the 440 to 700 cm⁻¹ region with a maximum near 500 cm⁻¹, but these strong bands exhibit much more detail than the broad features in hematite. Since again there is little evidence for a major spectral influence of water or OH on the spectra, we would conclude that these detailed and apparently characteristic spectral features are caused by Fe³⁺-O⁻ vibrational transitions specific to this cubic close packed (ccp) polymorph of hematite.

Spectra of our three goethite samples (Figure 2) show more and narrower features below 1000 cm⁻¹ than any of the anhydrous iron oxides in Figure 1. In addition, a strong OH vibrational feature near 3120 cm⁻¹ is apparent. These goethite spectra are consistent with previously reported spectra [e.g., Ryskin, 1974; Gadsden, 1975; Salisbury et al., 1991]. The bands near 900 cm⁻¹ and 800 cm⁻¹ have been interpreted as in-plane and out-of-plane OH bending transitions by Ryskin [1974]. Based on these assignments and the measurements of the anhydrous phases discussed above, we interpret the remaining strong absorption features near 450 and 600 cm⁻¹ as Fe³⁺-O⁻ vibrational transitions diagnostic of the goethite hcp crystal structure [Sherman et al., 1982; Klein and Hurlbut, 1985].

The spectrum of our lepidocrocite sample (Figure 2) is similar to the goethite spectra in that there are many narrow features at the lowest frequencies and a very strong OH vibrational feature near 2100 cm⁻¹, but there are important differences. First, the OH vibrational band is much broader in lepidocrocite than it is in goethite. Second, bands interpreted as the in-plane and out-of-plane OH bending transitions [Ryskin, 1974] near 1160 and 750 cm⁻¹ are shifted substantially between the y and a oxyhydroxide phases. And third, a complex of bands occurs near 420 to 600 cm⁻¹ that may be due to Fe³⁺-O⁻ vibrational transitions and which are shifted relative to the Fe³⁺-O⁻ band positions of other ferric oxides and oxyhydroxides. The difference between the lepidocrocite and goethite spectra must be due in large part to structural differences between these hcp and ccp polymorphs [e.g., Schwertmann and Taylor, 1989].

While we do not reproduce the data here, Nypont and Kage [1971] show a spectrum of akaganite (the β phase polymorph of goethite) that also shows a complex spectrum exhibiting many features similar to those seen for goethite and lepidocrocite but occurring at slightly different frequencies (Table 1). It appears that there are enough systematic structural differences between goethite and its structural polymorphs to allow each to have very distinct and diagnostic mid-IR spectral signatures.

The transmission spectrum of our final crystalline sample, maghemite, is shown in Figure 2. This spectrum is similar to a previously measured spectrum by Salisbury et al. [1991] in that it shows very few spectral features. Only one strong band is seen near 580 cm⁻¹ (17.2 μm) which, because of the lack of evidence for OH or water contamination, is most likely due to Fe³⁺-O⁻ vibrational transitions in the cubic magnetite structure. The steep drop-off in transmission at the highest frequencies is most likely caused by the general visible to near-IR opacity of maghemite (resulting from very strong Fe³⁺ and Fe²⁺ charge transfer transitions; e.g., Morris et al., 1985), although an additional contribution from the increasing effects of scattering cannot be ruled out.

None of the samples show strong absorption features associated with any of the low-level impurities (SO₄, SO₃, etc.) measured by Morris et al. [1985]. Weak features near 1000 to 1100 cm⁻¹ are most likely caused by the small amount of SO₂ in these samples, which is the most abundant impurity besides water. Other weak features near 1050 cm⁻¹ and 2900 to 3000 cm⁻¹ are probably caused by C-H vibrations from residual alcohol used in the sample preparation procedure.

Nonaqueous Ferric Oxides and Palagonites

Spectra of our nanophase ferric oxide and palagonite samples are shown in Figures 3 and 4. Also shown in Figure 4 are transmission spectra of other palagonite samples obtained previously by Rosash [1992] and Cripps and Bartholomew [1992]. The spectra of these materials are substantially different from those of the well-crystallized ferric oxides/oxyhydroxides. The primary differences are that these materials do not exhibit as many features as the crystalline materials and the spectral contrast of the features that do exist in the nanophase and some of the palagonite samples is much lower than for the well-crystallized samples. Most of the samples show distinct absorptions in the 450 to 1200 cm⁻¹ (8.3 to 22.2 μm) region. By comparison with the crystalline sample spectra, it is probable that bands occurring in the 450
Figure 1. Transmission spectra of well-characterized and well-crystallized submicron powders of the iron oxides hematite (HMS series) and maghemite (MHS series). Data are plotted in frequency on the left and wavelength on the right.
Figure 2. Transmission spectra of well-characterized and well-crystallized submicron powders of the iron oxyhydroxides goethite (GTS series) and lepidocrocite (LPS2). A magnetite (MTS4) spectrum is shown at the bottom. Data are in frequency on the left and wavelength on the right.

Many of the nanophase materials show a general decrease in transmission towards the higher frequencies. As in the well-crystallized samples above, this effect is most likely a
5302

BELL ET AL.: MID-IR TRANSMISSION SPECTRA OF FERRIC MATERIALS

Table 1. Major Features in Mid-IR Transmission Spectra of Ferric Materials

<table>
<thead>
<tr>
<th>Sample</th>
<th>Frequency of Strongest Transmission Features, cm⁻¹</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFH3</td>
<td>478, 548, 1408, 1630, 3402</td>
<td></td>
</tr>
<tr>
<td>MFH5</td>
<td>478, 542</td>
<td></td>
</tr>
<tr>
<td>MFH16</td>
<td>444, 525, 602, 3408</td>
<td></td>
</tr>
<tr>
<td>MFH15</td>
<td>447, 528, 1458, 1653, 3467</td>
<td></td>
</tr>
<tr>
<td>MNNH7373</td>
<td>467, 546, 1037, 1635, 3492</td>
<td></td>
</tr>
<tr>
<td>MHS3</td>
<td>446, 553, 636, 694, 879, 1049, 1093, 2927, 3350</td>
<td></td>
</tr>
<tr>
<td>MHS4</td>
<td>442, 553, 636, 694, 879, 1051, 1093, 1637, 2927, 3408</td>
<td></td>
</tr>
<tr>
<td>S6FN28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT64</td>
<td>577, 700a, 881a, 1051, 1092</td>
<td></td>
</tr>
<tr>
<td>MNNH14111</td>
<td>579</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mid-IR Transmission Features

Well-Crystallized Hematite (α-Fe₂O₃)

Strong, well-characterized bands are shown in boldface. An "a" after the frequency means that the band occurs as a shoulder with an approximate center at the given value. A "d" or a "t" after the frequency indicates that the feature is a doublet or a triplet. Only the frequency of the strongest absorption is listed for close doublets or triplets.

Notes: [1], Samples characterized by Morris et al. [1985]; [2], Samples measured and characterized by Salisbury et al. [1991], band minima determined by us; [3], Sample characterized by Nyquist and Kagel [1977]; [4], Sample characterized by Bell et al. [1993]; [5], Samples characterized by Morris et al. [1991]; [6], Additional sample characterization by Golden et al. [1994]; [7], Samples characterized by Morris et al. [1989]; [8], Sample prepared and characterized by D.G. Schulze.

manifestation of the increased effect of scattering at shorter wavelengths. All of the nanophase ferric oxide component in Barrholomew, been examined. After examining the S6FN28 spectrum and comparing it to the spectrum of the silica gel matrix alone (Figure 3), we have been unable to identify any features that are unique to the nanophase ferric oxide component in S6FN28. The 1000-1100 cm⁻¹ band in PH-1, 91-1, and the Pahala Ash samples comes from the presence of other silicate minerals in these naturally occurring materials [Ryskin, 1992; Craig and Bartholomew, 1992; Bell et al., 1993]. The band at 1643 cm⁻¹ (6.1 μm) in PH-1 may also result from silicate minerals like olivine or pyroxene [e.g., Nash et al., 1993], or from an H-O-H bending fundamental of structural water [Ryskin, 1974], all of which are known to exist in minor amounts in PH-1. A weak feature at a similar position can also be seen in the S6FN28 and silica gel spectra. Bands centered near 3400 cm⁻¹ in the 91-1 and Pahala Ash spectra result from the OH stretching fundamental of structural water in these samples. This feature is absent in PH-1 presumably because that palagonite sample had much of its water driven off by a thermal alteration event [Bell et al., 1993].

Discussion and Implications

Transmission spectra such as these cannot be used to directly deduce the expected remote-sensing signatures from ferric-iron-bearing materials on the Martian surface or in the airborne dust because the samples were measured in an environment (pressed KBr pellet) that is nothing like that which is likely to occur in nature. Specifically, these spectra do not reveal the same important and diagnostic effects of scattering...
Figure 3. Transmission spectra of well-characterized nanophase ferric materials. The TNA series nanophase iron oxide samples were produced in the lab by oxidation in air of trinuclear acetato hydroxy iron(III) nitrate [Morris et al., 1991], the S11-4-DGS synthetic ferrihydrite sample was prepared by D.G. Schulze, and the S6FN28 nanophase iron oxide sample was produced in the lab by calcining silica gel impregnated with ferric nitrate solutions in air at 550°C [Morris et al., 1989]. The transmission spectrum of S6FN28 is shown along with a spectrum of the silica gel material alone that has been offset by 30% relative to S6FN28. Data are in frequency on the left and wavelength on the right.
Figure 4. Transmission spectra of Mars-analog palagonites from the island of Hawaii compared to mid-IR Mars spectra from Mariner 7, Mariner 9, and the KAO. PH-1 < 20 μm is a spectrum of the finest siliceous fraction of a naturally occurring thermally altered palagonite tephra from Mauna Kea (Bell et al., 1993). Spectrum 91-1 is from a fine fraction (< 38 μm) sample of Pahala Ash collected near South Point and measured by Rose et al. [1992]. The two other Pahala Ash spectra are from a different sample of the Pahala Ash unit than 91-1 and were measured by Crisp and Barabashov [1992]. One of the spectra is from a subsample of Pahala Ash dried in the laboratory. The Mars spectra are plotted as brightness temperature, and representative spectra are shown from Mariner 9 IRES [Hanel et al., 1972]. Mariner 7 IRES (T.Z. Martin, Mariner 6/7 infrared spectrometer: Data set restoration, submitted to Journal of Geophysical Research, 1991), and 1988 Kuiper Airborne Observatory observations [Pohlack et al., 1990]. Data are in frequency on the left and wavelength on the right.
that will certainly be evident in real Martian data [see Salisbury et al., 1991; Salisbury and Wald, 1992; Nash et al., 1993]. However, this study has demonstrated that both well-crystallized and nanophase ferric oxides and ferric oxyhydroxides exhibit complex spectral features at mid-IR wavelengths. Well-crystallized ferric oxides/oxyhydroxides have strong, well-resolved absorption features primarily in the 400 to 1200 cm⁻¹ (0.3 to 25 μm) region. Analysis by Roush and Bell [this issue] of the Salisbury et al. [1991] mid-IR reflection spectra of crystalline hematite, goethite, and magnetite and the assumption of Kirchhoff's law (emissivity = 1 − reflectivity) indicates that the absorption features seen in our transmission spectra of well-crystallized samples should be detectable as high-contrast features in remotely sensed Mars emissivity spectra. The nanophase materials and palagonites, however, typically show much weaker mid-IR absorption features. By inference then, we can expect that emissivity features in the spectra of these poorly-crystallized materials both in the lab and potentially in the Martian environment will be subdued relative to those seen in the crystalline materials. In fact, the subtle nature of emissivity features in palagonite samples PH-1 and 9-1-1 has been demonstrated by Roush and Bell [this issue]. Thus, our transmission measurements provide a starting point for understanding how to interpret one aspect of existing and future Mars mid-IR observations in that they provide information on the variations in absorption band positions of ferric iron bearing materials as a function of composition and structural environment.

Figures 1-4 clearly indicate that ferric-bearing materials spanning a wide range of mineralogy and crystallinity exhibit many strong diagnostic absorption features in the mid-IR. To date, no conclusive evidence exists for the presence of any of the Fe⁶⁺-O⁻² feature in ground-based or spacecraft Mars data sets (Figures 4). Hane et al. [1995] conducted a preliminary examination of the Mariner 7 and Mariner 9 spectra and included hematite and goethite in their spectral searches. They favored montmorillonite as explaining the 9.5- and 21-μm features seen in the Mars data, although the presence of kaolinite and, to a lesser extent, hematite and goethite, could not be completely excluded. Torn et al. [1977] searched many of the Mariner 9 dust storm spectra for evidence of ferric oxides or oxyhydroxides. They found no conclusive evidence for hematite absorption features below 600 cm⁻¹, but were not able to rule out the possible occurrence of substantial amounts of goethite either mixed with or coating other materials. It is important to consider how the Martian atmosphere could potentially interfere with the interpretation of ferric mineral bands in mid-IR remote sensing observations. Radiative transfer models show that CO₂ absorbs all of the major mid-IR absorption features in the Martian atmosphere [e.g., Crisp, 1990]. Strong CO₂ features occur near 550 to 750 cm⁻¹, 2000 to 2300 cm⁻¹, and 3400 cm⁻¹. The two higher-frequency bands are not likely to cause problems in the detection of ferric materials, but the lowest frequency band (15 μm) will be a concern because of overlap with Fe⁶⁺-O⁻² vibrational transitions (Table 1; Figure 4). Because these absorption features occur primarily in the unattenuated wing at frequencies lower than the center of the CO₂ band, however, it will most likely be possible to model the predicted atmospheric spectral response and remove or at least unambiguously identify such features in spectra of surface materials and atmospheric dust [e.g., Toon et al., 1977; Pollack et al., 1990]. But because of CO₂ band saturation, it is unlikely that any useful surface spectral signatures could be detected from 630 to 700 cm⁻¹, except possibly from surface landers.

Experience with high spatial and spectral resolution ground-based and spacecraft visible to near-IR spectroscopic observations indicates that increasing the spatial and spectral resolution allows for more detailed characterization of spectral heterogeneity on Mars [e.g., Bell et al., 1990; Marzilli et al., 1993; Mustard and Bell, 1994]. Ferric minerals are an obvious and important component of the optical surface of Mars, and high spatial and spectral resolution ground-based or spacecraft mid-IR surveys like those of Moersch et al. [1993] or of the Mars Observer/Surveyor TES instrument [Christensen et al., 1992] will have a much higher chance of being able to detect and spectrally characterize the Martian surface. An understanding of the spectral contribution of ferric-bearing materials in the mid-IR will be an important aspect of the analysis and interpretation of these new data sets.

Conclusions

Our results suggest that high spatial resolution, moderate spectral resolution thermal infrared observations of Mars may provide significant insight into the ferric mineralogy of that planet. The subdued nature of the spectral features in the poorly-crystallized samples indicates that any future mid-IR remote sensing investigations must be able to achieve a high signal-to-noise ratio in order to detect and characterize such materials. Continued laboratory study of the reflectance and emissivity features of ferric-bearing materials as a function of composition and structural environment. In fact, the subtle nature of emissivity features in the spectra of these poorly-crystallized materials both in the lab and potentially in the Martian environment will be subdued relative to those seen in the crystalline materials. In fact, the subtle nature of emissivity features in the spectra of these poorly-crystallized materials both in the lab and potentially in the Martian environment will be subdued relative to those seen in the crystalline materials.

Acknowledgments

This research was supported by grants from the National Research Council/National Academy of Sciences and the NASA Planetary Geology Program. We thank Robert Bohn, Scott Sandford and Louis Allenwealth for assistance with and access to the NASA Ames Astrochemical Laboratory FTIR spectrometer facility. We also thank Joe Crisp for graciously providing digital versions of the Palos Ash palagonite transmission spectra used in the Palaos Palagonite Analysis Project. We also thank Paul Lacey, and Jack Salisbury for constructive reviews of an earlier draft of this paper. We note that the new transmission data presented here are available digitally to interested researchers. Please send an electronic mail request for more information to the first author.

References


Thermal emission measurements 2000–400 cm⁻¹ (5–25 μm) of Hawaiian palagonitic soils and their implications for Mars

Ted L. Roush
Department of Geosciences, San Francisco State University, San Francisco, California

James F. Bell III
National Research Council, NASA Ames Research Center, Moffett Field, California

Abstract. The thermal emission of two palagonitic soils, common visible and near infrared spectral analogs for bright soils on Mars, was measured over the wavelength range of 5 to 25 μm (2000 to 400 cm⁻¹) for several particle size separates. All spectra exhibit emissivity features due to vibrations associated with H₂O and SiO. The maximum variability of emissivity is ~20% in the short wavelength region (5 to 6.5 μm, 2000 to 1500 cm⁻¹), and is more subdued, <4%, at longer wavelengths. The strengths of features present in infrared spectra of Mars cannot be solely provided by emissivity variations of palagonite; some other material or mechanism must provide additional absorption(s).

Introduction and Summary

Insight into the availability of liquid water, and the duration, mode, and extent of weathering occurring throughout Mars’ history can be provided by understanding the mineralogy of the surface materials on Mars. Si and Fe are the two most abundant elements on Mars [Toon et al., 1977; Clark et al., 1977, 1982], but how these are combined into surface minerals has not been directly determined. The mineralogy is inferred from in situ analyses [Toon et al., 1977; Clark et al., 1977, 1982; Hargraves et al., 1977, 1979; Oyama et al., 1978; Basu and Riehdon, 1978, 1979; Huguenin, 1982; Quinn and Orenberg, 1993], and spectroscopic observations [reviewed in Soderblom (1992) and Roush et al. (1993), also see Pollack et al. (1977), Adams et al. (1985), Erard et al. (1991), Bell (1992), Marrard et al. (1993), and Mauaud and Bell (1994)].

Interpretations of spectroscopic observations of Mars reveal that the ferric mineralogy occurs in two distinct forms: (1) nanophase or truly nanophase Fe³⁺-bearing materials [Morris and Lauer, 1990; Morris et al., 1989] that spectrally resemble certain terrestrial palagonites [e.g., Singer et al., 1979; Evans and Adams, 1980; Singer, 1982; Morris et al., 1990; Bell et al., 1995; Roush et al., 1993]; and (2) well-crystalline ferric oxides [Pollack et al., 1977; Morris et al., 1985, 1989; Bell et al., 1990; Bell, 1992]. The available data indicate that the poorly-crystalline "palagonite-like" phases are spectrally dominant [e.g., Hunt et al., 1973; Toon et al., 1977; Soderblom, 1992; Roush et al., 1993] and that the well-crystalline ferric oxides cannot constitute an abundance of more than about 4 wt % [Bell, 1992]. Thus palagonites are commonly used as visible and near-infrared spectral analogs for bright soils on Mars. Hawaiian palagonitic soils contain abundant Si and Fe, and laboratory studies have revealed nanophase iron oxides as the coloring agent in certain samples [Morris et al., 1990; Bell et al., 1993; Golden et al., 1993].

Only transmission [Roush, 1992; Roush et al., 1993] and reflectance [Brucinostial, 1987; Roush et al., 1988; Crisp and Barthelemy, 1992; Bishop et al., 1993] measurements of palagonites have been reported in the mid- to far-infrared. Recently sensed thermal infrared data measure the energy emitted from planetary surfaces. For ideal conditions [see Christensen and Hartmann, 1993], when all emitted and reflected energy is measured, emittance is related to reflectance via Kirchhoff's law (emittance = 1 - reflectance). Such conditions are rarely met in nature and only occasionally in the laboratory.

On Mars the spectrally dominant phase resembles terrestrial palagonites hence, it is important to understand the mid- and far-infrared emissivity of such materials. We report the measured emissivity of a Mauna Kea palagonitic soil whose transmission spectra have been previously reported [Roush, 1992] and of a thermally altered volcanic tephra [Bell et al., 1993]. Both samples represent analogs for formation mechanisms capable of producing highly altered secondary weathering products on Mars. Such information can be used for interpretation of past (e.g., Mariner 6, 7, and 9) and future spacecraft (e.g., Mars Global Surveyor, MGS), as well as containing Earth-based [e.g., Pollack et al., 1990; Roush et al., 1991, 1992; Moersch et al., 1993], thermal infrared observations of Mars.

The measured thermal emission of the two palagonites exhibits complex spectral behavior within the range of 2000 to 400 cm⁻¹ (5 to 25 μm). Emissivity contrasts are subdued...
have been previously studied using reflectance spectroscopy. The prominent spectral contrast is observed for all samples and particle sizes in the 1250 to 2000 cm$^{-1}$ (5 to 8 $\mu$m) region with contrasts of up to 20% for the finer particle sizes.

All emissivity spectra exhibit features due to vibrations associated with molecular H$_2$O. Features associated with the Si-O Christianson frequency (CP) suggest that the samples studied here are chemically and mineralogically different. The observed behavior of the CP suggests some compositional segregation is associated with particle size separation. All spectra exhibit subtle emissivity features that may be associated with ferric oxides and/or ferric hydroxides. However, more definitive assignments must await emissivity measurements of fine-grained samples of these types of materials. Whether or not nanophase ferric oxides and/or ferric hydroxides are consistent with the spectral data also must await measurements of these materials.

Comparison of the palagonite emissivities to infrared measurements of Mars indicate that H$_2$O features are less obvious than those of terrestrial materials, supporting previous suggestions of dehydrated Martian surface materials [see Soderblom, 1992; Roush et al., 1993]. This suggests that mid-IR spectra can be used for readily identifying the presence of H$_2$O-bearing regions on Mars. The palagonites are capable of producing spectral features, associated with Si-O and Fe-O, at frequencies ($\nu$) consistent with the spacecraft data but cannot produce the magnitude of the features seen in the observational data. This suggests that (1) if materials similar to these palagonites are a major surface component on Mars, then high-quality (signal-to-noise) thermal observations are required to assess the relatively weak emissivity contrasts expected; and (2) the observed spectral variations of Mars are likely due to contributions of other materials or mechanisms, such as atmospheric aerosol and gas absorption. The implication for Mars is that compositional variability between similar materials is detectable and could be mapped on a global scale.

### Experiment

**Sample Description and Preparation**

Palagonitic soil. Palagonitic soil 91-1 [Macdonald et al., 1983] and was collected near South Point on the island of Hawaii. Samples from this site have been previously studied using reflectance spectroscopy [Singer, 1982; Brackenrath, 1987; Crisp and Bartholomew, 1992]. After the sample was dried at room temperature, the bulk sample passing through a 2-mm sieve was retained. Elemental abundances of the bulk sample, determined by X-ray fluorescence analysis, are provided in Table 1 along with the hygroscopic moisture percentage and the specific surface area. Subsequently, the sample was dry sieved into eight particle size classes using 0.02-, 0.045-, 0.09-, 0.15-, 0.25-, 0.5-, 1-, and 2-mm sieves. A histogram of the relative weight percentage of each particle size fraction is shown in Figure 1.

**Volcanic tephra.** Volcanic tephra PHI-1 from the Puu Huluhulu (Mauna Kea) cinder cone on the island of Hawaii has been previously studied in the visible and near-infrared and is a sample thermally altered by local magmatic activity [Bell et al., 1993]. The sample was wet sieved using from seven size fractions using 0.02-, 0.045-, 0.09-, 0.15-, 0.25-, 0.5-, and 1-mm sieves. The relative abundance of each particle size fraction are shown in Figure 1. Fe and water concentrations were determined by instrumental neutron activation analysis (INAA) and thermogravimetric analysis (TGA) and are listed in Table 1. Additional analytical techniques have been used to characterize the mineralogy of this sample [Bell et al., 1993].

**Thermal Emission Measurements and Calculations**

Emissivity spectra were acquired at the Arizona State University Thermal Emission Laboratory [Anderson et al., 1991]. Each particle size separate was poured into a copper sample cup and the powder was leveled with a teflon-coated laboratory spatula. The cup was placed in an oven and...
heated at 80°C overnight, a procedure that likely removes loosely bound water but at these low temperatures tightly bound and/or chemically adsorbed water should be retained [e.g., Bruckenthal, 1987]. The samples were transferred to the emission chamber, and particle size separates of sample 91-1 and PH-1 were heated at 70.0 ± 0.2°C and 80.0 ± 0.2°C, respectively, during data acquisition. Spectra were also acquired of a blackbody source at liquid nitrogen temperature and approximately 80°C. From these measurements the sample emissivity can be derived [Christewen and Harrison, 1993]. Each spectrum consists of 200 scans, obtained at a constant resolution of 4 cm⁻¹, over the entire spectral domain using a Fourier transform spectrometer. The rms peak-to-peak variability in the derived emissivity is ±4-6% in the 2000-1400 cm⁻¹ region and improves to ±0.4-0.6% at <1400 cm⁻¹, as illustrated in Figure 2.

To more readily identify emissivity features, we smoothed the original 4 cm⁻¹ data using a Gaussian filter deconvolution program with a distance weighting scheme and a 3-σ despiking algorithm. Figure 2 shows the results for various Gaussian half-widths. We have used a five-channel half-width at half maximum convolution on all subsequent data presented here because this compromise retains the overall broad features while eliminating high-frequency peak-to-peak variations. Figure 3 illustrates the emissivity spectra of all particle size separates of sample 91-1 and all spectra exhibit maxima at 1980, 1870, 1650, 1520, ~1200, and 675 cm⁻¹. Figure 4 illustrates the emissivity spectra of all particle size separates of sample PH-1 and in this case maxima occur near 1950, 1850, 1640, 1550, 1520, between 1250 and 1300, and 675 cm⁻¹.

As seen in Figures 3 and 4, the maximum variability of emissivity, ~10-20%, is observed at the highest v, 1500 to 2000 cm⁻¹, and is most pronounced for the finest grain sizes. At lower v the emissivity variation is subdued, typically <4%, but again, more pronounced in the finest grain sizes. The subdued emissivities observed at low v is consistent with model calculations of the palagonite emissivity [Pollock et al., 1990, Figure 12a] and the emissivity of Pahala Ash calculated from reflectance [Crisp and Bartholomew, 1992, Figures 4 and 5].

Discussion

Spectral Features

Emissivity maxima near 675 cm⁻¹ of both samples (see tick marks in Figures 3 and 6) are most likely associated with the Si-O bending fundamentals of silicate minerals and rocks [Farmer, 1974; Solomon et al., 1991; Nash et al., 1993], although some ferric oxides and oxyhydroxides exhibit features near this v [Bell et al., this issue] and identify only poorly defined corresponding features in the emissivity spectra near 460 and 550 cm⁻¹ (see tick marks in Figures 5 and 6). As indicated above, metastable iron oxides appear to be the coloring agents in certain palagonites [Morrison et al., 1990; Bell et al., 1993; Golden et al., 1993]. Bell et al. [this issue] show the transmission spectra of several metastable iron oxides. The spectra of these materials exhibit absorption minima at v similar to those of the crystalline iron oxides and oxyhydroxides. However, the
features are more subdued for the nanophase iron oxides when compared to the crystalline materials. Thus, by analogy to the crystalline materials, we would suspect that the nanophase iron oxides would produce more subdued features when measured in emission. This conclusion would be consistent with nanophase oxides being the coloring agent in the palagonites studied here.

Emissivity maxima near 1200 cm\(^{-1}\) (91-1) and between 1250 and 1300 cm\(^{-1}\) (PH-1) (see pointers on Figures 5 and 6) most likely correspond to the CF commonly associated with the silicate minerals and rocks. The CF position has been related to the polymerization and geometry of Si-tetrahedra and is observed at \(\nu > v\) for highly polymerized silicates than for poorly polymerized silicates (Cowen, 1969; Logan et al., 1973; Walter and Salisbury, 1989; Lucas, 1991; Nash et al., 1993). Based on these previous studies, the apparent difference of CF seen in the two samples here suggests that the Si-bearing component in 91-1 is less polymerized than in PH-1. The CF of both samples appear to shift to \(\nu > v\) with decreasing particle size (see Figures 5 and 6). Logan et al. [1973] reported a CF shift to \(\nu < v\) with decreasing particle size for individual minerals, yet no CF shift is seen in the spectra of Salisbury and Wald [1992].
The fine fraction of PH-1 is depleted in poorly polymerized silicates [Bell et al., 1993]. Thus the CF shift to >v with decreasing particle size is consistent with compositional rather than particle size changes. Although more complete analysis of 91-1 is required, we speculate that compositional rather than grain size effects are responsible for the observed CF shift.

Figure 5. Enlargement of the emissivity, in the 400 to 1650 cm\(^{-1}\) spectral domain, of sample 91-1 for grain sizes of (a) 1-2 mm, (b) 0.5-1 mm, (c) 0.25-0.5 mm, (d) 0.15-0.25 mm, (e) 90-100 \(\mu\)m, (f) 45-90 \(\mu\)m, (g) 20-45 \(\mu\)m, and (h) <20 \(\mu\)m. Each large tick on the vertical axis is 0.05, and each small tick is 0.01 units in emissivity. Vertical tick marks indicate spectral features discussed in the text, and inverted triangles indicate the position of the local maxima associated with the CF as discussed in the text.

Figure 6. Enlargement of the emissivity in the 400 to 1650 cm\(^{-1}\) spectral domain, of sample PH-1 for grain sizes of (a) 0.5-1 mm, (b) 0.25-0.5 mm, (c) 0.15-0.25 mm, (d) 90-100 \(\mu\)m, (e) 45-90 \(\mu\)m, (f) 20-45 \(\mu\)m, and (g) <20 \(\mu\)m. Each tick on the vertical axis is 0.01 units in emissivity. Vertical tick marks indicate spectral features discussed in the text, and inverted triangles indicate the position of the local maxima associated with the CF as discussed in the text.

The spectral features in the 1380 to 1560 cm\(^{-1}\) region (see tick marks in Figures 5 and 6) occur where carbonates, nitrates, and hydrocarbons all have fundamental modes. More detailed chemical and mineralogical analyses of these samples are required before a specific assignment can be suggested.

The 1650 (91-1) and 1640 (PH-1) cm\(^{-1}\) emissivity maxima (see tick marks in Figures 3 and 4) occur at v typical of the H-O-H bending fundamental, and Table 1 indicates water is associated with these samples. The presence of emissivity maxima in all spectra suggests that water is present in all grain sizes and may be enhanced in the finest grain sizes, as
indicated by an emissivity decreases $\gamma \geq 1650 \text{ cm}^{-1}$. This is consistent with the reported inverse correlation of water content and grain size for PH-1 [Bell et al., 1993]. Although Mars is a desiccated environment compared to the Earth, hydrous materials have been suggested as surface and atmospheric aerosol constituents based on observed spectral features [see Soderblom [1992] and Roush et al. [1993] for more details, and also Enard et al. 1991; and Marshie et al., 1993].

We provide no specific assignment for the emissivity maxima located in the 1500-1850 and 1850-1870 cm$^{-1}$ regions (see tick marks in Figures 3 and 4). Reflectance spectra of many silicates exhibit features at similar $\gamma$ to those observed here [Sallisbury; et al., 1991; Nash et al., 1993]. Sallisbury and Walter [1989] suggested such features represent overtone and/or combination modes of the lower $\gamma$ fundamental modes. Figures 5 and 6 are an expanded view of the emissivity in the 400-1650 cm$^{-1}$ region for 91-1 and PH-1, respectively, and on such figure are dotted lines at the predicted fundamental $\gamma$. There appear to be very weak emissivity maxima present in the 91-1 spectra at the predicted $\gamma$ but the correspondence for the PH-1 spectra is absent.

Comparison with Martian Data

There have been a variety of spectral observations of Mars in the infrared from spacecraft, Earth-based telescopes, and airborne telescopes [see Soderblom [1992] and Roush et al. [1993] for recent reviews]. All of these observations have measured the total flux from Mars, including (1) flux emitted from the surface; and (2) absorption and scattering due to Martian atmospheric aerosols and gases. Quantitative comparisons of these observations to the measured emissivities are beyond the scope of this paper. Here we provide a qualitative comparison to the observational data. As a result, we address whether the measured emissivity of the palagonites studied here produce spectral features at $\gamma$ consistent with similar features seen in the observational data, and whether or not surface emissivity alone can produce the intensity of features seen in the observational data.

Spectra of the Martian surface (5269-700 cm$^{-1}$) 1.0-14.1 um) were acquired during the 1969 fly-by of Mars by Mariner Mars 7 (MM7) [Herr et al., 1972] (also T. Z. Martin, Mariner 6/7 Infrared Spectrometer: Data set restoration, submitted to Journal of Geophysical Research, 1991) for a small region of the planet with the best spatial resolution sampling $\approx 255$ km$^2$ areas. Thousands of Martian spectra (2000-2000 cm$^{-1}$, 5-50 km$^2$) were acquired during the 1971-1972 Mariner Mars 9 (MM9) orbital mission [Hanel et al., 1972], providing global coverage, but the best spatial resolution came from areas sampling $\approx 10^8$ km$^2$. In 1988 infrared observations of Mars were obtained from the Kuiper Airborne Observatory (KAO) (1850-950 cm$^{-1}$, 5.4-10.5 km$^2$) [Roush et al., 1989; Pollack et al., 1990] for limited regions of the surface and sampled areas of $\approx 4 \times 10^8$ km$^2$. Figures 7, 8, and 9 compare the scaled emissivity of the coarsest and finest palagonite samples to a spectrum of a Martian bright region obtained by MM7, MM5, and KAO, respectively.

The MM7, MM5, and KAO spectra are all flat in the 1600 to 2000 cm$^{-1}$ region, unlike either palagonite sample. This suggests compositional differences between the palagonite samples and the materials comprising the surface and/or atmospheric dust of Mars. Because we believe that the $\approx 1650$ cm$^{-1}$ band indicates water associated with the palagonites, this difference may be related to the relative hydration state of the Martian surface materials. If this is the case, then the weaker features in the observational data suggest that the Martian surface materials are less hydrous than the palagonites studied here, and is consistent with previous interpretations, based on near- and mid-infrared spectra [see Soderblom 1992; Roush et al., 1993].

The observational spectra all exhibit a minimum centered near 1050-1075 cm$^{-1}$ with relative depths ranging from $\approx 1\%$ to $10\%$. Additionally, the MM9 spectra have a minimum near 670 cm$^{-1}$, with relative depths varying from $\approx 1\%$ to $6\%$. Based on analysis of the MM9 data by Toon et al. [1977] both of these band depths are variable and a function of emission angle. The three spectra shown by Toon et al. [1977, Figure 9] illustrates that the band depths are correlated with the atmospheric dust loading. As shown in Figure 10, Martian atmospheric gases alone exhibits
spectra do not appear capable of engendering the strength of the feature seen in the observational spectra. Some other material(s) (aerosols or gases) or mechanism(s) contribute to these features. For example, Pollack et al. [1990] reproduced the strength of the 1075 cm\(^{-1}\) feature in the KAO data using models that included contributions due to atmospheric dust and CO\(_2\) opacity.

**Significance to Future Remote Sensing Observations of Mars**

The MGS is intended to carry a duplicate of the ill-fated Mars Observer thermal emission spectrometer, which was designed to return data in the 1600-2000 cm\(^{-1}\) spectral domain and provide global coverage of Mars at a spatial resolution of 3x3 km [Christensen et al., 1992], representing an increase of \(\approx 2\) orders of magnitude in spatial resolution.

---

Figure 8. Comparison of Mariner 9 infrared intercon- 667 ter spectrometer (IRIS) spectrum of a bright region on Mars, -29.8° latitude, 306.4° longitude (solid line), obtained during heavy atmospheric dust loading, to 91-1 (dotted line), scaled to 270°K, and PH-1 (dashed line), scaled to 280°K, for both (a) coarse and (b) fine grain size samples.

Wavelength (\(\mu m\))

<table>
<thead>
<tr>
<th>Wavelength ((\mu m))</th>
<th>25.0</th>
<th>12.5</th>
<th>8.33</th>
<th>6.25</th>
<th>5.0</th>
</tr>
</thead>
</table>

---

Figure 9. Comparison of Keeler Airborne Observatory spectrum of Mars (solid line), -24° latitude, 132° longitude, to 91-1 (dotted line), scaled to 270°K, and PH-1 (dashed line), scaled to 280°K, for both (a) coarse and (b) fine grain size samples.
Figure 10. One-way transmission of the Martian atmosphere for nominal surface pressure (6.1 mbar) that includes absorptions due to CO$_2$, CO, H$_2$O, O$_3$, and O$_2$ gases but does not include any atmospheric dust absorption (after Crisp, 1990).

Figure 10. One-way transmission of the Martian atmosphere for nominal surface pressure (6.1 mbar) that includes absorptions due to CO$_2$, CO, H$_2$O, O$_3$, and O$_2$ gases but does not include any atmospheric dust absorption (after Crisp, 1990).

over existing data. Analyses of recent spectral imaging of Mars [Bell et al., 1990; Bilting et al., 1990; Endr et al., 1991; Bell, 1992; Mustard et al., 1995; Murchie et al., 1995; Mustard and Bell, 1994] illustrate how increasing spatial resolution can reveal compositional differences in areas once believed homogeneous. If MGS can achieve the same nominal spatial resolution, then our knowledge regarding the surface composition of Mars will surely improve.

Acknowledgments. This research was supported by the Mars Observer Participating Scientist Program and NASA's Planetary Geology and Geophysics Program via RTOP 151-01-60-01. We extend our gratitude to Phil Christensen, Paul Carpenter, P.R. Christensen, and P.W. Bell for their advice regarding the surface composition of Mars. We thank Wendy Calvin for her review that helped improve the manuscript.

References


ROUSH AND BELL: THERMAL EMISSION OF HAWAIIAN PALAGONITIC SOILS


Pollack, J.B., T.L. Roush, P. Witthoeft, J. Bregman, D. Wood, C. Bishop, G. Toon, D. Balcz, and R. Friedman, Thermal emission spectra of Mars (5.4-10.5 μm); Evidence for sulfates, carbonates, and hydrates, J. Geophys. Res., 95, 14,625-14,627, 1990.


J.P. Bell and T.L. Roush, NASA Ames Research Center, MS 245-3, Moffett Field, CA 94035-1000. (e-mail: roush@lanslom.arc.nasa.gov)

(Received March 1, 1994; revised August 23, 1994; accepted September 19, 1994.)
Page intentionally left blank
Hematite, pyroxene, and phyllosilicates on Mars:
Implications from oxidized impact melt rocks
from Manicouagan Crater, Quebec, Canada

Richard V. Morris, 1 D. C. Golden, 2 James F. Bell III, 3 and H. V. Lauer Jr. 4

Abstract. Visible and near-IR reflectivity, Mössbauer, and X-ray diffraction data were obtained
on powders of impact melt rock from the Manicouagan Impact Crater located in Quebec, Canada.
The iron mineralogy is dominated by pyroxene for the least oxidized samples and by hematite for
the most oxidized samples. Phyllosilicate (smectite) contents up to ~15 wt% were found in some
heavily oxidized samples. Nanophase hematite and/or paramagnetic ferrihydrite is observed in all
samples. No hydrous ferric oxides (e.g., goethite, lepidocrocite, and ferrihydrite) were detected,
which implies the alteration occurred above 250°C. Oxidative alteration is thought to have
occurred predominantly during late-stage crystallization and subsolidus cooling of the impact melt
by invasion of oxidizing vapors and/or solutions while the impact melt rocks were still hot. The
near-IR band minimum correlated with the extent of alteration.

Introduction

Oxidative alteration of impact melt sheets is reported at many terrestrial impact structures, including Manicouagan
(Phinney et al., 1978) and West Clearwater Lake (Phinney et al., 1978), both in Quebec, Canada, and the Ries basin (Pohl et al.,
1977; Newesey et al., 1986) in Germany. Oxidation apparently occurs shortly after the impact by an influx of oxidizing vapors
and/or fluids while the rocks are still hot but below solidus temperatures (Phinney et al., 1978). Hematite is reported as an
oxidative product at all those impact structures (Phinney et al., 1978; Phinney et al., 1978; Pohl et al., 1977). Because
meteoritic impact also occurs under oxidizing surface conditions on Mars, a number of studies (e.g., Newsom, 1980, 1986;
Kieffer and Simonds, 1980; Allen et al., 1982; Clifford, 1986; Elsener et al., 1989) have advocated that a significant fraction
of Martian soil may consist of eolian samples of hydrothermally altered (oxidized) impact melt sheets. Because of the identification of
hematite in visible and near-IR spectral data of Martian bright regions (e.g., Morris et al., 1990; Bell et al., 1990; Morris and
Lauer, 1990; Morris et al., 1993), this view is at least consistent with the above studies of terrestrial impact melts. The
mineral assemblages in terrestrial impact melt rocks that have undergone oxidative alteration are thus putative mineral
assemblages for the Martian surface, and their optical properties relevant to the mineralogic interpretation of
Martian spectral data. We report here compositional, Mössbauer, and spectral data (visible and near-IR) for powders
of impact melt rocks from Manicouagan Crater.

Sample and Methods

Samples. We studied 23 samples of powdered coherent rock from the Manicouagan impact melt sheet. All rock
samples were powdered in an alumina mortar and passed through a 50 µm sieve. Floran et al. (1976, 1978) report
chemical and petrographic data for 16 of these samples.

Analytical methods. A Cary-14 spectrophotometer configured with a 14-cm-diameter integrating sphere was used to
obtain visible and near-IR reflectance spectra. Elsener and Ranger Mössbauer spectrometers using 57Co(Rh) sources were
used to obtain iron Mössbauer spectra (Fe57Mö). Mössbauer spectra were fit to theoretical line shapes using an in-house

Copyright 1995 by the American Geophysical Union.
Paper number 94011500.
0148-0227(95)01500-05.

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 100, NO. E3, PAGES 3316-3338, MARCH 25, 1995
computer program (JSCFIT). Saturation magnetizations (Ms) were calculated from magnetization data obtained on a Princeton Applied Research model 151 vibrating sample magnetometer. Concentrations of major (Fe, Ca, Na, and K) and minor and trace (Sc, Cr, Rh, St, Ba, and rare-earth elements) elements were obtained by instrumental neutron activation analysis (INAA). A Scintag XDS 200 X-ray diffractometer using CuKα radiation was employed to obtain powder X-ray diffraction (XRD) patterns. Mg-saturated, oriented samples of clay-size fractions (<2 µm obtained by gravity sedimentation of an aqueous dispersion) with and without glycol solvation were used to determine if phyllosilicates (and smectite in particular) are present. More detailed information on analytical procedures is given by Morris et al. [1989, 1990], Golden et al. [1993], and Bell et al. [1993].

Results and Discussion

Mössbauer Mineralogy

Iron Mössbauer spectroscopy (FeMS), through the positions of spectral lines, is sensitive to both the oxidation state and mineralogy of iron. The relative intensities of spectral lines provide quantitative information about the relative distribution of iron among its oxidation states and iron-bearing mineralogies. Because the technique is sensitive only to iron (specifically the isotope 57Fe which has 2.2% natural abundance) and thus blind to iron-free phases (e.g., feldspar and quartz), the absolute modal abundance of iron-bearing phases cannot be determined. However, if the concentration of iron in the iron-bearing phases can be determined or estimated (e.g., from the mineralogical identification), the relative modal proportions of iron-bearing phases can be estimated. FeMS spectra (293 K) for samples having the highest and lowest proportions of ferric iron are shown in Figure 1. A total of nine components, not all present in all samples, were needed to fit these spectra. The Mössbauer parameters for the nine components are compiled in Table 1. These parameters are values averaged over all spectra where they could be determined without constraining peak positions. The number of spectra used to calculate averages and the maximum relative area observed for each component are also given. The mineralogy associated with six of the nine components is obvious from inspection of the spectra and values of their Mössbauer parameters. The hematite sextet (well-crystalline and unsplit) associated with these three components is not readily assignable just from positions of spectral lines.

Mössbauer spectra for Manicouagan impact melt rocks having lowest and highest values of Fe3+/Feo,.

In order to constrain mineralogical assignments for Fe3D2, Fe3D3, and Fe3D4, we outlined MAN-74-342A and 608A in air at 400, 500, 800, and 1000°C. For MAN-74-342A, Fe3D2 and Fe3D3 are not present by 400 and 800°C, respectively. For MAN-74-608A, magnetite and pyroxene are not present by 500 and 1000°C, respectively. The 1000°C spectra for both samples are shown in Figure 2. Both are characterized by a hematite sextet and an asymmetric ferrous doublet (Fe3D2). These spectra are very similar to that for MAN-74-177, and their quadrupole splitting is larger than that for Fe3D3. The mineralogy associated with these three components is not readily asssigned just from positions of spectral lines.
present in our highly altered samples, we assign FeOFe and Fe$_3$O$_4$ to phyllosilicates (smectite).

Possible assignments for Fe$_3$O$_4$ include pseudobrookite solid solutions from oxidation of U(III) in nanophase hematite (np-Hm) from oxidation/decomposition of silicate phases including mafic minerals [Kroub et al., 1991] and phyllosilicates [Waldon et al., 1982; Moskowitz et al. and Hargraves, 1982], and/or peramagnetic ferric iron (para-Fe$_3^+$) in structural sites of silicate/aluminosilicate phases. The asymmetry of the doublet suggests contributions from more than one source. Np-Hm is probably a major contributor to Fe$_3$O$_4$ because a sharp band near 450 nm, which would be indicative of Fe$_3^+$ at well-defined structural sites, is not present. Np-Hm does not have well-defined bands [Morris et al., 1989; Morris and Lauer, 1990]. Although the mineralogical assignments made above for Fe$_3$O$_4$, Fe$_3$O$_6$, and Fe$_3$O$_8$ are not definitive, we can reasonably rule out other potential mineralogies. Lepidocrocite (p-FeO(OH)) and superparamagnetic goethite (o-FeO(OH)) are not reasonable assignments for the ferric doublets because the phases have decomposition temperatures below 400°C [e.g., Morris and Lauer, 1981]. The oxides from crystalline goethite are also not present in any sample. The Mössbauer parameters for siderite (FeCO$_3$) are also not consistent with the data. The low sulfur content of these samples (<20 wt% [Floran et al., 1978]) implies that jarosites, sulfates, low-spin ferrous sulfides (e.g., FeS$_2$), and other minerals that contain iron and sulfur are not present in significant amounts.

Relative component areas for each sample are given in Table 2. These areas include a correction for different recoilless fractions for ferrous and ferric iron. We used a value of 1.21, which is the average value we calculated from the data of De Grave and Van Alboom [1991], as the ferric to ferrous recoilless fraction ratio [De Grave and Van Alboom, 1991].

Examination of the data in Table 2 shows that sample saturation magnetization ($I_s$) correlates with the relative area of magnetite (or titanomagnetite), so that its oxidation product is weakly magnetic hematite (or titanohematite) rather than strongly magnetic magnetites (or titanomagnetics). It may be that the relatively high temperatures for oxidative alteration at Manicouagan favored alteration of (titanomagnetics) to hematites rather than to magnetites. As reported by Allègre et

### Table 1. Average Mössbauer Parameters (293 K) for the Nine Components, Number of Samples Used for Averages, and Maximum Area Observed for Component

<table>
<thead>
<tr>
<th>Component</th>
<th>IS*</th>
<th>QS*</th>
<th>B*</th>
<th>Number of Samples</th>
<th>Maximum Area, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe$_3$O$_4$</td>
<td>0.38</td>
<td>0.00</td>
<td>51.6</td>
<td>14</td>
<td>67</td>
</tr>
<tr>
<td>Fe$_3$O$_6$</td>
<td>0.29</td>
<td>0.01</td>
<td>49.0</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Fe$_3$O$_4$</td>
<td>0.70</td>
<td>0.00</td>
<td>45.7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Fe$_3$O$_6$</td>
<td>0.38</td>
<td>0.70</td>
<td>12</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>Fe$_3$O$_4$</td>
<td>0.40</td>
<td>0.37</td>
<td>25</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Fe$_3$O$_6$</td>
<td>1.13</td>
<td>0.60</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Fe$_3$O$_6$</td>
<td>1.12</td>
<td>2.06</td>
<td>19</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Fe$_3$O$_6$</td>
<td>1.13</td>
<td>2.57</td>
<td>16</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Fe$_3$O$_6$</td>
<td>1.07</td>
<td>0.67</td>
<td>22</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

*IS = isomer shift; QS = quadrupole splitting; B* = hyperfine field. Isomer shifts are reported relative to Fe metal at room temperature.

*Calculated using a value of 1.21 as the ferric to ferrous recoilless fraction ratio [De Grave and Van Alboom, 1991].

Figure 2. Mössbauer spectra of MAN-74-342A and MAN-74-608A after calcining at 1000°C.
## Table 2. Percentage of Total Fe Associated With Iron-Bearing Phases (Mössbauer Mineralogy), Ratio of Fe\(^{57}/\)Fe\(^{55}\), Sample Saturation Magnetization \(J_s\), and Position of Near-IR Band Minimum for Manicouagan Impact Melt Rocks

<table>
<thead>
<tr>
<th>Sample</th>
<th>Hematite (hkHm)</th>
<th>np-Hm; pFe(^{57})</th>
<th>Phyllosilicate</th>
<th>Magnetite</th>
<th>Ilmenite</th>
<th>Pyroxene</th>
<th>Fe(^{57})/Fe(^{55})</th>
<th>(J_s) of Sample, Near-IR</th>
<th>Fe(^{57}) Chemistry</th>
<th>A m/Kg</th>
<th>Band, nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAN-74-15</td>
<td>42</td>
<td>30</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>18</td>
<td>0.81</td>
<td>0.74</td>
<td>0.34</td>
<td>935</td>
<td></td>
</tr>
<tr>
<td>MAN-74-74C</td>
<td>47</td>
<td>27</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0.89</td>
<td>-</td>
<td>0.32</td>
<td>845</td>
<td></td>
</tr>
<tr>
<td>MAN-74-83</td>
<td>29</td>
<td>28</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>36</td>
<td>0.63</td>
<td>-</td>
<td>0.14</td>
<td>930</td>
<td></td>
</tr>
<tr>
<td>MAN-74-95A</td>
<td>37</td>
<td>25</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>0.73</td>
<td>0.73</td>
<td>0.04</td>
<td>850</td>
<td></td>
</tr>
<tr>
<td>MAN-74-97</td>
<td>57</td>
<td>24</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0.86</td>
<td>0.78</td>
<td>0.08</td>
<td>840</td>
<td></td>
</tr>
<tr>
<td>MAN-74-125</td>
<td>36</td>
<td>40</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0.85</td>
<td>-</td>
<td>0.25</td>
<td>875</td>
<td></td>
</tr>
<tr>
<td>MAN-74-131</td>
<td>9</td>
<td>35</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>39</td>
<td>0.56</td>
<td>0.48</td>
<td>0.73</td>
<td>920</td>
<td></td>
</tr>
<tr>
<td>MAN-74-133</td>
<td>31</td>
<td>40</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>16</td>
<td>0.76</td>
<td>0.65</td>
<td>0.30</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>MAN-74-176</td>
<td>40</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>0.71</td>
<td>0.70</td>
<td>0.08</td>
<td>910</td>
<td></td>
</tr>
<tr>
<td>MAN-74-177</td>
<td>67</td>
<td>19</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0.92</td>
<td>0.84</td>
<td>0.06</td>
<td>830</td>
<td></td>
</tr>
<tr>
<td>MAN-74-217</td>
<td>37</td>
<td>25</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>0.66</td>
<td>0.62</td>
<td>0.08</td>
<td>910</td>
<td></td>
</tr>
<tr>
<td>MAN-74-229</td>
<td>13</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>0.35</td>
<td>0.34</td>
<td>0.03</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>MAN-74-243</td>
<td>15</td>
<td>32</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>48</td>
<td>0.51</td>
<td>0.46</td>
<td>0.33</td>
<td>980</td>
<td></td>
</tr>
<tr>
<td>MAN-74-342A</td>
<td>45</td>
<td>18</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0.88</td>
<td>-</td>
<td>0.05</td>
<td>855</td>
<td></td>
</tr>
<tr>
<td>MAN-74-390</td>
<td>21</td>
<td>23</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>45</td>
<td>0.52</td>
<td>0.54</td>
<td>0.69</td>
<td>940</td>
<td></td>
</tr>
<tr>
<td>MAN-74-396</td>
<td>28</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>0.56</td>
<td>0.55</td>
<td>0.05</td>
<td>920</td>
<td></td>
</tr>
<tr>
<td>MAN-74-407</td>
<td>17</td>
<td>22</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>58</td>
<td>0.41</td>
<td>0.41</td>
<td>0.12</td>
<td>940</td>
<td></td>
</tr>
<tr>
<td>MAN-74-506</td>
<td>17</td>
<td>32</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>40</td>
<td>0.58</td>
<td>0.64</td>
<td>0.64</td>
<td>910</td>
<td></td>
</tr>
<tr>
<td>MAN-74-512A</td>
<td>5</td>
<td>28</td>
<td>0</td>
<td>17</td>
<td>4</td>
<td>45</td>
<td>0.47</td>
<td>-</td>
<td>1.30</td>
<td>935</td>
<td></td>
</tr>
<tr>
<td>MAN-74-608A</td>
<td>9</td>
<td>21</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>64</td>
<td>0.32</td>
<td>-</td>
<td>0.73</td>
<td>980</td>
<td></td>
</tr>
<tr>
<td>MAN-74-689E</td>
<td>17</td>
<td>37</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>35</td>
<td>0.62</td>
<td>-</td>
<td>0.64</td>
<td>920</td>
<td></td>
</tr>
<tr>
<td>MAN-75-3</td>
<td>28</td>
<td>33</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>28</td>
<td>0.69</td>
<td>0.73</td>
<td>0.51</td>
<td>930</td>
<td></td>
</tr>
<tr>
<td>MAN-75-20</td>
<td>34</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>0.62</td>
<td>0.59</td>
<td>0.16</td>
<td>910</td>
<td></td>
</tr>
</tbody>
</table>
MORRIS ET AL.: HEMATITE, PYROXENE, AND PHYLLOSILICATES ON MARS

The amount of phyllosilicate in these samples is significantly larger than we observed previously in Hawaiian palagonitic soils [Golden et al., 1993; Morris et al., 1993]. Floran et al. [1978] and Allen et al. [1982] also report smectite in Manicouagan samples based on petrographic and X-ray diffraction data, respectively.

From their petrographic observations, Floran et al. [1978] concluded that smectite formed as a part of the oxidative alteration of silicate glass and mafic minerals (especially olivine). Our limited XRD data are consistent with this view. MAN-74-608A, which is the least oxidatively altered (lowest Fe$_3^+$/Fe$_{tot}$) sample we studied, has the least smectite. Because Mössbauer data are sensitive only to iron-bearing smectites and because the iron content of the smectites is not known and may be variable, smectite contents determined by Mössbauer analysis and by X-ray diffraction and by Mössbauer spectroscopy. alteration of silicate glass and mafic minerals (especially olivine).

The first set of columns is the average composition and its standard deviation for all 24 samples analyzed by Floran et al. [1978]. The second and third sets of columns are equivalent data for samples having the 12 highest and 12 lowest values of Fe$_3^+$/Fe$_{tot}$, respectively. The last set is for nine samples analyzed in this study.

![Figure 3. Comparison of Fe$_3^+$/Fe$_{tot}$ determined by chemical analysis and by Mössbauer spectroscopy.](image)

---

**Table 3. Compositional Data for Impact Melt Rocks from Manicouagan Crater**

<table>
<thead>
<tr>
<th>Element</th>
<th>All 24 Samples</th>
<th>Highest Fe$<em>3^+$/Fe$</em>{tot}$</th>
<th>Lowest Fe$<em>3^+$/Fe$</em>{tot}$</th>
<th>(This Study)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>SD</td>
<td>Average</td>
<td>SD</td>
</tr>
<tr>
<td>SiO$_2$</td>
<td>57.75</td>
<td>1.18</td>
<td>57.90</td>
<td>1.09</td>
</tr>
<tr>
<td>TiO$_2$</td>
<td>0.77</td>
<td>0.06</td>
<td>0.79</td>
<td>0.03</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>16.51</td>
<td>0.62</td>
<td>16.63</td>
<td>0.72</td>
</tr>
<tr>
<td>FeO</td>
<td>3.98</td>
<td>0.89</td>
<td>4.66</td>
<td>0.54</td>
</tr>
<tr>
<td>Fe$_2$O$_3$</td>
<td>2.29</td>
<td>0.74</td>
<td>1.69</td>
<td>0.35</td>
</tr>
<tr>
<td>MgO</td>
<td>0.11</td>
<td>0.02</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>MnO</td>
<td>3.50</td>
<td>0.53</td>
<td>3.25</td>
<td>0.41</td>
</tr>
<tr>
<td>CaO</td>
<td>5.84</td>
<td>0.95</td>
<td>5.34</td>
<td>0.35</td>
</tr>
<tr>
<td>Na$_2$O</td>
<td>3.85</td>
<td>0.79</td>
<td>3.96</td>
<td>0.27</td>
</tr>
<tr>
<td>K$_2$O</td>
<td>3.04</td>
<td>0.54</td>
<td>3.12</td>
<td>0.19</td>
</tr>
<tr>
<td>P$_2$O$_5$</td>
<td>0.22</td>
<td>0.03</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>1.73</td>
<td>0.36</td>
<td>1.84</td>
<td>0.31</td>
</tr>
<tr>
<td>Fe</td>
<td>4.58</td>
<td>0.28</td>
<td>4.37</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Concentrations, wt %

| Sc | 18.08 | 1.97 | 17.26 | 1.05 | 18.40 | 1.96 |
| Co | 79.9 | 28.8 | 77.4 | 25.5 | 76.2 | 24.6 |
| Ba | 1276 | 306 | 1459 | 201 | 1098 | 213 |
| La | 37.76 | 5.37 | 39.5 | 5.2 | 35.6 | 4.7 |
| Ce | 78 | 10 | 81 | 9 | 74 | 9 |
| Sm | 6.64 | 0.68 | 6.91 | 0.63 | 6.41 | 0.61 |
| Eu | 1.85 | 0.24 | 1.91 | 0.24 | 1.85 | 0.21 |
| Tb | 0.81 | 0.16 | 0.82 | 0.16 | 0.83 | 0.19 |
| Yb | 2.38 | 0.17 | 2.45 | 0.17 | 2.37 | 0.16 |
| Lu | 0.563 | 0.033 | 0.375 | 0.022 | 0.356 | 0.026 |

Concentrations, µg/kg
sufficiently high temperatures (500-800°C). This may explain, for example, the correspondence between the Mössbauer spectrum for MAN-74-177 and those for MAN-74-608A and 342A heated to 1000°C and the relatively lower proportion of smectite in MAN-74-177 compared to 342A (Table 2).

Despite large differences in mineralogy from hematite-dominated to pyroxene-dominated assemblages for Manicouagan impact melt rocks, their chemical composition is very homogeneous. As discussed by Floran et al. [1978], there are no statistically significant positional (vertical, lateral, and radial) chemical variations within the impact melt sheet. In Table 3, we used the same chemical data to show that the process of oxidative alteration also did not impart statistically significant variations in chemical composition. The average chemical composition for the 12 samples having the highest value of the Fe\(^{3+}/Fe\(_{\text{total}}\) (most oxidized) is the same as that for the 12 least oxidized samples. The average concentration of iron (expressed as Fe) for the two groups is essentially identical, for example. In Table 3 we also report the average composition determined from INAA data for nine additional Manicouagan samples. There is no statistical difference between these data and those reported by Floran et al. [1978].

Reflectivity Spectra

Reflectivity spectra for samples with the lowest and highest proportions of ferric iron are shown in Figure 4 (Mössbauer spectra in Figure 1). In agreement with Mössbauer mineralogy, the reflectivity spectra of the most oxidized and least oxidized samples (three each) are dominated by the spectral characteristics of hematite and pyroxene, respectively. The changes in spectral slope near 520 and 620 nm, the relative reflectivity maximum and the band minimum (or plateau) centered near 860 nm for the most oxidized samples are all characteristic of bulk-Hm (i.e., crystalline hematite) [e.g., Morris et al., 1985]; these samples are light pink to salmon in color. The two-band signature of olivine pyroxene or type-B clinopyroxene [e.g., Cloos and Gaffey, 1991] is clearly evident in the spectra of the least oxidized samples, which are light gray in color. The band minima range from 950 and -1800 nm (MAN-74-512A) to 1040 and -2100 nm (MAN-74-407). Our spectral data do not extend to long enough wavelengths to define well the minimum for the second band. The spectral features of bulk-Hm and pyroxenes are superimposed on a relatively featureless ferric absorption edge through the visible and near-IR. This ferric edge is attributed to chromophoric hematite, which has appropriate spectral properties [Morris et al., 1989; Morris and Lauer, 1990] and paramagnetic ferric iron observed in the Mössbauer data.

No bands attributable to electronic bands of ferrous and/or ferric iron in smectite are apparent in the spectra. Based on Mössbauer mineralogy, sample MAN-74-342A has the highest proportion of iron-bearing smectite. As discussed above, the spectral features resulting from iron are dominated by those for bulk-Hm. The bands at ~1400 and 1900 nm, which are combination and overtone bands of H\(_2\)O fundamental vibrations, probably result in part if not wholly from smectite. Note that the reflectivity spectrum of MAN-74-608A, which is the least oxidized sample and contains only a trace of phyllosilicates, has no discernible bands at 1400 and 1900 nm.

Figure 5 is a correlation of the position of the near-IR band and the ratio Hm/(Hm+Px) of these compositions determined from Mössbauer data. Figure 5 shows the correlation results from iron are dominated by those for bulk-Hm. The bands at ~1400 and 1900 nm, which are combination and overtone bands of H\(_2\)O fundamental vibrations, probably result in part if not wholly from smectite. Note that the reflectivity spectrum of MAN-74-608A, which is the least oxidized sample and contains only a trace of phyllosilicates, has no discernible bands at 1400 and 1900 nm.

Figure 5 is a correlation of the position of the near-IR band and the ratio Hm/(Hm+Px) of these compositions determined from Mössbauer data. Figure 5 shows the correlation results from iron are dominated by those for bulk-Hm. The bands at ~1400 and 1900 nm, which are combination and overtone bands of H\(_2\)O fundamental vibrations, probably result in part if not wholly from smectite. Note that the reflectivity spectrum of MAN-74-608A, which is the least oxidized sample and contains only a trace of phyllosilicates, has no discernible bands at 1400 and 1900 nm.

Figure 5 is a correlation of the position of the near-IR band and the ratio Hm/(Hm+Px) of these compositions determined from Mössbauer data. Figure 5 shows the correlation results from iron are dominated by those for bulk-Hm. The bands at ~1400 and 1900 nm, which are combination and overtone bands of H\(_2\)O fundamental vibrations, probably result in part if not wholly from smectite. Note that the reflectivity spectrum of MAN-74-608A, which is the least oxidized sample and contains only a trace of phyllosilicates, has no discernible bands at 1400 and 1900 nm.
hyperssthene (low-Ca pyroxene) < enstatite (high-Ca pyroxene) [e.g., Huang, 1977]. On this basis, we would expect low-Ca pyroxene to be preferentially in the least oxidized rocks, which is contrary to observation. However, differential alteration could still be consistent with the data if the rocks in the high-Ca pyroxene group had, prior to alteration, significantly more high-Ca pyroxene than low-Ca pyroxene. The normative calculations of Floran et al. [1978] indicate that this is in fact the case. For samples where both reflectivity and normative data are available, prealteration values of enstatite (enstatite-hypersthene) are 0.72 and 0.39 for high-Ca and low-Ca groups, respectively.

In summary, it is clear from reflectivity, chemical, and petrographic data that both high-Ca and low-Ca pyroxenes are present. Although these pyroxenes are characterized by different band minima, oxidation of each to hematite could produce overlapping mixing relationships between their band minima and that for hematite (model 1). However, it is also possible that only three band minima occur (980, 920, and 850 nm for high-Ca and low-Ca pyroxene and hematite, respectively) and that these phases, from a special set of circumstances, are optically dominant in samples with increasing whole-rock values of Hm/(Hm+Px). This is reasonable for hematite, but not necessarily for the pyroxenes.

As discussed next, examination of several samples with 900- to 920-nm band minima shows that likely candidates for both models are present among the samples. In Figure 7 are Mössbauer and reflectivity spectra for five samples having band minima between 900 and 920 nm. The Mössbauer data show that both pyroxene and (bulk) hematite are present. If the 800- to 1000-nm region is not considered, optical manifestations of bulk-Hm include a relative reflectivity maximum near 750 nm, and inflections near 520 and 620 nm. A broad band centered in the 1600- to 2300-nm region is a manifestation of pyroxene. If any of these hematite or pyroxene features are present, a corresponding feature in the 800- to 1000-nm region might be expected. The hematite features at 520, 620, and 750 nm are most evident in the spectrum of MAN-74-217, much less intense in 176 and 506, and virtually absent in 131 and 133. A ~2000-nm band is clearly present for MAN-74-506, and present but much weaker for the other four samples. For MAN-74-506, 131, and 133, the spectral contrast for the ~2000-nm band may be reduced relative to the 800- to 1000-nm band because magnetite (which is optically opaque [e.g., Morris et al., 1985]) is present and because the ~2000-nm band is normally less intense in pyroxene [Clowes and Gaffey, 1991]. Thus, optical contributions in the 800- to 1000-nm region from both hematite and pyroxene seem likely for MAN-74-217 and 176. Optical contributions from pyroxene alone seem to dominate the spectra of MAN-74-506, 131, and 133.

Application to Mars

Reflectivity Spectra

Spectral data from the Imaging Spectrometer for Mars (ISM) instrument on the Phobos 2 spacecraft are currently the highest spatial resolution data for Martian surface materials. In the 800- to 1300-nm region, these data show band minima extending from ~850 to ~980 nm [Murchie et al., 1993; Mustard et al., 1993]. In agreement with previous interpretations of the same band from lower spatial resolution data [e.g., Morris et al., 1989; Morris and Lauer, 1990; Bell et al., 1990], bands near ~850 nm were assigned by Murchie et al. [1993] to bulk (well-crystalline) hematite. Bands near 980 nm

![Figure 5. Position of near-IR band minimum as a function of Hm/(Hm+Px) as determined by Mössbauer spectroscopy. The ratio Hm/(Hm+Px) is ratio of the molar amount of Fe associated with hematite (bulk Hm) to the molar amount of Fe associated with hematite plus pyroxene.](image)

![Figure 6. Position of near-IR band minimum as a function of FeO/FeO+ by Mössbauer.](image)
were assigned to calcic pyroxene or possibly an intimate mixture of pigeonite and augite [Mustard et al., 1993]. There seems to be general agreement on the mineralogical assignment of hematite and pyroxene to the bands near 860 and 980 nm, respectively. Can unequivocal mineralogical assignments be made for band minima at intermediate wavelengths?

In Figure 8 are Phobos 2 ISM spectra [Marchie et al., 1993] whose band minima range from ~850 nm (hematite) to intermediate band positions near 920 nm. Mustard et al. [1993] consider that the origin of intermediate positions is small mixing between low-salbedo (optically pyroxene) material of volcanic origin and highly altered (optically hematite) bright dust or soils. Marchie et al. [1993] considers several models but favors the interpretation that these spectra indicate the presence of an additional ferric-bearing phase such as goethite (α-FeOOH), ferrhydrite (Fe₅O₇·3H₂O), or jarosite (K⁺Fe₃⁺[SO₄]₂(OH)₆). As discussed next, we advocate here an alternate hypothesis: that the intermediate band centers represent intimate hematite-pyroxene assemblages and/or pyroxene analogous to the Manicouagan impact melt rocks.

In Figure 9 we plot reflectivity spectra between 800 and 1300 nm for selected impact melt rocks from Manicouagan Crater. As in Figure 8, the spectra are arranged in order of increasing wavelength of the band minimum. If we did not have Mössbauer mineralogy and magnetic and chemical data for these samples, we would be in a similar position for their interpretation as we are for the Phobos 2 data. Namely, we would with a high degree of confidence assign hematite and pyroxene to the ~850 and 980- to 1000-nm band positions.

Figure 7a. Mössbauer spectra for five Manicouagan impact melt rocks having near-IR band minima in the region 900-920 nm.

Figure 7b. Same as Figure 7a for reflectivity spectra.
A number of studies [Newsom, 1980; Allen et al., 1982; Clifford, 1993] have advocated that a significant portion of Martian soil represents erosional products of hydrothermally altered impact melt sheets. For example, Clifford [1993] has calculated an equivalent global impact melt layer of 521 m. By analogy with Manicouagan impact melt rocks and in agreement with observations for Mars, oxidative alteration of Martian impact melt sheets above 2500°C and subsequent erosion could produce rocks and soils with variable proportions of hematite (both bulk and nanophase), pyroxene, and phyllosilicates as iron-bearing mineralogies. The Manicouagan samples show that this mineralogical diversity can be accomplished at constant chemical composition, which is also indicated for Mars from analyses of soil at the two Viking landing sites [Clark et al., 1982].

If this process is volumetrically important, these phases on the Martian surfaces would have been formed rapidly at relatively high temperatures on a sporadic basis throughout the history of the planet. This process is in contrast to slow, relatively continuous, low-temperature processes involving dissolution of ferrous-bearing 0.8−1.0−1.2−1.4 phases and precipitation of low-temperature and ferric-bearing phases such as goethite, lepidocrocite, and ferrihydrite [e.g., Banin et al., 1993; Burns, 1993]. In this case, the heat associated with repetitive meteoritic impact and volcanic processes would convert hydrous ferric oxides to their anhydrous forms (e.g., hematite) as observed for Hawaiian palagonitic iron associated with phyllosilicates. More definitive spectral evidence for phyllosilicates would be (Al,Mg,Fe)-OH transitions near 2300 nm [e.g., Clark et al., 1990a,b]. Unfortunately, our data do not extend beyond 2100 nm.

**Impact Processes on Mars**

Oxidative alteration of the Manicouagan impact melt rocks is thought to occur shortly after the impact by influx of oxidizing vapors and/or fluids while the rocks are still hot but below solidus temperatures [Floran et al., 1978]. For Manicouagan, the solidus is estimated at about 915°C [Simonds et al., 1978]. Based on the petrographic observations of Floran et al. [1978] and Phinney et al. [1978] for Manicouagan and West Clearwater Lake impact melt rocks, respectively, Ti-bearing (and relative large) hematite particles are formed by oxidative alteration of primary titanomagnetite. Essentially Ti-free hematite particles extending in size from micron-sized particles (bulk-Hm) to particles below the limit of optical microscopes (nanophase hematite) are formed from glass and mafic minerals. Because of their small size and high number density, the essentially Ti-free particles are likely the dominant ferric iron pigmenting phase. In addition to hematite, Floran et al. [1978] and Phinney et al. [1978] include hydrous ferric oxides as potential oxidative alteration products; however, our data (primarily the Mössbauer mineralogy) show that only hematite is present, at least for Manicouagan samples. An explanation for the absence of ferrihydrite and ferric oxyhydroxide phases (e.g., goethite and lepidocrocite) and presence of hematite is that the oxidative alteration took place at temperatures above 250−300°C when these phases are thermally unstable relative to hematite [Morris and Lauer, 1981]. The phyllosilicates (primarily smectites) must also have formed (from silicate minerals and glass, according to Floran et al. [1978]) above these temperatures.

Impacts have obviously been an important surface modification and alteration process throughout Martian history. As in Figure 8, the spectra are arranged according to band minimum from lowest to highest wavelength.
S ≠ NC: Multiple source areas for Martian meteorites

Allan H. Treiman
Lunar and Planetary Institute, Houston, Texas

Abstract. The Shergottite, Nakhlit, and Chassigny (SNC) meteorites are igneous rocks from Mars, inferred previously to have been propelled into space (and to Earth) by a single impact event. However, the sum of petrologic, chemical, and chronological data suggests two distinct sites of origin: S from one, NC from another. Differences between the S and NC include distributions of terrigenous and space weathering effects, hydrogen isotopes, and extinct sources of the SNCs. Formation of two meteorite source craters on Mars have been considered but rejected because of partially different histories.

Introduction: Rocks From Mars

Three small classes of basaltic meteorites, collectively called the SNCs (for Shergottites, Nakhliites, and Chassigny), are samples of the Martian crust, and so are of surpassing interest to the planetary sciences. Lacking Martian samples returned by spacecraft missions, the SNC meteorites have stood as proxy returned samples for petrologic and geologic studies (McSween, 1985). Unlike a sample return mission, however, the source sites of the SNCs are not known, so their utility is greatly diminished. There have been attempts to determine the Martian source of the SNCs, all of which have assumed or concluded that the meteorites came from a single source crater (Wood and Ashwal, 1981; Nyquist, 1983; Vickers and Malozemoff, 1987; Mengelis-Arck et al., 1992). Mengelis-Arck et al. (1992) located nine potential source craters in Tharsis, but none are consistent with constraints imposed by all of the SNCs. Multiple source craters on Mars have been considered but rejected (Bogard et al., 1984; Vickers and Malozemoff, 1987), partly on cratering and orbital dynamic bases and partly because all SNCs were inferred to have the same crystallization age.

However, differences between the S and NC meteorites are significant, and suggest that these two groups of meteorites experienced different histories from mantle differentiation to the time they landed on Earth. These differences are most consistent with the S originating at a different impact crater than the NC (Treiman, 1993a, 1994a). The inference of multiple source craters has broad implications for the geology and petrology of the SNC meteorites, and furthermore for impact cratering rates on Mars, impact mechanics, and meteorites.

The Martian Meteorites

The most convincing evidence that the SNC meteorites are from Mars is that they contain within them a trapped gas component identical in every measurable respect to Mars' current surface atmosphere [Bogard et al., 1984; Wimer and Pepin, 1988]. The Mars atmosphere is unique among gas reservoirs in the solar system, notably for its abundance ratios of N₂/N, ³⁵Cl/³⁷Cl, ³³⁵Ar/³⁷Ar, ³³⁵Ar/³²⁶Ar, etc. [Pepin, 1989], leaving essentially no doubt that the SNCs formed on Mars. Even lacking this comparison (based on Viking lander data), there is a strong case for the SNCs' Martian origin, based on their volatile element contents and oxidation states, indicating a large source planet, their young crystallization ages (1300 and ~180 Ma) and trace element fractionations, indicating recent and recurrent igneous activity, presence of hydrous magmatic minerals, indicating a hydrous mantle or crust, and their petrochemical and isotopic compositions, indicating an active hydrosphere within the last 1300 to ~180 m.y. (e.g., Wood and Ashwal, 1981; McSween, 1985; Gouding, 1992). All of bodies in the solar system, only Mars and the Earth could yield such rocks. A terrestrial origin is precluded by the oxygen isotope composition and geochemistry of the SNCs [Clayton and Mayeda, 1983].

The shergottites, abbreviated here as S, include basaltic and cumulate igneous rocks with abundant low-calcium pyroxenes. The shergottites, Zagami and EETA79001 meteorites are basaltic, with or without additional (cumulate) pyroxene [Stolper and McSween, 1979, McSween and Kearsavage, 1983]. EETA79001 contains two basalt lithologies, A and B, in igneous contact. The ALHA77005 and LERO8516 meteorites are websterites or lherzolites, cumulate rocks of olivine and low-calcium pyroxene with lesser augite and plagioclase [McSween et al., 1979; Zagami, 1989; Harvey et al., 1993; Treiman et al., 1994]. The EETA79001A lherzolite contains xenoliths and xenocrysts from a websterite or lherzolite, here called EETA79001X. For convenience, the Martian lherzolites are grouped here with the
Amphibole of igneous origin is present inside most S [Tremban, 1985, and unpublished data, 1994]. The nakhlite meteorites, abbreviated here as N, are Nahlas itself, Lafayette, and Governor Valdare [Busbey et al., 1980]. The N are augite-rich igneous rocks with significant proportions of olivine and dward little low-calcium pyroxene. They are interpreted to be cumulus rocks, formed from a Ca-rich basaltic magma [Harvey and McSween, 1992a; Tremban, 1993b]. Amphibole of igneous origin is reported in inclusions in nakhlite olivines [Harvey and McSween, 1992a].

The Chassigny meteorite, C here, is a diabase, interpreted as an igneous cumulate [Florin et al., 1978]. It is most abundant pyroxene is augite, some of which may also be cumulate, as may some of its less abundant low-Ca pyroxene [Mittlefehldt et al., 1994]. Plagioclase is not abundant. Olivines in Chassigny contain magnetic inclusions with common grains of water-bearing amphibole, and rare biotite [Florin et al., 1978; Johnson et al., 1991; Watson et al., 1993].

ALH84001 is a newly recognized Martian meteorite, previously classified as a diachroite [Mittlefehldt, 1994]. Data on ALH84001 are limited, although it is clearly not an S, N, or C. It will be discussed separately later.

SC: Evidence for Multiple Source Craters

The SNC meteorites have consistently been inferred or concluded to originate from a single impact crater on the Martian surface [Wood and Ashbol, 1981; Nyquist, 1982; McSween, 1985; Vickery and Ashbol, 1987; Mognini-Mark et al., 1992], although some studies have considered multiple crater origins [Hough et al., 1984; Vickery and Ashbol, 1987; Jones, 1989]. A single crater origin was considered most likely in that the SNC meteorites were inferred to have all crystallized at 1300 m.y. [Wood and Ashbol, 1981; Shih et al., 1982], and because impact ejection was considered to be an unusual or unique event [Wood and Ashbol, 1981; Nyquist, 1982]. The idea that normal impacts could have ejected the SNCs seemed to be directly refuted by the glaring absence of meteorites from the Moon. Since these early works, many lunar meteorites have been found in Antarctica and elsewhere, the crystallization ages of the S are now known to be different from those of NC [Jones, 1985], and it has been shown that normal meteorite impacts can eject rocks from Mars [Vickery and Ashbol, 1987]. But the possibility that the SNC meteorites came from multiple source craters is still rejected [e.g., Mognini-Mark et al., 1992].

My thesis here is that petrologic, chemical, and chronological evidence suggests that the SNCs originated at two distinct sites on Mars. The evidence is summarized in Table 1, in reverse chronological order, from which it appears that the S and NC have been distinct through their whole histories, from source mantle, through magma composition, through low-temperature aqueous alteration, through impact ejection, and through encounter with Earth.

The following discussions document the data and inferences summarized in Table 1, starting with the present and working backward through time (not the order of strongest argument). I have relied on synthesis and summary references where possible, so as not to further burden the reference list; apologies are offered for not citing every relevant contribution.

Terrestrial Residence Age

The terrestrial residence ages of meteorites, the time intervals since they fell to Earth, can be determined either by observation of a fall or by analysis of a meteorite for certain cosmogenic radioactive nuclides [e.g., Mittlefehldt et al., 1986]. As summarized in Table 1, the terrestrial ages of the S and NC overlap, but appear to have sampled different age populations. Thus, the ages are consistent with S and NC originating in separate asteroidal populations before coming to Earth.

<table>
<thead>
<tr>
<th>Property</th>
<th>S</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrestrial exposure, years</td>
<td>32 - 200,000</td>
<td>&lt;200</td>
</tr>
<tr>
<td>Cosmic ray exposure age, Ma</td>
<td>~2.5±0.5, 0.5</td>
<td>1±1</td>
</tr>
<tr>
<td>Shock effects</td>
<td>mesoblastic after plagioclase; abundant shock glass; implanted atmosphere gas; dismembered isotopic systems;</td>
<td>original plagioclase preserved; shock glass rare; no implanted atmosphere; no isotopic disturbances</td>
</tr>
<tr>
<td>Pre-terrestrial aqueous alteration:</td>
<td>aluminosilicate clay? poorly crystalline?</td>
<td>smectite-iron oxide well-crystalline</td>
</tr>
<tr>
<td>Crystallization age, Ma</td>
<td>~180</td>
<td>~1300</td>
</tr>
<tr>
<td>Magmas composition:</td>
<td>low Ca (orthopyroxene, pigeonite)</td>
<td>high Ca (augite)</td>
</tr>
<tr>
<td>Magmas procheimlalinity, incompatible</td>
<td>depleted versus CI</td>
<td>enriched versus CI</td>
</tr>
<tr>
<td>Magmas source: martile</td>
<td>depleted before 1300 Ma; depleted again at 180 Ma</td>
<td>depleted before 1300 Ma; enriched at ~1300 Ma</td>
</tr>
<tr>
<td>Magmas source: crust</td>
<td>no &quot;crystall&quot; components</td>
<td>two &quot;crystall&quot; components</td>
</tr>
</tbody>
</table>

See text for full explanation.
Sharpsite meteorites (in the broad definition above) have been observed to fall (Shergotty, Zagami), and have been collected at three separate sites in Antarctica. Residence ages for three of these are: ALHA77005, 200,000 years [Nishizumi et al., 1986, 1992; Freundel et al., 1986]; EETA79001, 12,000,000 years [Nishizumi et al., 1986; Fujii and Donnaper, 1988]; and LEW88516, < 50,000 years [Nishizumi et al., 1992]. On the other hand, all of the NC meteorites were found on fall or found shortly after fall and contain no products of terrestrial weathering [Treiman et al., 1985; Gooding et al., 1991; Treiman et al., 1993].

The terrestrial residence ages of the S and NC overlap, as meteorites from both groups have been observed to fall. But the absence of NC meteorites from Antarctica may be significant in suggesting that no (or few) NC meteorites fell over the past million years represented by stony meteorites in the Antarctic collections [Nishizumi et al., 1986]. If the historical proportion of S versus NC falls, two versus four, was the proportion that fell in Antarctica, the probability of finding three S and so NC meteorites from Antarctica is 1/27, or 0.037. This value may be of statistical significance; it is not likely to reflect collection bias, as many pyroxenite and dunite meteorites (like N and C, respectively) have been collected in Antarctica.

Cosmic Ray Exposure Age

The cosmic ray exposure age of a meteorite is the time interval that it spent being irradiated by solar and galactic cosmic rays; the age can be measured by abundances of some isotopes produced during irradiation, the cosmogenic nuclides. Exposure ages for the NC are ~11 Ma, much older than exposure ages for S, which are 2.8 or 0.5 Ma [e.g., Bogard et al., 1984; Nishizumi et al., 1986; Treiman et al., 1994]. The differences in exposure ages is significant, and has been interpreted in terms of ejection from the parent planet [Zinner, 1990] and ejection of parent meteoroids in space [e.g., Bogard et al., 1984], or a combination of both mechanisms [Treiman, 1993]. Separate source sites (and ejection events) for S and NC meteorites are different exposure ages, but the ages do not necessarily preclude a single source crater [Bogard et al., 1984; Vickers and Melosh, 1987].

Shock Metamorphoses

Effects of shock metamorphism are apparent in all of the SNC meteorites, presumably reflecting the trauma they suffered on ejection from Mars by meteorite impact. The S have uniformly been shocked to much higher levels than the NC, suggesting significantly different shock histories. Different shock histories are consistent with and probably required by the two-cratet scenario, but are also consistent with ejection as different objects from different fields around a single impact.

The shock experienced by the S was "strong to very strong," stages SS-SS of Steffler et al. [1991]. Petrographic manifestations of this shock include transformation of original plagioclase to maskelynite and to liquid; intense cracking, planar deformation elements, polygonization and even melting of pyroxene and olivine, formation of the high-pressure phases ringwoodite and majorite, and common melting of the bulk rocks [Redwine, 1985; Treiman et al., 1994]. These data are consistent with peak pressures of 25-55 GPa [Steffler et al., 1991]. This intense shock event has been held responsible for disturbance of radio-isotope chronometers [Dhitr et al., 1982], and for implantation of Martian atmospheric gases [Bogard et al., 1984; Wiens and Pepin, 1988; Ogil and Bagnone, 1982].

The age of the shock event is not known, but may be 15±5 m.y. for ALHA77005 [Auguste, 1989].

On the other hand, the NC meteorites are effectively "unshocked to weakly shocked," stages S1 to S3 of Steffler et al. [1991]. Petrographic manifestations of these low shock levels include the presence of original plagioclase, olivine with sharp to slightly unshocked extinction, absence of high-pressure phases, and rarity of shock effects [Shergotty et al., 1989; Eckert et al., 1980]. Greater shock pressures are apparent locally in the NCs, where pyroxenes and olivines may be granulated or shocked, and in which shock textures occur typically [Steele et al., 1993]. These data are consistent with overall peak shock pressures of approximately 15 GPa, with localized peak pressures up to 56 GPa. Compared to S, the NC show no indication that shock has affected their radiocarbon systems, nor that shock has had any effect on their trace gas contents [O'Leary, 1988; Drake et al., 1993].

Petrological Aquatic Alterations

Preterrestrial aqueous alteration products (hydrox silicates, oxides, and ionic salts) in the SNC meteorites are important here as telltale in the compositions of waters at their emplacement site(s) near the surface of Mars. The mineralogy and compositions of the alteration products are different in NC and S, suggesting that their parent magma was emplaced into terrains of different geochemistry or with different mechanisms of alteration. The geochemistry and mechanisms of alteration are not yet known, and will be a fruitful field of research. However, the differences in alteration materials between the S and NC are great enough to suggest distinctly different alteration histories, consistent with different emplacement sites and pre-igneous histories.

The presence of preterrestrial hydroxalteration products in achondrite meteorites has been difficult to establish, even though early students of the N suggested that their "distinctive" minerals might be preterrestrial [Boucke and Reid, 1975; Becter et al., 1976]. Strontigraphic proof that the SNC meteorites contained products of preterrestrial aqueous alteration was provided first by Gooding and Bagnone [1986] for the EETA79001 shergottite and by Gooding et al. [1991] and Treiman et al. [1993] for the N. The mineralogy of the alteration materials are summarized in Table II of Gooding [1992]. Treiman et al. [1993] provide additional data. Wentworth and Gooding [1993] found evidence for preterrestrial Fe-Mg phosphates in ALHA77005, but the remaining secondary minerals in it and LEW88516 could have formed during Antarctic weathering [Smith and Steele, 1984]. There have not been extensive searches for preterrestrial alteration products in Shergotty or Zagami.

All of the analyzed SNC meteorites contain ionic salts of preterrestrial or possible preterrestrial origin (see references above). The most common are alkaline earth minerals (both Ca and Mg) and carbonates (Ca, Mg, Fe, Mn). Silicate alteration materials are known only in the N and in the shergottite EETA79001, so the comparison of Table I is between those groups. The silicate alteration materials in EETA79001 are hydrous aluminous silicates of variable composition and unknown structure [Gooding and Muenow, 1986]; some grains are consistent with illite clay, while others are described as "S/Cl-bearing micabole" in that chemical analyses could not be rationalized in terms of...
known mineral stoichiometry. The compositional variations and inability to match known stoichiometries strongly suggests that these materials are poorly crystalline, although their crystallinity has not been measured. There are no iron oxide phases in the alteration assemblage.

The alteration materials in the N are coarser and more abundant than in EETA79001, and so are better preserved. In the N, olivines and pyroxenes have been altered to "sidestone," an intimate mixture of smectite clay (low Al, high Fe) and iron oxides (hematite and fayalite) [Gondop et al., 1991; Treiman et al., 1993]. These clays are "coarsely" crystalline, with individual crystallites to 5 μm across.

There are some data on the isotopic compositions of alteration materials in the SNC meteorites, which show a few systematic differences between the S and NC. Oxygen in water in the N has Δ18O significantly greater than oxygen in SNC subhedral silicates [Karlsson et al., 1992], suggesting that NC were altered by a water reservoir that was not equilibrated with the source region of the NC magmas. On the other hand, only one sample of S meteorite (one eucrite of Zagami) showed a similar enrichment; oxygen in water in Shergotty, EETA79001, and the other Zagami eucrite had Δ18O near that of the bulk silicates [Karlsson et al., 1992]. This difference in oxygen composition is echoed in heavy noble gas isotopes [Dohle et al., 1993], but not in deuterium/hydrogen ratios [Wasson et al., 1994].

Crystallographic Ages

Ages of igneous crystallization are among the most important data consistent with separate sources for the NC and S meteorites (Table 1). Crystallographic age measurements by K-Ar, Rb-Sr, Sm-Nd, and U-Th-Pb radiochronometers are summarized by McSween [1985] and Jones [1986]. For the S, McSween [1985] found no basis at that time to choose among 1300, 350, and 180 Ma as being the true ages of crystallization. However, the current understanding of SNC chronology makes it very unlikely that the S crystallized at the same time as the NC, consistent with origins at separate craters.

For the NC, these radiochronometers are all discordant within error at 1300 Ma [McSween, 1985]. Given the absence of strong shock effects (see above) or extensive metamorphism in the nakhlites, this age is almost certainly that of igneous crystallization [Berkley et al., 1981; Harvay and McSween, 1992b].

The radiochronology of the S is complex; it is accepted here that they crystallized from magma at ~180 Ma [Jones, 1986; 1989]. All of the S have at least one radioisotope chronometric age of ~180 Ma, which originally had been ascribed to shock metamorphism [Shah et al., 1982; McSween, 1985]. However, Jones [1986] argued that this view was inconsistent with the igneous chemical zonings retained in the minerals of most S (especially Zagami and EETA79001). In particular, the Zagami shergottite retains igneous zoning in all of its major minerals, and yields ages of ~180 Ma from Rb-Sr, Sm-Nd, and U-Th-Pb radiochronometers. This age for Zagami is almost certainly that of crystallization, and Jones [1986] argued that the ~180 Ma ages for other S also represent igneous crystallization. It remains clear, however, that some isotopic systems have been perturbed through shock or assimilation of foreign material [Jones, 1986, 1989]. Confirmation of Jones' interpretation came from Zagami [1989], who found a comparable crystallization age for EETA79001 and an age of shock metamorphism of 150-15 Ma, and from Chen and Wasserburg [1993], who report a single Pb-lead event at ~170 Ma for LEW88516. In the absence of serious challenges to Jones' [1986] arguments, it seems most likely that the S all crystallized from magma at about 180 Ma.

Magma Composition

Crystallization sequence (silicates). The sequence in which minerals crystallized in an igneous rock is a direct consequence of the composition of the rock's parent magma. The crystallization sequence is readily determined in petrographic examination, and is particularly useful for cumulate igneous rocks, which do not have the compositions of magmas. This discussion focuses entirely on silicate minerals and ignores clays and other oxide minerals. The different crystallization sequences in the S and NC meteorites, summarized in Table 2, indicate that they formed from magma groups that were fundamentally different and must have originated in source regions of different mineral proportions, as first noted by Stolper et al. [1979]. Subsequent studies, new meteorites, and additional samples of known meteorites have validated and extended Stolper et al.'s [1979] inference (Table 2), and it remains clear that N and C formed from similar magmas, and that the NC and S magma groups are fundamentally different (Table 2).

Among the S, the earliest crystallizing (and most abundant) pyroxenes are low-calcium varieties, either orthopyroxene or pigeonite. Low-calcium pyroxene was the earliest mineral to crystallize in the shergottites proper: Shergotty, Zagami, and the A and B lithologies of EETA79001 [Stolper and McSween, 1979; Steele and Smith, 1982; McSween and Jarviswick, 1983; McCoy et al., 1992; Treiman and Sutton, 1992]. In the other shergottite lithologies (ALHA77005, semilithic in EETA79001, LEW88516), olivine was the first silicate mineral to crystallize, followed by low-calcium pyroxenes [McSween et al., 1979; Steele and Smith, 1982; McSween and Jarviswick, 1983; Landberg et al., 1990; Harvey et al., 1993; Treiman et al., 1994].

Among the NC meteorites, the first pyroxene to crystallize was augite [Bunch and Reid, 1975; Boctor et al., 1976; Plasma et al., 1978; Berkley et al., 1980]. The position of olivine in radiogenic crystallization has been uncertain, but there is now agreement that olivine and augite crystallized together [Harvey and McSween, 1992a; Treiman, 1993b]. Low-calcium pyroxene (pigeonite) appears later in crystallization, in part as a replacement of olivine [Berkley et al., 1980; Harvey and McSween, 1992b]. In Chassigny, olivine was the earliest silicate phase to crystallize, followed most probably by augite and pigeonite [Farror et al., 1978; Louglo and Pan, 1989, Johnson et al., 1991]. Washaba and Crozets [1994] suggest that pigeonite crystallized before augite, but definitive textural relationships have not yet been found.

Trace elements. Although great effort has been expended in the analysis and interpretation of trace element compositions of the SNC meteorites, only the greatest comparison is adequate here: parent magmas of the S were depleted in incompatible elements (e.g., the light rare earth elements, Rb, U, Th etc.), and the parent magmas of the NC were enriched in these elements, as shown in Figure 1. The differences between incompatible element abundance patterns in S and NC are so great that no simple petrologic process could produce both groups at once. The differences in trace element abundances between S and NC are not required or predicted by their bulk compositional differences, and so are independent constraints on their origins.
Table 2. Crystallization Sequences For Silicate Minerals In The Martian Meteorites

<table>
<thead>
<tr>
<th>Meteorite Group</th>
<th>Sequeces of Crystallizing Minerals</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>OL → OL+OPX/PIG → OPX/PIG → PIG → AUG + AUG+PL</td>
</tr>
<tr>
<td>ALHA77005</td>
<td></td>
</tr>
<tr>
<td>LEW98316</td>
<td></td>
</tr>
<tr>
<td>EETA79001X</td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>OL → OL+DIO → OL+DIO+PIG + AUG+PL</td>
</tr>
<tr>
<td>Treimann</td>
<td></td>
</tr>
<tr>
<td>Chassigny</td>
<td></td>
</tr>
<tr>
<td>Eta Kallateye</td>
<td></td>
</tr>
<tr>
<td>Quesenator</td>
<td></td>
</tr>
<tr>
<td>Valdesa</td>
<td></td>
</tr>
</tbody>
</table>

For each meteorite group (S or NC), a generalized crystallization sequence is given left to right. Each meteorite is listed at the point where it enters the crystallization sequence. Mineral abbreviations are: OL = olivine, OPX = orthopyroxene (high Ca), PIG = pigeonite pyroxene (low Ca), AUG = augite pyroxene, Pl = plagioclase.

Magma Source Regions

The source regions for basaltic magmas can be studied only indirectly, because melting and subsequent fractionation can obscure many clues about them. However, the available evidence suggests that there was no direct petrogenetic relationship between the sources for the S and NC magmas, although some indirect relationship is likely.

Among the least ambiguous clues to magma source regions are radio-isotope ages and initial ratios of radiogenic to nonradiogenic isotopes at the time of igneous crystallization. These initial ratios are preserved through igneous fractionation, and so reflect the source(s) of the magma. The most detailed analysis of initial isotope values in the SNCs is by Jones [1989], who tried to find a coherent petrogenetic scheme to relate the SNCs' initial radio-isotope ratios for 206Pb/204Pb, 142Nd/144Nd (from 206Pb, 208Pb from 204Pb, and 147Sm/147Nd, and 143Nd/144Nd from 147Sm).

The S do not all have the same initial isotopic ratios; this heterogeneity has been interpreted as a mixing of the isotopic signatures of three reservoirs, one mantle and two crustal [Jones, 1989]. Although the NCs are enriched in incompatible elements (Re, Nd, U, Pb), they were derived from a source region that was depleted in these elements; at least part of the depletion must have occurred within a few tens of millions of years after planet formation [Nyquist et al., 1991]. The enrichment in incompatible elements must have happened immediately before the NC parent magmas were formed (1300 m.y.), and is the only geochemical event evident in the NC source region after planet formation and initial differentiation (see references in Jones [1989]). Specifically, there is no evidence that the NC parent magmas assimilated any crustal components that are recognizable isotopically [Jones, 1989].

The NC have essentially identical radiometric ages (Table 1) and initial isotope ratios [Jones, 1989]. Although the NC are enriched in incompatible elements (Re, Nd, U, Pb), they were derived from a source region that was depleted in these elements; at least part of the depletion must have occurred within a few tens of millions of years after planet formation [Nyquist et al., 1991]. The enrichment in incompatible elements must have happened immediately before the NC parent magmas were formed (1300 m.y.), and is the only geochemical event evident in the NC source region after planet formation and initial differentiation (see references in Jones [1989]). Specifically, there is no evidence that the NC parent magmas assimilated any crustal components that are recognizable isotopically [Jones, 1989].

The S do not all have the same initial isotopic ratios; this heterogeneity has been interpreted as a mixing of the isotopic signatures of three reservoirs, one mantle and two "crustal" [Jones, 1989]. The mantle reservoir could reasonably have been the NC source mantle after normal isotopic evolution to ~180 Ma. This mantle must have become further depleted in Nd immediately before generation of the S parent magmas, but without depletion of other incompatible elements [Jones, 1989]. However, this scenario may require additional reservoirs to accommodate abundances of 144Nd (from extinct 146Sm) and 206Pb [Jones, 1989; Nyquist et al., 1991].

Thus, it is reasonable that both the NC and S were derived from mantle materials that were similar before 1300 m.y., but thereafter experienced divergent geochemical processing. Derivation from similar mantle materials does not necessarily imply that the NC and S formed in the same geographcal spot. On Earth, mantle provinces of common differentiation age can extend for thousands of kilometers [Bell and Black, 1989], and there is no a priori reason to expect that mantle provinces on Mars might be significantly smaller.

Summary

Thus, the sum of petrologic, chemical, and chronologic data in Table 1 suggests that the S and NC followed entirely separate histories, from at least the inferred mantle differentiation event at ~1500 Ma to their arrival on Earth. Put bluntly, the S and NC have had very little in common since 1300 Ma, except for their
shared origin on Mars. No single property listed in Table 1 would conclusively show that the S and NC formed at separate sites on Mars, a reasonable explanation or rationalization could be found for each difference. But the intersection of these many reasonable explanations is so small, so unlikely, that formation at two separate sites on Mars seems the simplest explanation of the pervasive differences between the S and the NC.

**Single Crater Scenarios**

In most previous studies, the SNC meteorites have been inferred to come from a single impact event on the Martian surface. I have shown above that petrologic, geochemical, and chronological evidence suggests separate source craters for S and NC. However, single crater scenarios are themselves inconsistent with available chemical and physical inferences.

**Ejection of all SNC meteorites from Mars in a single impact event** at ~180 Ma is the scenario most favored in the literature [Wood and Ashwood, 1981; Albat, 1977; Bogard et al., 1997]. This scenario is no longer tenable because (as discussed above) the ~180 Ma event recorded in S is not shock but igneous crystallization [Jones, 1986, 1989, Jagoutz, 1989].

Ejection of the SNCs from a single crater at 11 Ma (the cosmic ray exposure ages for NC) might be possible if the crater ejection zone overlapped distinct S and NC terrains, or hit a vent zone of S over NC. This scenario is not tenable because, either, Vickery and Melosh [1987] showed that it requires a crater of ~285 km diameter on a young (~180 Ma) volcanic surface or completely obliterating a young volcanic center on an older (1300 Ma) volcanic surface. In the former case, there are no such craters on young surfaces [Vickery and Melosh, 1987, Mengistu-Mark et al., 1992]; the latter case is completely ad hoc. In addition, ejection and subsequent orbit evolution must somehow ensure that (1) of materials exposed to cosmic rays for ~11 Myr, only NC lithologies arrive at Earth, and (2) of materials exposed to cosmic rays for 2.5 Myr or less, only S lithologies arrive at Earth. Neither scenario seems likely.

**S ≠ NC: Implications of Multiple Source Craters**

If the SNC meteorites left Mars in more than one impact event, at least some current understandings of Mars and its surface processes must be revised. First and foremost, accepted cratering scenarios must be revised to permit smaller impacts capable of ejecting solid material from Mars, to allow greater current (or near past) rates of crater production on Mars, or both. Second, many craters on Mars become acceptable as S or NC source sites, by eliminating the restriction that all S and NC come from the same site. And third, the lack of ancient or nonvolcanic lithologies among Martian meteorites suggests some additional problems in the recognition of meteorites, in the physical properties of Martian surface materials, or in impact mechanics.

**Ejection Mechanics and Cratering Rate**

The scenario favored here, where S and NC meteorites hailed from different source craters on Mars, was evaluated by Vickery and Melosh [1987] and rejected as improbable. Indeed, under current models of impact ejection processes and cratering rates, it is quite unlikely that Mars would have experienced an impact event into ~180 Ma (the cosmic ray exposure age for NC) volcanic surfaces. This deficit in SNC-ejecting craters of proper ages and sizes can be ameliorated by postulating greater cratering rates (short- or long-term), smaller permissible crater sizes, or both. From the SNCs alone, it is not possible to separate the effects of cratering rate and crater size.

Ejection of meteorites from planetary surfaces is most likely by a surface spallation mechanism [Vickery and Melosh, 1987]. The size of the largest ejecta fragments produced by an impact is directly dependent on the size of the impact, and thus on the size of the impact crater produced. If ejecta need be no larger than 0.5 m diameter, current theory predicts that impacts producing craters of 212 km diameter are adequate, provided the impactor velocity is greater than about 10 km/s [Vickery and Melosh, 1987]. If the ejecta need be larger (e.g., to satisfy constraints of cosmic ray exposure) larger craters are required: ejection of 3-m fragments requires a crater of ~300 km diameter, and ejection of 15-m fragments requires a crater of 275 km diameter [Vickery and Melosh, 1987].

Current rates of crater production on Mars are poorly known, and may be significantly greater than suggested in commonly used models [Table 3]. Brown et al. [1992] and Mengistu-Mark et al. [1992] accept Nomade and Hillis's [1981] second model (NH-2 in Table 3) of relatively low crater production rates, which are comparable to the nominal rate for the Moon over the last 3 Gyr. An updated version of that model is given by Nomade [1983], cited below as NH-3. Tanaka et al. [1988, 1992] accept Harman et al.'s [1981] preferred cratering rate model (ET in Table 3), with approximately twice Moon's nominal crater production rate. Harman et al. [1981] recognized the possibility of greater cratering rates, up to twice their preferred rate. Recent analyses of crater populations on Earth and of asteroid and comet populations near the Earth suggest that cratering rates over the last 300 m.y. (at least) are 2-3 times the nominal average lunar rate [Grive, 1986; Shoemaker et al., 1990]. Further, the population of Earth-approaching asteroids and the abundance and sources of meteorites from the Moon suggest even higher rates of crater production [Rubinowitz, 1993; Warren, 1994a,b]. The 3×ET model of Tables 3-6 accommodates these inferred higher rates by assuming a crater production rate on Mars of 3 times Harman et al.'s [1981] preferred model.

To evaluate the consistency of these models of meteorite ejection and cratering rates, Table 3 shows predicted cratering rates and numbers of craters produced on Martian surfaces of various ages over the last 200 m.y., approximately twice the cosmic ray exposure ages of the NC. There is no a priori reason to begin counting meteorite-forming events at 11 Myr, to just cover and include the NC ejection event (Table 1), ejection-forming events on Mars at 20 m.y. could yield meteorites on Earth with only slightly lower probability than events at 11 Myr. (Figure 1 of Wetherill [1984]). Tables 4 and 5 assume that the given crater production models were followed over the past 250 m.y. and 1800 m.y., because the choice of crater production model influences the absolute ages of Martian surfaces inferred from crater counts. These time intervals were chosen to include the crystallization ages of the S and NC, respectively (180 Ma and 1300 Ma, Table 1); as before, there are no a priori reasons for limiting these time intervals to be the crystallization ages.

The assumed cratering rate has a large, compounded effect on the calculated number of potential S or NC source craters because increased crater production decreases the estimated absolute ages of surfaces based on crater count statistics. Thus, increased cratering rates yield more craters per unit time, and more surface area younger than a given age. The effects on this compounding effect is can be seen by comparing the HT and 3×ET models in Table 4, a tripling of the cratering rate yields
Figure 2. Potential Martian source areas for shergottite (S) meteorites, shown in black on map of Mars. Stippled areas are heavily cratered highlands, diagonal stripes show Valles Marineris. a) The 9IR crater production model (Table 3), potential sources restricted to basaltic units of upper Amazonian age. Includes units Aop, Aaa, Ats, Aa3, Aaa4, and Aatu of Scott et al. [1987] and Tanaka et al. [1988]. b) The 3x8IR crater production model, potential sources restricted to basaltic units of upper and middle Amazonian ages. Includes units in Figures 2a and Aa3, Aa2, Aaa4, Aaa3, Aatu, and Aai5 of Scott et al. [1987] and Tanaka et al. [1988].
approximately 10 times the number of likely source craters. As shown in Figure 2, the HT model implies that only surfaces of late Amazonian age (by crater densities) are <250 m.y. old (Tanaka et al., 1992); volcanic surfaces of this age are present only on a portion of the Tharsis plateau (Olympus Mons itself, and some flow fields surrounding the Olympus Mons and the Tharsis Montes) and as scattered patches in Acidalia and Arcadia Planitia (Scott et al., 1987). On the other hand, the 3xHT model implies that surfaces of middle and late Amazonian age <250 m.y. old, in addition to the above late Amazonian volcanic surfaces, one must include most of the remainder of Tharsis, and most of Amazonis and Arcadia Planitia (Figure 2) (Scott et al., 1987).

It is also assumed here that there is a high probability that at least some fragments from each impact event are transported to Earth, and thus may appear in the meteorite record (Pieters and Tanaka, 1993).

Craters: S, 1 NC. A scenario in which S and NC meteorites hailed from different source craters on Mars was considered explicitly by Vickery and Melosh (1987) and rejected because it would require a crater 212 km diameter for NC (impact of 0.5 m diameter) and a crater 250 km diameter for S (impact of 0.5 m diameter). The latter is much more unlikely (Table 4) and in fact no craters of this size are presently observed on Earth (Mouginis-Mark et al., 1982). Based on Table 4 and 5, the three crater production models predict in the last 20 m.y. the following numbers of potential source craters: model NH-2, 0.003 and 0.15; model N-83, 0.02 and 0.4; and HT 0.03 and 3.2. None of these models predict a high likelihood of separate craters for S and NC meteorites.

But production rates for SNC ejecting craters are strongly dependent on the details of the physical models involved. For instance, if one accepts a high crater production rate of 3xHT, one would expect production of 0.3 such 250 km S source craters and 2.2 such 212 km NC source craters in 20 m.y. (Tables 4, 5; Figure 2), within the range of plausibility.

Similarly, if one accepts HT crater production and a minimum crater diameter to eject the S of 216 km, one would expect production of 0.31 S source craters in 20 m.y. (Table 4), again within the range of plausibility. At this point, it seems more reasonable to accept higher cratering rates, like 3xHT, given evidence for comparably greater cratering rates on the Earth and Moon (Creevy, 1984; Shoemaker et al., 1990; Warren, 1994a,b). If this 3xHT crater production rate were coupled with ejection of S from craters of ≥25 km diameter, half the nominal size limit of Vickery and Melosh (1987), one would expect to find a likely source crater for S (Table 4).

Vickery and Melosh (1987) rejected this scenario for failing to account for the absence of meteorites from ancient and non-volcanic terrains of Mars. This point is significant and is considered below, but does not constitute a death-blow to the scenario.

Craters: S, 1 NC. A related scenario for SNC origin holds that each exposure age group (Table 1) represents a separate impact event on Mars, Case 1 of Vickery and Melosh (1987), and recommended by Jones (1989). This scenario is attractive in that the EETA79001 shergottite, which has a cosmic ray exposure age of 0.5 Ma, is in some ways distinct from the other S and EETA79001 meteorite with an igneous contact, contains xenoliths, contains unaltered Martian atmosphere, contains aluminosilicates among the terrestrial alteration materials, contains the high-pressure shock minerals ringwoodite and majorite, and contains strong evidence for a second assimilated crustal component (Fegley et al., 1984; McSween, 1985; Gooding and Mason, 1986; Jones, 1989). On the other hand, EETA79001 has essentially the same crystallization age, mineral chemistry and magma chemistry (bulk and trace element) and petrology as the other S, and its xenoliths are quite similar to the ALHA77005 and LEW88516 shergottites (McSween, 1985; Treiman, 1993b; Wadhwa et al., 1994).

This scenario is attractive in that it requires no modifications.

Table 3. Expected Production of Impact Craters on Mars, Last 20 m.y.: Craters Larger Than Given Diameter per 10^6 km^2

<table>
<thead>
<tr>
<th>Crater</th>
<th>Diameter km</th>
<th>NH-2</th>
<th>N-83</th>
<th>HT</th>
<th>3xHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.15</td>
<td>0.22</td>
<td>1.1</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.04</td>
<td>0.09</td>
<td>0.3</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.016</td>
<td>0.05</td>
<td>0.1</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.009</td>
<td>0.03</td>
<td>0.08</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>(0.003)</td>
<td>0.015</td>
<td>0.03</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>(0.0016)</td>
<td>0.009</td>
<td>0.02</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>(0.00035)</td>
<td>0.008</td>
<td>0.005</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>(0.0003)</td>
<td>0.004</td>
<td>0.005</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>


Table 4. Expected Production of Impact Craters on Mars, Last 20 m.y.: Craters Larger Than Given Diameter per 10^6 km^2

<table>
<thead>
<tr>
<th>Diameter km</th>
<th>NH-2*</th>
<th>N-83</th>
<th>HT†</th>
<th>3xHT‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.30</td>
<td>0.44</td>
<td>4.6</td>
<td>10.0</td>
</tr>
<tr>
<td>8</td>
<td>0.08</td>
<td>0.18</td>
<td>1.2</td>
<td>10.0</td>
</tr>
<tr>
<td>12</td>
<td>0.03</td>
<td>0.11</td>
<td>0.5</td>
<td>4.7</td>
</tr>
<tr>
<td>16</td>
<td>0.018</td>
<td>0.07</td>
<td>0.3</td>
<td>2.7</td>
</tr>
<tr>
<td>25</td>
<td>(0.006)</td>
<td>0.03</td>
<td>0.13</td>
<td>1.1</td>
</tr>
<tr>
<td>32</td>
<td>(0.003)</td>
<td>0.018</td>
<td>0.08</td>
<td>0.7</td>
</tr>
<tr>
<td>50</td>
<td>(0.001)</td>
<td>0.004</td>
<td>0.003</td>
<td>0.3</td>
</tr>
<tr>
<td>64</td>
<td>(0.0005)</td>
<td>0.003</td>
<td>0.002</td>
<td>0.2</td>
</tr>
</tbody>
</table>

See Table 3 for crater production models; values in parentheses may not be accurate. Areas of volcanically modified terrains drawn and calculated from Tanaka et al. (1992), including also unit Azas of lava flows covered with a thin veneer of channeled sediments (K. Tanaka, personal communication, 1994).

* Half of upper Amazonian: 2.0 x 10^6 km^2.
† Upper Amazonian: 4.1 x 10^6 km^2.
‡ Middle and upper Amazonian: 11.8 x 10^6 km^2.
to impact ejection theory. Vickery and Melosh [1987] calculated that the ejecta from each impact event need be no bigger than 0.5-m diameter and the ejecting craters need only be larger than 12-km diameter [Vickery and Melosh, 1987]. Production of so many suitable craters in the last 20 m.y. is likely given a 3xET crater production model, and barely tenable under the HT model, two craters from an expected 0.5 (Tables 4, 5). However, the scenario does require deriving all the petrologically and chronologically similar S from multiple craters, and may be inconsistent with the total mass of Mars material transferred to Earth [Vickery and Melosh, 1987]. It also fails to succeed with the absence of meteorites from ancient and nonvolcanic terrains of Mars, but this point is considered below. I do not fact this the scenario is probable, but find no cause to reject it outright.

Source Craters

It is not within the scope of this paper to exhaustively review potential source craters for the S and NC meteorites, especially in light of Mouginis-Mark et al.'s [1992] compilation of potential source craters. Mouginis-Mark et al.'s [1992] accepted the scenario that all of the SNCs came from a single source crater, and thus tried to find craters on surfaces of ~1800 Ma (the S) that could potentially have had access to 1300 Ma basaltic rocks (the NC). Their search included craters of >10 km diameter, recognizing that there were no craters of >50 km in the youngest (late Amazonian) volcanic terrain, such as are required in the models of Vickery and Melosh [1987]. Of the 25 candidate craters on Tharsis, none satisfied all constraints, particularly that it have access to S and NC lithologies, so Mouginis-Mark et al. [1992] listed their nine most likely candidates, with recommending and detracting features for each. However, if the S and NC meteorites did not come from a single crater, most of the most likely craters in Mouginis-Mark et al.'s [1992] become potential sources for one or the other meteorite group. Based on ages of surfaces into which they were emplaced, one might suggest that their craters 1, 3, or 7-9 might be the source for S (7 is most likely because it is largest), and craters 2 and 4-9 might be the source for NC [Mouginis-Mark et al., 1992].

Other Lithologies and Ages?

If the S and NC originated from different craters, ejection from Mars cannot have been caused by a unique impact, and similar impacts should have occurred on other surfaces on Mars. Where are the meteorites from those Martian surfacess? This question led Vickery and Melosh [1987], among others, to reject sceneries with multiple SNC source craters. However, given the evidence of Table 1, this question remains as a challenge to planetary geologists, impact dynamos, and meteoriticists. It should be noted that lunar meteorites, 10 total now (including paired samples), appear to be a representative sampling (by exposed area) of the major lunar surface unit highslands, mare basalts, KREEP-rich rock [Landstrom et al., 1994; Warren, 1994b].

To quantify the number of potential meteorite-ejecting impacts on Mars, one can compare the values of Tables 5 and 6, which show the numbers of craters of given sizes on volcanic surfaces ~1800 Ma and on the whole planet. The HT crater production model implies ~6 times as many meteorite-producing craters over the whole planet as on volcanic surfaces ~1800 Ma. Naively, one might expect approximately five times as many ancient and nonvolcanic Martian meteorites as young volcanic Martian meteorites. The HT crater production model implies ~3 times as many meteorite-producing craters over the whole planet as on the ~1800 Ma volcanic surfaces. Again naively, one might expect twice as many ancient and nonvolcanic Martian meteorites as young volcanic Martian meteorites.

The lack of Martian meteorites from ancient and nonvolcanic terrains admits many possible explanations. First, older and nonvolcanic surfaces might not contain coherent enough material to be ejected as 0.5-m or larger fragments [Jones, 1989]. Second, especially aeolian surfaces are likely to be weak. Older volcanic surfaces are, however, interpreted to be coherent and strong [Tanaka and Golombek, 1989; Mouginis-Mark et al., 1992].

The physical properties of older surface materials may preclude ejection of meteorite-sized fragments. Calculations of impact ejection by spallation [Vickery and Melosh, 1987] considered ejection only from coherent rock surfaces. Spallation as a meteorite ejection mechanism is likely to be less effective from a weak (low tensile strength) surface (equation 1 of Vickery and Melosh [1987]; Jones [1989]), e.g., loosely packed,

### Table 5. Expected Production of Impact Craters on Mars, Last 20 m.y.: Craters Larger than Given Diameter on Volcanic Surfaces Aged ~1800 Ma

<table>
<thead>
<tr>
<th>Diameter km</th>
<th>NH-2*</th>
<th>NH-3</th>
<th>HT</th>
<th>3 x HT**</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.1</td>
<td>1.7</td>
<td>27</td>
<td>186</td>
</tr>
<tr>
<td>8</td>
<td>0.3</td>
<td>0.7</td>
<td>7.0</td>
<td>48</td>
</tr>
<tr>
<td>12</td>
<td>0.13</td>
<td>0.4</td>
<td>3.2</td>
<td>22</td>
</tr>
<tr>
<td>16</td>
<td>0.07</td>
<td>0.3</td>
<td>1.8</td>
<td>12.4</td>
</tr>
<tr>
<td>25</td>
<td>(0.02)</td>
<td>0.12</td>
<td>0.76</td>
<td>5.2</td>
</tr>
<tr>
<td>32</td>
<td>(0.01)</td>
<td>0.07</td>
<td>0.47</td>
<td>3.2</td>
</tr>
<tr>
<td>50</td>
<td>(0.004)</td>
<td>0.02</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>(0.002)</td>
<td>0.01</td>
<td>0.12</td>
<td>0.8</td>
</tr>
</tbody>
</table>

See Table 3 for crater production models, values in parentheses may not be accurate. Areas of volcanically modified terrains drawn and calculated from Tanaka et al. [1988], including also unit Au of lava flows covered with thin veneer of channel sediments (K. Tanaka, personal communication, 1994).

### Table 6. Expected Production of Impact Craters on Mars in Last 20 m.y.: Craters Larger than Given Diameter on Whole Planet

<table>
<thead>
<tr>
<th>Diameter km</th>
<th>NH-2*</th>
<th>NH-3</th>
<th>HT</th>
<th>3 x HT</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>21</td>
<td>32</td>
<td>160</td>
<td>490</td>
</tr>
<tr>
<td>8</td>
<td>5.5</td>
<td>13</td>
<td>40</td>
<td>130</td>
</tr>
<tr>
<td>12</td>
<td>2.4</td>
<td>7.6</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>16</td>
<td>1.3</td>
<td>5.1</td>
<td>11</td>
<td>33</td>
</tr>
<tr>
<td>25</td>
<td>(0.44)</td>
<td>2.2</td>
<td>4.6</td>
<td>14</td>
</tr>
<tr>
<td>32</td>
<td>(0.23)</td>
<td>1.3</td>
<td>2.8</td>
<td>8.4</td>
</tr>
<tr>
<td>50</td>
<td>(0.07)</td>
<td>0.4</td>
<td>1.2</td>
<td>3.5</td>
</tr>
<tr>
<td>64</td>
<td>(0.04)</td>
<td>0.2</td>
<td>0.7</td>
<td>2.2</td>
</tr>
</tbody>
</table>

See Table 3 for crater production models.
heavily fractured, brecciated, or weathered surface material, as some of the shock energy will be expended compacting and deforming the material. Perhaps rocks of the Hesperian-age oxidized plains, containing extensive volcanic units on Mars, are too broken or weathered for impact ejection (K. Tanaka, personal communication, 1994). And third, perhaps materials from the old impacts do fall to Earth and are not recognized as Martian or meteoritic [Lindstrom et al., 1994]. The ALH84001 meteorite, of Martian origin, is an orthopyroxenite, and was misclassified and studied for eight years as a common diogenite (Mittlefleldt [1994], see below). Given the varied geology of Mars, we may reasonably expect Martian meteorites to include altered basaltic rocks, dunites, and carbonates, and possibly to include olivine, granites, or anorthosites. So none of these meteorites have been reported, notably from the blue ice areas of Antarctica, where terrestrial rocks are rare to absent (R.P. Harvey, personal communication, 1994). However, it is not clear that the scientific community would seriously consider a reported "sandstone from the sky" as possibly being a legitimate meteorite.

And ALH84001?

The ALH84001 meteorite, long thought to be a diogenite, has recently been reclassified as a unique Martian meteorite, not belonging to the S, N, or C groups (Mittlefleldt, 1994). ALH84001 is an orthopyroxenite, ruling out any affinity to NC by bulk magma composition and crystallization sequence arguments (Table 1 and above). Also unlike the NC, ALH84001 is strongly shocked (plagioclase converted to maskelynite). Affinities to the S, though stronger, are also questionable. By mineralogy, ALH84001 could be a cumulate member of S, similar to the EETA79001 (Table 1) (Mcllwain and Jasawich, 1983). However, the parent magma of ALH84001 was enriched in incompatible elements (Mittlefleldt, 1994), while S are depleted. In addition, ALH84001 contains abundant products of aqueous alteration, dominantly Fe-Mg carbonates, while the S contain only rare products of alteration materials, dominantly Ca-bearing carbonates and sulfates (Golding, 1992, and references therein). And unlike any SNC, ALH84001 appears to show evidence for multiple shock events: a late event that yielded maskelynite after plagioclase and caused cracking and fracturing; and an early event responsible for pervasive crush zones (fract gauge), subsequently annealed and recrystallized.

Although data on ALH84001 are limited and preliminary, it seems likely that it is much older than the other Martian meteorites. An age great is suggested by the presence of recrystallized crush zones (Mittlefleldt, 1994; Treiman, 1994b), which appear to record an early impact brecciation followed by annealing at high temperatures. This event was probably not the impact that caused ALH84001 from Mars; for the other SNCs, the ejecting impact events did not cause recrystallization, even of maskelynite [Duke, 1968], suggesting a very low temperature history. ALH84001 can best be compared with lunar "recrystallized cataclastic rocks" of highland derivation [Stoffler et al., 1980]. Following this analogy, ALH84001 could possibly be from the Martian highlands, a product of Martian magmatism more than 3 Gyr ago.

Acknowledgments: I am greatly indebted to J. Jones for his papers on meteorites and for many detailed discussions. Additional discussions with W. Kiefer, A. Vickery, M.M. Lindstrom, D. Mittlefleldt, P. Warren, and R. Harris (and many others) have helped clarify my thoughts. O. McColm assisted with access to some catalog literature.


References


Hartmann, W.K., et al., Chronology of planetary volcanism by


Multiple Source Areas for Martian Meteorites


Treiman, A.H., The Martian sources of the SNC meteorites (two, not one), and what one and can’t be learned from the SNC meteorites (abstract), Mars: Past, Present, and Future — Results from the ASUFT Program, pp. 49-51, Lunar and Planetary Institute, Houston, 1993a.


Treiman, A.H., Xenoliths in the EETA79001 shergonite: Geological and astronomical implications of similarities to the ALHA77005 and LEW88516 shergonites (abstract), Meteoritics, 28, 451, 1993a.


Allan H. Treiman, Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston TX 77058-1113. (e-mail: treiman@lpi.jsc.nasa.gov)

(Received February 28, 1994; revised July 14, 1994; accepted August 24, 1994.)
Simultaneous adsorption of CO₂ and H₂O under Mars-like conditions and application to the evolution of the Martian climate

Aaron P. Zent and Richard C. Quinn
SEIT Institute, Space Sciences Division, NASA Ames Research Center, Moffett Field, California

Abstract. The Martian regolith is the most substantial volatile reservoir on the planet; estimates of its adsorbed inventory have been based on simple measurements of the adsorption of either water or CO₂ in isolation. Under some conditions, H₂O can poison adsorbate surfaces, such that CO₂ uptake is greatly reduced. We have made the first measurements of the simultaneous adsorption of CO₂ and H₂O under conditions appropriate to the Martian regolith and have found that at H₂O monolayer coverage above about 0.5, CO₂ begins to be displaced into the gas phase. We have developed an empirical expression that describes our co-adsorption data and have applied it to standard models of the Martian regolith. We find that currently, H₂O does not sufficiently displace CO₂, implying that the adsorbate inventories previously derived may be inaccurate, not more than 3–4 kPa (30–40 mbar). No substantial increase in atmospheric pressure is predicted at higher obliquities because high-latitude ground ice buffers the partial pressure of H₂O in the pores, preventing high monolayer coverages of H₂O from displacing CO₂. The peak atmospheric pressure at high obliquity does increase as the total inventory of exchangeable CO₂ increases.

1. Introduction

The regolith, cap, and atmosphere constitute an integral part of the Martian climate system. Water and carbon dioxide exchange between these three primary reservoirs on a variety of timescales. Over astronomical timescales of 10⁵–10⁶ years, geologically significant amounts of volatiles can exchange, profoundly affecting the Martian climate.

The exchange of H₂O and CO₂ have always been modeled separately, and the exchange patterns predicted by numerical climate models have been described separately [e.g., Farmer and Domn, 1979; Toon et al., 1986; Fanale et al., 1983, 1986; Zent et al., 1987]. This practice omits any interactions between the two species from consideration. In particular, the competition for a finite number of adsorption sites in the Martian regolith, a phenomenon known as co-adsorption, may significantly alter the distribution of CO₂ from that predicted in single volatile models. In previously reported laboratory measurements, the presence of H₂O vapor severely poisons the CO₂ adsorptive capacity of commercial zeolites [Scott, 1968; Carter and Huxlin, 1974]. However, these measurements were made under conditions very different from those encountered in the Martian regolith.

The objective of this paper is to measure the simultaneous adsorption of CO₂ and H₂O under conditions that are appropriate to the Martian regolith, and to apply those measurements to a numerical model of the Martian CO₂ inventory, and quasi-periodic climate change on Mars.

We undertook this study because CO₂ adsorption measurements, made under conditions of zero relative humidity, have been used to estimate the inventory of exchangeable CO₂ in the Martian atmosphere + cap + regolith [e.g., Fanale and Cannon, 1971], and also to predict the evolution of the Martian climate caused by oscillation of orbital elements [e.g., Fanale and Cannon, 1974; Toon et al., 1980; Zent et al., 1987; François et al., 1990 (see Kieffer and Zent [1992] for review)]. Failure to account for a potentially first-order process such as adsorptive competition jeopardizes the conclusions reached in previous studies.

2. Background

Previous Climate Models

Previously published climate models rely on the assumption of adsorptive equilibrium of CO₂ between the atmosphere and regolith. Numerical representations of adsorptive equilibrium are based on empirical fits to data acquired under conditions of zero relative humidity [Fanale and Cannon, 1971, 1978; Zent et al., 1987]. One assumes that the isotherms so derived correctly describe the behavior of the entire Martian regolith. One can then solve the heat diffusion equation for annually averaged insulation and determine the temperature profile to any depth of interest. One must assume values for the thickness of the regolith, for its thermal conductivity, and for the planetary heat flux.

The temperature profile of the regolith as a function of latitude, combined with the atmospheric pressure, permit one to calculate how much CO₂ is adsorbed on the regolith as a function of longitude and depth. The current exchangeable CO₂ inventory is estimated as that which simultaneously satisfies the adsorptive equilibrium, the observed atmospheric pressure, and the absence of large polar CO₂ caps. By varying assumptions of the depth of the regolith and its specific surface area, a variety of CO₂ inventories consistent with observations can be generated.

The atmospheric pressure history over an obliquity cycle...
can then be found by recalculting the thermal gradient at each obliquity, and rebalancing the C02. If the atmospheric pressure exceeds the vapor pressure at polar temperatures, permanent polar caps are assumed to exist, and they are then allowed to fix the atmospheric pressure. The adsorbate inventory is recalculated in accord with the new atmospheric pressure, and the remainder of the C02 inventory is assumed to be in the quasi-permanent polar caps. This climate model approach was developed by Fanale et al. [1982] and the same scheme will be employed in this study.

Previous climate models predict that at obliquities slightly lower than at present (roughly 25°), permanent CO2 polar caps will form, trapping a significant fraction of the C02 inventory. At higher obliquities, the high-latitude regolith releases C02, but this is somewhat compensated by cooling of the low-latitude regolith, which takes up some of the C02. [Toon et al., 1980; Fanale et al., 1982]. The result is a slight increase in pressure as the obliquity increases (Figure 1).

The Role of H2O

In laboratory measurements of the simultaneous adsorption of water and C02, the adsorptive abundance of C02 was substantially depressed from its RGS = 0 levels. Carter and Husain [1974] passed a gas stream carrying C02 and H2O through a packed zeolite bed at 273 K, and examined the breakthrough curves of the two gases. The principle is identical to that used in gas chromatography, where the greater the adsorptive coverage on the bed material, the longer the gases take to appear at the end of the bed. They found that C02 breakthrough was significantly faster when water was present in the gas stream, indicating that the C02 adsorptive capacity of the zeolite material was being poisoned. Importantly, the breakthrough curve for water was unaffected by the presence of C02, indicating that water adsorption was unaffected by C02. The explanation of this asymmetry is that the lifetime of an adsorbed molecule on a surface is dependent on the partial molar enthalpy of adsorption

\[
\tau = \tau_0 \exp \left( \frac{Q}{RT} \right)
\]

where \(\tau_0\) is a characteristic frequency for the vibration of the adsorbed, characteristically of the order of \(10^{-14}\) s. In Table 1 the characteristic lifetime is shown as a function of \(Q\) for the adsorbate.

![Figure 1](image_url)  
**Figure 1.** Previous models of the evolution of the Martian climate with obliquity were calculated without regard to the effects of H2O. This example is after Zent et al. [1987].

298 K. Typical values for the molar enthalpy of adsorption for water and C02 are of the order of \(4.48 \times 10^4\) J mol\(^{-1}\) [Mooney et al., 1972] and \(1.96 \times 10^2\) J mol\(^{-1}\) [Zent et al., 1987], respectively. At 210 K, this would imply adsorbed lifetimes of \(7 \times 10^{-3}\) s for H2O and \(4 \times 10^{-3}\) s for C02, a difference of 6 orders of magnitude. C02 is present at roughly 800 times the number density of H2O in pore gases, and so collides and adsorbs more frequently, but the difference in adsorbed lifetime suggests that adsorbed H2O is essentially unaffected by the presence of C02 at Mars-like conditions. We shall see that our data support this conclusion.

Scott [1968] also reported that C02 was displaced from adsorption sites by the presence of H2O. In his study, co-adsorption for purification of He streams in gas-cooled nuclear reactors was studied experimentally at adsorbent temperatures of 298 K, by passing a He stream at 10-30 atm over zeolite beds, and metering in both water and CO2. He found that sorbed CO2 was irreversibly replaced by sorbed water and that CO2 loading was dependent on water concentation.

If C02 is also substantially displaced from Martian materials, then the current estimates of the adsorptive capacity of the regolith may be too high. The Martian regolith is at RG = 1 throughout the latitude-depth domain in which ground ice is stable, roughly poleward of 40° latitude, with notable outliers in high-altitude areas [Paige, 1992; Mellon and Jakosky, 1993]. Our objective in this paper is to determine how H2O and C02 compete for adsorption sites as a function of temperature and their respective partial pressures at Mars-like conditions, and then apply that relationship to a mathematical model of quasi-periodic climate variations throughout the Martian obliquity cycle.

3. Laboratory Effort

**Experiment**

In order to test the adsorptive capacity of Martian regolith materials, we measured co-adsorption of H2O and C02 under conditions appropriate to the Martian regolith, as well as the isotherms of the individual species in isolation.

The manifold used for the adsorption measurements (Figure 2) allows the temperature of the soil, and the partial pressures of C02 (99.996%) and H2O to be controlled. The partial pressure of C02 is controlled with a gas-regulating valve and measured with a capacitance manometer. The partial pressure of the H2O reservoir is controlled by submerging the H2O reservoir in a cryogenic bath. The temperature of the H2O reservoir is measured with a thermocouple temperature probe in direct contact with the ice. Doubly distilled H2O is then admitted to the system, and the H2O pressure is adjusted to the desired value.

<table>
<thead>
<tr>
<th>(Q, \text{ J mol}^{-1})</th>
<th>(t) at 298 K, s</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 \times 10^2)</td>
<td>(5.21 \times 10^{-14})</td>
</tr>
<tr>
<td>(1 \times 10^3)</td>
<td>(7.49 \times 10^{-14})</td>
</tr>
<tr>
<td>(5 \times 10^3)</td>
<td>(2.31 \times 10^{-13})</td>
</tr>
<tr>
<td>(1 \times 10^4)</td>
<td>(2.83 \times 10^{-12})</td>
</tr>
<tr>
<td>(2 \times 10^4)</td>
<td>(1.00 \times 10^{-10})</td>
</tr>
<tr>
<td>(5 \times 10^4)</td>
<td>(2.90 \times 10^{-7})</td>
</tr>
<tr>
<td>(1 \times 10^5)</td>
<td>(1.68 \times 10^0)</td>
</tr>
</tbody>
</table>
was degassed by placing it in the reservoir, freezing with liquid nitrogen, degassing, melting, and then refreezing. Four cycles of freezing, followed by degassing and melting, were carried out.

We used palagonite, collected from the summit of Mauna Kea, and sieved to <38 μm, as our Mars analog material. The palagonite is conditioned before each run by heating it overnight to 120°C under vacuum. The temperature of the soil is controlled with a second cryogenic bath and measured with a thermocouple probe in direct contact with the soil. The soil is kept at a higher temperature than the H₂O reservoir in order to keep the relative humidity in the soil chamber below unity.

The dead space in the soil sample is calculated by expanding He into the soil reservoir and noting the pressure drop. It is assumed that there is negligible He adsorption. It was determined that adsorptive equilibration takes at least four hours. A series of experiments was performed in which the only variable was the equilibration time. Analyses on all samples exposed 4 hours or longer yield the same adsorbate loading, within experimental error. For one of our 210 K measurements, we carried out the exposure for 12 hours to investigate kinetic barriers to equilibration. Since our soil samples were small (roughly 0.1 g), four hours was adequate time to charge them with their equilibrium load, even at 210 K.

After equilibration, the adsorbed gases are desorbed from the sample, trapped, and analyzed by gas chromatography. To trap the CO₂ and H₂O, the soil chamber is heated to 120°C for 2 hours, and the desorbed gases are captured in a U-trap cooled in a liquid nitrogen bath. Spectroscopic and differential scanning calorimetry studies of palagonite have shown that virtually all H₂O is desorbed at this temperature [Bruckenthal, 1987].

After the trapping step is complete, the trap is removed from the manifold, and desorbed volatiles are analyzed using a Varian 3400 gas chromatograph (GC) with thermal conductivity detector. To analyze a sample, the trap is heated to 120°C to minimize gas adsorption on the walls of the trap, and a sample is introduced into the GC column using a gas sampling valve also heated to 120°C. A 6 ft × 1/8 inch HayeSep N 60/80 mesh column, heated to 140°C was used to separate the gases. Peak elution areas are calculated using an onboard integrator.

**Results**

**Measurements.** We first measured zero relative humidity CO₂ isotherms in order to establish a baseline against which to compare the effects of H₂O. We measured the RH = 0 isotherms by simple pressure drop; the GC was not used. We used these isotherms to calculate a specific surface area of 94.8 m² g⁻¹ for the palagonite. Brunauer, Emmett, and Teller (BET) specific area determinations made using N₂ at 77 K were 116 m² g⁻¹. The data were comparable to previously measured data for CO₂ on palagonite [Zent et al., 1987], when corrected for differences in the specific surface area of the soil. Table 2 shows the data for the zero relative humidity CO₂ isotherms, which we will use as controls to examine the amount of desorption caused by H₂O.

<table>
<thead>
<tr>
<th>Adsorbed CO₂, g/g</th>
<th>Pressure, Pa</th>
<th>Temperature, K</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0179</td>
<td>15.3</td>
<td>215.8</td>
</tr>
<tr>
<td>0.0223</td>
<td>135.5</td>
<td>215.8</td>
</tr>
<tr>
<td>0.0265</td>
<td>146.0</td>
<td>215.8</td>
</tr>
<tr>
<td>0.0286</td>
<td>166.6</td>
<td>215.8</td>
</tr>
<tr>
<td>0.0302</td>
<td>391.3</td>
<td>215.8</td>
</tr>
<tr>
<td>0.0319</td>
<td>611.5</td>
<td>215.8</td>
</tr>
<tr>
<td>0.0077</td>
<td>72.0</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0089</td>
<td>197.8</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0110</td>
<td>384.3</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0128</td>
<td>571.8</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0144</td>
<td>720.1</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0154</td>
<td>962.2</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0169</td>
<td>1259.8</td>
<td>210.0</td>
</tr>
<tr>
<td>0.0037</td>
<td>257.0</td>
<td>242.0</td>
</tr>
<tr>
<td>0.0052</td>
<td>516.4</td>
<td>242.0</td>
</tr>
<tr>
<td>0.0064</td>
<td>724.9</td>
<td>242.0</td>
</tr>
<tr>
<td>0.0074</td>
<td>1000.6</td>
<td>242.0</td>
</tr>
<tr>
<td>0.0086</td>
<td>1281.5</td>
<td>242.0</td>
</tr>
</tbody>
</table>
We also made a few measurements of \( \text{H}_2\text{O} \) in isolation, with the gas chromatograph. The data gathered on \( \text{H}_2\text{O} \) in isolation are shown in Table 3. The soil temperature was 294.5 K (room temperature), and the temperature of the \( \text{H}_2\text{O} \) ice was 273.15 K, maintained by an ice-water bath. The partial pressure of \( \text{H}_2\text{O} \) was therefore 611.6 Pa, and the relative humidity in the soil chamber was 0.193. High adsorptive coverages, 40 mg g\(^{-1}\), were measured.

We next measured the co-adsorption isotherms at a variety of temperatures, partial pressures, and relative humidities (Table 4). At \( T_{\text{eff}} = 210 \) K, we found that \( \text{H}_2\text{O} \) has limited effectiveness in displacing \( \text{CO}_2 \) from its \( R_p = 0 \) coverage. In the first series of measurements, we buffered the ice at 184 K (\( R_H = 0.013 \)), in the second, at 206 K (\( R_H = 0.47 \)), and in a third at 209 K (\( R_H = 0.87 \)). In the final case the measured \( \text{H}_2\text{O} \) coverage is only about 7% of a complete monolayer, and the \( \text{CO}_2 \) adsorptive coverage is reduced by about 25% from its zero relative humidity coverage. The data are shown in Figure 3, along with the zero relative humidity data for comparison. This result is significant because these conditions are characteristic of much of the regolith. Further, the \( \text{CO}_2 \) adsorption data measured via gas chromatography at the lowest humidity nearly overlie the isotherms measured via pressure drop, giving us confidence in both techniques.

At higher temperatures \( (T_{\text{eff}} = 242 \) K) the presence of water begins to be felt. We measured co-adsorption isotherms at \( R_p = 9.3 \times 10^{-3} \) and 0.5. At \( R_p = 0.5 \) the measured \( \text{H}_2\text{O} \) coverage is about 0.25 monolayers, and the \( \text{CO}_2 \) coverage is depressed by roughly 50% from its dry coverage. The data are shown in Figure 4.

The adsorption of \( \text{H}_2\text{O} \) is virtually independent of the partial pressure of \( \text{CO}_2 \). This is what is expected from previously published work. Surface area analyses suggest that at \( R_H = 0.47 \), there is only about 6% of a monolayer coverage at this partial pressure. The more water in the system, the more error in the data.

Empirical fits. The \( R_p = 0 \) \( \text{CO}_2 \) data can be fit by least squares to a logarithmic form as follows:

\[
\rho_\text{\text{H}_2\text{O}} = A_\text{\text{H}_2\text{O}} \rho^\gamma \text{P}^\beta
\]

where \( A_\text{\text{H}_2\text{O}} \) is the specific surface area of the soil, 9.48 \times 10^6 m\(^2\) kg\(^{-1}\), and \( \gamma \), \( \beta \), and \( \beta \) are unknowns to be determined. We find a best fit to the data for \( \gamma = 0.2786 \), and \( \beta = -4.0711 \). Figure 5 shows the adsorption data and the isotherms fit to them.

We assume that the adsorptive coverage of \( \text{H}_2\text{O} \) is independent of the \( \text{CO}_2 \) partial pressure or coverage. This assumption is supported by our data, by the data from Carter and Hussain (1974), and by consideration of the relationship between \( \Delta H_f \) and adsorbed lifetimes. By taking the co-adsorption data described below, we are able to fit an isotherm of the same form as (2) to the \( \text{H}_2\text{O} \) data. The corresponding values for the \( \text{H}_2\text{O} \) isotherm constants are

\[
\rho_\text{\text{H}_2\text{O}} = A_\text{\text{H}_2\text{O}} \rho^\gamma \text{P}^\beta
\]
$ZENT AND QUINN: CO_2-H_2O CO-ADSORPTION$

Figure 5. CO_{2} data and isotherms fit to (2).

$8 = 0.7842, \gamma = 0.57904, and \beta = -3.209.$ The H_{2}O isotherms are shown in Figure 6, along with the data.

There are several difficulties with completing orderly isotherms for H_{2}O. One such problem derives from our use of cryogenic baths to control the partial pressure of H_{2}O in the experiment. There are only certain partial pressures available, limited by the supply of cryogenic chemicals, which have specified eutectics. Second, the technique we have employed assumes that all H_{2}O is desorbed from the soil and is vaporized in the LN_{2} trap prior to injection into the GC column. If either of these conditions fails to apply, the technique will underestimate the adsorbed H_{2}O.

The H_{2}O measurements in the co-adsorption data show a considerable standard deviation, which may arise from violation of any of these assumptions, or which may simply arise from hysteresis in the adsorption branch of the isotherm.

In deriving an empirical expression that will describe co-adsorption, we assume a modified form of the Langmuir isotherm. Studies of mixed gas adsorption are fragmentary because of difficulties with experimental and theoretical investigations [Jaroniec and Madey, 1988]. Discussion of the theoretical foundations of co-adsorption is beyond the scope of this paper. The problem is particularly bedeviling for heterogeneous surfaces, where many different functional groups are exposed to the adsorbates, and a variety of adsorption energies characterize individual adsorption and desorption events. Our approach is to derive our own empirical equation, which we use to fit to the data, and for subsequent interpolation. Although the derivation depends upon a number of unrealistic assumptions, if thought of as an empirical representation rather than a physical explanation, it serves well [Carter and Husein, 1974; Sircar and Kumar, 1983; Golden and Sircar, 1994]. We assume that only one molecule can adsorb at a given site; if $f$ represents the fraction of collisions that result in CO_{2} adsorption, then the adsorption rate is

$$f(1 - \theta - \theta_{w})$$

where $\theta_{c}$ is the fraction of adsorption sites already occupied by CO_{2} and $\theta_{w}$ is the fraction of sites occupied by H_{2}O. The adsorption rate depends on the lifetime of adsorbed molecules and the coverage, so at equilibrium,

$$f(1 - \theta - \theta_{w})$$

or

$$\frac{f_{n}C_{CO_{2}}}{4\nu \exp \left(\beta/T\right)} \frac{\theta_{c}}{1 - \theta - \theta_{w}}$$

if we let $n_{CO_{2}} = PV_{RT}$, per the ideal gas equation, and take the root mean square velocity of the molecules as

$$\nu = \left(\frac{8RT}{\pi m}\right)^{1/2}$$

then we can define a constant that is independent of the temperature and partial pressure,

$$\alpha = \frac{(PV_{RT})^{1/2}}{4R\nu}$$

so (5) reduces to

$$\frac{aP}{T^{1/2} \exp \left(\beta/T\right)} = \frac{\theta_{c}}{1 - \theta - \theta_{w}}$$

which can be solved for $\theta_{c}$, the fractional monolayer coverage of CO_{2}, as

$$\theta_{c} = \frac{(1 - \theta_{w})aP}{aP + T^{1/2} \exp \left(\beta/T\right)}$$

This equation can be linearized by rearranging, and taking the log of all terms,

$$\ln (1 - \theta_{c} - \theta_{w}) - \ln (\theta_{c}) + \ln (P) - \ln (T^{1/2})$$

$$= \frac{\theta_{c}}{T} - \ln (\alpha)$$

and solved for $\alpha$ and $\beta$. We set the left-hand side of (10) to $Y$ and set $X = 1/T$. In calculating the value of $Y$, we use the measured co-adsorption data as well as the zero relative
4. Climate Model

Description

The numerical model employed to extrapolate the laboratory results to the Martian climate is similar in many respects to previously published models of quasi-periodic climate change on Mars.

The regolith is divided into chunks, bounded by latitude and depth, from the pole to the equator and from the bottom of the regolith to the surface. Some thickness must be assumed for the powdered component of the regolith, which may be expected to follow the adsorption isotherms measured in the laboratory; this is in distinction to the megaregolith, which is composed largely of blocks and fractured basement which is insignificant in terms of specific surface area exposed to the atmospheric gases.

We assume that the heat diffusion equation describes the flow of heat through the Martian regolith:

$$\frac{dT}{dt} = \frac{K}{\rho c} \nabla^2 T$$

(11)

where $T$ is temperature, $t$ is time, $K$ is the thermal conductivity of the regolith (which must be assumed), $\rho$ is the regolith density (we assume 2000 kg m$^{-3}$), and $c$ is the heat capacity of the regolith. We use a temperature-dependent heat capacity applicable to dry silicates over the appropriate temperature range (Winter and Sarri, 1968). The upper boundary condition is the surface temperature appropriate to the annual average insolation at each latitude, and the lower boundary condition is

$$\frac{dT}{dz} = -\frac{Q}{K}$$

(12)

where $Q$ is the planetary heat flow (0.03 J m$^{-2}$ s$^{-1}$) and $K$ is the thermal conductivity of the regolith (0.8 W m$^{-1}$ K$^{-1}$). Given these assumptions, a characteristic temperature is calculated for each chunk of the regolith. We take that temperature as the average temperature of the bounding grid points for each chunk.

We next assume, per the laboratory data and literature, that the adsorptive coverage of $H_2O$ is independent of the partial pressure of $CO_2$. The equilibrium $H_2O$ coverage of each chunk is calculated according to (2). The annual average vapor pressure of $H_2O$ in the atmosphere is calculated by averaging the vapor pressure over the poles at 20 time steps throughout the Martian year. Because of the non-linearity of the vapor pressure curve, one cannot calculate the annual average vapor pressure from the annual average temperature. The vapor pressure expression is

$$P = a \exp(b/T)$$

(13)

where $a = 3.56 \times 10^{12}$ (Pa) and $b = -6141.7$ K. If the regolith is saturated at the temperature of a given regolith chunk, then the partial pressure of that chunk is assumed to be controlled by ice at the local temperature.

The adsorptive coverage of $CO_2$ can be found in the following way. The solution to (11) will provide the fractional monolayer coverage on a per kilogram basis, for a given chunk of the regolith, if $\theta_0$, $T$, and $P$ are provided. The first two terms are already calculated for each chunk. The pressure must be found iteratively as described below.

The calculation proceeds by assuming a value for the total amount of exchangeable $CO_2$ in the atmosphere + regolith system, and finding the atmospheric pressure $P$, such that

$$\sum CO_2 = P A_a$$

$$+ 2p, M_c \sum_{i=1}^{10} \sum_{l=1}^{10} V_{i,l}(1 - \theta_{o2}) aP + \frac{T_i}{T_i} \exp\left(\frac{bT_i}{T_i}\right)$$

(14)

where the final term is summed over each chunk of regolith through each depth $z$ and latitude $l$. The fractional monolayer coverage must be multiplied by the mass of a monolayer, $M_c$, and the density of the regolith $\rho_l$. The term $V_{i,l}$ is the volume of each chunk of regolith, and $A_a$ is the total surface area of Mars. The last term is doubled because we sum over only one hemisphere. If the atmospheric pressure $P$ is greater than the equilibrium vapor pressure over $CO_2$ for the temperature predicted for the cap, then quasi-permanent $CO_2$ caps are assumed to exist and to set the atmospheric pressure. The mass of the caps is then found from

$$M_c = \sum CO_2 = P A_a$$

$$- 2p, M_c \sum_{i=1}^{10} \sum_{l=1}^{10} V_{i,l}(1 - \theta_{o2}) aP + \frac{T_i}{T_i} \exp\left(\frac{bT_i}{T_i}\right)$$

(15)

Results

$CO_2$ inventory. Straightforward inspection of the data in Figures 3 and 4, which represent conditions throughout much of the regolith, suggests that $H_2O$ is inefficient in displacing $CO_2$ under current conditions. More elaborate numerical modeling bears out this impression.

Although the total inventory of exchangeable $CO_2$ ($CO_2-H_2O$)

is supplied to the model, it is possible to constrain its value. Given assumptions of the specific surface area of the regolith materials ($A_l$), the geothermal heat flux $Q$, and the cap
ZENT AND QUINN: CO₂-H₂O CO-ADSORPTION

Figure 8. The predicted variation of atmospheric pressure with obliquity for the baseline case where Z = 75 m and \( \Sigma \text{CO}_2 = 722.8 \text{ kg m}^{-2} \) (cf. Figure 1).

Figure 9. The full range of variation predicted by our model for the largest, baseline, and smallest exchangeable CO₂ inventory.

albedo, there are fixed combinations of \( \Sigma \text{CO}_2 \) and total regolith thickness Z for which the model predicts the observed conditions at the present orbital configuration. As baselines, we assume that the specific surface area of the Martian regolith materials is 17 m² g⁻¹, consistent with the conclusions of Ballow et al. [1978], and planetary heat flow is 0.03 m² s⁻¹ g⁻¹ [Fanale et al., 1982]. We use a CO₂ cap albedo of 0.6, invariant with the obliquity of the planet.

The criteria for accepting a combination of \( \Sigma \text{CO}_2 \) and Z are that the model must predict an atmospheric pressure of 750 Pa and produce no significant CO₂ caps. We find a range of plausible inventories, shown in Table 5. Unless the finely powdered regolith is deeper than 100 m on a global average, then the most likely exchangeable CO₂ inventory is in the range of 1-4 kPa, which is consistent with estimates of the exchangeable inventory made without consideration of co-adsorption [e.g., Zent et al., 1987]. The full range of possible exchangeable inventories is bounded by 750 Pa, which is what is observed in the atmosphere at present, and which implicitly implies no significant powdered regolith; the upper limit is 19 kPa, equivalent to 5035 kg m⁻², which is the appropriate value for 1 km of powdered regolith.

Climate evolution. At obliquities higher than at present, the atmospheric pressure depends on the assumed values of Z and \( \Sigma \text{CO}_2 \). We can select any combination of Z and \( \Sigma \text{CO}_2 \) that appear in Table 5 and explore the evolution of the system as the obliquity and eccentricity of Mars vary. Our baseline is Z = 75 m and \( \Sigma \text{CO}_2 = 722.8 \text{ kg m}^{-2} \), and we vary the obliquity of the planet through its full range of 10°-45° [Biller, 1996; Ward and Rudy, 1991]. Figure 8 shows the predicted atmospheric pressure history for our baseline case, as a function of obliquity. In comparison with Figure 1, we can see that including the physics of co-adsorption does not significantly alter the predicted climate history of Mars. At the highest obliquities, the atmospheric pressure increases to about 1000 Pa (10 mbar). As the obliquity of the planet decreases, so does the atmospheric pressure, as the high-latitude regolith cools and adsorbs more CO₂. The H₂O adsorbed load of the high-latitude regolith actually decreases, since more H₂O molecules find ice to be the lowest energy state.

We explore the effects of our baseline assumptions in Figure 9, where the most extreme combinations of Z and \( \Sigma \text{CO}_2 \) are assumed. Where 1 km of powdered regolith is assumed, and 19 KPa of exchangeable CO₂, we find that atmospheric pressures increase to 1440 Pa at \( \theta = 45° \). Quasi-permanent polar caps appear as the obliquity goes below 20.6°.

If we assume only 15 m of powdered regolith over the planet, and an exchangeable inventory of 1.1 kPa, the pressure increases to no more than 850 Pa at \( \theta = 45° \). There is simply insufficient CO₂ to affect a significant increase in the atmospheric pressure. Quasi-permanent polar caps form as the obliquity drops below 24°.

At lower obliquities the atmospheric pressure is buffered by the presence of polar caps, and the temperature of those caps controls the atmospheric pressure. We have assumed a CO₂ cap albedo of 0.6. The first appearance of a quasi-permanent polar cap as obliquity decreases and the poles cool also depends on the albedo of polar CO₂ ice. If we assume that the albedo of the permanent cap is 0.65, instead of 0.6, a permanent cap is established at 24° for the baseline case. The atmospheric pressure is lower at any given obliquity as the caps grow brighter, and the caps are more massive.

5. Discussion

H₂O Data

The validity of the results in this paper depends upon the assumption that (2), with the proper coefficients, accurately describes the adsorption of H₂O on Martian regolith materials. Unfortunately, the partial molar enthalpy of adsorption calculated from (2) is only \( 1.04 \times 10^4 \text{ J mol}^{-1} \), which is somewhat lower than is typically reported.
Fortunately, because we have been compelled to make repeated measurements of adsorption at the same soil temperature and CO₂ partial pressure, we have enough data to associate standard deviations with our H₂O measurements. Typically, adsorption measurements made at the same conditions have standard deviations less than 30%. While this is certainly less accuracy than we could wish for, it gives us confidence that our adsorption isotherms are not in error by more than a few tens of a percent. As we shall see below, our qualitative conclusion, that H₂O cannot substantially displace CO₂ from its adsorption sites, is robust in the face of that level of uncertainty.

The fundamental reason that co-adsorption does not alter the climatic evolution as dramatically as might have been expected is that ground ice acts to buffer the partial pressure of H₂O throughout much of the regolith. In numerical experiments in which ground ice was prevented from buffering the pore vapor pressure, increase in the atmospheric H₂O vapor pressure with obliquity caused higher H₂O adsorption, and significant CO₂ desorption. Without the buffering effects of the ground ice, the atmospheric pressure reaches 3-4 kPa at obliquities of 40°, as virtually all adsorbed CO₂ is eventually displaced.

As demonstrated by Anderson et al. [1967], at any relative humidity, the adsorptive coverage of H₂O on the Martian regolith decreases with decreasing temperature. Our adsorption data support this conclusion, as do the data of Anderson et al. [1978]. For example, we found no more than about 3 mg H₂O g⁻¹ soil at soil temperatures of 210 K, even when the relative humidity approached unity. By contrast, we measured 40 mg H₂O g⁻¹ soil at θ = 9.2 when the soil was at 294 K. In the middle- to high-latitude Martian regolith, where temperatures are usually below 210 K and where most adsorbed CO₂ is bound, the upper limit on H₂O coverage is extremely low. Assuming that the surface area occupied by an adsorbed H₂O molecule is of the order of 10.5 Å², and using our sample specific area of 94.8 m² g⁻¹ soil, we estimate that a monolayer of adsorbed H₂O weighs 27 mg g⁻¹ soil. Since the upper limit on adsorbed H₂O is around 3 mg g⁻¹ at 210 K, we find that no more than about 10% of a monolayer is likely at these low temperatures. Since displacement is linear with the fractional monolayer coverage of a competing adsorbent, the effect of the water on the total CO₂ inventory is no more than 10%. It is for this reason that an uncertainty in H₂O coverage of 30%, or 3% of a monolayer, does not affect our conclusion that H₂O is inefficient in absent in displacing CO₂ into the vapor phase.

**Extent of Regolith**

A second major issue that arises from this study is the uncertainty in the total mineral surface area exposed to the Martian atmosphere. The total area is usually represented as the product of the specific surface area of the regolith and its uniform depth across the planet; however, many deposits with different volumes and characteristic specific surface areas probably cover the surface. It is important to determine the correct structure for modeling the Martian regolith, since most models of volatile inventory and climate history were all built on the assumption of adsorptive equilibrium between the atmosphere and 10⁻¹⁰ m of fine-grained, powdered regolith. It is interesting to compare our understanding of the regolith structure on the Moon to the thickness of the regolith postulated for Mars.

The structure of the lunar regolith, at least at the Apollo sites, is reasonably well understood. Active seismic experiments were undertaken to sound the shallow stratigraphy of the lunar crust in both the highlands and the mare. The salient conclusion from our point of view is that the fine-grained component of the lunar regolith is only 5-15 m thick. At Apollo 14, in what is thought to be Imbrium ejecta, the uppermost layer detected in seismic experiments was 8.5 m thick, and is interpreted as the highly gardened material that was found across the lunar surface [Cooper et al., 1976]. It is this powdered material that is analogous to the material put in the adsorption apparatus.

**Comparison With Viking Gas Exchange Experiment**

We can make one comparison with data in order to assess the utility of our palagonite sample as a Mars-analog material. The Viking spacecraft performed soil analyses in the course of the Viking Gas Exchange Experiment. When soil from the VLO site was humidified, after equilibrating with the atmosphere at 291 K, 9800 nmol of CO₂ cm⁻² were released [Oyama and Berdahl, 1977]. The model assumed by Ballou et al. [1978] in analyzing this experiment is that the H₂O displaced all of the adsorbed CO₂. Based upon the equations determined in this experiment, we predict the release of 12,850 nmol of CO₂ cm⁻², assuming full desorption of the CO₂ equilibrated at conditions of 700 Pa and 291 K. The discrepancy between our prediction and the observed CO₂ displacement is about 30%, again suggesting the degree of uncertainty in our empirical representations.

**6. Summary**

We have measured the simultaneous adsorption of water and carbon dioxide on fine-grained palagonite at conditions appropriate to the Martian regolith. We find that water competes poorly for adsorption sites at low temperatures and pressures characteristic of the Martian regolith. Therefore, previous upper limits estimated for the adsorbed CO₂ inventory on Mars, 3-4 kPa, remain valid. The most likely values depend on the total exposed silicate surface area. We use 720 kg m⁻² and 75 m of powdered regolith as a baseline. At higher obliquities, the atmospheric pressure history is dependent on the total inventory of CO₂. When the total inventory is large (19 kPa) and the powdered regolith is deep (1 km), the atmospheric pressure goes up to around 1440 Pa and θ = 45°. If the total inventory is assumed to be very small (1.1 kPa) and the powdered regolith shallow (15 m), atmospheric pressures do not increase significantly with obliquity.

Buffering of pore gases by hard-frozen permafrost is primarily responsible for limiting the effectiveness of H₂O in displacing adsorbed CO₂. The maximum H₂O coverage decreases with temperature, and hence CO₂ displacement becomes less effective at lower temperatures.
Major uncertainties in the system include the utility of palagonite to represent the Martian near-surface materials, and the total surface area of regolith material exposed to the Martian atmosphere, and hence the exchangeable CO$_2$ inventory. Future investigations will focus on these issues.

Acknowledgments. This research was supported by the Mars Surface and Atmosphere Through Time program under RTOP 155-01-60-02 and the Planetary Geology Program under RTOP 155-01-60-01. The authors would like to thank Christopher McKay for helpful discussions. We would also like to thank Fraser Fanale and an anonymous reviewer for constructive reviews.

References


Page intentionally left blank
Evidence of ancient continental glaciation in the Martian northern plains

Jeffrey S. Kargel,1 Victor R. Baker,2 James E. Boggs,3 Jeffrey F. Lockwood,4 Troy L. Péwé,5 John S. Shaw,5 and Robert G. Strom7

Abstract. Whorled ridges, spaced about 2-6 km and forming lobate patterns with lobe widths of about 150 km, occur at many locations in the northern plains of Mars, commonly in close association with sinuous troughs that contain medial ridges. These landforms resemble moraines, tunnel channels, and eskers found in terrestrial glacial terrains, such as the relict continental North America. Some Martian landscapes may have formed by disintegration of continental glaciers that covered much of the northern plains into the early Amazonian (i.e., late in Martian geologic history). Meltwater processes apparently were important in the collapse of these hypothesized ice sheets, hence, the glaciers apparently were wet based in part. Whereas striking similarities exist among areas of the northern plains and some glaciated Pleistocene terrains on Earth, there are also important differences; notably, drumlin fields, such as those in many glacial landscapes on Earth, are rare, absent, or not yet resolved in images of the Martian northern plains. Another major difference is that postglacial fluvial and other water-related modifications (especially erosion) of Pleistocene terrains are substantial, but similar modifications are not observed in the northern plains; a virtually complete and sudden decline in the activity of liquid surface water following glaciation in the northern plains seems to be implied. The climatic implications of the hypothesized Martian glaciers and their decline are unclear. We investigate two possibilities, alternatively involving a relatively warm paleoclimate and the modern Martian climate. The hypothesized ice sheets in the basins within the northern plains (generally at elevations lower than -1 km) suggest a relationship of these frozen bodies of water with former regional lakes or seas, which may have formed in response to huge discharges of water from Martian outflow channels. This possible relationship has been modeled. Glaciers may have evolved from seas by their transition to development of outflow channels and erosional redistribution of sea ice. The transition to glaciation may have taken several million years if the climate was very cold, comparable to today’s, or tens of thousands of years if the climate was as warm as modern Antarctica. A glaciated sea may have involved an extended period of glaciolacustrine and ice shelf processes.

1. Introduction

The hypothesis that Mars was glaciated carries important implications for its geologic, hydrologic, climatic, and possible biological evolution. Glaciation may have contributed to development of surficial landscape features (Claypool, 1982; Tricart 1988); formation of volcanic features (H. F. F. P. 1975, unpublished; Hodges and Moore, 1979; Luchinna, 1981; Barratton, 1987; Chapman, 1994); and erosion and deposition in the northern plains (Carr, 1986; Luchinna et al., 1986; Baker et al., 1991; Scott and Underwood, 1991; Kargel et al., 1992). Others have hypothesized large amounts of ground ice or ice sheets of continental extent (Carr and Schaber, 1977; Allen, 1979; Rosacker and Judson, 1981; Mouginis-Mark, 1983; Squyres and Carr, 1986; Squyres et al., 1987; Christensen, 1989; Costa, 1989; Jönn, 1991; Luchinna, 1993; Costard and Kargel, 1994). In this report we examine aspects of the Martian northern plains where landscapes which we interpret to be glacial are widespread. It is difficult with available data to rule out proposed alternate explanations for individual landforms (witness the recent debate over photogeologic interpretations of sinuous ridges in the Martian southern hemisphere [Kargel and Strom, 1992; Metzger, 1992; Ruff, 1994]). Notwithstanding this unavoidable uncertainty, physical analogies of Martian landforms and entire landscapes with terrestrial glacial terrains offers a strong case for the hypothesis of ancient Martian continental glaciation.

Our conclusions are ideally consistent with those of Chapman [1994], who mapped the area of eastern Utopia Planitia near Elysium, where features occur that he interprets to have formed by volcano-ice (glacier?) interactions.
Specifically, we and Chapman [1994] have concluded that (1) suites of landforms are morphologically consistent with ancient glacier activity, (2) large, thick ice sheets existed in the northern plains, (3) the ice sheets principally occupied low ground and enveloped high terrain up to a level of about 1 km elevation, and (4) the ice sheets were active during the Early Amazonian (i.e., relatively late in Martian history).

2. Observations

2.1. Thumbprint Terrain

"Thumbprint terrain" (TT) is among the most widespread terrain types in the northern plains. Named for its resemblance in Viking Orbiter images to fingerprints, TT consists of parallel, on occasion, or nested sets of regularly spaced curvilinear ridges or aligned hills. The ridges are 0.5-2.5 km wide and 1-40 km long. The characteristic spacing of thumbprint ridges is 2-6 km. Whorled lobes of TT are 75-150 km wide. The heights of thumbprint ridges are not known exactly, but they probably occur in the entire range from about 10 m to 200 m high.

Thumbprint terrain is distinctive, although it occurs in a variety of gravitational morphologic types, two typical examples are shown in Figures 1a and 1b. Some unusual types, which occur mainly in Isidis Planitia, may represent a variant of the same or a completely different phenomenon (Figures 1c and 1d).

Twenty-two areas of TT, between 3,000 and 420,000 km² each, have been identified in the northern plains, generally a few hundred kilometers from the cratered plains boundary and at elevations between 0 and -2 km (Figure 2). In addition, two small patches of TT occur in the Hellas Basin. TT has not been found in the cratered highlands, in wrinkle-ridge plains (Hesperian ridge plains), or on the Tharsis and Elysium plateaus. Thus, all known occurrences of TT are restricted to topographic basins [Lockwood et al., 1992].

Thermal inertia, derived from Viking orbiter infrared thermal mapper measurements, relates to the effective grain sizes of surface rocks [Christensen and Moore, 1992]. All known areas of TT in the northern plains have moderate or high thermal inertias (0.006 to 0.012 cm² Hs⁻¹ K⁻¹), values that are inconsistent with air fall dust deposits [Edgett and Christensen, 1991]. Sixteen of the 22 areas of TT have high thermal inertia values (similar to or higher than that of the boldestrown Viking 2 landing site), which probably indicates materials that are significantly coarser than Martian dunes [Edgett and Christensen, 1991]. The other 6 areas have moderate thermal inertias, consistent with coverage by sand; however, these values also could be produced by partly block-covered surfaces that have fewer boulders than the Viking 2 site.

2.2. Associated Landforms

Scott and Underwood [1991] recognized that TT is associated with branching troughs and medial ridges in Utopia Planitia (Figure 3a). Scott [1983] had interpreted similar troughs in Aris Vallis as meander relics. Significantly, TT is closely associated with troughs in at least nine other areas of the northern plains; where image resolution is sufficient to resolve small details, some troughs at each location have medial ridges. Figures 3b and 4 show a similar landscape in Arcadia Planitia. TT ridges in Arcadia Planitia are 5-40 km long, they are spaced 4-24 km apart, and they form whorled lobes (conceal to the NE) -100 km wide. TT in Arcadia Planitia merges southward into an anastomosing system of troughs (Figures 3b, 4). The troughs, 2-6 km wide, are fairly well integrated, although some are isolated. TT-associated troughs in Arcadia Planitia, as in several other areas of the northern plains, contain medial ridges (Figures 1b, 3a, 4b, 4c), which range from 400 to 1,200 m in width and are probably a few tens of meters high. Medial ridges are 1 to 40 km long; they exhibit abrupt variations in width and, like the troughs, many terminate abruptly. Some medial ridges exhibit low-order branching (Figure 4c). Smooth plains (smooth at resolutions of tens to hundreds of meters) appear to extend the southern ends of the troughs in Arcadia Planitia. Embayment of troughs by smooth plains is the rule in other occurrences of this type of landscape.

Other features in Arcadia Planitia include mounds 1-5 km across; some are flat topped, others are rounded, and many possess basal scarps or terraces (Figure 4b). Similar mounds occur near other areas of TT and, as in Arcadia Planitia, may occur within sinuous troughs (Figures 1b, 4b). Some mounds connect segments of medial ridges.

This association of TT, sinuous troughs, medial ridges, mounds, and smooth plains occurs widely in the northern plains but apparently nowhere else on Mars. As discussed below, we favor a glacial interpretation for its origin. Accordingly, we refer to this type of landscape as the TT-associated glacial landscape (TTAL).
Figure 1. Different forms of thumbprint terrain (TT), which is interpreted as types of moraines or glaciotectonic ridges. Solar illumination is approximately from the bottom of each image. (a) This example of TT consists of relatively continuous ridges with inverted V-shaped or rounded profiles. In places, the topographic expression of TT is very slight and is mainly discernible as bright arcs. The prominent ridge at top center (small arrow) occurs where many smaller ridges converge from two sides, which suggests that it may be an interlobate moraine. Mosaic of Viking printer (VO) frames 057B52 to 057B57, western Utopia Planitia, latitude 50.4°N, longitude 289.0°W. North given by large arrow (upper left). Scene is 125 x 189 km. (b) Another form of TT (upper right) is characterized by curvilinear alignments of disconnected hills and short ridges with rounded tops. This scene shows another type of ridge (interpreted as possible eskers), relatively long and narrow compared to TT ridges, that trend approximately orthogonally to the TT; some of these narrow ridges occur in shallow troughs (interpreted as possible tunnel channels), and others connect mounds (possible kames). This type of narrow ridge has otherwise been interpreted as ice-shelf features [Lucchitta et al., 1986], linear pingos [Scott, 1983], pressure ridges at the edges of a mud ocean or giant mud flows [Jons, 1991], or barrier beaches [Parker et al., 1993]. VO frame 608A06, southwestern Utopia Planitia, latitude 36.5°N, longitude 277.8°E. North up. Scene width 102 km. (c) Another possible form of TT consists of long linear or arcuate ridges whose crests bear summit pits or grooves; the ridges locally have flat tops. VO frame 146523, Isidis Planitia, latitude 16.6°N, longitude 277.8°E. (d) Yet another possible variation of TT is distinguished by lineaments or arcs of disconnected hills and short ridges, many with summit pits. Also present are disorganized groups of hills and short ridges, recognized as possible TT only by a close association with and gradation into a more obvious type of TT. VO frame 146513, Isidis Planitia, latitude 16.1°N, longitude 277.2°E. Scene width ~32 km.
Apparently like their Martian counterparts, terrestrial tunnel channels characteristically contain or merge into eskers. The scales, sinuosities, and integration of terrestrial and Martian tunnel channels and channel-esker complexes are similar in many cases, including (1) single, isolated channels and eskers and (2) complex networks of anastomosing, rectilinear, and low-order dendritic channels and eskers [Armstrong and Tipper, 1948; Farrand, 1969; Tipper, 1971; Weight, 1975; Shaw and Healy, 1977; Ehlers, 1981; Krüger, 1983; Christiansen, 1987; Boyd et al., 1988; Attig et al., 1989; Hebrard and Amor, 1990; Mooers, 1990; Shaw and Grevil, 1991; Shaw et al., 1992; Brennand and Sharp, 1993; Kargel, 1993; Brennand, 1994; Brennand and Shaw, 1994]. Figure 5 shows an example of anastomosing tunnel channels on Earth, some of which contain medial eskers. Tunnel channel-esker complexes are commonly attributed to progressive channelization of subglacial floods during the collapse of glaciers; alternatively, eskers may represent a rejuvenation of subglacial flow under a more normal hydraulic regime than that which caused the formation of channels [Brennand and Sharp, 1993].

Sinuous eskerlike ridges have been described from the southern hemisphere of Mars, including the south polar region, the Hellas Basin, the Argyre Basin, and highlands adjoining Argyre. Widely variable interpretations have been applied to those ridges [Kargel and Strom, 1992; Kargel, 1993; Metzger, 1993; Ruff, 1994]. Although some of these ridges are associated with channels [Kargel, 1993], most are not, and none are closely associated with TT, which generally seems to be lacking in the southern hemisphere. Thus, it appears that the sinuous ridges of the southern hemisphere of Mars, though probably glaciogenic, may have formed under different circumstances than TT-associated sinuous ridges in the northern plains.

TT resembles the form and scale of common types of recessional moraines deposited by Pleistocene ice sheets across central North America [Geological Society of America, 1959; Wright and Frey, 1963] and northern Europe [Fyfe, 1990]. Recessional moraines in the Great Lakes region form whorled lobes ~75-150 km wide (Figure 6), the same as Martian TT (Figure 3). Moraines in this region are spaced 3-25 km apart, averaging a little greater than (but overlapping) the spacing of TT ridges. Individual moraines in the Great Lakes region are 1-25 km wide, also averaging a little greater than (but overlapping) the widths of TT ridges. The cross profiles of Great Lakes moraines include sharp-crested, rounded, and flat-topped varieties, and some also have summit depressions (kettles) produced by melting of buried ice, analogous to the variety of cross profiles observed for Martian TT ridges. The height of these moraines is commonly 10-60 m. Some moraines illustrated in Figure 6 have nearly constant widths and heights for tens of kilometers, whereas others have discontinuous, "beaded" longitudinal profiles, both analogous to some types of Martian TT. These, like most moraines, are composed mostly of poorly sorted sediment and contain a large proportion of boulders and cobbles.

Several other types of ice-marginal deposits on Earth, including sublacustrine moraines, structures formed by subglacial deformation of water-saturated till, and ice-thrust ridges, also form spatially periodic wheel patterns similar to Martian TT [Hoopy, 1957; 1960; Burnett and Holdsworth, 1974; Sanders, 1981; Sorensen, 1983; Aber et al., 1989; Zilliacus, 1989; Beauty and Prichommet, 1991]. This general similarity of forms produced by distinct glacial processes broadly supports the hypothesis of glaciation on Mars, but it leaves the exact mechanism of formation of TT unclear. Although we seek a unified interpretation of TT on Mars, it is entirely possible that different potentially glacier-related
Figure 3. Generalized geologic/geomorphologic maps of two terrains of possible glacial origin (see text for detailed interpretations). (a) Northwesternmost Utopia Planitia showing the association between whorled ridges of TT and a system of sinuous to rectilinear troughs, many of which contain medial ridges. (b) Part of Arcadia Planitia showing three major terrains. Compare with corresponding image mosaic (Figure 4a). Smooth plains embay troughs at their southern and western extremities (left) and mantle old impact craters (top center). The dominant landforms illustrated here modify old cratered plains, which postdate the ancient period of heavy bombardment. Mapped from Viking Orbiter images 776A32, 776A34, and 810A39.
processes may have operated at different times and locations to produce the different forms of TT.

Terrestrial recessional moraine complexes commonly form by several processes, resulting in composite ice-marginal deposits that locally include lodgement tills, push moraines, glaciotectonic thrusts, flow tills, crevasse fill, sediment diapirs, kames, eskers, and deltas. Whatever the exact mechanism(s) of formation of TT, the best terrestrial analogs involve liquid water in the proglacial environment and/or in subglacial till. Terrestrial glaciers that are completely frozen to their beds commonly cause little erosion and do not generally produce notable recessional moraine complexes. One may infer that thumbprint terrain on Mars was formed with the involvement of liquid water as well as ice, if indeed the moraine interpretation generally is correct.

The smooth plains in the NPGA are interpreted as proglacial lacustrine sediments. On Earth, glaciolacustrine or glaciomarine plains commonly embay or mantle tunnel channels and eskers (Armstrong and Tipper, 1948; Ehlers and Linke, 1989). One cannot rule out the possibility that other types of deposits may be components of the Martian smooth plains, but the interpretation of these plains as lacustrine sediments is consistent with the suggested importance of water in the formation of TT, troughs, and medial ridges.

The mounds in Arcadia Planitia (Figure 4b), Utopia Planitia (Figure 1b), and several other areas of NPGA may be kames. Terrestrial kames, commonly similar in shape to the Martian mounds, are characteristically associated with eskers, tunnel channels, moraines, and other features that were formed by ice stagnation (Zolli, 1965; Kräger, 1983; Ehlers...
and Linke, 1989; Mooers, 1990; Pwé and Reger, 1993). In some places on Earth, eskers originate directly from kames, and the eskers and kames occur together within tunnel channels [Mooers, 1990]; similar associations may be shown in Figures 1b and 4b. A basal scarp or terrace, such as those of some Martian kame-like mounds, can be eroded into kames by waves in a proglacial lake [Zoltai, 1965], or a basal scarp can reflect the shape of the ice-walled lake basin in which a kame was deposited [Pwé and Reger, 1993].

Regional geomorphologic mapping of glacial landscapes, recently with the aid of orbital imagery, has provided important insights for the sources and motion of glaciers and for glacial processes [e.g., Aylsworth and Shills, 1989; Boulton and Clark, 1990; Clark, 1994]. Viking imagery generally lacks the high spatial resolutions and lateral continuity necessary for comparable regional mapping, although certain features, especially thumbprint terrain, are bold enough that fairly complete maps are possible for some regions. The general trends of curvature of TT ridges in Utopia and Arcadia Planitiae can be interpreted as markers of the northerly sources of ice and of an advance generally toward the highlands. If the smooth plains of the NPGA are lacustrine deposits, then lakes may have been impounded between the ice sheet on one side and the highland/lowland boundary on the other; this would be analogous to the late Pleistocene glacial Lake Agassiz, which was impounded between the Laurentide ice sheet to the north and basement rocks and glacial sediment deposits to the south [Teller and Clayton, 1983]. Idilis Flumina and other basins within the northern plains (notably near Deuteronilus Mensae) contain enclosed whirls of TT, possibly indicating progressive decay of isolated bodies of ice in depressions within the northern lowlands.

The close association of channel-ridge features with TT and the other features described here is a strong point favoring the glaciation hypothesis. However, these associations do not cover the full spectrum of glacial landscapes on Earth, nor are the observed associations always identical with the most common occurrences on Earth. Two notable differences are (1) the apparent absence of large drumlin fields in the Martian northern plains, and (2) the usual convexity of TT with respect to channel-ridge features on Mars, contrasted by the usual concavity of terrestrial moraine sets with respect to tunnel channel-esker complexes. We don't know of any theory of glacial landscape genesis that would prevent moraine sets from being convex toward tunnel channels and eskers. The common association of TT with channel-ridge complexes on Mars indicates that this relationship is important. Perhaps it is indicating that subglacial outflows occurred first, the glaciers then receded or partly floated from the area of tunnel channels and eskers, and then the moraine sets formed during a phase of steadier retreat and more normal hydrologic regime; this could represent a possible reversal of the usual sequence on Earth. The apparent absence of drumlins in the Martian northern plains is discussed below.

3.2. Absence of Candidate Drumlin Fields in the Northern Plains

Drumlins are elongate, streamlined hills produced by glacial action; they are commonly 500 to 1000 m long, 100 to 500 m wide, and tens of meters high. Large fields of drumlins have not been observed in the Martian northern plains, though they may exist in Hellas [Kargel et al., 1991; Kargel and Strom, 1992]. It is difficult to rule out the presence of drumlin fields in the northern plains, considering that there are relatively few Viking images with the high resolutions and low sun inclinations that would be necessary for their resolution. There are some good Landsat images of swarms of drumlins on Earth, but these are exceptional and generally are acquired under conditions of very low solar inclinations. Terrestrial drumlins commonly are more difficult to resolve in satellite images than are dunes of
Figure 6. Map showing the pattern of ice-marginal recessional moraines associated with the retreat of the Laurentide ice sheet from the Great Lakes region.

comparable width, because drumlins commonly have lower height:width ratios than many dunes, and drumlins usually do not have steep edges near the angle of repose. Nevertheless, the apparent absence of large fields of drumlins in the northern plains may be real and significant. Swarms of drumlins in many glaciated terrains on Earth are variously interpreted to have formed either by (1) catastrophic subglacial meltwater floods that overtopped tunnel channel rims and extended in vast sheets across the landscape [e.g., Shaw and Kvill, 1984; Shaw et al., 1989; Shoemaker, 1990, 1992] or (2) plastic deformation of water-saturated subglacial till [Aylsworth and Shills, 1989; Boyce and Eyles, 1991].

If terrestrial tunnel channels and moraines were invariably associated with drumlins, a similar association might be expected on Mars. But this is not the case: for example, tunnel channels in northern Germany are not associated with drumlins [Ehlers, 1981]. The north German channels are exceptionally deep and are cut into easily erodible sedimentary rocks. The lack of drumlins in this region of Europe can be explained utilizing the subglacial flood theory of drumlin formation [e.g., Shaw and Kvill, 1984] if deep channels contained the subglacial floods. Thus, without sheet floods, drumlins were not formed. It may be that large Martian channels also carried flood discharges, accounting for the absence of drumlins between the channels. Alternatively, under the deforming-till theory of drumlin formation [e.g., Boulton, 1987], the interchannel areas may have lacked the large amounts of till or other un lithified sediment that would be necessary for the formation of drumlins: the formation of Martian drumlins by this mechanism on a scale that would have been visible in Viking orbiter images would have required the prior existence of a widespread blanket of un lithified till-like sediment at least several tens of meters thick. As yet another variation of this interpretation of drumlins, the theory of moulding of terrestrial drumlins by ice streams or surges [e.g., Clark, 1994] would imply that the Martian glaciers in the area of the 'T' may not have exhibited the surge behavior thought to be necessary for drumlin formation; a lack of surge behavior could be caused by a general condition of a frozen glacier bed in a wide zone up to the glacier terminus and a restriction of basal melting to isolated regions well behind the terminus in areas of thick ice or elevated geothermal gradient.
3. Models of Martian Glaciation

Implications of basal associations of glacial terrains on Mars. As described above, putative glacial terrains in the northern hemisphere are concentrated near an elevation of -1 km, seeming to imply a connection with former bodies of ground liquid water (Lauretta et al., 1986; Parker et al., 1989; 1993; Scott et al., 1992; Chapman, 1994). Although some interpreted glacial terrains occur at high elevations in the southern hemisphere, glaciations in all of these regions also may be related to pondings of water in major depressions (Argyre, Hellas, and the South Polar Basin). Atmospheric transfer of moisture from basins and precipitation over uplands can explain evidence of glacia tion in the Charming Montes of Argyre (Kargel and Strom, 1992), on Maela Planum near Hellas [Kargel et al., 1991], and in mountains near the South Pole (Kargel, 1993). There are no hard constraints available on the rate of such transfers, hence, no constraints on prevalent atmospheric temperature and humidity. In the case of the putative glacial landscapes of Dorsa Argenta (Kargel, 1993), the cause may have been gravity flows of water and ice from the South Polar Basin and subsequent drainage into Argyre. Thus, ponded water may have been instrumental in the origins of all known glacial terrains on Mars. It is unclear whether these bodies of water may have induced major changes in the thermal regime of the Martian climate (Baker et al., 1991; Kargel and Strom, 1992) or merely supplied the water necessary for glaciation.

The glacial interpretation of these vast and widely distributed landscapes on Mars cannot be divorced from the theoretical problems of the formation of large Martian glaciers. We have considered two very different groups of models for the origin of the NPGL and the hypothesized glaciers. One model is roughly based on an Earth-like hydrologic cycle that included the growth and collapse of the Laurentide ice sheet. Another group of models is based on a unique Martian history that is directly linked with the glacial origin of outflow channels and the formation of large lakes or seas in the northern plains.

Laurentide Ice Sheet analog. The growth and collapse of Pleistocene ice sheets may have had important respects in common with the growth and demise of Martian ice sheets. The Laurentide Ice Sheet is the area of glacial Lake Agassiz and the Great Lakes [Geological Society of America, 1959; Krüger, 1983; Teller and Clayton, 1983; Ehlers and Látko, 1989] is a possible partial analog for the ice sheets in the northern plains. A possible scenario of glacial geologic events in Arcadia Planitia follows.

1. An ice sheet was deposited or accumulated over the area of Arcadia Planitia that now is occupied by troughs and TT. The ice sheet was mobile for long enough to entrain a substantial quantity of rocky debris (which later was deposited as eskers, moraines, and glaciolacustrine plains).

2. Subglacial geothermal meltwater (produced by slow basal melting caused by the normal geothermal gradient and heat flow) or subglacial meltwater (produced by seasonal atmospheric or solar heating of the glacier's surface) was released at the glacier bed and caused erosion of tunnel channels during an early period of increasing discharge, probably during stagnation of the ice sheet.

3. Eskers were deposited in tunnel channels some time after maximum water flow during ice stagnation, and kames were deposited in these pits and collapse depressions as water discharge reduced.

4. Smooth plains were deposited in a proglacial lake.

5. TT recessional moraines or glaciolacustrine ridges were formed.

6. The lake and the remnants of the stagnated glacier evaporated, leaving “freeze-dried” glacial landscapes intact with little opportunity for degradation since then.

As an alternative to stage 2, the sources of liquid water might have been extraglacial. Catastrophic discharges of ground water from the highlands may have formed the outflow channels and then pooled between the ice sheet and the highland/outflow escarpment. Hydraulic pressure at the front of the ice sheet may have lifted the ice, resulting in “reverse outflows” from the lake into the glacier-covered area along the base of the ice sheet [after the model of Shoemaker, 1992], thereby producing tunnel channels and eskers.

Freezing of Martian lakes and their transition to glaciers. Besides the apparent absence of drumlins in the northern plains, another important difference with respect to terrestrial glacial landscape analogs is that postglacial fluvial modification of Martian glacial terrains is nil, despite the fact that hundreds of millions or billions of years have elapsed since their formation. If the Martian glaciers, like the Laurentide ice sheet, ablated largely by surface melting, then implying warm conditions, then why are the glacial landforms virtually pristine at the resolution of available images? This important aspect will be discussed below, for now it is sufficient to say that the differences between glacial terrains on Earth and Mars warrant consideration of other glacier models for Mars.

The concentration of putative glacial landforms near an elevation of -1 km in the northern plains and especially their apparent absence at high elevations may indicate a connection with paleolakes or seas, as suggested by Kargel and Costard [1993] and Chapman [1994]. Of course an ice-covered lake or sea is not a glacier. But is it possible that a frozen lake/sea may evolve into a glacier and thereby preserve a similar geographic relationship as previously held by the lake?

The thermal evolution of a Martian lake is considered here in two stages. First, we model the time required for the lake to start to freeze, which turns out to be a fairly short period. Second, we model the progressive solidification of the lake, including the redistribution of ice caused by sublimation and precipitation. These models result in the grounding of lake ice and the eventual formation of substantial surface topography on the ice mass, so that glacial flow at a finite rate would be inevitable. Our purpose is not to model any specific Martian glacier, but simply to illustrate our point that a frozen lake/sea may evolve into a glacier and thereby preserve a similar geographic relationship as previously held by the lake.

If the water discharged by outflow channels had already started to freeze by the time it entered the lake basin, the surface freezing of the lake would begin as soon as it formed. However, if the lake water originated from a deep aquifer, it may initially have been warm, so that some period of time would be required before freezing would begin. This period can be calculated for reasonable assumptions of the pertinent conditions. We consider two widely differing cases. In case “a,” the annually averaged surface temperature on Mars is 200 K, initial water temperature is 274 K, and water depth is 100 m. In case “b,” Martian surface temperature is 250 K, initial water temperature is 293 K, and water depth is 1000 m. In both cases we assume for the sake of simplicity that there is complete mixing (by gravity flows and wind-driven waves and currents) and we neglect evaporative enthalpy loss; hence, cooling of the lake can be approximated as the integrated loss of thermal energy from the lake at a rate very roughly estimated by the difference of the upward and downward blackbody radiative fluxes. Present climatic conditions in the middle latitudes of the northern plains correspond...
approximately to case a, where freezing would commence in about 19 days. Freezing would begin in case b after about 18 years. These are upper limits, because (1) any evaporative heat loss would increase the rate of cooling, (2) development of density stratification could allow commencement of freezing of the upper thermal boundary layer before the entire lake mass had cooled to the freezing point, and (3) an atmosphere with optically thin spectral windows could radiate more effectively than assumed. Hence, a Martian lake would not have had a liquid surface for long, unless the ambient surface temperature was very close to or above the freezing point. Once freezing had commenced, the lake ice would have gradually thickened. We considered the detailed thermal evolution of a Martian lake that was big enough to explain the scale and distribution of putative glacial landforms described above. The scenario modeled below includes the progressive thinning of lake ice owing to the combined effects of surface cooling, normal low levels of geothermal heating, sublimation of ice from the warmer half of the lake, and cold trapping of vapor on the colder half. The lake basin is assumed to be 3000 km wide and sinuosity, reaching a depth of 2000 m in the middle. So that the modeled topography roughly reflects an idealized representation of the northern plains, we set the initial lake shore at an elevation of -1000 m, so that the center of the lake bed is at -3000 m. This situation roughly corresponds to a Martian paleocharacter that extends from about latitude 30°N to latitude 80°N. To put the amount of water in perspective, consider the idealized case that this seaway extends around the globe in this latitude belt; the volume of water is \(5 \times 10^{14} \text{m}^3\), corresponding to a global layer of water 340 m deep, consistent with other estimates of Mars’ surface and subsurface water inventory (Carr, 1996). The volume of water could be substantially less if the northern plains had several large lakes instead of a single seaway. The freezing of the lake is modeled under two sets of climatic conditions, described below. Geothermal heat flow is modeled as the Earth’s global average heat flow today (Fowler, 1990), which is roughly equal to Mars’ heat flow 2 b.y. ago (Stevenson et al., 1983), a possible age of Martian glacial deposits. Volcanic activity may have contributed to geothermal heating, local inputs of geothermal heat may have exceeded the levels considered here.) The effects of surface cooling are modeled by the solution to the Stefan problem (Ferzado and Schubert, 1982); however, unlike this solution, which involves the thickness of the ice layer increasing throughout time at a rate inversely proportional to time, in the present model the effects of constant sublimation from the ice surface eventually results in a steady state ice thickness (e.g., the southern part of the lake in model 2, see Figure 8). The thickness of the ice also depends on the input of geothermal heat or the addition of precipitated ice. These effects are all included. It is also assumed that the ice is in hydrostatic equilibrium wherever ice overlies water. However, the effects of large-scale glacial creep are not considered. Model 1 corresponds approximately to the modification (cooling) of current conditions (Zurek et al., 1992) caused by closure of the northern plains by a large ice dome in the southern ablation areas and a bright ice mantle in the northern areas of deposition. The annual average surface temperature is assumed to vary linearly across the lake from 190 K at its southern (temperate latitude) edge to 160 K at its northern edge. Initial rates of net sublimation vary linearly from nothing at the center to 0.5 mm/yr at the northern edge. As ice eventually recedes from the warmer, southern half of the lake, net sublimation rates are highest, part of the moisture source dries up, so that deposition rates in the colder, northern half increase in response. The linear variation of the rates of sublimation and accumulation is an assumption rather than a result of a rigorous climate model, which has not yet been developed for the pertinent conditions and which would be expected to produce a more complex pattern of warming and cooling. Model 2 represents relatively mild conditions comparable to much of Greenland and Antarctica today. Annual average surface temperatures are at most near the freezing point of water, and cold trapping of vapor on the colder half of the lake from 250 K at its southern edge to 200 K at its northern edge. Initial rates of net sublimation vary from 0 to 0.5 mm/yr at the lake’s southern edge to 0 at its northern edge. As in model 1, as ice recedes in the warm ablation zone of the lake, deposition in the northern depositional zone slows. The results of models 1 and 2 are shown in Figures 7 and 8, respectively. The complete solidification of the model Martian lake requires about 10^11 years under cold climatic conditions (Figure 7). Extensive redistribution of ice by sublimation and recondensation takes millions of years. At the end of 5 m.y., the ice has redistributed itself in such a way that glacial creep likely would be significant. However, this glacial ice is cold-based almost throughout its bed, so that ice-rock interactions might be fairly minor. Normal low levels of geothermal heating may induce basal melting if any part of the glacier exceeds several kilometers in thickness, so that formation of interesting landforms are not excluded by a cryogenic climate such as today’s. High local levels of geothermal heating, e.g., caused by subglacial volcanism, could cause basal melting beneath very thin ice sheets. In the case of a relatively warm, Antarctic-type environment (Figure 8), sublimation and recondensation occur more rapidly, so that substantial topography develops on the grounded part of the ice even within a few tens of thousands of years. A warm climate results in a glacier that discharges into an ice-covered lake. The lake does not completely solidify, and large parts of the glacier are wet-based and other parts consist of an ice shelf that is grounded at both ends. Glacial flow and meltwater processes probably would produce interesting landforms, especially near the grounding lines. 3.4. Subglacial Flooding Mechanisms and Paleoecological Implications The glacial hypothesis implies that an ice sheet once existed where none is now. The wide distribution of TT (Figure 2) suggests either one very large or several smaller ice sheets. The formation of one or more ice sheets may have been a response to the inundation by water of the northern plains (Baker et al., 1991). Glacier formation may have been either a direct response due to the progressive solidification of large lakes and grounding of the ice (Figures 7-8) or an indirect response due to the development of a warmer, wetter Martian climate involving widespread snowfall and ice accumulation. Hence, the hypothesis of existence of ancient ice sheets in itself does not strongly constrain the paleoecological conditions. Requirements are (1) a large mass of surface water or ice, (2) severity or location of the lake is assumed to be a hydrologically closed system. Initial rates of net sublimation vary linearly across the lake from 0.5 mm/yr at the southern edge to 0 at the northern edge. If the lake is assumed to be a hydrologically closed system.
low latitudes would indicate that the Martian environment (but not necessarily the climate) was once very different from today's. Considerable liquid water apparently was involved in the formation of the NPGA. The climatic implications, if any, depend on the mechanism whereby meltwater was generated. The Laurentide and Fennoscandian ice sheets are possible analogs for ancient Martian ice sheets; modern Martian polar caps and polar layered deposits are another model. Liquid water could affect the beds of both types of ice sheets, yet the water would be generated by different mechanisms; hence, very different paleoclimates would be implied. Any model must be reconciled with the nearly pristine condition of the Martian glacial terrains (when viewed at resolutions on the order of 100 m per pixel). The rapid disintegration of Earth's late Pleistocene ice sheets was prompted by a warming climate. If Pleistocene glacial terrains are acceptable analogs of the Martian terrains,
then a similar, relatively warm climate may be implicated in deglaciation on Mars. However, it is not clear whether a warm climate is necessary for deglaciation or whether warming simply was the particular condition that triggered the demise of Pleistocene ice sheets. It is possible that, given a continuing supply of moisture regardless of surface temperatures, ice sheets tend to thicken until subglacial geothermal melting occurs. This idea is supported by the occurrence of basal meltwater accumulations ("lakes") beneath some of the coldest and thickest parts of the East Antarctic ice sheet [Oswald, 1975]. If basal meltwater accumulates, megafloods or ice streaming (or surging) may result. In the most extreme cases (as with the Laurentide and Fennoscandian ice sheets), this water-modified dynamic behavior may spur deglaciation.

If deglaciation on Mars was caused by a warm climate, then there must have been very rapid climatic cooling and drying following deglaciation. Otherwise, even a few tens of thousands of years of sublimation deglaciation would have taken too much time, hundreds of thousands of years of sublimation deglaciation. However, the warmer climate would have left a mark (such as glacial features that are partly blanketed by postglacial sediment, overprinted by normal rainfall runoff valleys or gullies, or otherwise degraded); no such degradation is observed. Perhaps deglaciation occurred rapidly in a modest greenhouse climate [Fanale et al., 1992] that was further modified by a brief period of high obliquity [Kieffer and Zent, 1992; Mellon and Jakosky, 1993], which resulted in warm summers at high and middle latitudes; no sooner had the ice melted than cold, dry conditions set in again and effectively freeze-dried the glacial terrains.

Alternatively, Mars always may have been too cold for significant surface melting. If the modern polar layered deposits are a better analog for the ancient Martian ice sheets, and geothermal basal melting provided the water, or if the catastrophic eruption of highland aquifers and outflow channels provided the water, then no particular paleoclimatic conditions are implied. Ablation of the ice sheets and exposure of subglacial water-formed landscapes may have been accomplished by the slow sublimation of glacier ice under exceedingly cold, dry conditions or relatively warm, dry periglacial conditions; once the terrain beneath such an ice sheet was exposed, there would have been no liquid water to degrade the landscapes.

Regardless of the Martian paleoclimatic, the floods that formed putative Martian tunnel channels, eskers, kames, and other features may have analogs in Earth's Pleistocene history. Subglacial floods, some of enormous magnitude, and consequent formation of vast networks of tunnel channels, eskers, moraines, kames, and other glacial features, were common during the late evolution of the Laurentide and Fennoscandian ice sheets [Wright, 1973; Shaw and Scotese, 1984; Attig et al., 1988; Shaw, 1989; Shaw et al., 1989; Sharpe and Cowan, 1990; Shoemaker, 1990, 1992; Raths et al., 1993; Brennand, 1994; Brennand and Shaw, 1994]. The glaciated and isostatically depressed Hudson Bay area may have filled with subglacial water and then quickly emptied, perhaps repeatedly [Shoemaker, 1992], causing rapid formation of extensive glacial landscapes. Alternatively, such landscapes may have been formed by rapid ice advances promoted by subglacial till streams (mobile subglacial layers of water-saturated clay- and silt-rich sediment) or deformable till beds [Begét, 1986; Alley et al., 1989; Clark, 1994] (instead of or in addition to basal meltwater floods). Either way, on Mars as well as on Earth, meltwater probably was the crucial agent that modified ice-sheet behavior, thus resulting in the formation of the observed landscapes.

3.5. Alternative Glacier-Ice Compositions

The ambiguities of possible paleoclimatic implications of putative Martian glaciers as proposed above are substantial. An added uncertainty pertains to the composition of the glacier ice. Hitherto, we have assumed that Martian glaciers were made of ordinary water ice. Glaciers on Earth are composed of water ice because $H_2O$ is the only abundant atmospheric constituent that can exist at the surface of Earth as a relatively soft and abundant solid near a state of equilibrium with the atmosphere. The fact that $H_2O$ can also...
3.6. Chronology of Glaciation

The NPGA modifies plains that have mapped ages of late early Hesperian, late Hesperian, and early Amazonian, thus providing an upper limit on the age of the hypothesized glacial period(s). We have confirmed Hesperian ages of two areas of NPGA on the basis of crater counts, which show typical Hesperian values of N4 (the number of craters larger than 4 km in diameter per million km²). However, there are strong indications that these terrains were greatly modified well into the Amazonian (Figure 9). Four regions where glaciation is thought to have been an important process have N1-N1.41 (number of craters between 1.0 and 1.41 km in diameter per million km²) that are far less than would have formed in the time since the origins of these terrains. By contrast, areas that have not been glaciated, according to our interpretation, have approximately the densities of small craters that would be expected according to their geologic ages (determined from the densities of large craters). Hence, unglaciated terrains yield almost concordant stratigraphic ages.

![Figure 9](image-url)

Figure 9. Crater densities (numbers of craters per unit area) and possible absolute ages of four terrains that are thought to have been glaciated (Arcadia, Utopia, and Arsia Mons) and four terrains that provide no morphologic indications of ever having been glaciated (three areas of the Tharsis plateau and an area of Sols Planae). N4 is the number of craters larger than or equal to 4 km in diameter per million square kilometers and is equated with the geologic terrain age. N1-N1.41 is the number of craters between 1.0 and 1.41 km in diameter per million square kilometers; this parameter is modeled as the age of erosional or depositional modification (glaciation, if it occurred), as described in the text. “All craters” means that the graphed data do not include any attempt to morphologically classify the craters as to their state of preservation. If a terrain has not experienced a substantial erosional or depositional event, then the stratigraphic ages given by N4 and N1-N1.41 should be concordant (near the bold sloping lines, which represent the production size-frequency cratering curve). The crater production function, which changed at the end of “late heavy bombardment” (approximated here as the end of the early Hesperian), and the densities of these crater densities in terms of stratigraphic ages is from Tanaka [1986]. Proposed translations of crater densities into absolute ages are given on the right-hand axis.
ages for N4 and N1-N1.41, whereas putative glacial terrains yield sharply discordant ages.

A simple interpretation of Figure 9 is that some process obliterated a large fraction of small craters during the early Amazonian while preserving large ones in the few putative glacial terrains. We assume, for the sake of simplicity, that (1) the terrains of interest formed during a single discrete event, (2) a single later event erased craters smaller than 1.41 km completely but did not erase craters larger than 4 km, and (3) no other events have altered the crater size distributions. It follows that the relative model geologic age of terrain formation is given by N4 and that the model age of modification is given by N1-N1.41. The choice of these diameters is partly from the requirements that the numbers of craters must be statistically meaningful and the resolution limits of the imagery must support the diameters of craters counted (counts of craters smaller than 1 km in diameter generally would be unreliable).

The choice of diameters for craters that would and would not be susceptible to obliteration by continental glaciation can be supported on the basis of crater dimensions. Figure 10 shows the dimensions of several component features of impact craters. Obviously, large craters are more difficult to erode or bury completely than small ones. If the depth of a crater, its erosion and burial was typically 100-200 m, comparable to the depths of Martian tunnel channels and the heights of thumbprint ridges (and typical of the average amount of rock removed by the Pleistocene glaciations on Earth [White, 1972, 1988; Bell and Laine, 1982]), then kilometer-sized craters would generally be obliterated while craters larger than 4 km would be preserved, although in degraded form.

If a terrain has not experienced a substantial erosional or depositional event, then the stratigraphic ages given by N4 and N1-N1.41 should be concordant; otherwise N1-N1.41 should be younger than N1. The validity of these ages as lower limits depends on an assumption that other geologic processes subsequent to the putative glaciation did not erode or bury many craters. According to the cratering time scale of Neukam and Wise [1991], the early Amazonian spanned the period from 1800 m.y. to 700 m.y. ago; according to Figure 5, glaciation may have occurred any time during this period. Thus, glaciation seems to have occurred relatively late in Martian history, long after heavy bombardment. The early Amazonian age suggested for these widespread glacial events is consistent with the stratigraphic age of thick deposits in eastern Utopia Planitia around the western flank of Elysium [Chapman, 1994].

The duration of glaciation is very poorly constrained. Analogies with the Laurentide ice sheet suggest a glacially active period longer than 10^5 years. Alternatively, if the ice sheet was analogous to polar layered deposits or if they formed as the evolved remnants of frozen lakes, the ice sheet's construction and sublimation may have required tens of millions of years; however, the primary glacial landscapes formed by wet-based glaciers may have formed during a few weeks to a few years in a given area (the typical durations implied by episodes of catastrophic flooding). A very long lifetime of the ice sheets, at least millions of years, is suggested by the fact that numerous volcanic landforms thought to have been produced by volcanic interactions with an ancient ice sheet are preserved in eastern Utopia Planitia [Chapman, 1994]; it is improbable that all this volcanic activity would have occurred when an ice sheet was present unless it was there for a long time.

4. Conclusions and Possible Observational Tests

Many Martian terrains in the northern plains appear to have been formed by active, wet-based, continental glaciers.

A puzzling characteristic of Mars' putative glacial landforms is that they are so well preserved, although they must be at least hundreds of millions of years old. This and other differences of the Martian glacial terrains with respect to...
terrestrial glacial terrains suggests that an unelucidated aspect of Martian glaciation was involved. The postglacial climate of Mars may have been extremely cold (perhaps similar to today’s climate), thus preventing erosion; or else warm, humid conditions during the postglacial period would have had to have been transient (10^3-10^4 years).

In sum, the hypothesized Martian glaciation does not, at the current level of understanding, require a unique set of palaeoclimatic conditions or glacier composition. However, the capabilities of Mars missions planned for the next 10 years could allow (1) the study of the detailed morphology, sedimentology, and petrography of ancient glacial landforms and deposits, and (2) the determination of the composition, thickness, and climatic conditions of origin of the existing Martian polar caps. In so doing, we may hope to obtain the desired tight constraints on the Martian paleoclimate during the hypothesized ancient era of glaciation and deglaciation. If terrestrial glacial landforms are any guide at all to Mars, the glacial interpretations of Mars imply certain specific characteristics of the landforms and sediment that constitute these landscapes. Unpiloted Mars space missions of the near future may contribute substantially to our understanding of Martian glaciation (Table 1). These missions will include the orbital Mars Global Surveyor spacecraft (launches in 1996 and 1998) and the Mars Surveyor landers (which will include small rovers with limited mobility, starting with Pathfinder in 1996-1997 and new launches about every 2 years). Global imaging at moderate resolution should improve coverage of the northern plains, thus allowing a more complete mapping of thumbprint terrains and putative tunnel channels and eskers.

High-resolution imaging of representative areas may show some of the fine details of putative glacial features. Global laser altimetry may contribute additional knowledge of the morphology and large-scale topography of putative glacial landforms and terrains. Many putative glacial terrains and landforms are expected to be very bouldery, and this should show up in high-resolution images and as high values in maps of thermal inertia generated from thermal emission spectrometer data. The proposed Mars Polar Pathfinder lander may tell us the composition of fines and the abundance of dust in the upper meter of the existing northern Martian polar cap and the thickness of the ice cap at the landing site; hence, we

### Table 1. Possible Observational Tests by Spacecraft

<table>
<thead>
<tr>
<th>Putative Feature</th>
<th>Characteristics Not Readily Discernible in Viking Data</th>
<th>Types of Additional Analysis Needed; Expected Result if the Glaciation Hypothesis is Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eskers</strong></td>
<td>Sorted fluvial sediments; sand, gravel, boulders; cross-bedding</td>
<td>MOC on MGS*: rare boulders; beds observable; eskers forms. Laser altimeter on MGS: eskers morphology; cross low divides. TES on MGS†: High thermal inertia, unless dust-covered. Rover: Sand, gravel, and cobbles; cross bedding coomon.</td>
</tr>
<tr>
<td><strong>Tunnel channels</strong></td>
<td>Fluted bedrock or boulder bed. Some channels trend downhill, others uphill.</td>
<td>Laser altimeter on MGS: Downhill and uphill trends. TES on MGS†: High thermal inertia, unless dust-covered. Rover: Gravelly, bouldery bed; bedrock fluted, water-polished.</td>
</tr>
<tr>
<td><strong>End moraines and till plains</strong></td>
<td>Poorly sorted and some sorted sediments; clay, silt, sand, gravel, and matrix-supported boulders; some faceted and striated; diverse lithologies. Outwash plains and channels.</td>
<td>MOC on MGS: many 1- to 5-m boulders; moraine morphologies; meltwater features such as small channels and outwash plains. Laser altimeter: moraine morphologies. TES on MGS†: High thermal inertia, unless dust-covered. Rover imaging: Fine sediment, gravel, and matrix-supported boulders of diverse lithologies; a few percent of boulders striated and faceted; some melt-water-deposited boulders and lenses possible; small diamicton structures possible. Rover micropetrography and APXS‡: diverse mineralogic and chemical compositions of lithic clasts.</td>
</tr>
<tr>
<td><strong>Glaciolacustrine plains</strong></td>
<td>Mainly clayey-silt, some outsized dropstones; sediments draped over topography up to level of possible wave-cut terraces; varved deposits.</td>
<td>MOC on MGS: Very smooth; if wind- or water-etched, bedding may be visible and lateral extensive; iceberg-ploughed furrows and dropstone boulders possible. Laser altimeter: Very smooth, originally horizontal surface (now may be gently warped by rebound). TES on MGS†: Low to moderate thermal inertia. Rover imaging: Very smooth plain; fine-grained mudstones; dropstone boulders and glaciolacustrine gravel lenses possible. Short drill core: fine laminae and varves. Returned sample: diverse mineralogic and chemical comositions of lithic clasts. Stable isotopes: Authigenic precipitates indicate equilibrium near 273 K (lower if exotic ice composition).</td>
</tr>
</tbody>
</table>

* MOC—Mars Observer Camera (MGS).
† TES—Thermal Emission Spectrometer (MGS).
‡ APXS—Alpha-proton X-ray backscatter analyzer (Pathfinder and Mars Surveyor landers).
may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Petrographic analyses of short drill cores taken from sublacustrine moraines, Generator Lake, Baffin Island, north-central Canadian Arctic Archipelago, demonstrate that tillic facies of several percent of the boulders and cobbles may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Petrographic analyses of short drill cores taken from sublacustrine moraines, Generator Lake, Baffin Island, north-central Canadian Arctic Archipelago, demonstrate that tillic facies of several percent of the boulders and cobbles may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Petrographic analyses of short drill cores taken from sublacustrine moraines, Generator Lake, Baffin Island, north-central Canadian Arctic Archipelago, demonstrate that tillic facies of several percent of the boulders and cobbles may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Petrographic analyses of short drill cores taken from sublacustrine moraines, Generator Lake, Baffin Island, north-central Canadian Arctic Archipelago, demonstrate that tillic facies of several percent of the boulders and cobbles may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Petrographic analyses of short drill cores taken from sublacustrine moraines, Generator Lake, Baffin Island, north-central Canadian Arctic Archipelago, demonstrate that tillic facies of several percent of the boulders and cobbles may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Petrographic analyses of short drill cores taken from sublacustrine moraines, Generator Lake, Baffin Island, north-central Canadian Arctic Archipelago, demonstrate that tillic facies of several percent of the boulders and cobbles may soon know whether the present northern polar cap is likely to be flowing and, therefore, whether it constitutes a true ice cap [Fisher, 1993].

Proof or disproof of the Mars glaciation hypothesis beyond all reasonable doubts may require a traverse to a putative esker or moraine by a rover (preferably one more mobile than those planned for Mars Surveyor) or perhaps analysis of a short drill core from a proposed glacial lake deposit. For instance, a rover could show (or not show) that a putative esker, true to this interpretation, consists of fairly well sorted fluvial sediments rather than basaltic or glacially formed flows, or some other lithology that would be predicted on the basis of another hypothesis of origin. A rover's mission included a drive to a putative esker, true to this interpretation, consists of a variety of unconsolidated features, some of which may reveal varves, which may demonstrate the presence of a glacial lake deposit. Some observers might include analyses of stable, radiogenic, and radioactive isotopes in evaporite or cryogenic salts and authigenic pre-

Victor R. Baker, Department of Geosciences, University of Arizona, Tucson, AZ 85721.
James E. Beget, Department of Geology and Geophysics, University of Alaska, Fairbanks, AK 99775.

KARGEL ET AL.: ANCIENT GLACIATION IN MARS NORTHERN PLAINS


Victor R. Baker, Department of Geosciences, University of Arizona, Tucson, AZ 85721.
James E. Beget, Department of Geology and Geophysics, University of Alaska, Fairbanks, AK 99775.
Low-temperature and low atmospheric pressure infrared reflectance spectroscopy of Mars soil analog materials

Janice L. Bishop
Departments of Chemistry and Geological Sciences, Brown University, Providence, Rhode Island

Carl M. Pieters
Department of Geological Sciences, Brown University, Providence, Rhode Island

Abstract. Infrared reflectance spectra of carefully selected Mars soil analog materials have been measured under low atmospheric pressures and temperatures. Chemically altered montmorillonites containing ferricydrite and hydrated ferric sulfate complexes are examined, as well as synthetic ferricydrite and a palagonitic soil from Haleakala, Maui. Reflectance spectra of these analog materials exhibit subtle visible to near-infrared features, which are indicative of nanophase ferric oxides or oxyhydroxides and are similar to features observed in the spectra of the bright regions of Mars. Infrared reflectance spectra of these analogs include hydration features due to structural OH, bound H\textsubscript{2}O, and adsorbed H\textsubscript{2}O. The spectral character of these hydration features is highly dependent on the sample environment and on the nature of the OH, structural OH features observed in these materials at 2.2-2.3 pm, and 6 pm are reported here in spectra measured under a Martian-like atmospheric environment. In spectra of these analogs measured under dry Earth atmospheric conditions the 1.9-\textmu m band depth is 8-17%; this band is much stronger under moist conditions. Under Martian atmospheric conditions the 1.9-\textmu m feature is broad and barely discernible (1-3% band depth) in spectra of the ferricydrite and palagonitic soil samples. In comparable spectra of the ferric sulfate-bearing montmorillonite the 1.5-\textmu m feature is also broad, but stronger (6% band depth). In the low atmospheric pressure and temperature spectra of the ferricydrite-bearing montmorillonite this feature is sharper than the other analogs and relatively stronger (6% band depth). Although the intensity of the 3-\textmu m band is weaker in spectra of each of the analogs when measured under Martian conditions, the 3-\textmu m band remains a dominant feature and is especially broad in spectra of the ferricydrite and palagonitic soil. The structural OH features observed in these materials at 2.2-2.3 \textmu m and 2.75 \textmu m remain largely unaffected by the environmental conditions. A shift in the Christiansen feature towards shorter wavelengths has also been observed with decreasing atmospheric pressure and temperature in the midinfrared spectra of these samples.

1. Introduction

Spectral experiments on particulate, ferric-containing analogs to Martian soil have been performed to facilitate interpretation of remotely acquired spectra. These analog materials include chemically altered montmorillonites that contain nanophase ferric oxyhydroxides or ferric sulfate complexes, a palagonitic soil and ferricydrite. Since reflectance spectra of such minerals and soils are environmentally sensitive, accurate spectral measurements of these samples at Mars analogs requires measurement under Martian atmospheric conditions. The results presented here encompass spectral experiments of laboratory analog materials to Martian bright region soils measured in an environmental chamber under controlled atmospheric and temperature conditions. As the atmospheric pressure and temperature are reduced in these experiments the partial pressure of H\textsubscript{2}O in the sample chamber is decreased and the samples are dehydrated. The presence of smectites in bright region surface material on Mars has been suggested based on analyses of chemical measurements from Viking [Toulmin et al., 1977; Baird et al., 1977] and is further supported by identification of smectites in some of the SNC meteorites [Gouging et al., 1991; Toulmin et al., 1994]. Spectroscopic analysis in the visible to near infrared have shown that chemically altered, ferric-enriched smectites prepared in the laboratory (e.g., ferricydrite-bearing montmorillonites) exhibit important similarities to the soils on Mars [Bishop et al., 1993a; Bishop, 1994]. Visible absorption bands centered near 0.91 \textmu m are found in ferricydrite and ferricydrite-bearing montmorillonites (J. L. Bishop et al., Reflectance spectroscopy of ferric sulfate-bearing montmorillonites as Mars soil analog materials, submitted to Icarus, 1994; hereinafter referred to as submitted manuscript) and also in the bright soils of Arabia on Mars [Mecord et al., 1982; Murckie et al., 1993a]. Montmorillonites containing hydrated ferric sulfate
complexes exhibit a ferric absorption feature near 0.88 μm (J. L. Bishop et al., submitted manuscript, 1994), which is also observed in the surface material at Lasse Planum [McCord et al., 1977; Murchie et al., 1993b]. Geochronal and spectral analyses of palagonitic soils indicate that they may have formed on Mars as well [Allen et al., 1981; Singer, 1982]. Spectral measurements of palagonitic soils indicate that these materials have a ferric band centered near 0.86 μm and a spectral brightness of about 30-40% reflectance, near that of soils on Mars in the bright regions [Singer, 1982; Morris et al., 1990].

Smectites, ferrhydrite, and palagonitic soils all contain structural OH bonded to metal cations, adsorbed H₂O, and bound H₂O. The infrared absorption features due to OH and H₂O have been measured under variable environmental conditions in smectites [Presn, 1975; Bruckenshal, 1987; Cayne et al., 1989a;; Johnston et al., 1992; Bishop et al., 1994] and in palagonitic soils [Bruckenshal, 1987]. Preliminary experiments involving a variety of Mars analogs under differing environmental conditions examined the influence of exposure history on water content and absorption features due to H₂O [Bishop et al., 1993b]. These experiments showed that reflectance spectra measured under dry conditions of ferrhydrite, ferric sulfate-bearing montmorillonite, and palagonitic soils contain sufficient bound H₂O to retain a strong 3-μm band similar to that observed in spectra of Mars. Spectra measured under similar conditions of naturally occurring smectites (on Earth) and also many ferric-bearing smectites exhibit a weaker 3-μm band.

Low-temperature reflectance spectra of hydrated minerals have exhibited changes in the band shape of the hydration features [Clark, 1981a]. Reflectance spectra of Martian soil analog materials dehydrated under reduced atmospheric pressure have also shown variations in the spectral character of the hydration features [Bishop and Pieters, 1990b]. In the low atmospheric pressure experiments, as well as in some of the low-temperature experiments, these changes are due to dehydration of the sample. Additional experiments involving low-pressure and low-temperature reflectance spectra found changes in the hydration features and in the Christiansen feature of Mars soil analog materials [Bishop and Pieters, 1994]. The effects of the environment on water in clays, oxides, and salts are manifested in the spectral hydration bands measured via reflectance spectroscopy. Environmental influences on the spectral hydration features are especially important in interpreting observations of Mars where subtle spatial variations in the strengths of cation-OH and H₂O absorptions have been observed in telescopic spectra [Bell and Crisp, 1994] and in spectra measured by the French near-infrared imaging spectrometer ISM on the Phobos 2 spacecraft [Bibring et al., 1990; Arnold, 1992; Murchie et al., 1992, 1993a]. For these reasons, reflectance spectra are measured in this study of Mars analog materials under the pressure and temperature conditions of temperate regions on Mars.

2. Background

2.1. Spectral Properties of Mars

Soderblom [1992] and Roush et al. [1993] have recently produced reviews of the spectroscopic properties of the surface of Mars. The spectral properties of the Martian atmosphere [e.g., Rosenvoile et al., 1992; H₂O ice in the north polar cap [e.g., Clark and McCord, 1982], and CO₂ ice in the south polar cap [e.g., Calvin, 1990] have also been studied. Reflectance spectra of the bright areas on Mars are characterized by ferric crystal field transitions in the extended visible region [Adams and McCord, 1969; Singer et al., 1979; Sherman et al., 1982; Bell et al., 1990]. These spectra include absorptions near 0.5 μm and 0.9 μm, an inflection near 0.6 μm, and a reflectance maximum at 0.7-0.8 μm. The absorption features in the spectra of bright regions on Mars have been assigned to octahedrally coordinated Fe³⁺ transitions [Sherman et al., 1982; Burns, 1993] and exhibit band centers ranging from 0.85 to 0.92 μm [McCord et al., 1977; Singer, 1982; Bell et al., 1990; Murchie et al., 1993a,b]. This implies heterogeneous ferric-containing phases or possibly heterogeneous ferric mineralogy in the Martian bright regions.

The shape of the 3-μm absorption band indicates a prevalence of bound molecular H₂O in Martian surface materials [Hoek et al., 1977; Singer et al., 1990; Blaney and McCord, 1989; Erard et al., 1991]. Since CO₂ in the Martian atmosphere obscures the 2.7-2.8 μm region, features due to structural OH species cannot be detected in currently available spectra of the surface materials. A weak 2.2-μm feature, characteristic of structural OH in amorphous clays, has been observed in telescopic spectra of Mars [Bell and Crisp, 1993] and in ISM spectra of some areas on Mars [Arnold, 1992; Murchie et al., 1993a,b]. This suggests that aluminous clay silicates commonly found on Earth occur in some of the bright regions of Mars as well. Variation in the strengths of the 2.2-μm and 3-μm features in spectra of bright and dark regions on Mars [Murchie et al., 1993a] indicates heterogeneity in the type, crystallinity or quantity of silicates present in the surface material, as well as variations in the water content.

2.2. Atmospheric Properties of Mars

Surface temperatures on Mars span the range 130 to 290 K [Kieffer et al., 1977]. Spring temperatures vary from 170 to 230 K in north temperate latitudes, and from 200 to 250 K in south temperate latitudes [Davies, 1979]. Atmospheric pressures on the surface of Mars were measured by the Viking landers and vary from 800 Pa (6.7 mbar) to 1000 Pa (9.5 mbar) [Hess et al., 1979]. From mass spectrometry on the Viking landers, Owen et al. [1977] report a CO₂(g) abundance in the Martian atmosphere near 95%. Climatic models indicate that atmospheric CO₂ pressures vary from 0.1 to 15 mbar [Farquhar et al., 1982]. Atmospheric water vapor on Mars varies diurnally and regionally across the planet from about 1 to 100 perceptible microns [Farmer et al., 1977]. This water vapor abundance indicates the depth of water precipitation that would result if all of the water in the atmosphere precipitated on the surface [Carr, 1981].

2.3. Spectral Experiments Under Low Temperatures

Clark [1981a] presented reflectance spectra of montmorillonites mixed with ice at low-temperatures. A decrease in the intensity of the 1.4-μm and 1.9-μm features and a decrease in the albedo were observed as the temperature was lowered [Clark, 1981a]. This is explained by low-temperature adsorption isotherms for montmorillonite, which show decreases in the amount of adsorbed water under reduced temperatures [Anderson et al., 1967]. Reflectance experiments
involving H₂O or CO₂ frosts have observed temperature-dependent and grain size-dependent spectral behavior [Kieffer, 1970; Clark, 1981b; Lucey and Clark, 1985]. The bandwidths and intensities of the spectral features varied with grain size and temperature, but the band positions remained constant in these studies.

2.4. Spectral Properties of Mars Analogs

Reflectance spectra are shown in Figure 1 for a group of Mars soil analogs from 0.3 to 3.5 μm. These spectra of ferrihydrite, ferrihydrite-bearing montmorillonite, ferric sulfate-bearing montmorillonite and a palagonitic soil from Haleakala, Maui, were measured using a biconical configuration in the Nicolet Fourier transform interferometer (FTIR) under H₂O-purged conditions from 1.2 to 3.5 μm, then scaled to bidirectional spectra measured from 0.3 to 1.2 μm using the reflectance experiment laboratory (RELAB) spectrometer at Brown University. Composite observational spectra of Olympus Mons, Mars [Mustard and Bell, 1994] are shown in Figure 1 for comparison. The telescopic spectrum was scaled to the ISM spectrum at 0.77 μm, and a model was used to remove the strong atmospheric bands. Absorption features due to OH and H₂O are observed in these spectra near 1.4, 1.9, 2.2 or 2.3, 2.75, and 2.8-3.1 μm. The visible and near-infrared spectral features of the altered montmorillonites and ferrihydrite are described in detail by [Bishop et al., 1993a, submitted manuscript, 1994] and of palagonitic soils by [Singer, 1982], [Brueckenthal, 1987], [Morris et al., 1990] and [Golden et al., 1991].

2.4.1. Ferrihydrite. Ferrihydrite is a nanophase, reddish brown ferric oxyhydroxide. The structure contains Fe³⁺ in tetrahedral and octahedral sites bound to OH, or H₂O, and the formula Fe₇₋₈(OH)₆H₂O has been assigned based on X-ray powder diffraction [Eggleston and Fitzpatrick, 1988]. Reflectance spectra of ferrihydrite exhibit subtle ferric absorption features in the extended visible region including a reflectance maximum at 0.80 μm and a reflectance minimum at 0.92-0.93 μm [Bishop et al., 1993a]. Infrared spectra of ferrihydrite include weak features due to structural OH at 1.41 μm, 2.3-2.4 μm, and 2.75 μm, while stronger and broader features due to water are present near 1.95 μm and 3 μm [Bishop et al., 1993a].

2.4.2. Ferric-bearing montmorillonite. Montmorillonite belongs to the smectite mineral group, which is characterized spectrally by features due to adsorbed H₂O, bound H₂O, and structural OH. Chemisorbed water in smectites is bound to the interlayer surfaces or cations and is a necessary part of the smectite structure, while physisorbed water is adsorbed on the interlayer or grain surfaces with longer (looser) bonds at lower energies (longer wavelengths). This chemisorbed and physisorbed water will be informally referred to as "bound" and "adsorbed" water. The term "structural OH" indicates the OH molecules octahedrally coordinated to Al and other cations in the smectite layers.

Spectral features in smectites due to bound water are observed at 1.41 μm, 1.91 μm, 2.9 μm, and near 6 μm [Prout, 1975; Carisii et al., 1983, Bishop et al., 1994], due to adsorbed water at 1.46 μm, 1.97 μm, 3.0 μm, and near 6 μm [Prout, 1975; Carisii et al., 1983, Bishop et al., 1994], and due to structural OH at 1.41 μm, 2.3-2.5 μm, 2.75 μm, and 10-12 μm [Farmer, 1974; Bishop et al., 1994]. Montmorillonites doped with ferric salts (e.g., ferrihydrite-bearing and ferric sulfate-bearing montmorillonite) have shown ferric absorptions near 0.5 μm and near 0.9 μm that are dependent on the chemical conditions of the doping procedure [Bromer et al., 1993; Bishop et al., 1993; L. Bishop et al., submitted manuscript, 1994]. The spectrum of ferrihydrite-bearing montmorillonite (Figure 1) exhibits a ferric...
spectra, a reflectance maximum at 0.78 μm, and a rounded shoulder near 0.6-0.65 μm. Palagonitic soil. This sample (Haleakala crater, Maui) produced the largest effect on the Christiansen feature.

2.5. Christiansen Feature

The Christiansen frequency is defined as the frequency at which the real part of the refractive index of individual particles equals that of the surrounding medium [Henry, 1948; Connell, 1969]. For powdered silicates, the Christiansen frequency occurs at the frequency of maximum emissivity and maximum transmission [Salsbury, 1993]. This feature is characterized by a sharp decrease in α-factor in reflectance spectra and occurs where the real index of refraction, n, is near unity and the extinction coefficient, k, is decreasing strongly.

Spectral experiments [Logan et al., 1973] have shown shifts in the Christiansen feature of quartz and other minerals as a function of particle size, sample texture, background temperature, and atmospheric pressure. In this Logan et al. study, emission spectra exhibited shifts in the Christiansen feature toward longer wavelengths with decreasing particle size and towards shorter wavelengths with increasing atmospheric temperature and pressure. In a more recent study involving reflectance spectra of fine-grained olivine, the Christiansen feature was observed to broaden with decreasing particle size and perhaps shift toward shorter wavelengths as well [Hayes and Mustard, 1994].

3. Experimental Procedures

3.1. Sample Preparation

3.1.1. Ferrhydrite. The ferrhydrite was synthesized by gradually adding potassium hydroxide to a ferric nitrate solution until a pH of 7-8 was reached, according to the procedure for 2-line ferrhydrite described by Schwertmann and Cornill (1991). The resulting moist sample was rinsed with H_2O over a porous frit, lyophilized, and dried sieved to <55 μm. Additional information about this ferric oxyhydroxide and ferric sulfate complexes [Bishop, 1994; J. L. Bishop et al., submitted manuscript, 1994]. A suspension was formed from the SWY montmorillonite and H_2O and repeated washing with dilute HCl produced an H-exchanged form of the montmorillonite.

3.1.2. Ferric-bearing montmorillonites. SWY-1 montmorillonite (The Clay Minerals Society, Source Clays Repository) was chemically altered in the laboratory to produce a large collection of montmorillonites containing a variety of interlayer ferric oxyhydroxide and ferric sulfate complexes [Bishop, 1994; J. L. Bishop et al., submitted manuscript, 1994]. A suspension was formed from the SWY montmorillonite and H_2O, and repeated washing with dilute HCl produced an H-exchanged form of the montmorillonite. The ferrhydrite-bearing and ferric sulfate-bearing montmorillonites included in this study were prepared by adding 0.5 M ferric or ferric sulfate solutions at pH 3 to the H-montmorillonite. Dilute NaOH was added dropwise until pH 5 and pH 3.5, respectively. The suspensions were then centrifuged and lyophilized to form a fine powder, which was dry sieved to <55 μm. Additional information about the ferrhydrite-bearing montmorillonite (# 25) and ferric sulfate-bearing montmorillonite (# 117) used in these experiments can be found in the work by Bishop (1994) and J. L. Bishop et al. (submitted manuscript, 1994).

3.2. Reflectance Spectra

3.2.1. Nicolet. Reflectance spectra were measured relative to a rough gold surface (Infragold) using a Nicolet 740 FTIR in a H_2O and CO_2-purged environment. A PbSe detector was used from 0.9 μm to 3.2 μm and a detector for hydrous sulfate (DTGS) detector from 1.8 μm to 25 μm. The sample chamber, including a remotely controlled sample insert with multiple sample positions, was purchased from SpectroTech. Bulk powdered samples are measured horizontally in this system. Spectra are averaged from two locations on each of two replicates for each kind of sample.

3.2.2. RELAB. Additional spectra were measured under ambient conditions relative to H20 in the visible and near-infrared using a photomultiplier tube with the bidirectional RELAB spectrometer at Brown University (see J. L. Bishop et al., submitted manuscript, 1994 for analyses of these spectra). A detailed description of this instrument is provided by Mustard and Pieters (1988). The bidirectional spectrometer allows for variable illumination (i) and emergent (e) angles.
Spectra included here were measured at $i = 30^\circ$ and $e = 0^\circ$. The sample dish is rotated during the measurement to eliminate orientation effects.

### 3.2.3. Environmental chamber

Reflectance spectra were also measured under controlled atmospheric pressure and temperature conditions using an environmental chamber attachment (for the Nicolet 740 FTIR) built by Connecticut Instruments. Reflectance spectra were measured using a liquid nitrogen-cooled mercury-cadmium-telluride (MCT) detector in the range 1.5-20 μm for each sample under ambient pressure and then again at several increments under reduced atmospheric pressures. Special care was taken to deflect the IR beam during pressure adjustment to avoid additional heating of the sample. Dry pressure was pumped down to 1 bar, then to approximately 40% relative humidity giving a torrs (-13 mbar) at 1-bar atmospheric pressure. The partial pressures and temperatures.

Pressure adjustment to avoid additional heating of the sample.

### 4. Results

Reflectance spectra are shown of ferrihydrite, ferrihydrite-bearing montmorillonite, ferric sulfate-bearing montmorillonite, and a palagonitic soil under variable atmospheric and environmental conditions in Figure 2. Spectra of each sample were measured under ambient atmospheric air pressure at 2.5 μm, under 1 mbar CO$_2$ pressure at near -40°C, and under 1 mbar CO$_2$ pressure at near -1 W6°C. The spectra shown in Figure 2, panels 2a, 3a, 4a, and 5a, extend from 5000 to 2500 cm$^{-1}$ (1.8 to 4.0 μm); the spectra in Figure 2, panels 2b, 3b, 4b, and 5b, extend from 1800 to 1250 cm$^{-1}$ (5.3 to 8.0 μm); and the spectra in Figure 2, panels 2c, 3c, 4c, and 5c, extend from 1400 to 800 cm$^{-1}$ (7 to 12 μm).

In Figure 2, panel 2a, the broad feature at 1.95 μm and the shoulder at 2.5 μm observed in the ambient ferrihydrite spectrum are reduced in intensity in the spectra measured under low-pressures and temperatures. The spectra measured under low-pressure and temperature conditions also show a weak 2.75-μm feature that is not resolvable in ambient spectra. In Figure 2, panel 3a, the spectra of ferrihydrite-bearing montmorillonite measured under low-pressure and temperature conditions show little change in the features at 2.2 and 2.75 μm due to structural OH, but suppression of the features at 1.9, 2.5 and 3 μm due to adsorbed and bound H$_2$O is observed. Similarly, the spectra of ferric sulfate-bearing montmorillonite in Figure 2, panel 4a, show little change at 2.2 and 2.75 μm; however, significant loss in intensity at 1.9, 2.5 and 3 μm is observed. In spectra of the palagonitic soil shown in Figure 2, panel 5a, the 1.9-μm feature is essentially removed, the 2.2-μm band is enhanced, and the broad 2.5- and 3-μm features are weakened in spectra measured under low-pressure and temperature conditions.

Water vapor fine structures are seen near 6.1 μm is seen in spectra of each of these samples in Figure 2, panels 2a, 3a, 4a, and 5b. In the ferrihydrite spectra (Figure 2, panel 2b) this feature nearly disappears under reduced atmospheric pressures and temperatures. In Figure 2, panels 3b, 4b, and 5b, this water vapor feature in the spectra of the ferrihydrite-bearing montmorillonite, the ferric sulfate-bearing montmorillonite, and the palagonitic soil is substantially reduced upon decreasing the atmospheric pressure, and reduced again slightly upon lowering the temperature.

For spectra of the montmorillonite and the palagonitic soil shown in Figure 2, panels 3c, 4c, and 5c, a shift in the Christiansen feature was observed towards shorter wavelengths as the atmospheric pressure and temperature were lowered. In the ferrihydrite spectra (Figure 2, panel 2c) there is an absorption feature near 1250 cm$^{-1}$ (7.5 μm) that disappears under dry conditions, thus complicating detection of shifts in the Christiansen feature for ferrihydrite. The CO$_2$ feature was required to achieve the lower temperatures in these experiments, but did not appear to influence the position of the Christiansen feature (e.g., no change in the Christiansen feature was observed as a function of the CO$_2$ pressure).

Band depths have been calculated using the method of Clark and Roush (1984) for the spectra shown in Figure 2. For the ferrihydrite spectra band minima of 1.95 μm, 2.32 μm, and 2.95 μm were used. For the spectra of
Figure 2. Reflectance spectra of ferrihydrite (panel 2), ferrihydrite-bearing montmorillonite (panel 3), ferric sulfate-bearing montmorillonite (panel 4), and a palagonitic soil (panel 5) measured under variable atmospheric conditions are shown from: (a) 5500 cm⁻¹ to 2500 cm⁻¹ (1.8 pm to 4.0 pm), (b) 1800 cm⁻¹ to 1250 cm⁻¹ (5.5 pm to 8.0 pm), and (c) 1400 cm⁻¹ to 800 cm⁻¹ (7 pm to 12 pm). The spectra shown here were measured under ambient pressure and temperature (solid curve), under reduced atmospheric pressure (patterned curve), and under reduced atmospheric pressure and temperature (dashed curve).

Infrared Features due to Bound and Adsorbed H₂O

Reducing the atmospheric pressure and temperature in these experiments decreases the water vapor in the sample.
Table 1. Band Depths for Hydration Features

<table>
<thead>
<tr>
<th>Feature, μm</th>
<th>1.91</th>
<th>1.93</th>
<th>1.95</th>
<th>2.21</th>
<th>2.32</th>
<th>2.95</th>
<th>6.08</th>
<th>6.14</th>
<th>6.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferrihydrite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient</td>
<td>14</td>
<td>8</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-44 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferrihydrite-bearing montmorillonite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient</td>
<td>12</td>
<td>12</td>
<td>84</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-47 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferric sulfate-bearing montmorillonite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient</td>
<td>17</td>
<td>11</td>
<td>89</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-42 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palagonitic Soil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient</td>
<td>8</td>
<td>1</td>
<td>90</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-38 °C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All values are in percent.

Changes in the spectral hydration features in these experiments depended on the nature of the bonds binding H₂O molecules to the mineral surfaces or interlayer cations. For example, Mg-exchanged montmorillonites exhibit greater decreases in the intensities of the 1.9-μm and 3-μm features than Fe-exchanged montmorillonites, because the interlayer Mg cations are bound to more H₂O molecules [Bishop et al., 1994]. Additional spectral experiments using a Wyoming bentonite showed decreases in the intensities of the features near 1.4 μm and 1.9 μm in the temperature was lowered [Clark, 1981a].

Variation in the behavior of the hydration features in spectra of the ferrihydrite, ferrihydrite-bearing and ferric sulfate-bearing montmorillonites and palagonite soil measured under low atmospheric pressure and temperature conditions (Figure 2) indicates differences in how water is present in these samples. Detailed descriptions of the spectral character of these Mars analogs as a function of atmospheric conditions follow. First, however, there is one trend shared by spectra of all of these samples: reduction in the 2.5-μm shoulder under reduced pressures and temperatures. This feature at 2.5-μm in the ambient spectra of ferrihydrite, montmorillonite, and palagonite is thus apparently due to H₂O adsorbed in a similar manner in each of these samples.

The spectral character of the hydration features in chemically altered montmorillonites is dependent on the ferric interlayer components present. The spectra of the ferrihydrite-bearing montmorillonite in Figure 2, panels 3a and 3b, show a decrease in the intensity of the 3-μm and 6-μm features when the atmospheric pressure was reduced to 1 mbar, but no additional intensity loss was observed upon reducing the pressure further or lowering the temperature. In contrast the spectra of the ferric sulfate-bearing montmorillonite in Figure 2, panels 4a and 4b, show continual loss of intensity of the 3-μm and 6-μm features as the pressure and temperature are lowered. This indicates that the adsorbed H₂O molecules are removed from the ferrihydrite-bearing montmorillonite under relatively mild dehydrating conditions and that additional H₂O in this sample is tightly bound (chemisorbed). For the ferric sulfate-bearing montmorillonites there must be additional H₂O molecules in the system that are more tightly bound than the chemisorbed H₂O. The spectra of both the ferrihydrite-bearing montmorillonite and the ferric sulfate-bearing montmorillonite in Figure 2, panels 3a and 4a exhibit subtle features near 3400 cm⁻¹ (2.95 μm) and 3200 cm⁻¹ (3.12 μm) when measured under dehydrated conditions. However, spectra of these samples measured in a moist environment show a broad 3-μm band lacking distinguishable features [J. L. Bishop et al., submitted manuscript, 1994].

The H₂O stretching and bending combination band near 1.9 μm loss intensity primarily as a result of dehydrating under reduced atmospheric pressure, with only secondary loss in intensity from dehydration under reduced temperature, in spectra of the ferrihydrite-bearing and ferric sulfate-bearing montmorillonites. The band depths decreased from 12% to 6% for the ferrihydrite-bearing montmorillonite spectra and from 17% to 6% for the ferric sulfate-bearing montmorillonite spectra. Although the band minimum is near 1.9 μm in
spectra of both samples, this feature is much broader toward longer wavelengths in spectra of the ferric sulfate-bearing montmorillonite, especially noticeable under dehydrated conditions. This long-wavelength shoulder must be linked to the long-wavelength broadening of the 3-μm band due to the interlayer sulfate complexes in this sample.

These experiments indicate that bound H₂O in the ferrhydrite-bearing and ferric sulfate-bearing montmorillonites exhibit spectral features near 1.91, 2.9, 3.1, 5.9, and 6 μm. Adsorbed H₂O and H₂O coordinated to the sulfate groups exhibit broad spectral features near 1.97, 3, and 6 μm that hinder identification of the features due to bound H₂O. Water-coordinated to the sulfate groups in the ferric sulfate-bearing montmorillonites is more firmly bound than the adsorbed H₂O, and therefore, broad features near 1.97 and 3 μm can be observed in spectra measured under lower pressures and temperatures. Analyses of these ferric sulfate-bearing montmorillonites, including Mössbauer spectroscopy, differential thermal analysis and X-ray diffraction, suggest that small particles of the mineral Schwenktsite and hydrated ferric sulfate complexes surround the montmorillonite grains (J. L. Bishop et al., submitted manuscript, 1994).

The broad combination band centered near 1.95 μm in spectra of ferrhydrite (Figure 2, panel 2a) loses intensity (14% to 5% band depth) and shifts towards 1.8 μm as the pressure and temperature are lowered. The broad 3-μm band due to H₂O stretching vibrations in these spectra decreases in intensity as well (94% to 80% band depth), and a band minimum near 2.95 μm can be detected under dehydrated conditions. The 6.0-μm H₂O bending feature (Figure 2, panel 2b) disappears as the pressure and temperature are reduced, and another feature near 1540 cm⁻¹ (7.5 μm) becomes weaker under lower pressures and finally appears as two features at 1400 cm⁻¹ (7.1 μm) and 1300 cm⁻¹ (7.7 μm) in the low-temperature spectra. It is unclear whether these features at 7.1 and 7.6 μm were present, but masked, in the ambient spectra, or whether a change in the ferrhydrite structure is responsible for these features. Another feature near 1400 cm⁻¹ (6.7 μm) broadens and shifts slightly toward shorter wavelengths as the pressure and temperature are decreased. Reflectance spectra of hematite (Salisbury et al., 1991) exhibit absorption features near 6 and 6.5 μm that may be related to those observed for ferrhydrite. Absorption features in reflectance spectra of goethite are observed near 5.5 and 6 μm (Salisbury et al., 1991).

The 1.9-μm H₂O combination band weakens such that it is barely discernible in spectra of the palagonitic soil measured under reduced pressures and temperatures. The broad 3-μm band exhibits continued H₂O loss as the pressure and temperature are decreased (90% to 79% band depth), and a band center is observed near 2.05 μm under low-pressures and temperatures. The H₂O bending feature near 6 μm decreased in intensity, as well, with decreasing pressure and temperature.

In summary, the ferrhydrite and palagonite soil contain a large amount of adsorbed H₂O and bound H₂O that is not constrained to claylike interlayer sites. Both forms of water are manifested in the spectra of these materials as a broad, strong (90-94% band depth) feature near 3 μm under ambient atmospheric conditions. Upon dehydration under Marslike conditions, ferrhydrite and palagonitic soil exhibit a very weak or absent feature near 1.9 μm and a weakened, less broad 3-μm band (although intensity is lost this feature remains very strong). In contrast, spectra of the ferrhydrite-bearing and ferric sulfate-bearing montmorillonites exhibit primarily a sharpening of the 1.9-μm and 3-μm bands when measured under Marslike conditions. These materials contain H₂O bound in specific interlayer sites resulting in features near 1.9 μm, 2.95 μm, and 3.1 μm under dehydrated conditions. Additional H₂O molecules coordinated to sulfate complexes, or adsorbed in the interlayer region or along grain surfaces give broadening near 1.97 μm and 3 μm.

5.2. Infrared Features due to Structural OH

The fundamental OH stretching feature at 2.75 μm and the combination band at 2.2 μm in spectra of the montmorillonites in Figure 2, panel 4a, remains unchanged as a result of the decreases in atmospheric pressure and temperature. The band depth of the 2.2-μm structural OH combination band is ~12% in spectra of the ferric sulfate-bearing montmorillonite and ferric sulfate-bearing montmorillonite. The fundamental OH bending features at 920 cm⁻¹ (10.9 μm), 880 cm⁻¹ (11.4 μm), and 850 cm⁻¹ (11.8 μm) in spectra of the montmorillonites in Figure 2, panels 3c and 4c, were also unaffected by the environmental changes. Octahedrally coordinated OH groups in smectites are responsible for these features (Farmer, 1974). In montmorillonites, primarily Al cations occupy these octahedral sites, giving rise to the features observed at 2.2, 2.75, and 10.9 μm. Small amounts of Fe and Mg substitute for Al in these sites are responsible for the features at 11.4 and 11.8 μm (Farmer, 1974) and the long-wavelength shoulder of the 2.5-μm band (Bishop et al., 1993a).

Spectra of ferrhydrite measured under low-pressures and temperatures (Figure 2, panel 2a) exhibit a fundamental OH stretching feature at 2.75 μm that is masked by the strong 3-μm water band in spectra measured under ambient conditions. The strength of the combination band at 2.2 μm is slightly enhanced in spectra of ferrhydrite under dehydrated conditions. The OH groups are octahedrally coordinated to Fe instead of Al and vibrate at a lower energy (longer wavelength) because the O is more tightly bonded to the Fe cation. This feature observed at 2.3 μm is similar to that observed in nanoterrane, a smectite that has primarily Fe in the octahedral site (Farmer, 1974). However, nontronite has strong crystal-field bands in visible spectra (Singer, 1982; Burns, 1993), which are not observed in the spectra of ferrhydrite. Additional mid-IR OH bending features are seen in nanoterrane near 820 cm⁻¹ (12.2 μm) (Farmer, 1974; Stubican and Roy, 1961), but not seen in the spectra of ferrhydrite in Figure 2, panel 2a, or in Bishop et al. (1993a). Ferrhydrite has only nonaqueous crystalline order (Eggler and Fitzpatrick, 1988), which may explain why the features observed in spectra of nanoterrane near 0.6, 0.9, 2.3, 2.75, and 12 μm (Singer, 1982; Burns, 1993; Farmer, 1974; Stubican and Roy, 1961), due to octahedrally coordinated Fe and OH species, appear only weakly if at all in the spectra of ferrhydrite.

The spectra of the palagonitic soil in Figure 2, panel 5a, exhibit an OH combination band at 2.2 μm despite the fact that the fundamental OH stretching and bending features are not distinguishable in Figure 2, panel 5c. The 2.2-μm band strength increases from 1% to 6% in spectra of the palagonitic soil as the atmospheric pressure and temperature were reduced. This effect is more noticeable in the broad 3-μm band is probably due to a fundamental OH stretching feature at 2.75 μm that is unresolved. The presence of a feature near 2.2 μm and possible feature near 2.75 μm indicate a phyllosilicate
component in this soil having Al in octahedral sites coordinated to OH groups. Since smectites have been found in palagonitic soils from Hawaii [Golden et al., 1993], it is likely that the palagonitic soil studied here has partially weathered to montmorillonite.

5.3. Christiansen Feature

Under low atmospheric pressure and low-temperature conditions the Christiansen feature is observed to broaden and shift towards shorter wavelengths in these experiments. The Christiansen feature occurs at 1245 cm\(^{-1}\) (8.0 \(\mu\)m) in both the montmorillonite samples under 1-bar air and 23°C (Figure 2, panels 3c and 4c). This feature shifts gradually to 1270 cm\(^{-1}\) (7.9 \(\mu\)m) to 1200 cm\(^{-1}\) (8.3 \(\mu\)m). The shifts in the Christiansen feature experienced here parallel previous observations by Logan et al. [1973] of shifts in the Christiansen feature in emission spectra of silicates toward shorter wavelengths with decreasing atmospheric pressure and temperature. The shifts towards shorter wavelength in the Christiansen feature observed here, however, were not accompanied by increased spectral contrast, as was observed by Logan et al. [1973].

Possible explanations for the shifts in wavelength of the Christiansen feature in the low-temperature reflectance spectra shown here include apparent changes in the optical properties of the bulk materials as a function of the temperature or a temperature gradient, as a function of dehydration, or as a function of effective sample particle size or texture. There is no precedent for dehydration of a material affecting optical properties. Decreasing the particle size in fine-grained olivine, however, has been observed to broaden the Christiansen feature in reflectance spectra with an apparent shift towards shorter wavelengths [Hayes and Mustard, 1994]. In the case of fine-grained soils, the sample texture is a complex mixture of tiny grains clustered together as aggregates. If water molecules adsorbed on the grain surfaces play a role in the ability of these small particles to cluster together, then dehydrating the soils may influence the sample texture and result in an effectively smaller particle size. These and other possible environmental effects on fine-grained soils need to be further evaluated.

6. Conclusions

The Mars soil analogs examined in these experiments are compositionally consistent with the chemical analyses of Viking and exhibit spectral features in the visible part of the spectrum that are similar to features observed in the bright regions on Mars. Reflectance spectra of these analog materials were measured under reduced atmospheric pressure and temperature environments comparable to the temperate zones on Mars. These experiments showed that the structural OH features near 2.2-2.3 \(\mu\)m and 2.75 \(\mu\)m in spectra of the palagonitic soil and ferrhydrite became enhanced when measured under Mars-like atmospheric conditions. The spectral features due to structural OH in ferrhydrite-bearing and ferric sulfate-bearing montmorillonite remained unchanged as a result of altering the atmospheric pressure and temperature. The spectral features due to adsorbed H\(_2\)O in these samples were lost or weakened under low-pressure and temperature conditions. Therefore broad features centered at 2.9 \(\mu\)m in the ferrhydrite and palagonitic soil are due to bound H\(_2\)O, and features at 1.9 \(\mu\)m, 2.9 \(\mu\)m, and 3.1 \(\mu\)m in the low-pressure and temperature spectra of the ferrhydrite-bearing and ferric sulfate-bearing montmorillonites are due to bound H\(_2\)O. The ferrhydrite-bearing montmorillonite contains adsorbed H\(_2\)O that is lost under relatively mild dehydrating conditions and bound H\(_2\)O that is a necessary part of the structure and is resistant to dehydration under low atmospheric pressure and temperatures. In addition to adsorbed and bound H\(_2\)O, the ferric sulfate-bearing montmorillonites contain H\(_2\)O molecules coordinated to sulfate groups, which are loosely bound to the montmorillonite system.

Spectra of the palagonitic soil when examined under atmospheric pressures and temperatures of temperate regions on Mars, exhibit enhanced near-infrared features, suggesting the presence of montmorillonite. This is further evidence that palagonitization is linked to smectite formation and that palagonitization of basaltic volcanic glass on Mars may be a mechanism for smectite formation in the surface material there.

7. Applications to Mars

Based on chemical composition, visible reflectance spectra, and infrared reflectance spectra acquired under low atmospheric pressures and temperatures, these samples are promising analog materials for the bright regions on Mars. The ferrhydrite-bearing montmorillonite or a mixture containing ferrhydrite and a silicate phase may be present in bright regions such as Azalih where the ferric absorption occurs at 0.91-0.92 \(\mu\)m. The ferric sulfate-bearing montmorillonite may be present in bright regions such as Lunae Planum where the ferric absorption occurs near 0.88 \(\mu\)m. The nature of the ferric and silicate components vary greatly in palagonitic soils; however, many exhibit hematite-like absorptions and may be present in bright regions having a ferric absorption near 0.85 \(\mu\)m.

The experiments presented here show the sensitivity of the spectral hydration features in soils to the atmospheric environment of the sample, and the importance of measuring analog soils under the appropriate environmental conditions. Under Marslike atmospheric conditions, the palagonitic soil exhibited a structural OH feature with a 6% band strength near 2.2 \(\mu\)m. A feature near 3.2 \(\mu\)m has also been observed by multiple researchers in some of the Martian bright regions [Arnold, 1991; Murchie et al., 1992, 1993a; Bell and Crisp, 1995]. Spectra of ferrhydrite-bearing montmorillonite and ferric sulfate-bearing montmorillonite samples are characterized by a stronger 2.2-\(\mu\)m feature (11-12% band strength). Spectra of ferrhydrite include a broad feature near 2.3 \(\mu\)m, which may be related to a feature in this region observed in some bright regions on Mars. Under Marslike atmospheric conditions (1) the 1.9-\(\mu\)m feature is substantially weaker in spectra of each of the samples, especially for the palagonitic soil and ferrhydrite, where the 1.9-\(\mu\)m feature is barely discernible, (2) the feature near 2.2-2.3 \(\mu\)m is stronger than the 1.9-\(\mu\)m feature in spectra of each of these analog materials, and (3) a strong 5-\(\mu\)m band is present in the spectra of ferrhydrite, palagonitic soil, and ferric sulfate-bearing montmorillonite. The results presented here are an important step in constraining interpretations of the hydration features in Mars bright region spectra.
References


Page intentionally left blank
Longitudinal dunes on Mars: Relation to current wind regimes

Pascal Lee and Peter C. Thomas
Center for Radiophysics and Space Research, Cornell University, Ithaca, New York

Abstract. Longitudinal dunes are extremely rare on Mars, but constitute a substantial fraction of terrestrial desert dunes. We report finding isolated examples of longitudinal dunes on Mars and relate their occurrence to expected sand transport regimes. Terrestrial longitudinal dunes form in bimodal and multimodal transport regimes. General circulation models and streak data indicate that bimodal and multimodal transport of sand should be very rare on Mars. Thus the dearth of longitudinal dunes on Mars is consistent with their apparent formation conditions on Earth.

1. Introduction and Summary

Although dunes are prominent features on the surface of Mars that are easily related to common terrestrial physical processes, the relative proportions of dune types on Mars are very different from those on the Earth. The scarcity of longitudinal dunes on Mars compared to their relatively common occurrence on Earth suggests some fundamental differences in the Martian eolian system that may have to be taken into account by models that relate eolian features to present or past climates, sediment sources, or to properties of the dune materials themselves.

This paper uses the characteristics of the few reported examples of Martian longitudinal dunes and the expected wind regimes to investigate the significance of their scarcity. We first briefly review the relationship of dune formation to wind regime and sand supply on the Earth, with particular emphasis on bimodal and multimodal wind regimes and dune forms. We review terrestrial longitudinal dunes and emphasize the findings that their formation requires bimodal or multimodal transport. Then the circumstances of occurrence of dunes on Mars, and particularly of longitudinal ones, are summarized, with a description of two newly found occurrences of longitudinal dunes in the north polar region. The longitudinal dunes occur in restricted areas at 73°N, 57°W and 71.4°N, 52.5°W and appear to be seifs spawned from the asymmetric elongation of barchan horns. We then evaluate the likely sand transport regimes on Mars using results from the Ames general circulation model, observations of wind streaks on Mars, and Viking Lander data. We show that the models and data available allow for only very limited multimodal transport on Mars, if any. We then focus on the relationship of the longitudinal dunes in the north polar erg to the current wind regime. The possibility of dune stabilization by salt cementation is also considered, although, on the basis of the good coincidence of many bedforms with current winds, we suggest that the dunes were formed by recent wind regimes.

We conclude that the scarcity of longitudinal dunes on Mars is entirely consistent with the unimodal transport regimes expected and observed for most of the planet.

2. Wind Regimes and Dune Forms

2.1. Wind Regimes and Sand Transport

Eolian dunes are bedforms in loose materials transported in saltation by the wind. Most dune forms are distinctive enough that they provide excellent evidence of the process of saltation even when seen with marginal resolution on another planet. Although dunes include a wide variety of sizes and shapes and occur in many geological, topographic, and climatic settings, there are useful generalities relating dune forms to the strength and variation of sand transport. These generalities should apply to dune forms on Mars and many of the worst complicating factors intervening on Earth are absent on Mars: vegetation and other effects of moisture.

The wind regimes to which dune forms can be compared are often described by the drift potential (DP), i.e., the total sand flux in mass/unit length/time calculated from wind measurements, and its relation to the resultant drift potential (RDP), which is the calculated net transport. The DP is the scalar sum of the calculated transport; RDP is the magnitude of the vector sum of the transport. Works such as those by Fryberger [1979] and Wason and Hyde [1983] describe the techniques for calculating drift potentials and their diverse applications. The calculation of the drift potential depends upon use of a relation of the mass transport rate, \( \dot{m} \), of loose sand to the wind shear velocity, \( V_s \). Many expressions have been derived from theory, wind tunnel testing, and field measurements to relate sand transport rates to shear velocity (e.g., Bagnold, 1941; Kanomura, 1951; Knudt, 1965; Lettau and Lettau, 1971; White, 1979). Summary comparisons are given by Greeley and Iversen [1985] and Sarre [1987]. The expression of White [1979] contains the elements essential to models of sand transport:

\[
\dot{m} = A \rho V_s (V_u - V_s) + \rho V_s \dot{V}_u / g (1)
\]

where \( A \) is a constant, \( \rho \) is atmospheric density, \( V_s \) is the shear velocity, \( V_u \) is the impact threshold shear velocity,
and \( g \) is gravitational acceleration. \( A \) is in \( \text{m}^2 \text{s}^{-2} \) units. \( V_* \) of course varies with particle size and density. For purposes of comparison, one particle size and density are assumed, and often, the most easily moved size (\( \sim 150 \mu \text{m} \) on Mars [Greeley and Iversen 1985]). We can simply write the sand transport quantity, \( Q \), as

\[
Q = A(V_1 - V_2)(V_1 + V_2) t
\]

(2)

where \( A \) is a constant and \( t \) is the time this particular wind condition is active. \( Q \) is the magnitude of a vector; its orientation is taken from measurements or a model. Over a period of time, a locale is subject to a variety of winds, and we sum the various \( Q \)'s to get the DP, and sum the vectors to get the magnitude and orientation of the RDP. The transport variability, TV, is then given by

\[
TV = \frac{\text{RDP}}{\text{DP}}
\]

(3)

This parameter is used by Fryberger [1979] and Wasson and Hyde [1983] to indicate the variability in sand transport directions and to relate quantitatively the wind regime to dune morphologies.

The value of \( TV \) can vary from 0 to 1. We illustrate some combinations of winds that give different values of \( TV \) to provide context for discussion below. Two equal transport vectors, 90° apart, give a value of \( TV \) of 0.71. Without at least one sand-transporting wind component more than 90° from another, the value of \( TV \) cannot fall below 0.71. Figure 1 shows the calculated sand transport from a set of winds, assumed to blow for equal periods of time, that vary from weakest to strongest by a factor of about 1.5. In plot A the winds range from 1.35 times \( V_* \) to 2.07 times; in plot F from 0.77 times to 1.18 times \( V_* \). The only change from A to F is the proportional decrease in all the wind magnitudes. As the winds decrease, the effective transport regime goes from multimodal (low values of TV) to unimodal (high values of TV) because some winds are below threshold and the others have decreasing transport capability. Simply by scaling the winds shown in A by 0.6, the transport changes to that in example E. Fryberger [1979] presents sand rose plots and drift potential statistics, along with Landsat images and dune morphology maps for a variety of terrestrial sand seas that can be compared to these idealized examples.

### 2.2. Sand Supply and Sand Transport

The most commonly used measure of the availability of sand is the equivalent sand thickness (EST). The EST is the average depth of sand if the dunes were spread out evenly. This quantity is difficult to measure accurately, but is fairly simple to estimate approximately because of the general geometric rules that apply to piles of sand formed by saltation accumulation. Measurement of dune widths, lengths, and spacing from aerial or orbital images can give reasonable estimates of sand volumes. Lancaster and Greeley [1990] discuss applications of EST measurements to Martian dunes. As discussed by Wasson,

<table>
<thead>
<tr>
<th>Wind</th>
<th>Drift Potentials</th>
<th>RDP</th>
<th>RDP/DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>0.61</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td>0.60</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>0.89</td>
</tr>
</tbody>
</table>

Figure 1. Model drift potentials (DPs) and resultant drift potentials (RDPs) for a four-wind system. The only difference among the examples is the gradual reduction in all winds from A to F; the ratio of wind speeds is the same throughout; as winds decrease, fewer components are above threshold, and transport by those that are above threshold decreases (see equation (1)). Simply by scaling the entire wind system by 0.6 from A to F, the resultant transport has gone from a strong multimodal regime to a weak, unimodal one. Note these are not sand roses; wind and transport arrows are directed downwind.
Drift potential has been used to describe the vigor of sand transport. Its relation to dune form or to EST is complex, though it may in some areas provide some additional discrimination. It is sufficient to emphasize here that strong winds do not mean large accumulations of sand.

2.3. Dune Forms and Wind Regimes

Apart from any influence from vegetation or local obstacles, dune forms are mostly related to the directional variability of effective winds and to the equivalent sand thickness; wind strength and particle size appear to have minimal influence [Wasson and Hyde, 1983]. Although morphologic and genetic classifications of dunes abound, it is most useful here to consider four groups: barchan, transverse, longitudinal, and star [see McKee, 1979; Lancaster, 1982, 1989; Lettau and Lettau, 1977]. Subtypes of course are recognized, and supplement the overall classification (e.g., longitudinal dune types of simple, compound, and complex) [Lancaster, 1982]. Good summaries of the range of dune morphologies on the Earth are found in the works by McKee [1979], Breed and Grow [1979], Fryberger [1979], and Fryberger and Goudie [1981]. Comparisons of terrestrial dune types with wind regimes on regional and continental scales have been made by Fryberger [1979], Wasson and Hyde [1983], Tsoar [1989], and Lancaster [1989]. These studies all show that dune types correlate very strongly with wind regime: barchans occur in the most unimodal winds; transverse dunes in slightly less unimodal; longitudinal dunes form along or close to the net transport vector of substantially bimodal or multimodal winds, and star dunes form in the most complex, multimodal winds. To illustrate the generality of these trends, we have replotted results from Wasson and Hyde [1983] and from Fryberger [1979] in Figure 2. Wasson and Hyde plot RDP/DP versus the equivalent sand thickness. Their data include dunes in Africa, Australia, South and North America, and show good distinction of dune types based on the combination of wind regime and sediment thickness. Note that Wasson and Hyde's data are given with 1 sigma ranges. Fryberger's data come from Africa, Asia, Australia, and North America, and are given as total ranges of RDP/DP for three groups of dunes, with no distinction of sediment thickness.

We emphasize that correlation of dune types with wind regimes involves considerable interpretation and extrapolation. The interpretation rests on classification of forms that are detected in the field and in aerial or orbital images with varying resolutions and lighting. The extrapolation resides primarily in applying wind data recorded at the margins of sand seas to their interiors (maps by Fryberger [1979] and Fryberger and Ahlbrandt [1979] are instructive). Extrapolation is also involved in assigning dune forms to present winds or to past winds, a problem for the very largest forms and for those that may be stabilized. In general, however, we regard the
correlation between forms and winds to be good enough for simple, broad comparisons.

2.4. Longitudinal Dunes and Their Significance

The terminology of "linear" and "longitudinal" has caused much confusion and grief. For the most part, linear is taken as a morphological term, longitudinal as defining a form that parallels a resultant transport vector. Linear and longitudinal dunes are dealt with in detail by Lancaster [1982] and Toor [1989].

Longitudinal dunes cover half to two thirds of the Earth's sand seas. They are relatively less common in Asia and in the Americas than in Africa, but are completely absent in only a few sand seas [Lancaster, 1982].

The formation of longitudinal dunes has been a subject of controversy. Several kinds of such dunes are often distinguished (e.g., lee dunes, vegetated-linear dunes, and self dunes [Toor, 1989]), each kind involving apparently a distinct mode of formation. Lee dunes are easily interpreted, and vegetation is absent from Mars, so our task reduces to comparing self-like dunes on the two planets. Self dunes are often simuous, kinked or convoluted in plan view, and have sharp-edged crests when fresh. Closely spaced selfs sometimes interact and their wiggly crest-lines may merge in Y-shaped junctions. Self dunes on Earth are commonly found in association with barchans or barchanoid dunes, and can be derived from the asymmetric elongation of barchan horns under a bidirectional wind regime [Renno, 1941; Lancaster, 1980; Toor, 1989]. Self dunes may also form under more complex, multimodal wind regimes (see recent review by Cooke et al. [1993]). For this work we are chiefly concerned with the findings that longitudinal (self) dunes appear to require bimodal or multimodal wind regimes to form.

3. Martian Dunes

3.1. Overview

The major masses of dunes on Mars occur in a partial ring of ergs around the north polar deposit at 70-80°N, and in intracrater dune fields 60-80°S. Dunes also are abundant in low-latitude chasma, and are found scattered in other topographic traps, chiefly in craters and near wrinkle ridges. Their global distribution can be related to both wind patterns and to source regions [Thomas, 1982; Greeley et al., 1992; Ward and Doyle, 1985; Thomas and Gierasch, this issue].

The Viking imaging coverage is more than adequate to classify the major dune types and locations on Mars. While some regions are seen only at resolutions of 200 m/pixel, and with less than optimal viewing, the considerable coverage of dunes that exists at well under 100 m/pixel allows confidence in the finding that Martian dunes are overwhelmingly crescentic and transverse [Breed et al., 1979a,b; Toor et al., 1979; Greeley et al., 1992].

3.2. Previous Study of Mars Longitudinal Dunes

A few examples of longitudinal dunes on Mars have now been found, one set at 59°S, three others at 70-73°N. All are small local concentrations.

Toor et al. [1979] reported dune morphologies that reflect up to three local wind directions at several places in the north polar erg. Two examples of elongate barchans were documented (73°N, 57°W and 77°N, 97°W), where the inferred effective winds differ by about 40°. Other locations show variations in effective winds typically of 60°. All examples are interpreted by Toor et al. to show domination of the dune forms by one wind regime and minor modification by others. The dunes reported exhibit dominantly transverse characteristics and true longitudinal dunes were not reported in these areas.

Edgett and Blumberg [1994] have identified perhaps the clearest examples of "linear" dunes so far located on Mars at 59°S, 343°W. The dunes are about 4-5 km in length, and occur between two masses of transverse dunes inside a 70-km-diameter crater. The longitudinal dunes apparently are derived from the upwind transverse dunes and are reformed downwind into another mass of transverse dunes. These are similar to longitudinal dunes emanating from transverse dune masses on the Earth. The identification of only one set of longitudinal dunes in the many dune fields in Noachis suggests that Edgett and Blumberg that local effects such as the topographic funneling of winds through breaches in the crater rim and on the crater floor are responsible for their restricted occurrence. They further note that the forms might also represent the result of long-term changes in the winds that are moving the main dune mass from a position in the southern side of the crater, towards the north. Edgett and Blumberg [1994] have also found star dunes at 8.8°S, 270.9°W, in a valley formed by overlapping crater rims.

3.3. Longitudinal Dunes in the North Polar Erg

We have found two sites of longitudinal dunes in the north polar erg. The first is at 73°N, 57°W [Lee et al., 1993], a few tens of kilometers south of one of the occurrences of elongated barchans reported by Toor et al. [1979]. The dunes are best resolved in Viking Orbiter frames 525B11-14, obtained at $L_0 = 38'$ at pixel scales of 29 m (Figure 3). They are clustered within a mass of transverse and barchanoid dunes near the southernmost extent of the north polar erg. The longitudinal dunes occur over an area of the order of 10$^5$ km$^2$ and are grouped in sets of subparallel linear ridges. The longest dunes are about 3 km in length, while dune width is approximately 90 m. Spacing between the dunes, measured orthogonally from crest to crest, is typically 150 m, ranging from less than 100 m in the descent parts of the field to about 300 m in the field's more open portions. In several instances, the longitudinal dunes display what appear to be kinks and Y-shaped junctions. The general wiggly or convoluted outline of the dunes suggests that they are of the self type. Many of the longitudinal dunes originate from barchans or barchanoid ridges by the asymmetric elongation of individual horns (Figure 3b). Many have completely lost any transverse character, and their connection with the transverse dune forms is only inferred by working backwards through the continuum of forms.

The second site is near (71.4°N, 52.5°W) (Figure 4), in the immediate vicinity of a 4-km impact crater, approximately 120 km southeast of the field discussed above. The dunes are imaged at 26 m/pixel in Viking Orbiter frame 467B13, acquired at $L_0 = 21'$. The
Figure 3a. Barchan-derived longitudinal dunes. The dunes occur within a field comprising many dune forms (individual barchans and barchanoid ridges) at 73°N, 57°W. Viking Orbiter image 325B13, acquired at $L = 38^\circ$ (resolution = 28.8 m/pixel).

Figure 3b. Enlargement of the central area of Figure 3a, showing the elongation of barchan horns into seifs.
longitudinal dunes occur as sets of linear ridges that run roughly parallel to a NW-SE-trending dune shadow streak. The dunes are best seen downwind from the crater itself, immediately off the southwestern edge of the streak. The longitudinal dunes are associated with transverse bedforms and grade into the barchanoid ridges amassed near the downwind end of the streak. Some longitudinal dunes are at least 1 km long. Dune width is of the order of 90 m. Typical dune spacing is approximately 150 m. These dimensional characteristics are similar to those of the seifs at (73°N, 57°W).

3.4. Summary of Mars Longitudinal Dune Occurrences

Longitudinal dunes occur on Mars only as small isolated groups and largely in areas otherwise dominated by transverse dune forms. There is no occurrence of longitudinal dunes on Mars remotely similar to the regional fields of longitudinal dunes on Earth that can cover well over 105 km2. The Martian longitudinal dunes do occur in areas where multiple effective winds modify barchan and transverse dune forms, produce streamers of dark material at widely divergent angles from dunes, and cause braided dune patterns. True longitudinal dunes, though, are very rare, and are only poorly resolved.

4. Modality of Sand Transport on Mars

4.1. Circulation Models

The major Martian winds are driven by seasonally dependent heating and by mass flux to and from the polar caps. Diurnal slope winds can also be significant. The rapid radiative response of the atmosphere and the low thermal inertia of the surface help produce the most severe seasonal changes in any observed planetary atmosphere [Magalhães, 1987]. Seasonal changes on Mars are not only large, they present a strong asymmetry between north and south because southern summer occurs near perihelion, leading to much more vigorous circulation in southern summer than during northern summer. Flux of CO2 to the polar caps is also asymmetric, and the different mass transferred into and out of each polar region each year drives winds of different strengths in the two hemispheres.

Figure 4. Longitudinal dunes at 71.4°N, 52.5°W. Viking Orbiter image 487B13, acquired at L = 21° (resolution = 26 m/pxel). The main region of longitudinal dunes is bounded on three sides by transverse dunes, and on the fourth grades into the dune-free region of the streak extending from the 4-km impact crater. Note that the transverse dune orientations vary by about 30° in this image.
General circulation and other models of the winds on Mars have had considerable (but not total) success in predicting directions and times of winds as shown by wind streaks and dune orientations [Greeley et al., 1993; French and Gienasch, 1979; Megalhanes, 1987], and by cloud motions [Kohn, 1983]. However, they rarely predict surface shear stresses that exceed threshold for the most easily moved particles on Mars. As noted by Greeley et al. [1992], Edgett and Blumberg [1994], and others, this situation likely means either that sand particles on Mars are less dense than thought, or that local enhancement of winds provides enough of an increase to the general circulation to allow the threshold shear to be exceeded. Inasmuch as large dust storms on Mars involve particles that are probably more difficult to set in motion than are sand-sized particles, it is clear that threshold velocities are being exceeded in the present epoch.

General circulation models (GCMs) [Pollack et al., 1990; Haberle et al., 1993] predict very seasonally peaked distributions of strong winds on Mars. Maximum surface calculated stresses occur around L = 280° (southern summer, the season of most large dust storms). High surface stresses also are predicted at latitudes 90°-70°N and 5° during seasons of maximum polar sublimation or deposition (spring and fall).

We use the Ames GCM results to evaluate predictions of multimodal transport regimes on Mars. The model is described by Pollack et al. [1990], Haberle et al. [1993], and Greeley et al. [1993]. It calculates atmospheric motion accounting for solar input, optical depth, topography, albedo, thermal inertia, and CO2 condensation on the surface. Assumptions are made about atmospheric and surface properties (such as the ratio of boundary layer thickness to surface roughness). The data used here were calculated for twelve 10-day intervals distributed through the Martian year (L = 20, 44, 71, 99, 127, 156, 178, 209, 247, 281, 306, and 349°), recorded in 15-hour time steps. The output was supplied by D. Blumberg and includes surface stress components in 9° × 7.5° cells. Nearly all the stresses are below threshold for the most easily moved particles on Mars [Greeley et al., 1993], and thus it might be argued that the models have no applicability to sand transport. However, the models do reproduce the directions of significant fractions of bright and dark streaks [Greeley et al., 1993] and of some cloud motions [Haberle et al., 1993]. In addition, the high stress areas correlate well with areas of higher block abundance [Christensen, 1986]. This relation suggests that the models offer reasonable predictions of the locations of the highest wind stresses capable of eroding dust coverages. We adopt the assumption that extreme effects are hidden by cell size smoothing (cells are 440 km by 530 km at the equator) and time step size (1.5 hours), and that local enhancements of the general circulation cause threshold to be exceeded over a significant fraction of the planet at some time during the typical year.

We model the drift potential by scaling the stress results to shear velocity, as the stress is proportional to the shear velocity squared. We assume several different threshold values and thus in effect we scale all the ratio of shear velocities to threshold velocity calculated from the GCM results. This scaling allows for uncertainties in our knowledge of the threshold shear velocity, calculated surface stresses, and amount of any local enhancement of the general circulation. The uniform scaling of course assumes that our uncertainty in the relation of shear velocity to threshold shear velocity is the same for all areas of the planet. The technique is obviously crude but is quite adequate for tabulating the general predictions of the GCM for changes in wind regimes at different points during the Martian year. A choice of a high threshold results in very little calculated transport; a very low value results in more transport, and, as shown in section 2, should result in higher order modality in the transport regime, i.e., lower values of TV.

The output for one typical day’s run (16 time steps; the 10-day total for each L bin is very similar to the typical day) during each of the 12 periods was summed for DP and RDP at each cell location, using (2). The values of TV for nearly all of these daily transport regimes were over 0.5, indicative of strongly unimodal winds, though a few were less than 0.6. The net transport for each of these 12 representative days was then summed for the whole year with the assumption that each wind regime operated for the same time period. Because of the slightly uneven spacing of the runs available, some minor bias may have been introduced. Orientations of the net transport vector (RDP), its magnitude, and the values of TV were saved for each of the 1000 cell locations. We have varied the threshold velocity in four steps from 0.43 to 0.76 of the inferred nominal threshold shear velocity; a scaling of 1.0 would correspond to a nominal shear stress of 550 × 10⁻⁵ N m⁻².

Predicted net transport directions for the four scaled threshold values are shown in Figure 5. Figure 6 presents the distribution of TV values with latitude for one of the four assumed threshold values. The average value of TV for each latitude is also plotted. The latitude dependences of average TV for the four assumed threshold values are compared in Figure 6b. Several important characteristics are evident from Figures 5 and 6. (1) Strict application of the nominal 550 × 10⁻⁵ N m⁻² applied to the GCM results yields virtually no effective transport. Reducing the threshold by 44% from the maximum (scaling reduced from 0.76 to 0.43) results in a very great increase in the fraction of the planet predicted to be subject to effective sand transport. The generally widespread distribution of colline features on Mars, and their frequent alignment with present observed and calculated winds, suggests that a model in Figure 5a is more realistic than that in Figure 5d. Note that the orientation of the net transport changes in some areas as the threshold allows more components to be effective. (2) Even with the lowest assumed threshold, transport regimes are mostly unimodal. For the lowest threshold considered, over 95% of the area is calculated to have a TV of over 0.5, and 75% a TV greater than 0.8. Regardless of how one might change thresholds or slightly modify model inputs, the association of most of the vectors in Figure 5a with winds of a very specific season emphasizes that multimodal sand transport on Mars is likely to be very restricted geographically in the absence of local modification of the general circulation. (3) The two polar margin regions and low northern latitudes are predicted to...
Figure 5. Net annual transport directions calculated from output of the Ames GCM for four scalings of the nominal threshold velocity (see text): threshold scalings are: (a) 0.43, (b) 0.54, (c) 0.65, and (d) 0.76. (A scaling of 1 would correspond to a shear stress of $550 \times 10^4$ N m$^{-2}$.) Vectors shown only for cells with transport (winds above threshold). As a ratio of threshold to modeled winds increases, areas with transport of sand decrease.
be the most likely areas of multimodal transport. The affinity of the polar margin regions for multimodal regimes is also predicted by the model of French and Gierasch [1979], in which strong prograde winds associated with fall CO₂ mass transport to the pole and retrograde spring winds associated with mass loss from the caps imply multimodal transport.

Although the GCM predicts that sand transport under a multimodal wind regime is most likely to occur at high latitudes where indeed the few examples of longitudinal dunes are found, the net transport vectors predicted for the north polar locations above about 60°N are considerably different from what the dune forms, wind streaks, and even clouds suggest. The discrepancy and its significance for the north polar longitudinal dune interpretations are discussed below in section 5. Suffice it to say here that the GCM predicts mostly unimodal winds on Mars, with multimodal regimes likely possible in only restricted locations.

4.2. Wind Streak Data

The vast majority of wind streaks on Mars are interpreted to belong to one of four categories: (1) dust deposits in the lee of craters (bright streaks; Type Ib); (2) erosional scars in dust in the lee of topographic features (dark streaks; Type II); (3) plumes of dark material blown from dune fields (splotch streaks; Type II); and (4) accumulations of windblown frost (frost streaks). The interpretations are based on a variety of morphologic, time variability, and photometric data, as well as modeling of wind flow over obstacles and of atmospheric stability [Greeley et al., 1992].

The orientations of bright streaks (Type Ib) and dark Type II streaks that emanate from intracrater dunes...
overwhelmingly follow the Hadley cell flow of southern summer between latitudes ±40° [Thomas et al., 1984; Maksimishin, 1987]. The bright streaks cover large areas with consistent wind patterns and with virtually no hint of multiple lobe streaks emanating from any single crater. The most variable of these streaks occur on regional slopes where dunes are essentially absent. Two aspects of the Type II streaks, interpreted to be saltating material blown off intracrater dunes, also suggest very unimodal transport. First, they show a consistent global pattern of flow that is largely coincident with the southern summer Hadley cell flow. Local variations in directions are small except for some biographical topographic effects [see Thomas et al., 1984, Figure 6b]. Second, most of the streaks are long compared to their widths. The large length/width ratios of many of these dark streaks suggest that once the sand is blown off the source deposit, it continues to move in a very uniform direction. Effective multimodal winds would be expected to form other streaks from the source deposits or to smear out the streaks. The boundaries of many Type II streaks are smaller streaks parallel to the main streak [Thomas et al., 1984], and indicate only one locally effective wind regime. In some instances the source deposits are resolvable as crescentic dunes with orientations similar to that of the associated dark wind streak (which can extend for well over 100 km).

Type II streaks at high southern latitudes mostly reflect southern spring flow, which is retrograde off the seasonal polar cap, to the northwest [Thomas et al., 1979]. There are a few local streamers from dune fields between 40° and 50°S that show more than one effective wind, but these seem to be a minor component of the streak population and vary in orientation by less than 50°. At high northern latitudes, observational evidence for spring flow is minimal except very near the polar layered deposits [Thomas, 1981]. In both polar regions prograde flow outward from the pole during fall and winter is recorded by frost streaks [Thomas et al., 1979]. Winds that occur during times of frost cover might be expected to be ineffective on materials covered by frost. However, the strong variegation of the seasonal frost caps [James et al., 1979] suggests the presence of areas within the caps that are minimally covered or perhaps entirely swept free of frost. This might provide the opportunity for wind to move sand also during the fall and winter seasons.

Crater-related wind streaks (Types Ib and Id) do record multiple winds at several locations on Mars. In the Daedalia region (30°S, 120°W), there are multiple dark (Type Id) streaks from single craters spanning orientations that diverge by as much as 100° [Veverka et al., 1977; Thomaz and Veverka, 1979]. These winds, however, probably would have TV's of over 0.7 because there are additional winds within the 100° fan. They also are not in areas that have many sand dunes. Some areas of Hesperia (28°S, 245°W) show nearly 180° differences in the winds that form Type Ib bright streaks and those that produce Type Id dark streaks [Veverka et al., 1981]. The wind regime cannot be classified as multimodal, but because the flow regime responsible for the streaks may occur only in association with the crater topography, a single location might not experience multimodal sand transport. This area, too, has a dearth of visible sand dunes. In many areas where slope winds are expected, there may be a daily reversal of wind directions, which could support multimodal transport. However, the slopes are not areas that accumulate large dune fields.

4.3. Viking Lander Data

Viking Landers operated for several seasons at 22°N, 48°W, and 48°N, 226°W. Arvidson et al. [1983] reported that winds above threshold for undisturbed surfaces were rare or absent at the lander locations. Hodographs of lander wind data indicate considerable daily and seasonal variation, but still with dominating winds such that any transport is essentially unimodal. Drifts observed at the Viking Lander 1 site apparently have suffered substantial erosion, but roughly align with crater-related wind streaks seen from orbit. The VL-1 image also show small streaks extending from rocks that roughly coincide in direction with the streaks observed from orbit [Logan et al., 1977]. Azimuths of these features range from about 180° to 220° and the GCM net transport vector for this area is close to 200° for all assumed thresholds (Figure 4). Wind directions recorded by the VL-1 are strongest toward an azimuth of about 200°-220° [see Haberle et al., 1993]. The Viking Lander 2 site did not have the large range of wind markers seen at the VL-1 site. On the basis of lander meteorology data, GCM predictions for the higher latitude VL-2 site appear to be somewhat less accurate than for the VL-1 site [Haberle et al., 1993].

4.4. Summary of Modality of Expected Sand Transport on Mars

Overall, the GCM results and the wind streak data suggest that multimodal transport of sand should be uncommon on Mars. Viking Lander observations are also consistent with this conclusion. The dearth of longitudinal dunes on Mars can be attributed to this relative lack of bimodal or multimodal transport.

5. Wind Regime of North Polar Longitudinal Dunes

Can the occurrences of longitudinal dunes in the north polar erg be ascribed specifically to the transport regimes modeled in section 4, or is the situation more complex? Because most dunes in the region are transverse, the longitudinal dunes would seem to be local exceptions to the general regime.

5.1. Dunes, Streaks, and GCM Results

Dune forms shown in Figure 3a are plotted in Figure 7 to show the distribution of major dune types: transverse (barlochoid) ridges, barichan dunes, barichans with elongated horns, and longitudinal (seif) dunes. Also plotted are the orientations of some of the most unambiguous wind markers in the area: the aforementioned dunes, the dune shadow streak that extends over 20 km from the large, partially buried impact crater, and dark streaks emanating from dunes and from the crater rim. In general the region shows the effects of winds toward between 90° and 135° azimuth. Differences in orientations
of barchan forms, elongate horns, and nearby longitudinal dunes are nearly all less than 20°. Some elongate barchan horns suggest winds with azimuths of 140-145°. The distal segment of the dune shadow streak may show winds with azimuths of about 145°. Wind directions of about 170° azimuth are indicated by a few dark streaks from the degraded crater rim.

Figure 8 shows the orientations of north polar dunes and streaks to provide context for Figure 7. The overall patterns are discussed in Thomas and Gierasch [this issue]. The flow inferred from dunes and streak orientation in the region of the longitudinal dunes is somewhat exceptional for the north polar erg in that winds there do not tend to confine dunes to a latitude belt, but allow material to flow south (Figure 8). Topography (poorly known) may influence the flow in this region as it is a broad trough loading south [Thomas and Gierasch, this issue]. Figure 9 compares the north polar dune and streak orientations with the transport vectors predicted from the GCM for different times of the year. The GCM results for the region of 74°N, 54°W where the longitudinal dunes occur are quite different from the winds inferred from the orientation of dunes and streaks; net transport is predicted to be from the east, while dunes and streaks suggest northwesterly winds. Interestingly, however, the pattern shown by the GCM for slightly lower latitudes would match those inferred for the dune areas very well (Figure 9b). Haberle et al. [1993] have noted that the GCM predicts a much longer period of flow of easterly winds in the northern spring than the cloud data show [Kadok, 1983]. They attribute this difference to the GCM not being able to fully account for changing dust loading during the run period. The GCM results for high southern latitudes are more consistent with the domal wind markers there, and with the French and Gierasch [1979] model. Inversion of the French and Gierasch model for northern latitudes of 60°-70° would suggest a bimodal regime with winter winds toward the southeast and spring winds toward the southwest. Winds to the southwest, however, are rarely shown by surface features except above latitude 80°N [Thomas and Gierasch, this issue]. We must assume that the asymmetry of dust loading reduces the effectiveness of the spring winds in the far north (as suggested by Haberle et al. [1993]), and that the winter, off-pole winds are stronger at latitudes 60-75°N than the Ames GCM output predicts.
Do these results invalidate those of section 4.1? That section’s primary conclusion was that the GCM predicted little multimodal transport on Mars. If anything, the GCM has given a more multimodal result for the high northern latitudes than is suggested by eolian features. Thus the conclusion of the likely paucity of multimodal transport on Mars remains firm.

The stable transport mode (type and size of dune) varies greatly over scales of only a few kilometers in the area where the north polar longitudinal dunes occur; this is most dramatically shown by the dune field shadow southeast of the large degraded crater (Figure 7). The variability in winds and sand supply induced by the modest topography of this crater’s rim (probably only tens of meters high) were sufficient to eliminate dunes entirely from a zone over 20 km long, even though dune forms at its margins show that sand must have moved through the area that lacks dunes. The change in transport that formed the longitudinal dunes from the barchan or the transverse dunes was probably much subdued.

All indications are that the north polar longitudinal dunes formed from a slight modification of winds with at least two distinct effective directions, but which differed by less than 30°. Some additional modification may have occurred under other winds, but their action was only poorly preserved in the eolian forms. Inasmuch as the north polar longitudinal dunes are intimately associated with elongate barchans, they almost certainly would fall in the very upper part of TV values applicable for longitudinal dunes in Figure 2.

5.2. Relic Features?

In this study we have compared Martian dune forms to present-day winds. But could the dunes instead have been formed by winds of a different climate? The dunes may effectively be relic in several ways: they could have been physically stabilized; they could still be active but unable to change rapidly enough to reflect current winds; or they could be inactive simply because winds are below threshold.

On the Earth, dune immobilization by cementation (the only form of stabilization that might occur on Mars) usually involves dissolution and reprecipitation of salts [e.g., Verspagen, 1968]. Terrestrial dunes made of salt-rich (in particular carbonates) and clay-rich sediments exposed to moisture are those most likely to experience cementation. Once lithified, the dunes may preserve their form over geological timescales, especially if the bedforms are spare significant weathering.

Several lines of evidence point to the availability of abundant salts and clays on Mars, the basic ingredients of typical terrestrial dune cements. Geochemical inventories conducted at the Viking Lander sites suggest that sulfates and chlorides are likely present, possibly along with carbonates and nitrates [Baird et al., 1976; Toumain et al., 1977]. Salts are also invoked as possible bonding agents responsible for the cementation of the many clods and blocks observed at the Viking Lander sites [Clark et al., 1982], and also for the pelliculation of fine-grained silt and clay to allow these materials to be transported by winds as sand-sized particles. From theoretical arguments based on the range of possible thermodynamic conditions prevailing on Mars and on the physical chemistry of salts, Clark and Van Hart [1981] suggested that halite (NaCl), magnesium and sodium sulfates [(Mg,Na₂)₂SO₄], and magnesium and calcium carbonates [(Mg,Ca)₂CO₃] are likely abundant on Mars. From a synthesis of remote-sensing observations, Jakosky and Christensen [1986] inferred the existence of a global duricrust on Mars formed perhaps by the mobilization of salt ions in a layer.
of water adsorbed within the regolith. These authors pointed out that the close similarity in the respective concentrations of S and Cl at the two well-separated Viking Lander sites is consistent with the existence of an indurated saline crust of global extent. By analogy with weathering processes operating in salt-rich regions on the Earth, salts were also considered as viable and effective weathering agents on Mars [Malin, 1974]. Salt-rich aqueous solutions have also been invoked in the preterrestrial alteration of several SNC meteorites (unusual chondrites thought to have originated from Mars [McSween, 1985, 1994] [Gooding et al., 1991; Treiman et al., 1993; Westworth and Gooding, 1994]). Numerous intracraterr and some extracraterr albedo features on Mars have also been interpreted in terms of salt or clay-rich playa deposits, some of which were apparently processed by winds into lunatees and panna [e.g., Lee, 1993; Williams and Zimbleman, 1994; Greeley and Williams, 1994]. The features provide additional circumstantial evidence in support of an abundant availability of salts and clays on Mars.

With salt- and clay-rich sediments available, the cementation of Martian dunes could occur if a solvent is available. Liquid water is not currently stable at the Martian surface but water may be adsorbed in a monolayer around near-surface soil particles [Fanale and Cannon, 1974]. In Antarctica, a monolayer of unfrozen, adsorbed water is capable of transporting dissolved ions through soils even at temperatures below freezing [Eglington and Anderson, 1973]. From these authors’ ion diffusion rates, Jakosky and Christensen [1986] indicated that salts could migrate several meters in the Martian upper surface layer over a timescale of the order of 10^5 years, a process which could be facilitated by the diffusion of water into and out of the regolith on a similar timescale.

Near-surface crusts could thus form on Mars over a period of 10^5 years, provided, of course, sediments are not disturbed over shorter periods. Wind regimes on Mars, however, experience major astronomically driven changes over timescales of 10^6 years. The season of perihelion controls which season experiences the strongest winds and undergoes a cycle of period 51,000 years. The dunes and streaks in the vicinity of the north polar longitudinal dunes appear to match present fall and winter flow, near southern summer and perihelion, the epoch of highest wind stresses. At the other extreme in the cycle, winds in this region would be expected to be dominated by northern spring and summer winds, toward the southwest. Because the alignment of the largest wind feature in the area, the dune shadow streak, with the present-day wind system is unlikely to be fortuitous, we believe most smaller bedforms near the streak probably also relate to winds that are recent and that have been effective over timescales very short compared to the 51,000-year cycle. Moreover, because dune lithification would appear to require quiescent, diffusive conditions over periods of 10^5 years to be effective, the observed bedforms are likely poorly stabilized.

Thus, although there is the potential for stabilization of dunes on Mars, the indications of current widespread activity and coincidence of many bedforms with current winds argues that it probably has a minor influence at most. The question of what bedforms might lag behind changing winds has been dealt with by Thomas [1981, 1982]. While some of the larger dune masses might lag, the overall patterns suggest minimal record of winds from the opposite phase of the 51,000-year cycle of perihelion season.
6. Conclusion

The scarcity of longitudinal dunes on Mars is entirely consistent with the unimodal transport regimes expected and observed for most of the planet. In fact, the scarcity of longitudinal forms on Mars can be taken as reinforcing the generalization of their formation under bimodal or multimodal transport regimes. Spacecraft missions such as Mars Global Surveyor will help amplify our knowledge of present and past eolian action on Mars by providing high-resolution views of dune morphology and global seasonal coverage of evidence of changes that show the orientation and frequency of activity of effective winds.

Acknowledgments. This research was funded in part by NASA grants NAGW-5374 and NAGW-5375. Discussions with A. Bloom, D. Blumberg, K. Edgett, P. Gierasch, R. Greeley, and J. Veverka were helpful. GCM run results were kindly supplied by D. Blumberg on short notice. R. Sullivan and W. Ward provided very helpful and detailed reviews. We also thank Mary Roth for help in preparing the manuscript.

References

Baird, A.K., P. Toulmin, kindy, Greeley, and consistent with the present and past eolian action on Mars by providing high-resolution views of dune morphology and global seasonal coverage of evidence of changes that show the orientation and frequency of activity of effective winds.

References

Baird, A.K., P. Toulmin, kindy, Greeley, and consistent with the present and past eolian action on Mars by providing high-resolution views of dune morphology and global seasonal coverage of evidence of changes that show the orientation and frequency of activity of effective winds.

References

Baird, A.K., P. Toulmin, kindy, Greeley, and consistent with the present and past eolian action on Mars by providing high-resolution views of dune morphology and global seasonal coverage of evidence of changes that show the orientation and frequency of activity of effective winds.


McSween, H.Y., Jr., What we have learned about Mars from SNC meteorites, Meteoritics, 29, 757-779, 1994.


P. Lee and P.C. Thomas, Center for Radiophysics and Space Research, Cornell University, Space Science Building, Ithaca, NY 14853-6801. (e-mail: lso@instrou.m.ti.cornell.edu)

(Received March 9, 1994; revised November 28, 1994; accepted January 13, 1995.)
Page intentionally left blank
Polar margin dunes and winds on Mars

Peter C. Thomas and Peter J. Gierasch
Center for Radiophysics and Space Research, Cornell University, Ithaca, New York

Abstract. The approximately concentric arrangement of layered deposits and dune fields at the two Martian poles may reflect a nearly steady state dispersal of material from the polar deposits. Data on effective surface winds from high resolution Viking Images combined with theory of local winds suggest that the northern dunes are in part confined to a latitude band by winds generated by their own low albedo. Dispersal of the dark sand from the southern polar region is not subject to this kind of feedback because the irregular topography prevents areal accumulations sufficiently extensive to produce winds.

Introduction

Both polar regions of Mars have crudely concentric arrangements of frosts and large sediment deposits: perennial frost caps, layered deposits, dune fields, and a variety of "mantling" deposits. However, the frost and sediments show some significant differences between the poles, among them the different perennial ice (water in the north; CO$_2$ in the south), and the contrast of a widespread (though incomplete) annulus of dunes about the northern layered deposits with scattered dune fields in isolated depressions in the south. Because the layered deposits at both poles appear to have suffered considerable erosional reduction from their maximum volumes, and the dunes appear to have their sources within the layered deposits, the arrangement of the dunes may reveal some of the mechanisms for the climate and geological influences on the evolution of the polar deposits [see Murray et al., 1972; Spyres, 1979; Howard et al., 1982; Thomas, 1982; Plaut et al., 1988; Thomas and Weltz, 1989]. The fundamental geological problem at issue here is the cause of the narrow latitudinal banding of dunes near the north polar deposit, and the much wider zone of intracrater dunes near the south polar cap.

In this study we first briefly review previous work on winds and the polar dunes. Next we present the mapping results for the north and south polar regions and note the geologic settings that may influence winds affecting the dune materials such as albedo contrasts and topography. Then terrestrial sea-breeze theory is reviewed and applied to the dark Martian dune fields that may provide strong surface temperature gradients because of their low albedo.

Finally, our conclusion that sea-breeze effects may be important in confining the north polar dunes is discussed as part of the overall evolution of the polar cap sedimentary systems. For this work we have presented only a few schematic drawings of the eolian features in Figure 1; complete sets of images and background on their interpretation for wind directions and eolian processes can be found in the works of Breed et al. [1979]; Tsoar et al. [1979]; Greeley et al. [1992], and many others.

Previous Studies of Polar Dune Winds

Tsoar et al. [1979] found that most of the north polar dunes had morphologies that show the effects of multiple wind directions. They suggested cyclonic winds due to temperature contrasts of cold, ice-covered areas and ice-free areas would be significant in controlling dune activity. Thomas [1982] mapped latitudinal variations of flow on-pole and off-pole that might help confine the dunes in both polar regions. Ward and Doyle [1983] suggested that changing east-west wind components at about 80°N contributed to dune occurrence and orientations. Howard [1980] mapped a very consistent pattern of summer flow within the north polar deposits (mostly above 80°N) retrograde out from the pole (from the northeast). Tsoar et al. [1979] additionally noted that winds in the polar regions were not simply latitudinally dependent and might fall into regional wind regimes.

An axisymmetric model of winds at the Martian poles by French and Gierasch [1979] indicated that the highest winds in the polar regions formed spiral patterns outward from the pole, prograde in the fall during condensation of the annual frost cap (to the southeast in the northern region), and retrograde (to the southwest) in spring. These calculations showed good agreement with south polar wind indicators mapped by Thomas et al. [1979]. The general circulation model (GCM) calculations reported by Greeley et al. [1993] show strong fall and winter winds from the west at 80°N, and moderate winds from the east in spring. The GCM does not appear to give significant north-south wind components. Parish and Howard [1993] used three-dimensional modeling to find that katabatic winds off the polar deposits could be very effective, but their intensity and directions could be substantially altered in summer by the increased insolation on slopes.

The models account for many aspects of the wind flow, but have not quantitatively addressed confinement wind regimes. We address this problem with a sea breeze wind model in conjunction with a digital map of features.
THOMAS AND GIERASCH: POLAR MARGIN DUNES AND WINDS ON MARS

A B C D E F G

Figure 1. Examples of eolian features mapped with effective wind orientations. (a) Crater related wind streaks. These may be dark or light and are formed by enhanced erosion or deposition of dust downwinds of craters or other topographic obstacles. (b) Splotch streaks; also called type II streaks. Dark material deflated from deposits (splotches) in craters. The orientation of the splotches is also a wind indicator. (c) Crescentic dunes. (d) Echo dunes. These form upwind of topographic obstacles. (e) Framing dunes. These large, transverse form are on the upwind margins of dune fields. (f) Dune field shadows. These are areas lacking dunes in the lee of obstacles caused by diversion of sand flow and increased complexity of wind patterns. (g) Intracrater dune fields. These usually form on the downwind side of topographic depressions. Some are discrete dunes, but many are continuous, thick mounds, with transverse bedforms on top.

together with published data on the dune fields and other eolian features. We study both polar regions since the differences in geological setting may indicate the relative effects of climate and geology on the disposition of surface sediments in the two polar regions.

Methods

For this study we desired a digital data set of wind direction markers that spans different sizes of features (hence possibly different formation timescales), and that had been collected in a uniform manner. Part of the strategy was to make as complete a search as possible, but to deal with only the unambiguous markers. We have examined the higher resolution Viking images of polar regions (latitudes greater than 60°) for eolian features that indicate surface wind directions such as dunes (crescentic, longitudinal, transverse dunes), wind streaks in dust, sand, and frost, and wind shadow patterns in dune fields that indicate effective wind directions. Of particular importance are the sizes of some of the features that may crudely indicate the length of time the effective winds would have had to operate. We have not mapped every dune form that has an interpretable effective wind, but only the representative examples from each image, to give a clear representation of the geographic coverage of types and wind directions rather than a complete catalog of all features. Ambiguous features were avoided. Table 1 lists the features mapped and the typical sizes and some crude estimates of possible relative effective timescales based on parameters from White [1979] and limits discussed in Thomas [1982]. Because the Martian climate is expected to show hemispherical reversal of some seasonal effects on the timescale of changing season of perihelion (51,000 years full cycle; Arvidson et al. [1979]; Kieffer and Zent [1992], the larger eolian forms might be the result of a range of climate conditions and effective wind regimes. The data have been entered into a digital file and in conjunction with global wind streak data reported earlier [Thomas, 1981; Thomas et al., 1984] allow a much more complete picture of the winds that have done geologic work near the Martian polar regions.

Results: North Polar Region

Figure 2 shows cylindrical projection maps of several of the eolian features in the north polar region: wind streaks, crescentic dunes and associated streaks, and larger dune features such as framing dunes, echo dunes, and dune field shadows (see Figure 1). As with the data of Howard [1980], Toor et al. [1979], the streaks in latitudes 80°-85° show consistent retrograde winds off the cap. The pattern becomes more complex below 80°N, and as shown in many other studies [e.g., Thomas et al., 1984], the pattern in mid-northern latitudes becomes geostrophic winds merging into the global, southern summer flow. The crescentic dunes show a complex pattern of flow to the southwest at higher latitudes and flow to the northeast at some longitudes in the regions below 80°N. The only region of apparent through-flowing material is in the longitude range 20°-100°E. This material, in fact, feeds into the large low albedo region of Acidalia [see also Thomas et al., 1984; Christensen, 1988]. Many of the crescentic dunes in this region, where the dune material is freely moving from the pole, are elongate barchans and even linear dunes (P. Lee et al., Longitudinal dunes on Mars, submitted to Journal of Geophysical Research, 1994) and indicate moderately different wind directions for long-term gentle winds, and shorter term, stronger flow. Differences in wind directions inferred from the asymmetric and elongate barchans can be nearly 90°, but are usually 45° or less. The echo dunes, framing dunes, and dune field shadows represent eolian forms that require much larger timescales to form than do individual crescentic dunes (see Table 1 and Figure 1). Reorientation of a crescentic dune would take 1 to 2 orders of magnitude less mass transport than the rearrangement of a field of dunes such as shown in Figure 1c, or the orientation of the large framing dunes or dunes formed in echo patterns around large topographic obstacles. These forms also show the same patterns of flow that include significant on-pole winds, as do the erosional outliers of polar layered deposits shown in Figure 2d.

The observed flow patterns suggest that the dunes are confined by regional winds, that polarward flow is substantial and has persisted over at least moderate geologic...
Table 1. Eolian Features and Possible Timescales

<table>
<thead>
<tr>
<th>Feature</th>
<th>Applicable Scale</th>
<th>&quot;Measurement&quot; Method</th>
<th>Timescale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bright and dark &quot;dust&quot; streaks</td>
<td>micrometers to centimeters</td>
<td>periodic images</td>
<td>days</td>
</tr>
<tr>
<td>Dark streaks from dunes</td>
<td>millimeters to centimeters</td>
<td>periodic images</td>
<td>seasonal</td>
</tr>
<tr>
<td>Dune field bedforms; individual dunes</td>
<td>100-200 m</td>
<td>saltation models</td>
<td>10^4-10^5 years</td>
</tr>
<tr>
<td>Echo, framing dunes; dune field shadows</td>
<td>1-30 km</td>
<td>saltation models</td>
<td>10^4-10^5 years</td>
</tr>
<tr>
<td>Intracrater dune fields</td>
<td>10-500 m</td>
<td>saltation models</td>
<td>10^4-10^5 years</td>
</tr>
<tr>
<td>Latitudinally distributed dunes</td>
<td>100-1000 km</td>
<td>saltation models</td>
<td>10^4-10^5 years</td>
</tr>
<tr>
<td>Polar layers</td>
<td>tens of meters</td>
<td>dust and ice deposition models</td>
<td>10^4-10^5 years</td>
</tr>
<tr>
<td>Polar layered deposits (PLD)</td>
<td>1-5 km</td>
<td>dust and ice deposition models</td>
<td>10^7 + years</td>
</tr>
<tr>
<td>Erosion of troughs in PLD</td>
<td>100 m to 1 km</td>
<td>guess (crater densities)</td>
<td>10^3 years</td>
</tr>
<tr>
<td>PLD extent</td>
<td>hundreds of kilometers</td>
<td>guess (crater densities)</td>
<td>10^8 years</td>
</tr>
</tbody>
</table>

Features are arranged approximately in order of increasing volume; see Figure 1 for examples. Applicable scale is the dimension of material that has to be moved to form the feature. For some features, such as dust streaks, individual dunes, and layers this is the thickness of material that must be deposited or removed. For features that require lateral rearrangement to form, the scale is a horizontal motion distance. Measurement method is the means by which the timescale for change has been estimated. Timescale is the time that the features may require under present Martian conditions to reform; see Thomas (1982) and Thomas et al. (1992) for background. These numbers are meant as only order of magnitude, relative values except for the first two rows.

Results: South Polar Region

Wind markers in the south are very different from those in the north. Dunes are largely confined locally in craters or by wrinkle ridges on plains. Although the surface area covered by dunes is smaller in the south than in the north, the volumes of sand may be just as large as in the north because of deep fill in many of the southern craters (Thomas, 1982; see also Lancaster and Greeley, 1990). Material blown from the crater deposits (splotches) forms many streaks (see Figure 1b) in the south, but only a few in the north. Likewise, the concentration of dunes in mounds inside craters in the south, and the lower image resolution in this region result in very few interpretable individual bedforms in the southern data compared to those in the north.

Figure 3 shows the maps of southern wind markers; note that the latitude range is greater than in Figure 2. The dark, crater-related streaks, likely formed by erosion of thin dust layers and observed to be the most ephemeral of these markers, show a complex pattern with significant on-pole flow (Figure 3a). This on-pole flow is distinct from that in the north in three major ways: it is retrograde, it is shown by the shorter-lived features, and the on-pole wind markers are at lower latitudes than those in the north.

The streaks (Figure 3b) are probably the next most rapidly modified features and show consistent retrograde flow from the polar regions. The streak orientations within craters, likely representative of a more substantial deposit and longer timescale of modification, also show consistent flow off-pole (Figure 3c).

The orientation of bedforms on dune masses show more variation than do the streaks from splotches (Figure 3d). However, those that could show a southwest or northeast flow require different winds. Such different effective winds could have occurred at different seasons, or different epochs, or could result from local characteristics of flow within the topographic trap (usually craters) of the dune mass.
Figure 2. Wind markers in the north polar region of Mars. Data digitized from Viking Orbiter Images. 

(a) Dust streaks: crater-related streaks. (b) Crescentic dunes and associated streaks. (c) Framing dunes, echo dunes, and dune field shadows. (d) Tapered outliers of polar deposits. (e) Frost streaks in annual CO₂ cap. (f) Wind marker azimuth as a function of latitude. Dashed lines denote boundary of flow from the pole (90°-270°) and flow toward the pole.

The dune field azimuths recorded from the high-resolution images provide strong support for the use of splotch orientations as representing winds over time periods longer than the splotch streaks. Some of the dune fields that are well positioned against crater walls show that the effective winds have had to operate over times adequate to move very large quantities of sand.

For complete comparison of north and south wind markers, frost streaks formed in the annual cap are shown in Figures 2e and 3f. Both show prograde, off-pole flow.
Figure 3. Wind markers in the south polar region. (a) Dark, crater-related streaks. (b) Splotch streaks: dark material blown from crater deposits. (c) Crater splotches: these are interpreted as occurring on downwind side of craters. (d) Bedforms on dune fields in craters. These are transverse ridges on a continuous, often deep mantle of material (sand?). Azimuths shown here are interpreted as winds with components to the west; all have 180° ambiguity. (e) Azimuths of dune fields; these are interpreted from morphology of dune field within crater and are similar to splotches in Figures 3b and 3c. (f) Frost streaks in the southern seasonal frost cap.
and were the basis for much of the French and Gierasch model of polar winds.

Geological Settings of Polar Dune Deposits

Before our testing of sea breeze influences on the dune field locations, we discuss some aspects of the geological settings. As noted by several works [Breed et al., 1979; Howard et al., 1982; Thomas, 1982] there is strong morphologic evidence for dune sources in the layered deposits, and color measurements support this inference [Thomas and Weitz, 1989; Herkenhoff and Murray, 1990].

The reexamination of the images for this study has reinforced these data by the absence of indications that any of the dunes in the major chaumas and reentrants are moving on-pole. The digital mapping did not include many of the common but subtle streaks emanating from scarps in the polar layered deposits (PLD's) that add to the evidence of recent dune material erosion from the polar layers. The restriction of the on-pole flow to dune areas at and beyond the margins of the northern PLD makes it very difficult to invoke present conditions for surface motion of the dune material to latitudes well above 80°N.

Although the north polar sand seas are often classed as being a ring about the pole, Lancaster and Greeley [1990] have emphasized that in terms of volumes of material, the sand seas are fragmented, and are dominated by the material between longitudes 110° and 250°; half the north polar sand is probably in the longitudes 130° to 230°. These dunes are in fact not simply south of the polar deposits; they are substantially contained between the polar deposits and icy outliers.

The complexity of the north polar dunes is well illustrated by maps of Toor et al. [1979] and Lancaster and Greeley [1990]. The plots in Figure 4 include approximate outlines of the major dune volume concentrations from Lancaster and Greeley [1990] and show that the regions of significant dune cover occur on latitudinal scales of about 4°-6°, or about 7°-3° (120-180 km) from margins of the dune field to the central regions of maximum dune cover and the region of apparent reversal of wind flow. Dune fields in the south polar region are all less than 70 km in extent; most are less than 30 km across. The south polar dunes are also at lower latitudes and are not confined by ice outliers, but by local depressions such as craters.

Regional topography is poorly known in the Mars polar regions; the digital data provided by the U.S. Geological Survey [1993] indicates a possible wide depression leading from about 75°N and longitudes 20°-90°W toward the south; elevations for most of the dune areas are about 0.5 to 3 km below the reference datum. In the south elevations average 2.5-5 km above the reference at 75°S. In

![Figure 4. Schematic diagram of major deposits at north and south poles. Outlines of polar deposits after Tanaka and Scott [1987]; Thomas et al. [1992]. (a) North pole: layered deposits, areas of major dune volumes modified from Lancaster and Greeley [1990], wind directions indicated by crescentic dunes, and ice outliers. (b) South pole: layered deposits and associated pitted deposits, dunes, and intracrater dune field orientations.](image-url)
THOMAS AND GIERASCH: POLAR MARGIN DUNES AND WINDS ON MARS

both poles the layered deposits may reach 1-2 km above the surrounding plains.

Temperature observations from the Viking Orbiter infrared thermal mapper (IRTM) have been reported for the polar areas by Keegan et al. [1991], Paige and Keegan [1991], and Paige et al. [1994]. Data of Paige et al. [1994] and data reported by Toner et al. [1979] for the region 70°-90° and longitude 130°-230° suggest temperature contrasts of as much as 20 K with the ice to the south over distances of a few tens of kilometers. Average maximum summer daytime temperature contrasts are probably about 5-10 K; ice temperatures were approximately 220 K.

The thermal inertia of dunes on Mars is compatible with particle sizes expected to saltate on Mars [Edgett and Christensen, 1991, 1994]. The north pole dunes have inertias that may be lower than most other Mars dunes, but are still consistent with saltating particles [Keegan et al., 1991]. The inferred approximate particle sizes means that these materials should be some of the most easily moved materials on Mars [White, 1979; Greeley et al., 1992].

Sea Breeze Theory and Extension to Mars

Haberle et al. [1979] carried out a zonally symmetric general circulation computation of Martian polar cap winds and found that strong winds of 20 to 30 m s⁻¹ developed in relatively narrow zones near the edge of the seasonal CO₂ cap. Here we shall argue that terrestrial sea breezes expected to be found near albedo contrasts in the north polar region.

The sea breeze on Earth is well understood. Dalu and Pielke [1989] and data reported by Dalu and Pielke [1979] for albedo contrasts of 0.3 and 0.75 km, we obtain approximately A = 10⁴ s⁻¹ for the Earth, which is of the same order as the Coriolis parameter f. A value of this same order is probably also appropriate for Mars. For example, a drag coefficient can be estimated from bulk drag considerations by writing

\[ D \Delta u = -C_{Dp} \rho u \]

where \( D \) is the boundary layer depth and \( C_{Dp} \) is a dimensionless drag coefficient, which should be the same on Mars or the Earth if the surface roughness characteristics are not very different. This expression leads to an estimate for the boundary layer damping rate of \( \lambda = C_{Dp} U D \), where \( U \) is a typical wind speed. Wind speeds are higher on Mars than on Earth [Zurek et al., 1992]. With \( C_{Dp} = 0.003 \) [e.g., Stull, 1988], \( U = 20 \) m s⁻¹, and a midlayer depth \( D = 0.75 \) km, we obtain approximately \( \lambda = 10^4 \) s⁻¹. In addition, the radiative relaxation rate on Mars for a layer of thickness 1.5 km is about \( 10^4 \) s⁻¹.

The north polar dunes have depth D is the depth of the heated layer, f is the Coriolis parameter, \( \lambda \) is the frictional or thermal damping rate (these are assumed to be the same), and N is the Brunt frequency. For Mars parameters we shall show that the same physics is valid, and we shall show that the scale \( D \) is about 100 km, in good agreement with the width of the zones of dune activity near albedo contrasts in the north polar region.

The notation is standard. The drive for the circulation is a forcing temperature perturbation \( T(x,z,t) \). At the surface it is equal to the departure of the surface temperature from the mean reference value \( T_0 \). We are primarily interested here in the steady case with \( \partial_t = 0 \), with the flow driven by a mean temperature difference produced by surface albedo variation. Diffusion, radiation, and frictional drag are all formulated in terms of a constant linear damping coefficient \( \lambda \). Parameter \( \Phi \) is the pressure perturbation divided by the mean density. Note that a wind parallel to the isotherms, \( v \), is produced only through the effect of the Coriolis parameter \( f \).

The depth of the circulation is determined by the vertical distribution of \( T(x,z,t) \). This involves boundary layer heating mechanisms and depends on the density of the atmosphere because of its influence on the thermal inertia. On the Earth the effective depth for boundary layer heating effects is typically a few hundred meters, and is verified by direct observations [see Stull, 1988]. On Mars there are not yet any observations of the thermal boundary layer thickness, but theoretical treatments indicate that it is a bit larger than 1 km [Huffman et al., 1973; Pollack et al., 1979]. Thus an important difference between Mars and the Earth is that the boundary layer on Mars is about 5 times deeper.

The appropriate damping rate \( \lambda \) is found to be about \( 10^4 \) s⁻¹ for the Earth, which is of the same order as the Coriolis parameter \( f \). A value of this same order is probably also appropriate for Mars. For example, a drag coefficient can be estimated from bulk drag considerations by writing

\[ D \Delta u = -C_{Dp} \rho u \]

which should be the same on Mars or the Earth if the surface roughness characteristics are not very different. This expression leads to an estimate for the boundary layer damping rate of \( \lambda = C_{Dp} U D \), where \( U \) is a typical wind speed. Wind speeds are higher on Mars than on Earth [Zurek et al., 1992]. With \( C_{Dp} = 0.003 \) [e.g., Stull, 1988], \( U = 20 \) m s⁻¹, and a midlayer depth \( D = 0.75 \) km, we obtain approximately \( \lambda = 10^4 \) s⁻¹. In addition, the radiative relaxation rate on Mars for a layer of thickness 1.5 km is about \( 10^4 \) s⁻¹.

The simplest way to parameterize the sea breeze model is to reproduce the governing equations from Dalu and Pielke [1989]. Let \( x \) be the horizontal coordinate parallel to the temperature gradient that forces the circulation, \( y \) be along the "coast" or albedo boundary, and \( z \) be upward, with \( x \), \( y \), \( z \) forming a right-handed system. Let the corresponding velocities be \( u \), \( v \), \( w \). Because the circulation is much wider than its depth (about 100 km compared to about 1 km) the vertical force balance is hydrostatic. The model is

\[ \begin{align*}
\partial_t u + \frac{f}{\rho} \partial_x \Phi &= -\lambda_4 u, \\
\partial_t v + \frac{f}{\rho} \partial_y \Phi &= -\lambda v, \\
\partial_t w &= \frac{\partial}{\partial x} (\Lambda^x w) + \frac{\partial}{\partial y} (\Lambda^y w) + \frac{\partial}{\partial z} (\Lambda^z w) \quad (2) \\
\partial_t T &= T_x u + T_y v + \frac{\partial}{\partial x} (\Lambda^x T) + \frac{\partial}{\partial y} (\Lambda^y T) + \frac{\partial}{\partial z} (\Lambda^z T)
\end{align*} \]

The north polar dunes have depth D is the depth of the heated layer, f is the Coriolis parameter, \( \lambda \) is the frictional or thermal damping rate (these are assumed to be the same), and N is the Brunt frequency. For Mars parameters we shall show that the same physics is valid, and we shall show that the scale \( D \) is about 100 km, in good agreement with the width of the zones of dune activity near albedo contrasts in the north polar region.
(which is part of the reason that this thickness is predicted) [Goody and Belton, 1967]. Thus both momentum and heat transfer considerations are consistent with \( \lambda = 10^4 \) s\(^{-1}\). The uncertainty in this parameter is probably a factor of 3 or so.

Because the Coriolis parameter at high latitudes on Mars is approximately \( f = 10^{-4} \) s\(^{-1}\), the same size as \( \lambda \), there will be appreciable turning of the sea breeze wind by Coriolis effects.

Rottman [1983] and Dalu and Pielke [1989] argue that the sea breeze on Earth is well described by the linear model, and it remains to verify that this will also be true on Mars. Relative to the frictional drag, nonlinear advection terms are of the order of \( u(\partial u/\partial x) \), where \( u \) is the sea breeze wind itself. With \( u = 10 \) m s\(^{-1}\), \( \lambda = 10^4 \) s\(^{-1}\), and \( L = 100 \) km, we find \( u(\partial u/\partial x) = 1 \), which indicates that nonlinear terms should have important effects. The same conclusion is reached in the terrestrial case, where both \( u \) and \( L \) are about a factor of 3 smaller. Nevertheless, the linear theory is a good approximation on Earth, and we shall assume that it also is on Mars.

In summary, the mechanisms that produce the sea breeze on Earth should carry over to Mars and produce local wind systems where there are sharp gradients of the surface temperature (scales smaller than 100 km). The most significant difference between Mars and Earth is that the depth of these flows is greater on Mars and, as a result, the width is also greater. Equation (1), from the results of Dalu and Pielke [1989], is used to estimate the width of the circulation. The Brunt frequency is given by

\[
N^2 = g \frac{\partial T_s}{\partial z} \left( \frac{\Delta T_x}{T_b} \right)
\]

With the acceleration of gravity \( g = 380 \) cm s\(^{-2}\), specific heat \( c_p = 0.7 \times 10^9 \) erg g\(^{-1}\) K\(^{-1}\), and assuming an isothermal mean temperature, we find \( N = 10^4 \) s\(^{-1}\). Using (1) with \( D = 1.5 \) km and \( \lambda = f = 10^4 \) s\(^{-1}\) gives \( L = 100 \) km.

The wind speed, according to Dalu and Pielke, is given by

\[
u = \frac{g \Delta T_x}{N} = \frac{380}{N} \text{m s}^{-1} \frac{\Delta T_x}{T_b}
\]

where \( \Delta T_x \) is the horizontal temperature contrast at the surface. This can be estimated from

\[
\frac{\Delta T_x}{T_b} = (1-A) S
\]

where \( T_b \) is the surface temperature and \( S \) is the flux of insolation. Letting \( \Delta A \) be the horizontal albedo contrast and linearizing (5) give

\[
\frac{\Delta T_x}{T_b} = \frac{\Delta A}{4(1-A)}.
\]

Maps of north polar region albedos [Paige et al., 1994] show dune area albedos of about 0.15 and adjacent areas with albedos of about 0.3 and higher on ice outliers. Using these albedos in (6) yields fractional temperature differences of about 5%, which implies wind speeds from (4) of about 20 m s\(^{-1}\) (Figure 5). These temperature differences are consistent with the measurements reported by Keegan et al. [1991] and Paige et al. [1994].

These winds probably do not have sufficient strength to initiate salivation and transport, but because the sea breeze mechanism is linear to a first approximation, the winds are additive and will augment other circulations that might be present. These include the tides, the mean circulation of the polar vortex, baroclinic instabilities, slope winds and local vortices [Zurek et al., 1992]. The local wind systems produced by albedo-induced temperature contrasts probably act in conjunction with other systems to produce the episodes of strong wind that dominate sediment transport. The topographic contrast of the polar deposits and surrounding plains may enhance the confinement by off-pole flow [Parish and Howard, 1993]. The mean direction of the sediment transport will be strongly biased toward that of the albedo-induced flow because maximum winds stresses will occur when all components are aligned. The difficulty of matching predicted wind stress with motion of materials is a familiar problem for Mars [Getaney et al., 1992, 1995; Edgett and Blumberg, 1994]; while many of the patterns of wind flow and eolian features match aspects of circulation models, the magnitudes of the calculated surface stresses are often somewhat less than those necessary to initiate motion. There is no doubt that motion occurs, and is occurring, so modeled maximum wind stresses should provide indications of the locations and orientations of the maximum wind transport.

**Summary and Discussion**

The principal conclusions are as follows:

1. The largest accumulations of dunes in the north polar region are confined by on- and off-pole winds.
2. The albedo contrast of the dunes, especially with icy areas, can generate significant winds into the areas of dunes, thereby providing confining wind components.
3. Although south polar dunes face more topographic obstacles to dispersal from the PLD, the rougher topography prevents areal coverage by low albedo materials sufficient to generate sea breeze type winds that might help confine the dune materials.
Self-regulated dispersal of sand from the poles is only a small part of the evolution of the polar deposits, as the sand is volumetrically small compared to the dust ice. Lancaster and Greeley, 1990; Herkenhoff and Murray, 1990; [Thomas and Weitz, 1989]. However, the observations of the dunes emphasize that the layered deposits are a small part of the evolution of the polar deposits, as the times for Mars, deosits are unlikely to be applicable to their formation, though they could theoretically apply to conditions of some of the unconformities or particular events recorded in the deposits [see Howard et al., 1982].

There are many other differences between north and south that we have not dealt with here, such as the differences in absolute elevations, polar ices, extent of residual caps, and entomorphology in deposits associated with the layers such as the pitted terrain in the south.

A better understanding of the mechanisms of the polar winds and their relation to overall deposit evolution could come with proposed Mars Global Surveyor orbital instruments. Many of the problems with modeling wind arise from lack of knowledge of topography and details of thermal inertia of both of which could be vastly improved with suggested instruments. Interpretations of geologically effective winds on all size scales and timescales would profit from more higher-resolution images.

Acknowledgments. This work was supported in part by NASA grants NASG-3274 and NASA-1997. Technical assistance was provided by B. Caetlin, R. Fox, B. Kite, C. Moos, M. Roth, and B. Boettcher. Helpful comments were provided by T. Paris and an anonymous reviewer.

References


Squireys, S. W., The evolution of dust deposits in the Martian north polar region, Icarus, 40, 244-261, 1979.


Tanaka, K. L., and D. H. Scott, Geologic map of the polar
Geology and landscape evolution of the Hellas region of Mars

Kenneth L. Tanaka and Gregory J. Leonard
U.S. Geological Survey, Flagstaff, Arizona

Abstract. Hellas basin on Mars has been the site of volcanism, tectonism, and modification by fluvial, mass-wasting, and eolian processes over its more than 4-b.y. existence. Our detailed geologic mapping and related studies have resulted in the following new interpretations. The asymmetric distribution of highland massifs and other structures that define the uplifted basin rim suggest a formation of the basin by the impact of a low-angle bolide having a trajectory heading S60°E. During the Late Noachian, the basin was infilled, perhaps by lava flows, that were sufficiently thick (>1 km) to produce wrinkle ridges on the fill material and extensional faulting along the west rim of the basin. At about the same time, deposits buried northern Malea Planum, which are interpreted to be pyroclastic flows from Amphitrites and Peneus Paterae on the basis of their degraded morphology, topography, and the application of a previous model for pyroclastic volcanism on Mars. Peneus forms a distinctive caldera structure that indicates eruption of massive volumes of magma, whereas Amphitrites is a less distinct circular feature surrounded by a broad, low, dissected shield that suggests generally smaller volume eruptions. During the Early Hesperian, a ~1- to 2-km-thick sequence of primarily fine-grained, eolian material was deposited on the floor of Hellas basin. Subsequently, the deposit was deeply eroded, except where armored by crater ejecta, and it retreated as much as 200-300 km along its western margin, leaving behind pedestal craters and knobby outliers of the deposit. Local debris flows within the deposit attest to concentrations of groundwater, perhaps in part brought in by outflow streams along the east rim of the basin. These flows may have deposited ~100-200 m of sediment, subduing wrinkle ridges in the eastern part of the basin floor. During the Late Hesperian and Amazonian, eolian mantles were emplaced on the basin rim and floor and surrounding highlands. Their subsequent erosion resulted in pitted and etched plains and crater fill, irregular mesas, and pedestal craters. Local evidence occurs for the possible former presence of ground ice or ice sheets ~100 km across; however, we disagree with a hypothesis that suggests that the entire south rim and much of the floor of Hellas have been glaciated. Orientations of dune fields and yardangs in lower parts of Hellas basin follow directions of the strongest winds predicted by a recently published general circulation model (GCM). Transient frost and dust splotches in the region are, by contrast, related to the GCM prediction for the season in which the images they appear in were taken.

1. Introduction

The Hellas region of Mars consists of diverse topography, geology, and surface characteristics, yet it is dominated by one particular feature, the Hellas impact basin. Although the basin was formed during early bombardment, it is relatively well preserved and makes up the deepest and broadest (>9 km relief, ~2000 km across) depression on Mars (Figures 1 and 2). Volcanism and channel dissection have modified large expanses of the northeastern and southern parts of the basin rim. Both of these modified areas and the floor of the basin are marked by wrinkle ridges. Also, a large deposit of proposed eolian origin covers the basin floor; this area is presently the site of origin of major dust storms [Martin and Zurek, 1993] and of the highest wind stresses on Mars predicted by general circulation models (GCMs) [Greeley et al., 1993].

In spite of the great storehouse of geologic information contained in the Hellas region, comprehensive mapping of the region has been undertaken only recently at 1:5,000,000 scale (G.J. Leonard and K.L. Tanaka, work in progress, 1995). This study focuses on the interpretation of the geologic and eolian history of the Hellas region. We address several questions and issues of regional to global geologic and climatic significance, including (1) What accounts for the asymmetric spatial distribution of basin massifs? (2) What was the nature of volcanism on the south rim of Hellas? (3) How were large and small channels generated on the basin rim? (4) What is the origin of basin-fill materials, and are they the source for present-day dust storms? (5) What was the overall history of deposition and erosion in the region, and is there a relation to planetwide resurfacing history? (6) How do the observed spatial distribution and orientations of interpreted, present-day eolian streaks and landforms (such as dust and frost splotches, dunes, and yardangs) compare with predicted surface wind stresses and directions based on the GCM of Greeley et al. [1993]? We address these questions by way of geomorphologic interpretation, stratigraphic relations, topographic measurements, crater counts, and application of...
Figure 1. Geographic nomenclature of the Hellas region of Mars. Numbers index image figures in this paper; HP, Hadriaca Patera; TP, Tyrrhena Patera; PP, Peneus Patera; AP, Amphitrites Patera. Photomosaic base derived from Mars Digital Image Mosaic.
Figure 2. Topographic map of the Hellas region of Mars; contour interval 1000 m [U.S. Geological Survey, 1991].
blocks of relatively resistant crustal material. Concentric graben (outside the study area) associated with the impact region resulted in densely Noachian. A southeast-trending low-angle impact probably bombardment formed impact craters throughout the study area. Produced circular caldera created Amphitrites, a less distinct circular feature surrounded by a broad, low, dissected shield. Runoff and (or) groundwater flows was emplaced south and west of Amphitrites and Peneus Planitia, and their sediments likely debouched into... The stratigraphy of the Hellas region is defined by geologic and geomorphologic history of the Hellas region. Heavy bombardment of the early crust in the Hellas region resulted in densely cratered terrain during the Early Noachian. A southeast-trending low-angle impact probably formed the Hellas basin cavity and surrounding uplifted rim material. Much of the rim comprises disrupted, large massif blocks of relatively resistant crustal material. Concentric graben (outside the study area) associated with the impact formed soon thereafter. In the Middle Noachian, continued bombardment formed impact craters throughout the study area. Ejecta, fluvial sediment from dissection of basin mantles, and possibly coeval mantles and volcanic deposits began infilling topographic lows, including the Hellas basin, highland areas outside the basin rim, and intermassif and intercrater areas of the basin rim. Such highland modification by local eolian, fluvial, and volcanic activity continued into the Late Noachian. Hellas basin became infilled, probably by lava flows that were apparently sufficiently thick to cause lithospheric sagging leading to wrinkle-ridge formation, normal faulting at Hellenous Mones, and lava effusion along the edge of the basin interior. Pyroclastic volcanism formed low-relief highland volcanoes on the outer parts of the basin and are commonly oriented along co-seasonal trends of seasonal frost and dust splotches form throughout the basin region and are commonly oriented along co-seasonal trends of transient, seasonal frost and dust splotches form throughout the basin region and are commonly oriented along co-seasonal trends of seasonal frost and dust splotches form throughout. The generalized results of our mapping are shown in Figure 3 and from crater counts (Tables 1–4). Because of the apparent significant obliteration of smaller craters over much of geologic time (as discussed later), emplacement ages are assigned more confidently from the densities of larger craters (generally >5 km in diameter). The stratigraphy of the Hellas region is defined by geologic mapping; we draw from our preliminary draft of a geologic map at 1:5,000,000 scale based on Viking images (G.J. Leonard and K.L. Tanaka, work in progress, 1995). Previous regional geologic mapping of note includes the Mars quadrangle series at 1:5,000,000 scale, based on Mariner 9 images [e.g., Potter, 1976; De Hon, 1977; Peterson, 1977], and the global Viking series at 1:15,000,000 scale [ Greeley and Crowell, 1987; Tanaka and Potter, 1990]. Also, detailed geologic maps at 1:500,000 scale from high-resolution images have been produced for the Hadriaca and Tyrrhena Paterae areas [Greeley and Crow, 1990; Crow et al., 1992]. Our mapping differs from the earlier work in that we used the Viking versus the Mariner data set, and from the later work in that we examine Hellas from a regional perspective versus global and local context. Stratigraphic assignments are derived from observed overlap and embayment contact relations recorded by geologic mapping (Figure 3) and from crater counts (Tables 1–4). Because of the apparent significant obliteration of smaller craters over much of geologic time (as discussed later), emplacement ages are assigned more confidently from the densities of larger craters (generally >5 km in diameter). The generalized results of our mapping are shown in Figure 3, which shows only units interpreted to be of eolian origin. These map units are briefly discussed in this section; detailed discussion of the formational and modification histories of some of the units is presented in following sections.
Figure 3. Generalized geologic map of the Hellas region of Mars (from G.J. Leonard and K.L. Tanaka, work in progress, 1995). Line indicates postulated low-angle impact trajectory for the Hellas basin-forming bolide. Note symmetry about trajectory of (1) distribution of basin rim unit and (2) planitia and planitia flank materials.
Table 1. Cumulative Crater Densities and Inferred Surface Ages of Geologic Units in the Hellas Region of Mars

<table>
<thead>
<tr>
<th>Geologic Unit* or Feature</th>
<th>Crater Density$^1$</th>
<th>Area, km²</th>
<th>Surface Age$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N(2))</td>
<td></td>
<td>(N(2))</td>
</tr>
<tr>
<td>Smooth material and interior mantles</td>
<td>275±26</td>
<td>54±12</td>
<td>407,368</td>
</tr>
<tr>
<td>Subdued ridged unit</td>
<td>362±23</td>
<td>163±23</td>
<td>301,139</td>
</tr>
<tr>
<td>Valles material$^3$</td>
<td>239±72</td>
<td>109±49</td>
<td>45,947</td>
</tr>
<tr>
<td>Flank flows of Tyrrhena Patera$^4$</td>
<td>39±170</td>
<td>87±33</td>
<td>80,576</td>
</tr>
<tr>
<td>Dissected rim unit, Malea Planum</td>
<td>347±26</td>
<td>133±16</td>
<td>512,346</td>
</tr>
<tr>
<td>Dissected rim unit, east rim$^4$</td>
<td>400±69</td>
<td>132±27</td>
<td>188,976</td>
</tr>
<tr>
<td>Interior deposit</td>
<td>345±36</td>
<td>136±23</td>
<td>266,698</td>
</tr>
<tr>
<td>Degraded interior deposit</td>
<td>250±22</td>
<td>83±13</td>
<td>519,018</td>
</tr>
<tr>
<td>Ridged plains material, Malea Planum</td>
<td>384±30</td>
<td>159±19</td>
<td>440,090</td>
</tr>
<tr>
<td>H scrutica Patera$^5$</td>
<td>572±75</td>
<td>175±41</td>
<td>103,089</td>
</tr>
<tr>
<td>H scrutica Patera$^5$</td>
<td>816±100</td>
<td>223±52</td>
<td>80,895</td>
</tr>
<tr>
<td>Tyrrhena Patera$^5$</td>
<td>666±161</td>
<td>313±111</td>
<td>25,544</td>
</tr>
<tr>
<td>Basin ridged unit</td>
<td>406±57</td>
<td>186±23</td>
<td>295,244</td>
</tr>
</tbody>
</table>

$^*Most units shown in Figures 3 and 4; some from Crown et al. [1992] and G.J. Leonard and K.L. Tanaka (work in progress, 1995).
$^1Crater density N(D) equals number of craters larger than diameter D (in kilometers) per million square kilometers.
$^2Relative ages based on Tanaka [1986] and reflect both material and modification ages (see text); N, Noachian; H, Hesperian; A, Amazonian; E, Early; M, Middle; L, Late; for material units, equivalent stratigraphic positions of epochs are shown, where "Lower" position corresponds to "Early" age and "Upper" to "Late."
$^3Data from Crown et al. [1992].
$^4Data derived from Table 1 and Figure 7 of Plescia and Saunders [1979].

The oldest (Early Noachian) material in the region consists of high-standing crust of the Hellas rim, known as the basin rim unit. This unit is marked by a high density of craters tens of kilometers in diameter as well as numerous rugged massifs and mountain ranges. Embayed by the basin rim unit and surrounding Hellas basin are Middle to Upper Noachian highland material. Locally, this unit is dissected by small valleys. Within Hellas basin is an annulus of Upper Noachian basin ridged unit. On the northeast rim of Hellas, H scrutica and Tyrrhena Paterae are made up mostly of Upper Noachian patera and patera flank materials. Ridged plains material was emplaced south of Hellas on Malea Planum beginning at the end of the Noachian and on Hesperia Planum at the beginning of the Hesperian. Volcanism in the Hesperian produced shield and dissected deposits on the south rim of Hellas; these deposits form patera material of the Amphitrites Formation [Tanaka and Scott, 1987]. Within the Hellas basin, a voluminous eolian interior deposit was emplaced during the Early Hesperian; where deeply eroded, this unit is mapped as degraded interior deposit (Figure 4). Upper Hesperian and Lower Amazonian rim and interior mantles and smooth material overlie and embay, in places, older material on the Hellas floor and rim (Figure 4). In the northeastern and southwestern parts of the basin and basin rim, Late Hesperian to Early Amazonian channel dissection formed the dissected rim unit, which includes small valleys as well as a few large, deep channels that emanate from large depressions. Locally, this unit may have been produced by volcanism and ground-ice interaction, because late-stage lava flows from Tyrrhena Patera are coeval with and adjacent to source areas of some of the channels. Emplacement of the subdued ridged unit (Figure 3), a plains material that covers and

Table 2. Binned Crater-Density Boundaries for Martian Epochs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Late Amazonian</td>
<td>&lt;20</td>
<td>&lt;10</td>
<td>&lt;5</td>
<td>5-19</td>
<td>2.5-9</td>
<td>5-13</td>
<td>6-12</td>
</tr>
<tr>
<td>Middle Amazonian</td>
<td>20-75</td>
<td>10-38</td>
<td>5-19</td>
<td>2.5-9</td>
<td>5-13</td>
<td>6-12</td>
<td>12-19</td>
</tr>
<tr>
<td>Early Amazonian</td>
<td>75-200</td>
<td>38-100</td>
<td>10-50</td>
<td>5-13</td>
<td>6-12</td>
<td>12-19</td>
<td>12-19</td>
</tr>
<tr>
<td>Late Hesperian</td>
<td>200-375</td>
<td>100-188</td>
<td>50-94</td>
<td>25-47</td>
<td>13-23</td>
<td>12-19</td>
<td>12-19</td>
</tr>
<tr>
<td>Early Hesperian</td>
<td>375-600</td>
<td>185-300</td>
<td>94-150</td>
<td>45-75</td>
<td>23-38</td>
<td>12-19</td>
<td>12-19</td>
</tr>
<tr>
<td>Noachian</td>
<td>&gt;600</td>
<td>&gt;300</td>
<td>&gt;150</td>
<td>&gt;75</td>
<td>&gt;38</td>
<td>&gt;19</td>
<td>&gt;19</td>
</tr>
</tbody>
</table>

*See Table 1; data determined for specified diameter bins.
TANAKA AND LEONARD: GEOLOGY AND LANDSCAPE OF HELLES REGION, MARS 5413

Table 3. Binned Crater Densities for Helles Basin Map Units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Crater Density*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth material and</td>
<td></td>
</tr>
<tr>
<td>interior mantles</td>
<td>E4</td>
</tr>
<tr>
<td>Subdued ridged unit</td>
<td>E4/E4</td>
</tr>
<tr>
<td>Interior deposit</td>
<td>500±10</td>
</tr>
<tr>
<td>Dissected rim unit,</td>
<td></td>
</tr>
<tr>
<td>Malea Planum</td>
<td>64±11</td>
</tr>
<tr>
<td>Ridged plains material,</td>
<td>100±15</td>
</tr>
<tr>
<td>Malea Planum</td>
<td>74±16</td>
</tr>
</tbody>
</table>

*See Table 1; data determined for specified diameter bins.

embays ridges of the eastern outcrops of the Helles basin ridged unit, is also possibly related to this dissection. Patches of dunes and transient splotches (Figure 4) appear to be relatively recent, and probably active, deposits within the central and northwestern parts of Helles basin and on various parts of the rim.

3. The Helles Impact Structure

The topographic expression and distribution of structures around the Helles region compose the signature of the impact event that produced the basin. These and other related features provide clues for understanding the geometry of the impact, which in turn provides a basis for better analysis of the styles and degrees of geologic and degradational processes that have occurred in the Helles region. The topographic rim of Helles is marked in part by many rugged massifs of the basin rim unit that were uplifted as a result of the Helles impact. Shadow and photoclinometric measurements of isolated massifs yield heights between 1 and 5 km (Figure 5). Also, differential uplifts of the basin rim are bounded by escarpments, particularly along the southeast basin rim, where an extensive region of uplift occurs. Here a lobe of the basin rim unit extends for about 1500 km (Greeley and Guest, 1987). By contrast, the northeast rim has only a narrow, discontinuous band of rim material, which averages about 300 km in width. One explanation for the broad, southeastern region of uplift is that it had been the site of an earlier impact basin and thus was more susceptible to tectonic deformation (Wichman and Schultz, 1989). Recently, Leonard and Tanaka (1993) proposed that this region resulted from a southeast-trending (S60°E), low-angle impact of the basin-forming bolide (see Figure 3).

Laboratory studies and observations of impact craters on planetary bodies indicate that oblique impacting projectiles produce features that are symmetric about (normal to) their trajectory path and asymmetric along their range path (Gault and Wedekind, 1978; Moore, 1969; Schultz, 1994). The Helles impact structure displays both types of features. Features symmetric across the proposed trajectory path include (1) broad concentric troughs at distances 1800 to 2500 km north and west of the basin (outside of study area), which formed at about the time of the impact; these features have been modeled as resulting from flow of the asthenosphere in Table 4. Binned Crater-Density Ages for Helles Basin Map Units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Crater Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth material and</td>
<td></td>
</tr>
<tr>
<td>interior mantles</td>
<td>EA</td>
</tr>
<tr>
<td>Subdued ridged unit</td>
<td>EA/MA</td>
</tr>
<tr>
<td>Interior deposit</td>
<td>EAM/A</td>
</tr>
<tr>
<td>Dissected rim unit,</td>
<td>MA</td>
</tr>
<tr>
<td>Malea Planum</td>
<td>MA</td>
</tr>
<tr>
<td>Ridged plains material,</td>
<td>EA</td>
</tr>
<tr>
<td>Malea Planum</td>
<td>EAM/A</td>
</tr>
</tbody>
</table>

Abbreviations defined in Table 1.
Figure 4. Map of scoured deposits and features in the Hellas region of Mars. The approximate boundary shown for the rim mantle north of Perseus Fasna is uncertain due to the lack of high-resolution images necessary to define the mantle's extent, which is likely to be much more than what is shown. Other units discussed in text. Lines show contacts of general geologic units in Figure 3.
Figure 5. Shadow-height and photoplotometric measurements (in meters) in the Hellas region. Triangles indicate heights of basin-rim massifs; circles show depths of channels; square shows scarp height of basin interior deposits. Some massif heights represent averages of closely spaced massifs that had values that varied by less than 400 m. Measurement precisions vary from <100 m to a few hundred meters depending on pixel size, lighting geometry, and definition of shadow boundary.
response to the transient impact cavity [Wickman and Schultz, 1989]; (2) prominent scarps that bound the southeastern region of uplift; areas in front (west) of the scarps appear to be structurally lower (Figure 6); (3) large areas between the narrow and broad parts of the rim (see Figure 3) that show a marked absence of massifs (Figure 5); and (4) large volcanic massifs surrounded by vast volcanic plains that occur just outside the basin rim, also between the narrow and broad parts of the rim (Figure 3). Asymmetric features along the range path include (1) higher (generally by ~1–2 km) and more densely concentrated massifs in the large southeast rim lobe (inferred downrange direction) than in the narrow northwest rim margin (see Figure 5); (2) deeper excavation and steeper walls on the uprange profile (Figure 2); and (3) an upsurge zone of subarcuate fractures and rim collapse, particularly the extensional faulting that formed the scarps associated with Hellespontus Montes along the west rim of Hellas.

Although the patterns of ejecta and crater morphology produced by oblique impacts have been simulated by laboratory experiments of projectiles impacting fine-grained materials, deformation of the lithosphere by oblique impacts at the scale of multiring basins has not been simulated by such experiments or predicted by theory. However, on the Moon, where basin ejecta is commonly better preserved than on Mars [Edgett, 1989], the asymmetry of ejecta deposits for a dozen Late-stage effusive activity (Late Hesperian) at Tyrrhena Patera produced a field of lava flows that cover the plains from Tyrrhena to near the east flank of Hadriaca.

4. Volcanism

Two and possibly three significant volcanic areas occur in the Hellas region as (1) paterae and associated plains on the northeast rim of Hellas basin; (2) paterae and associated plains on the south Hellas rim; and (3) an early floor of Hellas basin [Pettit, 1976]. The northeast rim volcanics and channels have been described and discussed extensively [Greeley and Crown, 1990; Crown et al., 1992; Crown and Greeley, 1993]; we will only briefly summarize those results here. The southern volcanics, which are also extensively channelled, have not been analyzed previously, and the possibility that the basin floor is filled by lava is examined here.

4.1 Northeast Rim Volcanics

This volcanic area includes Hesperia Planum and the broad, low shields of Hadriaca and Tyrrhena Paterae. Crater counts indicate that Tyrrhena Patera formed during the Late Noachian, and Hadriaca during the Late Noachian and Early Hesperian (Table 1). Hesperia Planum (ridged plains, by definition), are Lower Hesperian [Scott and Carr, 1978; Tanaka, 1986] and are thought to be composed of broad, thin, low-viscosity lava flows [Greeley and Spudis, 1981]. The morphologies and flank slopes and lengths of flows from the paterae indicate that they consist mainly of early pyroclastic flows followed by late-stage lava flows and modest caldera collapse. This volcanic evolution may be consistent with early volatile-rich magmas followed by volatile-poor magmatism; alternatively, the pyroclastic deposits could have resulted from hydromagmatic eruptions through the near-surface interaction of magma and ground water in a permeable crust [Greeley and Spudis, 1981; Greeley and Crow, 1990; Crown and Greeley, 1993]. Late-stage effusive activity (Late Hesperian/Early Amazonian) at Tyrrhena Patera produced a field of lava flows that cover the plains from Tyrrhena to near the east flank of Hadriaca.

4.2 South Rim Volcanics

Peneus and Amphitrites Paterae are at the crest of the south rim of Hellas and are surrounded by the ridged plains of Malea Planum (Figure 1). North of the patera, the ridged plains are dissected by narrow channels of Axius Valles. The entire volcanic area covers, in the study area alone, about 106 km2; more ridged plains extend south of the study area [see Tanaka and Scott, 1987]. The paterae reach an elevation of about 4 km, and the surrounding plateau slopes away down to 1 km southeast of Amphitrites Patera and to ~2 km at the edge of Hellas Planitia (Figure 2).

Peneus Patera (Figures 7a, 7b) is 100 x 120 km across and elongate in a northeastward direction. The patera is made up of a 10- to 30-km-wide zone of concentric normal faults and grabens that encircle a central floor, which is probably hundreds of meters below the surrounding plateau. The southeastern part of the floor is deformed by wrinkle ridges. The high elevation, ring faults, and central depression of the patera within ridged plains material suggest that Peneus is a major volcanic caldera formed by multiple, voluminous eruptions of magma and subsequent collapse of evacuated magma chambers [Tanaka and Scott, 1987].

Amphitrites Patera (Figure 7a), centered 250 km east of the center of Peneus, has a significantly different morphology from its neighbor. Unlike Peneus, Amphitrites has only a few
Figure 7a. Oblique view of Peneus (PP) and Amphitrites (AP) Paterae. Peneus Patera, 100-120 km across, is slightly elongate along a northeasterly trend and bounded by a zone of concentric normal faults and grabens. The floor is marked by large wrinkle ridges; one (arrow) may account for prominent bench (B) on southeast side of patera. Amphitrites Patera, 120 km across, forms a ridge and scarp structure surrounded by a 300-km-diameter low shield dissected by narrow channels; it is bordered to the south by crater Barnard. Note possible ring graben (G) and collapse pit (P). A few wrinkle ridges (W) radiate from Amphitrites Patera, thus diverting channels. South and west of Peneus Patera, local sets of similarly trending and spaced wrinkle ridges crosscut one another (mosaic of Viking images 94A74-76; centered at -57.00°S, 305.00°E, 240-250 m/pixel).

ring faults; within the summit region, these faults locally bound a shallow depression. Furthermore, the patera has traditionally been defined by a 120-km-diameter circular ridge and scarp (which is partly buried by the rim of crater Barnard) [e.g., Tanaka and Scott, 1987]; however, our mapping has established the outline of a low-relief shield nearly 300 km across that surrounds the circular structure. Channels radiate from the circular structure and terminate at the shield’s base. Locally, the channels are highly deformed, degraded, or masked due to wrinkle ridges, mantling deposits, and ejecta from crater Barnard. On the west flank of the shield, a 20-km-long pit may be a collapsed vent structure (Figure 7a). The broad form, low relief, and heavily dissected and degraded morphology of the Amphitrites shield suggest that it is mainly made up of pyroclastic rocks. In contrast to Peneus, eruptions of Amphitrites were probably of much lower volume, leading to shield buildup, local dissection, and little collapse.

The basin-rim slopes of Malea Planum, south of the paterae, are extensively dissected by Axius Valles, a collection of narrow, sinuous channels having very little tributary development (Figure 8). Discontinuous lobate tongues, possibly degraded lava, lahars, or other volcaniclastic deposits, are also common. Because of the highly degraded appearance of the slopes, Tanaka and Scott [1987] interpreted the material there to be pyroclastic rocks, similar to those of Hadriaca and Tyrrhena Paterae. Average slopes from Peneus and Amphitrites Paterae across the dissected ridged plains range from about 0.2-0.9° over distances of 400-500 km (the slope values are subject to a high degree of error). These
(2) the deposit occurs at the lowest elevations on Mars, where magma buoyancy considerations favor effusive volcanism [e.g., Wilhelms and Baldwin, 1989]; (3) a few flows appear to emanate from an impact crater along the basin rim (at lat 31° S, long 200.5°), indicating that local effusive eruptions have occurred within the basin; and (4) the unit nowhere appears highly eroded, which would be expected if the unit were made up of a more friable material (as discussed later, deep eolian etching has occurred in the overlying interior deposit). Thickness of the deposit is uncertain, but was probably sufficient to produce the wrinkle ridges.

5. Formation of Wrinkle Ridges

The most common structures at map scale in the Hellas region are wrinkle ridges. Wrinkle ridges are generally interpreted to be compressional folds and (or) thrust-fault structures [e.g., Golombek et al., 1991; Watters, 1993]. As...
Figure 9a. Northwestern Hellas Planitia, showing basin-ringed unit embaying basin-rim unit (along top of image). Interior deposit superposed on basin-ringed unit has largely been removed; traces include crater pedestals (P), some of which are knobby, and knobby outliers (K) below distinctive scarp of interior deposit (I). Connected ridges in lower half of image may be networks of compound crescentic dunes (see terrestrial example in Figure 9b). (Viking image 620447; centered at 33.82°S, 303.18°E, 247 m/pixel).

Figure 10. Southern Hellas Planitia, showing rugged, hummocky (H; deeply eroded) areas of the interior deposit and basin-ringed material (B); smooth material (S) buries the contact between these large units. A scarp (large white arrows) outlines the edge of the dissected rim unit (D). Although wrinkle ridges do not appear to deform the interior deposit, some scarps in the deposit may be controlled by underlying ridges. Small, pristine channels (small white arrows) are deflected around wrinkle ridges and cut dissected rim unit and interior deposit but are buried by smooth material. Mantling on basin-ringed material has partly buried crater rims (black arrows) and resulted in pedestal craters (P). (Viking image 363504 and 06; centered at -51.50°S, 295.50°E, -245 m/pixel).

Figure 9b. Top center of image shows network of compound crescentic (transverse) dunes in the Takla Makan Sand Sea, China. Dune sizes and pattern similar to those of ridges in Hellas Planitia (Figure 9a). Dominant wind direction is transverse to longer ridge axes. (Part of Landsat image 1129-04311-7).
The Hellas Planitia wrinkle ridges have a wide array of orientations; some preference to radial orientation is apparent. Compressional loading stresses would have developed within the basin due to its infilling by lavas (of which the thickness and mass are unknown) [Solomon and Head, 1979]. In addition, near-global compressional stress apparently occurred during the Late Noachian [Tanaka, et al., 1991]; as a result, large horizontal compressional stress would have developed in the basin ridged unit. Although global compressive stress would tend to preclude ring-graben development, some normal faulting at Hellas Planitia Moosiees on the west rim of Hellas occurred during the Late Noachian (based on the crater age of Wichman and Schultz [1989]).

These ring grabens are consistent with a basin-loading stress model produced by at least 1.2 km of basalt [Wichman and Schultz, 1989]. Finally, the stress model predicts magma ascent outside the basin load [Solomon and Head, 1979], which is consistent with the possible vent along the north edge of the basin ridged unit mentioned previously. This model is favored because it predicts basin-rim normal faulting, basin compressional stresses, and magma ascent along the basin rim, all of which are observed in and around Hellas and may be coeval.

Complex patterns of wrinkle ridges are also well recognized on Hesperia and Malea Plana [Chicarro, et al., 1985; Watters, 1993]. Many of these wrinkle ridges have circular patterns interpreted as resulting from subsidence over crater structures [Watters, 1993]; similarly, Amphitrites Patera is outlined by a circular wrinkle-ridge structure that may be controlled by a buried, circular collapse structure (Figure 7a). In addition, a few systems of wrinkle ridges radiate north-south and east-west from Amphitrites and Peneus Paterae (Figure 7a), and others appear concentric to Amphitrites (Figure 8); these relations suggest some influence of the paterae on the local stress field.


This section discusses observations used to infer past eolian and fluvial activity in the Hellas region of Mars. In some cases, other processes and factors may also account for the observations. (For example, many features in the Hellas region have been interpreted as glacial or periglacial in origin [Kargel and Strom, 1992]; this hypothesis is discussed in section 6.8.) Relative ages (Tables 1 and 4) are deduced from crater densities shown in Tables 1 and 3. As mentioned earlier, because of the apparent significant obliteration of smaller craters over much of geologic time, emplacement ages are more confidently assigned from the densities of larger craters (generally >5 km in diameter). Also, because craters superposed on a geologic unit are commonly obliterated up to a particular size range, we have derived a table of density values for crater-size bins (using a V2 geometric progression) for each Martian series (based on the scheme of Tanaka [1986]) for craters larger than 2 km in diameter (Table 2). (Recall that "series" are time-stratigraphic units and "epochs" are corresponding periods.) The results of this method for our crater counts of Hellas units are shown in Tables 3 and 4.

6.1. Noachian/Hesperian Highland and Patera Modification

The earliest modifications in the Hellas region affected the rugged basin rim unit. The unit is intensely cratered, particularly on the southeast side of the basin (Figure 6), which has one of the highest densities of craters >32 km in diameter on the planet (e.g., Tanaka, 1986). That area, compared to other highland areas, has a relatively low abundance of craters in the few tens of kilometers size range when compared to the abundance of larger craters [Watters, 1977]. This result may be due to post-Hellas impact obliteration and to the rugged topography, which would favor rapid degradation of smaller craters on the flanks of steep massifs and which would provide deep intermassif basins in which smaller craters could be buried by later deposition. Local channeling and eolian mantling (some of which is post-Noachian) contributed to the degradation of the massifs and to the deposition of intermassif plains material throughout the basin rim unit.

Highland material that mantles and embays the basin rim unit also shows significant erosion. In contrast to the craters on the basin rim unit, which mostly have preserved rims, most craters <20 km in diameter in the highland material appear "rimless" (Figures 12-14). The term "rimless" refers to craters that have little to no outer raised rim but have a distinctive inner rim scarp and crater floor that is below the level of the adjacent surface [Groen and Schultz, 1993]. In addition, some
craters of all sizes on highland material have well-preserved rims and ejecta. These observations appear independent of latitude (and thus the effects of ice-creep [see Squyres and Carr, 1986]), but they may indicate that craters on highland material were highly eroded during the Noachian or Early Hesperian [Tanaka, 1986; Wilhelms and Baldwin, 1989; Craddock and Maxwell, 1990, 1993] and that craters on the basin rim unit were better preserved. The latter deduction suggests that highland materials are poorly indurated and fine grained, whereas the basin rim unit is likely made up of more resistant rock. In western Terra Tyrrhena, valleys are rare but mantles are present, which indicates that eolian mantling and deflation may have muted crater rims (see Figures 4 and 13; also compare with rimless craters in northeastern Arabia Terra [Grant and Schultz, 1990; Moore, 1990]).

Also during the Noachian or Early Hesperian, much of eastern Terra Tyrrhena and a few intercrater basins in Promethei Terra were dissected by arrays of small valleys (Figure 14). The valleys debouch into small intracrater and intercrater basins that appear to be partly filled by deposits. Although rimless craters are common in the dissected material, the valleys appear to postdate crater-rim degradation, because the valleys cut the material that mutes the craters. Where valleys do cross crater rims, they either cut perpendicularly across the rim (thus removing only a small section of rim), or, as commonly happens in northeastern Arabia Terra, they cut a small section of rim (thus removing only a small section of rim).
in rare cases, they cut along the rim to below the level of the floor, which results in a raised crater floor. Most rimless and degraded craters show extensive gully development on the inner rim, suggesting that fluvial erosion contributed to rim degradation [see Grant and Schultz, 1994].

Valley erosion is also prevalent on the flanks of the four paterae in the study region. On Hadriaca and Tyrrhena Paterae, deeply etched and exposed layers demonstrate erosion by wind, water, and (or) mass wasting of pyroclastic deposits [Greely and Crown, 1990; Crown et al., 1992]. The valleys on Tyrrhena are distinctly flat-floored, whereas those on Hadriaca are V-shaped. This difference may be controlled by the presence of an upper layer(s) on Tyrrhena that has a relatively weak base along which scarp retreat progresses. Amphitrites Patera (Figure 7a) is surrounded by a low but intensively gullied shield that grades into southern Malea Planum, which is also dissected (by Axis Vales) all the way to Hellas Planitia (Figures 8 and 10). These valleys are more irregular in form than those of Hadriaca and Tyrrhena; this difference may have been caused by the lack of distinctive or continuous layering and by the prevalence of topographic obstacles produced by wrinkle ridges and possible flows, as well as by subsequent eolian mantling and deflation in the Malea Planum area (discussed below). Fenus Patera (Figure 7), on the other hand, does not cap a discernible shield, and valleys originate only on its sloping north edge.

Crater counts indicate that the erosion of the paterae and associated plains has been long lived. At decreasing crater sizes, crater densities show decreasing relative age (Tables 1 and 4), from Late Noachian to Middle Amazonian. At northern Malea Planum, where the largest crater dataset exists, the density of craters >16 km in diameter yields a Late Noachian age. Some of these larger craters appear embayed or dissected, and thus the age represents an intermediate age between the underlying basin rim unit and overlying dissected material, which is probably at least several hundred meters thick (given the embayment relations with large craters and knobs). Craters between 4 and 16 km across yield Hesperian ages, which probably reflect the age of the dissected material. At diameters <4 km, crater densities yield Amazonian ages, which probably date the eolian modification (which likely affected the upper tens to couple hundred meters of the patera units).

6.2. Lower Hesperian Basin Infilling and Modification

On the floor of the Hellas basin, a thick interior deposit covers about 10^6 km^2 (Figure 3). The deposit overlies Upper Noachian to Lower Hesperian ridged plains material, and it is embayed by fluvial and eolian materials of Late Hesperian and Amazonian age. Although its thickness has not been precisely determined, it probably exceeds 1 km over much of its extent. The upper part of one of the tallest erosional scarps within the deposit has a height (determined by shadow measurements) of about 700 m (Figure 5); the lower part of the scarp was not in shadow and the base of the deposit is at some unknown depth. Least eroded areas occur mostly in the eastern part of the deposit (Figure 4), and most of these areas appear armored by crater ejecta (Figure 15). Erosional landforms within the deposit include steep scarps, hummocky and knobby terrain, and deeply etched margins and hollows (Figures 15 and 16).

The crater-density distribution for the interior deposit is difficult to interpret because of very deep erosion in most places. In addition, relict surfaces may be preserved by ejecta from larger craters, causing preferential survival of larger craters. These effects result in a low-sloping crater density distribution, in which H(16) yields a Late Noachian age and N(2-2.8) as Early to Middle Amazonian ages (Table 4). Highly degraded hummocky and knobby areas of the interior deposit yield Hesperian to Middle Amazonian ages, indicating that crater obliteration has been long lived. Based on the Late Noachian age for the underlying basin ridged unit, the preferential preservation of larger craters, and the relative age of younger deposits (as well as modified surfaces of the interior deposit), we assign an Early Hesperian age to the unit (the Late Noachian crater density is suspect because of the erosional preservation of surfaces armored by the ejecta of large craters). The cratering result, along with the observed erosional features, may indicate that initial erosion as much as several hundreds of meters deep during the Early or Late Hesperian was followed by relatively modest erosion (typically, tens of meters deep) through the Late Hesperian and Amazonian.

Erosion of the interior deposit probably occurs primarily by eolian deflation. This conclusion is based on the friable nature of the deposits (which is attested to by deep erosion of
complex morphology, including hummocky (H), knobby (K), and relatively smooth (SM) areas; plateaus at top capped by west-trending ridges that may be yardangs (Y). Relatively low areas are covered by north- to northeast-trending patterns of dunes (D). Western border of interior deposit is marked by prominent scarp (arrows); crater (C) below scarp apparently is filled with outlier of interior deposit. (Viking images 363532 and 36; centered at ~ 35.75'S, 292.50'E, 250 m/pixel).

The deposit except where armored by crater ejecta, the paucity of possible fluvial channels, the intense wind stresses (Greeley et al., 1993), and the dunes (Figures 4, 9, 15, and 16) that are common in Hellas Planitia. The materials of the interior deposit are likely to be fine grained and poorly consolidated, which is in agreement with an interpretation made from Viking IR TM data (Moore and Edgett, 1993). Deposition of fines that likely resulted from eolian deflation probably has led to the development of a sandy, possibly indurated surface lag deposit (Moore and Edgett, 1993), which may have contributed to dune deposits in the basin and surrounding highlands. Locally, the interior deposit may be degraded by collapse and mass wasting. A smooth flow deposit originates from a shallow depression (Figure 15), which may have formed due to liquefaction and collapse, similar to outflow channels on the east rim of Hellas and elsewhere on Mars (Carr and Cloos, 1981; Nummedal and Prior, 1981). Some debris aprons found mostly along steep scars of the interior deposit may have formed with the assistance of interstitial ice. Thus, ground ice and water probably were present locally in the deposit; water and ice could have been supplied from frost and from the runoff of various channels that debouched onto the basin floor.

The interior deposit is not only deeply eroded, it also appears to have undergone significant lateral retreat in the northern, western, and southern parts of Hellas Planitia, where it may have once covered virtually the entire extent of the basin rimmed unit (see Figure 3). Along the edge of proposed retreat, the interior deposit is marked by a fairly continuous prominent scarp, which is only absent along the eastern edge of the deposit. Beneath this scarp, the basin rimmed unit is superposed by several pedestal craters (e.g., Figures 9 and 16) and partly exhumed craters that appear to have been once completely buried (vicinity of latitude 33° S, longitude 298°). Dunes, knobby outliers of the interior deposit, and younger mantle material are common below this scarp (Figures 9 and 16). Also, the edge of dissected patera flank material on northern Malea Planum forms a distinctive scarp, perhaps marking the location where the flank material once abutted the preretreat edge of the interior deposit (Figure 10). The total area of retreat may have been as much as 0.3 x 10^6 km^2.

Given the likelihood that the interior deposit is fine grained and poorly indurated, what is its possible origin? Some obvious sources include clastic sediment from valleys that feed into Hellas basin and pyroclastic fines generated by highland deposits on the Hellas rim (see Table 5). However, these sources probably have provided only a small part of the volume of the interior deposit (which at one time amounted to ~1-3 x 10^6 km^3 or more prior to its erosion). Small valleys elsewhere on Mars—generally have filled only small highland basins; much more clastic material can be provided by outflow channels. However, the outflow channels that debouched into Hellas postdate the interior deposit and emplaced into the basin only about 5% or less equivalent of sediment volume of the deposit (see Table 5). The highland patera appear to have generated a similar volume of material, based on the 1-2 km maximum heights of the paterae (Blake et al., 1981; U.S. Geological Survey, 1991; Mouginis-Mark et al., 1992) and the mapped extents of their surrounding deposits, which probably are hundreds of meters thick (Figure 3; Greeley and Crown [1990], Crown and Greeley [1993]). However, the deposits of Tyrrhena and Hadriaca Paterae, given their eruption dynamics, are more likely to be pyroclastic flows rather than air-fall deposits (Greeley and Crown, 1990; Crown and Greeley, 1993); similar arguments can be made for Peneus and Amphitrites Paterae (see above). These materials are deeply dissected, and the dissected material on the south rim of Hellas would have been the largest source for fluvial transport. This dissection may have overlapped in time with formation of the interior deposits. Overall, erosion of rim materials may have contributed only as much as 18% of the original volume of the interior deposit (Table 5).

By default, the majority of the deposit must therefore originate from beyond the rim of Hellas, which requires eolian transport. Eolian deposits of various ages have been proposed in various highland and lowland areas of Mars [e.g., Schultz and Luce, 1988; Grant and Schultz, 1980, 1993; Moore, 1990]. Possible eolian deposits of similar volume include the Medusae Fossae Formation (~4 x 10^6 km^3; Scott and Tanaka...
a sequence of friable materials along the highland/lowland boundary between Elysium and Olympus Montes [Scott and Tanaka, 1986; Greeley and Guest, 1987], and the polar layered deposits (~3 x 10⁶ km³), assuming an average thickness of 1 km and their areal extent as given by Tanaka et al. (1988). Given the Early Hesperian age of the interior deposit, the eolian material may have been derived from the fine fraction of eroded highland materials [Williamson and Baldwin, 1989; Craddock and Maxwell, 1993], fretted and knobby terrain development along the highland/lowland boundary [Tanaka, 1986], and times produced by impacts [e.g., McKinnon and Tanaka, 1989].

Apparently Hellas basin provided a sink for these materials. The depth of the Hellas basin is accompanied by elevated surface pressures within the lower parts of the basin, thereby allowing CO₂ to condense at relatively higher temperatures [Lunner et al., 1993]. Presently, CO₂ frost covers the entire basin floor during midwinter in the southern hemisphere [Briggs et al., 1997]. Any available dust particles act as nuclei for condensing CO₂, leading to the precipitation of these dust/frost particles onto the Hellas floor. In this way, large volumes of available dust could have been deposited within Hellas. This depositional mechanism may also contribute a significant fraction of volatiles within the interior deposits.

6.3. Hesperian-Amazonian Rim Mantling and Channeling

During the Late Hesperian/Early Amazonian, many small and a few large channels dissected the eastern and southern rim areas of Hellas basin (Figures 7a, b, and 17). On the east rim (Figure 17), Niger, Dao, and Harmakhis Valles may have resulted from catastrophic outbreaks related to volcanism or intrusion at Hadriaca Patera and volcanic flows from Tyrrhena Patera [Squyres et al., 1987]. Reull Vallis originates from ridged plains material of southern Hesperia Planum and therefore could also have an origin partly related to volcanism and perhaps controlled by basin structure. Its terminus is abrupt near the head of Harmakhis Valles, where it appears to be partly buried by an adjacent debris apron (Figure 17), indicating that at one time Reull Vallis may have been continuous into the head of Harmakhis Vallis. In total, the eastern valleys cover about 0.05 x 10⁶ km², and deeper sections locally exceed 1 km (Figure 5). The extensive dissection of the valleys may be in part due to the nature of the dissected material, which may include eolian mantles [Greeley and Guest, 1987].

All of the valleys except Reull terminate in eastern Hellas Planitia, incising into the dissected rim unit and the interior deposit (Figure 11). Much of the deposits of Reull Vallis may make up parts of the dissected rim unit (Figure 17), whereas the other valleys would have deposited sediment onto eastern Hellas Planitia. Such deposition in this area would explain the subdued appearance of wrinkle ridges and the density of craters 2-8 km across (Table 4). However, the densities of craters larger than 8 km yield chiefly Late Noachian to Early Hesperian ages, which suggests that the basin rim unit underlies the subdued ridged unit. Given the 0.3 x 10⁶ km³ extent of the subdued ridged unit and a maximum valleys volume of about 0.05 x 10⁶ km³, the average depth of burial is about 130 m, which is consistent with burial of most craters up to a few kilometers in diameter [Pike and Davis, 1984] and partial burial of wrinkly ridges (measured elsewhere on Mars at up to 150-200 m in height [Golombek et al., 1991]). This interpretation is also consistent with the lack of fan development at the mouths of the channels, which indicates that deposition was in the form of broad sheets [Ort, 1994]. Locally, viscous flow of the deposit is indicated where the subdued ridged unit forms a lobate scarp and embays rim material north of the mouth of Dao Vallis.

The south rim of Hellas also has numerous small channels and at least one large fretted channel, Mad Vallis (Figure 6). The small channels here are generally deeper and less numerous than their eastern rim counterparts and are locally infilled with relatively higher albedo material that in places choked off

<table>
<thead>
<tr>
<th>Table 5. Estimated Volumes of Material Eroded From the Hellas Rim</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit or feature</strong></td>
</tr>
<tr>
<td>Valles material</td>
</tr>
<tr>
<td>Dissected rim unit, Malea Planum</td>
</tr>
<tr>
<td>Tyrrhena</td>
</tr>
<tr>
<td>Hadriaca</td>
</tr>
<tr>
<td>Amphitrites</td>
</tr>
<tr>
<td>Peneus</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

*This work.

†Note that east rim valleys deposits postdate the age of the Hellas interior deposit. Therefore, contributions from observable eroded rim materials are 177,879 km³, which is only ~18% of volume of the interior deposit assuming minimum volume of deposit of 10² km³.

†Johnson et al. [1991].
†Plavcan and Saunders [1979].
†Crowe et al. [1992].
Figure 17. Harmakhis Vallis (H) and lower extremity of Reull Vallis (R) on eastern rim of Hellas. Rim material locally is dissected by degraded-appearing valleys (V) that emerge from smooth plateau deposits (S). These deposits in turn are capped in places by etched appearing mesas (M). Smooth, lobate, channelled material (L) may be deposit of Reull Vallis; note how deposit surrounds and channels cut inliers of rim material (I). Note debris aprons (A) along walls of valleys and bases of massifs and layering in south wall of Reull Vallis (arrows). (Viking images 329S25 and 26; centered at ~ 40.00°S, 265.00°E, 235 m/pixel).

channels completely. Their proximity to Amphitrites and Peneus Paterae, coupled with an observed paucity of channels on the opposite side of the local basin rim, suggest that they too may have formed by groundwater sapping or catastrophic outbreaks related to local intrusions and volcanism. Surface runoff, by contrast, should have produced channels on both sides of the high south rim.

On the upper slopes and rim of the Hellas basin, smooth, flat, irregularly shaped mantles are locally preserved (Figure 4). These mantles (tens to hundreds of meters thick) occur (1) in shallow topographic traps (such as in highland material as discussed previously (Figure 13)), within Amphitrites and Peneus Paterae, and on smooth plains surrounded by wrinkle ridges; (2) on broad plains, such as northern Malea Planum (Figures 7b and 8) and south of Harmakhis and Reull Valles (Figure 17), (3) within large centers (Figure 12), and (4) as pedestals beneath impact ejecta (Figures 8 and 10). Because pedestal centers of fairly high density occur on Malea Planum and small craters of moderate density are superposed on some of the mantles within craters, at least some of the mantles are probably Hesperian [see Tanaka and Scott, 1987]. However, other mantles appear uncratered (at 200–300 m/pixel resolution) and are likely to be Amazonian. These young mantles commonly appear etched, such as the material that thinly blankets the northern parts of Peneus Patera (Figure 7b) [Zimbelman et al., 1989] and the dissected slopes of northern Malea Planum. Some of the crater-fill mantles have pits hundreds of meters deep (Figure 12; see also Thomas [1984]).

Pits and etched features indicate the poorly indurated, friable nature of these mantles.

6.4. Hesperian-Amazonian Interior Mantling

Along the south edge of the floor of Hellas, in a topographic low between the Lower Hesperian interior deposit and the north edge of Malea Planum, smooth material of relatively low albedo forms a nearly continuous blanket (Figure 10). The blanket is made up of broad (>70 km across), rolling mounds having gently sloping margins that embay or feather into adjacent units. The blanket appears to be covered locally by dunes, and thus it may be made up of thick sequences of dune material in places. Elsewhere, the unit is pitted and etched, suggesting that the deposit is made up of friable material, likely dust or silt. Its crater density indicates a Late Hesperian or Early Amazonian age. In addition, possible eolian mantles that appear wind sculptured and similar in age to the blanket form generally sharp-edged, etched, irregular mesas near the center and the northwest edge of the Lower Hesperian interior deposit.

6.5. Amazonian Debris Aprons

Many steep slopes in the map area have lobate debris aprons that extend for several to more than ten kilometers from their bases. These lobes are on mantles of the basin rim unit (Figure 17), on channel walls of the dissected material (Figure 17), and on a few high mesa walls in the interior.
depôt. These features have been interpreted as talus; local pitting suggests that the materials may have been ice-rich [Crown et al., 1992]. The lobate form, steep slopes, and pitting of the aprons indicate emplacement from ice creep, as in terrestrial rock glaciers [Squyres, 1979; Crown et al., 1992].

6.6. Recent Wind Features: Dunes, Yardangs, and Splotches

Occurrences of dunes in the Hellas region are indicated by (i) linear, arcuate, and reticulate patterns of low ridges in the central, northern, and western parts of the Hellas floor (Figures 4, 9, and 15), and (ii) individual ridges (perhaps climbing dunes) on the top and flanks of some irregular masses within the Hellas basin (Figures 15 and 16). Also, dark dunes occur within craters west of the Hellas basin [Thomas, 1984]. A few large occurrences of reticulate ridge sets interpreted to be dune fields occupy the central and northwestern parts of the basin (Figures 9a and 16). The size and morphology of these features are similar to networks of compound crescentic dunes (also termed akle dunes) observed in the Takla Makan Sand Sea, China (Figure 9b), as well as in many other terrestrial locations [Breed and Crow, 1979]. Crescentic dunes include transverse types, which generally infer a specified wind regime based upon slipface or ridge-axis orientation, where ridge axes develop transverse (normal) to the resultant direction of effective winds.

By contrast, yardangs are ridges formed by wind fluting and are prevalent in some areas of Mars, where friable materials and strong winds exist [e.g., Ward, 1979]. In north-central Hellas Planitia, possible yardangs cup a large plateau (Figure 16). The wind direction inferred from these features is ambiguous, as no blunt or tapered ends are readily discernible at existing image resolutions. However, most yardangs on Mars are more elongate than those on earth and may be carved into ridges with very low width-to-length ratios [Ward, 1983], thereby making it more difficult to detect tapering. Another possible set of yardangs occurs in the center of Hellas Planitia and is marked by numerous small (c1 to 10 km long) knobs, many of which are elongate roughly north-south (Figure 18).

Within this set, tapering is readily apparent, with tapered ends predominantly pointing north-northeast. Some Viking images reveal seasonal, surficial, bright and dark splotches in various areas of the Hellas region, which have been interpreted to be eolian frost and sand deposits [Thomas et al., 1979; Thomas, 1984]. High-albedo, high-latitude splotches are likely CO₂ frost streaks. Other features attributable to wind include light and dark albedo streaks that trail from crater rims and other topographic obstacles (Figure 19). Some of these features are elongate and linear, indicating modification by a predominant, directional wind.

In addition to the local transport of sand and dust indicated by this discussion, net dust removal from Hellas and accumulation in equatorial regions is presently occurring [Moore and Edgett, 1995; Christiansen, 1986]. Saltation and dust devils are primary mechanisms that can account for lifting dust into the atmosphere, and the high predicted wind stresses in Hellas favor this activity [Greeley et al., 1992, 1993]. The identification of possible dune fields both within and outside the Hellas basin shows that saltation probably does occur or had previously occurred in this region, which coincides with observations that Hellas is a major source region for regional and global dust storms on Mars [Mansin and Zurek, 1993].

Figure 18. Knobby terrain in central Hellas Planitia represents highly degraded remnants of the Lower Hesperian interior deposit. The edge of a large crater (C) arms the deposit beneath it. Elongate knobs (K) are interpreted as yardangs and show definite tapering to the north-northeast. This orientation is consistent with high wind stresses predicted by a Mars general circulation model [Greeley et al., 1993], as well as with inferred wind directions from nearby dunes (D) and splotches. Note smooth material (S) embaying the knobby material, and the superposed pedestal craters (P). These features indicate mantling and stripping of relatively unconsolidated, likely eolian, material. (Viking image 363812; centered at 41.5°S, 300.4°W, 247 m/px).

Depositional cycles of 10⁵–10⁶ years under present astronomical and climatic conditions may reverse this transfer and result in cyclic formation of dust deposits as much as a few meters thick [Christenson, 1986; Greeley et al., 1992].

6.7. Theoretical Considerations Of Modern Sand and Dust Transport in Hellas

Many of the features described in the previous section show orientations indicative of wind direction. In this section we compare these inferred wind directions with those directions predicted by a recent Mars general circulation model (GCM) [Greeley et al., 1993; see also Skupnek, 1989]. The Greeley et al. [1993] OCM is a three-dimensional model that includes the effects of varying global topography, albedo, thermal inertia, polar CO₂ condensation, and dust opacity. Each cell in the 25 by 40 cell computational grid is 7.5° latitude by 9.0° longitude, and the model runs cover only six brief, discreet Lₐ (Martian austral longitude) intervals. Although the runs were spaced fairly evenly at ~60° Lₐ intervals, the model integrations at these Lₐ intervals covered only ~5° of Lₐ. Therefore, observed features were compared with the nearest model (seasonal) interval to when the image was taken, which in some cases differed by as much as 20° to 30° of Lₐ.
High-albedo, high-latitude splotches (likely CO$_2$ frost streaks) correlate well with predicted southern hemisphere autumn and winter winds coming off the annual polar cap [Thomas et al., 1979]. This relation fits for frost streaks observed along the south and southeast Hellas rim, as well as a couple of occurrences in the Hellas basin (see Table 6).

Most other ephemeral features correlate only marginally (53% in agreement) with GCM-predicted surface stresses of closest $L_{st}$(seasonal) interval, however the degree of agreement varies greatly from place to place (Table 6). In particular, the west rim and basin show excellent agreement between observed and predicted vectors, whereas all other areas have poor to marginal agreement (27% to 50%). The same ephemeral features also show marginal positive correlation (47%) with the GCM as a function of the maximum stress vector (maximum wind-stress direction regardless of season). The west and southeast rim vectors show the best GCM agreement (100% and 87%, respectively) as a function of maximum wind stress; remaining areas show poor to fair correlation.

The marginal correlation of observed transient feature orientations in places with those predicted by the GCM is not surprising and may be mainly due to the limited spatial and temporal resolutions of the model that do not detect rapid, local variations in wind patterns. In contrast to the relatively larger dune and yardang fields in Hellas, most splotches and streaks are much smaller than the spatial resolution of the model, where local factors such as topography, structure, and albedo, as well as seasonal atmospheric changes, can greatly influence winds. Even small-scale structure (including joint sets) can influence local wind direction and speeds [Warf, 1979].

The orientations of yardangs and dunes are generally controlled by the directions of the strongest seasonal winds, so the high correlation between the Hellas features and the GCM is not surprising. In addition, we observe an anticorrelation between the landform orientations and GCM predictions for other Martian seasons, even though these stresses are of considerable strength. One possibility is that these features are relics, preserving orientations governed by previous wind regimes. Also, we have not attempted to investigate whether the Hellas dunes show some control by secondary winds, which is entirely possible, particularly where dune patterns are complex (e.g., Figure 9a).

6.8. Discussion of Proposed Glacial Modification

Periods of net deposition and erosion are indicated by our stratigraphic and geomorphologic interpretations of the Hellas region. We attribute this history mainly to eolian and fluvial processes and, to a lesser degree, mass-wasting and ground-ice processes. Kargel and Strom [1992], on the other hand, proposed a late (Amazontian) glaciation event in Hellas. Their cited evidence for this event includes (1) erosional scouring and channel formation on northern Males Planum, (2) east-trending ridges (interpreted as transverse and recessional moraines) and other related features interpreted as ice-regression landforms (such as eskers and drumlins) on the floor of Hellas, and (3) possible lacustrine plains that may have been material deposited in glacial meltwater. In addition, Kargel and Strom [1992] argue that the overall pattern and morphologic associations of these features attest to a glaciation event. Our detailed study, however, has led us to...
Figure 20. Wind vectors in the Hellas region as inferred from oriented dust and frost splatches and from yardangs and dune forms. Arrows indicate downwind direction; bidirectional arrows indicate ambiguous wind directions. Solid arrows, dust and frost splatches; open arrows, dunes; circles, yardangs. Correlation of these wind vectors with GCM predictions is shown in Table 6.
Table 6. Wind Vectors in the Hellas Region of Mars as Inferred From Dust and Frost Streaks, Yardangs, and Dune Forms and Their Correlation With Predicted Wind Stress Vectors From the Mars GCM of Greeley et al. [1993].

<table>
<thead>
<tr>
<th>Viking Image Location (latitude, longitude)</th>
<th>Feature</th>
<th>$L_0$</th>
<th>Wind Vector</th>
<th>GCM Correlation†</th>
<th>$L_0$ Maximum Stress</th>
</tr>
</thead>
</table>

**North Rim**

- 620A06 18°S, 315° dark streak 53.2 N50E  -  
- 620A05 22°S, 315° light streak 53.2 N40E-S40W  -  
- 620A09 18°S, 308° light/dark streak 53.2 N30E-S30W  -  
- 395S06 23°S, 294° light/dark streak 96.3 N35W-S35E  +  
- 397S10 20°S, 287° light streak 97.2 N30W  +  
- 397S05 23°S, 279° light streak 97.2 N30W-S30E  +  

**Northeast Rim**

- 365S44 20°S, 239° light streak 83.1 N20W-S20E  -  
- 365S42 23°S, 257° light streak 83.1 S20E  -  
- 365S43 19°S, 255° light streak 83.1 S20E  +  
- 365S39 24°S, 251° light streak 83.1 S20E  +  
- 478A18 24°S, 259° light/dark streak 346.1 N8  +  
- 625A36 26°S, 261° light streak 55.4 N50E-S50W  +  
- 478A09 29°S, 209° light streak 346.1 N20E  +  
- 555A05 30°S, 260° light/dark streak 23.0 N60E-S60W  +  
- 625A20 34°S, 264° dark streak 55.4 S40W  -  
- 097A60 43°S, 260° light/dark streak 127.9 N70E-S70W  +  

**Southeast Rim**

- 097A75 53°S, 249° dark streak 137.9 N30E  -  
- 310B26 65°S, 255° dark streak 31.5 N65E-S65W  +  
- 504B51 63°S, 273° light streak 28.8 N20W-S20E, S20W  +  
- 504B50 65°S, 278° light streak 28.8 N30W-S30E  +  
- 504B53 61°S, 274° light streak 28.8 S20E  -  
- 504B55 60°S, 282° light/dark streak 28.8 N20W-S20E  +  

**South Rim**

- 431B41 63°S, 222° light streak 353.6 N30E-S30W  -  
- 094A75 59°S, 227° light/dark streak 126.5 N10W-S10E  +  
- 466B53 65°S, 303° dark streak 10.9 N45E-S45W  -  
- 431B40 61°S, 306° light/dark streak 353.6 N45E-S45W  -  
- 578B10 52°S, 303° light streak 61.8 N25E-S25W  +  

**West Rim**

- 510A54 43°S, 314° dark streak 2.3 N70E-S70W  +  
- 510A50 46°S, 321° light/dark streak 2.3 N40E-S40W  +  
- 584A28 40°S, 315° dark streak 37.2 N60E-S60W  +  
- 547A49 38°S, 318° light/dark streak 20.2 N70E-S70W  +  

**Hellas Basin**

- 620A47 34°S, 303° light/dark streak 53.2 N60E-S60W  +  
- 584A31 43°S, 306° dark streak 37.2 N40E-S40W  +  
- 584A12 46°S, 305° dark streak 37.2 N30E-S30W  +  
- 584A13 48°S, 293° dark streak 37.2 N45E-S45W  +  
- 578B14 48°S, 303° light streak 61.8 N40E-S40W  +  
- 584A16 46°S, 394° dark streak 37.2 N20E-S20W  +  
- 545B36 51°S, 281° yardangs 47.3 N70E-S70W  -  
- 365S24 47°S, 285° yardangs 82.3 N40E-S40W  -  
- 365S34 34°S, 293° yardangs 82.3 N70E-S70W  -  

*Note: $L_0$ is the latitude of the feature, and the wind vector is given as N°E or S°W. GCM Correlation indicates the correlation with the predicted wind stress vectors from the Mars GCM.
these conclusions for each point of evidence of the glacial hypothesis, and we provide further arguments that pose serious problems to the glacial interpretation. Severe problems to the glacial interpretation of the southern edge of the interior deposit are marked by a prominent scarp where we would expect the proposed glacial deposits to be more continuous. The eastern side of the interior deposit is the only part of the deposit that does not have a bounding scarp, and thus shows the only definitive continuity between Hellas basin materials and rim-derived materials (which we interpret as emplaced by local valley). The ridges thought to be recessional moraines, including a string of mesas [Kargel and Strom, 1992, Figure 4] that compose a group of plateaus on the Hellas interior deposit, appear to us as erosional remnants of the original surface material of the interior deposit that are locally armored by crater ejecta (e.g., Figure 15). Other ridges may be structural (Figure 10). The proposed "drumlins" in western Hellas are knobs elongate in a northerly direction. An alternative explanation is that the elongation is due to wind fluting, which forms yardangs. The knobs are associated with low mesas and small pedastal craters, some of which have elongate knobs appear covered by transverse dunes oriented perpendicular to the knobs (see Viking images 363512 and 14). Thus, a yardang interpretation is consistent with apparent prevailing wind direction, as inferred by local dune orientation and GCM predictions and the flat nature of the knobs (see Viking image 363512 and 14).

We agree with Kargel and Strom [1992] that the northeastern ridged plains of Hellas Planitia may have been covered by a temporary lake, for which we have cited evidence in the form of Late Hesperian/Early Amazonian deposition. However, we have discussed previously, the likely source of the water, based on temporal and spatial relations, were the large valleys on the east rim of Hellas. If the water source was from a melting glacier within Hellas Planitia, we might expect to see source channels coming off the interior deposit; none are observed.

Only local areas provide some equivocal geomorphologic evidence for ice sheets. First, a sinuous ridge system about 100 km long in southwestern Hellas Planitia (see Viking image 363511) may be interpreted as eskers [Kargel and Strom, 1992]. However, inverted stream topography produced by evaporite cementation of stream gravels and subsequent exhumation [Maizels, 1990; Rice and Mollard, 1994] is
and Mollard, 40 km across in central images that were yet unavailable on CD-ROM. We also are grateful to Carol Breed (USGS), who provided us with expert consultation on dune forms and the interpretations to the glacier hypothesis, we find other serious difficulties with that proposal. Overall, the assemblage of landforms and crater chronology in Hellas basin indicate a landscape evolution dominated by eolian processes and punctuated by brief and less important fluvial events. Mass wasting, ground-ice processes, and perhaps glaciation played local, minor roles.

Acknowledgments. The research in this paper was stimulated largely by discussions at the wrap-up Mars Surface and Atmosphere Through Time workshop in Houston in November 1993, in which planetary geologists, astronomers, and atmospheric scientists expressed a keen interest in understanding the geology of Hellas basin. We appreciate the careful reviews provided by John Grant, Jeff Kargel, Baerbel Blasius, and Klaas Hofman. We have assigned the interior deposit an Early Hesperian age based on crater density and stratigraphic relations. Also, the glacial hypothesis requires the rough equivalent of the entire estimated Martian water budget [Kargel and Strom, 1992].

References


TANAKA AND LEONARD: GEOLOGY AND LANDSCAPE OF HELLAS REGION, MARS


(Received March 29, 1994; revised October 13, 1994; accepted October 27, 1994.)
The circum-Chryse region as a possible example of a hydrologic cycle on Mars: Geologic observations and theoretical evaluation

Jeffrey M. Moore, Gary D. Clow, Wanda L. Davis, Virginia C. Gulick, David R. Janko, Christopher P. McKay, Carol R. Stoker, Aaron P. Zen

Abstract. The intersection and superposition relationships among channels, chaos, surface materials units, and other features in the circum-Chryse region of Mars were used to evaluate relative age relationships and evolution of flood events. Channels and chaos in contact (with one another) were treated as single discrete flood-carved systems. Some outflow channel systems form networks and are inferred to have been created by multiple flood events. Within some outflow channel networks, several separate individual channel systems can be traced to a specific chaos which acted as flood-source area to that specific flood channel. Individual flood-carved systems were related to widespread materials units or other surface features that served as stratigraphic horizons. Chryse outflow channels are inferred to have formed over most of the perceivable history of Mars. Outflow channels are inferred to become younger with increasing proximity to the Chryse basin. In addition, outflow channels closer to the basin show a greater diversity in age. The relationship of subsequent outflow channel sources to the sources of earlier floods is inferred to disfavor episodic flooding due to the progressive tapping of a juvenile near-surface water supply. Instead, we propose the circum-Chryse region as a candidate site of past hydrological recycling. The discharge rates necessary to carve the circum-Chryse outflow channels would have inevitably formed temporary standing bodies of H2O on the Martian surface where the flood-waters stagnated and pooled (the Chryse basin is topographically enclosed). These observations and inferences have led us to formulate and evaluate two hypotheses: (1) large amounts of the sublimated H2O off the Chryse basin flood lakes precipitated (snowed) onto the flood-source highlands and this H2O was incorporated into the near surface, recharging the H2O sources, making possible subsequent deluges; and (2) ponded flood-water in Chryse basin drained back down an anti basinward dipping subsurface layer accessed along the southern edge of the lake, recharging the flood-source aquifers. H2O not redeposited in the flood-source region was largely lost to the hydrologic cycle. This loss progressively lowered the viability of the cycle, probably by now killing it. Our numerical evaluations indicate that of the two hypotheses we formulated, the groundwater seep cycle seems by far the more viable. Optimally, 7/4 of the original mass of an ice-covered cylindrical lake (albedo 0.5, 1 km deep, 100-km radius, draining along its rim for one quarter of its circumference into substrata with a permeability of 3000 darcies) can be modeled to have moved underground (on timescales of the order of 10^3 years) before the competing mechanisms of sublimation and freeze down choked off further water removal. Once underground, this water can travel distances equal to the separation between Chryse basin and flood-source sites in geologically short (10^3 year-scale) times. Conversely, we calculate that optimally only ~40% of the H2O carried from Chryse can condense at the highlands, and most of the precipitate would either collect at the base of the highlands/lowslands or sublimate at rates greater than it would accumulate over the flood-source sites. Further observations from forthcoming missions may permit the determination of which mechanisms may have operated to recycle the Chryse flood-waters.

Introduction

There has been renewed interest in the idea of Martian hydrologic cycles largely brought on by T. S. Parker, V. R. Baker, and their colleagues, who explored the possibility that Mars once had large bodies of standing water and large continental-style glacial ice sheets located in topographically confined lowlands [e.g., Parker et al., 1989; Baker et al., 1991; Kargel and Strom, 1992]. Other studies have pointed to evidence for large lacustrine basins on Mars which are postulated to be very young in the martian stratigraphic sequence [Scott and Chapman, 1991; Scott et al., 1992].
These investigations imply that large amounts of water were cataclysmically released, late in Martian history, at least in part by mechanisms which restore lake-forming water to its source by some form of a hydrologic cycle. In a separate study focused on the sediments of ancient, sapping channel-fed lakes, Goldspiel and Squyres [1991] and Goldspiel et al. [1993] concluded that unless the upper several hundred meters of the Martian highlands were composed of very fine-grained noncohesive material, it would be necessary to recharge the associated up-channel aquifers to produce the amount of water needed to form the observed landforms.

The concept of a Martian hydrologic cycle (as envisioned for example by Parker et al. [1989] and Baker et al. [1991]) has been criticized [Carr, 1991] in part because it is very difficult to understand the physical mechanisms that would repeatedly allow vast (ocean-sized) amounts of liquid water to suddenly appear and then just as suddenly disappear several million years later. It is the lack of recognized mechanisms (exclusive of radical deglaciations from the modern Martian environment) to drive a Martian hydrologic cycle that makes many investigators inclined to question the possibility of such a cycle.

Over time a general consensus has developed that Mars has had large flood events in its history (see reviews in Carr [1981], Baker [1982], Mars Channel Working Group [1983]). Of the several channel types observed on Mars, those classified as outflow channels [Sharp and Malin, 1975; Pieri, 1976; Mars Channel Working Group, 1983] are considered the best evidence for catastrophic flooding. Outflow channels are characterized as generally large features that start full-bore from localized sources (usually chaotic terrain) [Sharp and Malin, 1975]. They are widest and deepest at the head and decrease in size distally. Generally, they have few, if any, tributaries. The interiors and down-flow plains of some outflow channels appear to be scoured. This assemblage of landforms bears a remarkable resemblance to catastrophic flood terrains on Earth [Baker and Milton, 1974]. On Mars, outflow channel systems are thought to have formed by catastrophic release of groundwater [Mars Channel Working Group, 1983] perhaps from outflow aquifers [Carr, 1979]. Once released, the water would have flowed down the regional slope, locally diverted by topographic barriers, carving the channels much in the way the sudden draining of Pincosone Lake Missoula carved the Channeled Scabland in the American Northwest [Baker and Milton, 1974; Mars Channel Working Group, 1983]. Most outflow channels on Mars are located around and disgorge into the topographically enclosed Chryse Planitia basin.

In this paper we present some geologic observations for evaluating the postulated role of an ancient catastrophic flood on Mars. For example, the Chryse outflow channels into Circum-Chryse flood deposits are inferred to have formed over most of the perceivable history of Mars. The oldest channels cut through the intercrater plains and the youngest channels intersect the Chryse smooth plains. Although determining the relative age of the youngest channels is clear, the unambiguous relative age of the oldest channels is more difficult. For example, although Margaritifer Vallis, one of the oldest channels, appears to have been exhumed from beneath intercrater plains material (Figure 1a), it is possible that Margaritifer Vallis formed contemporaneously with the surrounding terrains [M.S. Parker, personal communication, 1994]. Ravi Vallis (Figure 1c) has clearly formed in intercrater plains material. Though its lower reach appears to be buried by the Chryse smooth plains material, the possibility that the distal end of Ravi Vallis is truncated by Hydroses Chaos cannot be ruled out. However, detailed stratigraphic relations of the older channels to the intercrater plains and the Chryse smooth plains is not clear in some cases, what is clear is that some channels postdate and some channels predate the Chryse smooth plains.

In this study, no attempt was made to determine whether any of the outflow channels drain the circum-Chryse region of Mars. It is possible that some outflow channels drain the circum-Chryse region of Mars. This has important implications for determining the age of the circum-Chryse region of Mars.

Figure 2 shows the results of our relative age dating of outflow channels located in the Circum-Chryse drainage basin region of Mars. From these results, two trends become apparent from the placement of the circum-Chryse drainage basin. (1) The relative age of outflow channels tends to decrease towards the north and with...
Figure 1. The transection and superposition relationships among channels, chaos, surface materials and other features were used to evaluate the relative ages of the flooding events in the circum-Chryse region. (a) Channel (Margaritifer Vallis) being exhumed from beneath, and thus predates, intercrater plains material, which is the older of the two stratigraphic horizons used in this study. (b) Channel (Ravi Vallis) eroded into intercrater plains material but whose distal end is buried under Chryse smooth-plain material. Occurring between the two stratigraphic horizons, it is identified as an intermediate-aged flood feature. (c) Channel (Tiu Vallis) that has eroded the Chryse smooth-plain material and that thus classified as the youngest flood features.

decreasing elevation. (2) Outflow channels in the north tend to show a greater diversity in age. These trends do not change even if the outflow channels are divided into two groups: those that predate and those that postdate the formation of the Chryse smooth plains.

Another observation from this study is that several outflow channels contain multiple flood source sites within their boundaries. Although multiple episodes of flooding within the outflow channel have been noted by previous studies (e.g., Chapman and Zorn, 1989), it is not possible to determine unambiguously whether these episodes represent several discrete catastrophic flooding events within a channel separated by significant geological periods or whether these episodes reflect the waning stages of a single flood event and the progressive downsizing of a groundwater reservoir. If these episodes are separated by geologically significant periods, then recharge of aquifers is probably necessary for subsequent flood events. An example will illustrate this point. Ravi Vallis/Aramatun Chaos system is an outflow channel system that may have experienced multiple episodes of fluid flow (Figure 3), because it contains several regions of chaotic terrain within its boundaries. Chaotic terrain forms where groundwater is cataclysmically released from the subsurface.

In the first episode, fluid from Aramatun Chaos carved the main channel of Ravi Vallis. In succeeding events, fluid derived from two additional chaos sites reoccupied portions of the inner channels and modified previously collapsed areas. The area of chaotic terrain associated with subsequent events is smaller. The first preserved event that carved Ravi Vallis was much larger than any succeeding event within the channel system. This relationship is consistent with all other well-preserved channel systems. If groundwater needed to form later floods was not present during the oldest observed flood event, then recharging of groundwater reservoirs was needed for the younger floods. The older and larger floods may have wiped out evidence for any previous smaller flood events. However, it is not inconceivable that multiple and subsequently smaller flood source sites within an outflow channel represent progressive topping of a large groundwater reservoir or series of aquifers located at different elevations. If this was indeed the case, then a hydrological cycle may not have been required for each flood-source site within a given channel.

A final significant observation derived from image analysis is that the morphology of the circum-Chryse region is inconsistent with rainfall as a significant contributor of water to surface modification. Rainfall and its runoff set to intensely dissect the landscape into dense dendritic drainage networks. Nothing of this sort is observed in the circum-Chryse region. However, water transported to the surface of the study area in the form of snow cannot be ruled out.

Two major inferences are formulated from the observations. Water may have been repeatedly remelted to the near-surface reservoirs, as evidenced by the extended history of channeling in the region where subsequent channels tend to start interior to the sites of earlier flood-source sites. The discharge rates necessary to have carved the circum-Chryse outflow channels, which have been estimated on occasion to exceed 10⁶ m² s⁻¹,
Figure 2a. A sketch map of the circum-Chryse region illustrating the relative age groups assigned the outflow channels within this area.

Figure 2b. Plot of outflow channels as a function of age group along the vertical axis. Along the horizontal axis is plotted the source and terminus of each channel as a function of radial distance from the center of Chryse Planitia. This plot illustrates that outflow channels become younger with increasing proximity to the Chryse basin and that outflow channels closer to Chryse show a greater diversity in age.
and upwards to $10^7$ m$^3$ s$^{-1}$ [e.g., Baker, 1982; Baker et al., 1991], would have formed temporary standing bodies of water where floodwaters stagnated and pooled. The site of this pooling would have been Chryse basin, which is topographically enclosed and lowest along its boundary with the highlands to the south (U.S. Geological Survey, 1989). Such ponding in Chryse and elsewhere on the northern plains (and the consequences of such ponding) has been discussed by Lucchitta et al. [1986], Parker et al. [1989], Roto and Tanaka [1991], and Baker et al. [1992]. The possible fate of this water is discussed in the next section.

**Hypotheses**

These observations and inferences have led us to propose the following two alternative hypotheses, which will be referred to as the airborne precipitation cycling hypothesis, and the groundwater seep cycling hypothesis. Throughout this study we conservatively assume that the past Martian environment was similar to that of the present. Both hypotheses require some mechanism to mobilize the H$_2$O stored in the flood-source region aquifer, making possible its discharge as a new Chryse-bound flood. A good candidate mechanism is regional geothermal heat associated with the episodic upwelling of magma [Marsurshky et al., 1977; Baker, 1982]. Baker et al. [1991] proposed that catastrophic flooding was triggered by massive magnetic emplacement events, such as those associated with Tharsis volcanism. These events would have produced enormous hydrothermal systems capable of drawing in water from deep and adjacent supplies toward the thermal anomalies. Baker et al. [1991] explained that sudden cataclysmic outflows of groundwater were driven by hydrothermal convection, and likely assisted by rapid exfoliation of CO$_2$. In our study, we hypothesize that H$_2$O not
redistributed in the flood-source region was largely lost to the hydrologic cycle. This loss progressively lowered the vitality of the cycle, perhaps by now killing it.

The airborne precipitation cycling hypothesis postulates that large amounts of the $H_2O$ that sublimated off the temporary postflood lakes within the Chryse basin precipitated (snowed) onto the flood-source highlands. This scenario differs from the canonical view that the $H_2O$ went without pause to the polar sinks. In the airborne precipitation cycling hypothesis, strong northerly winds (which are inferred from modern wind streak orientations) transported the lake-derived $H_2O$ into the Chryse basin so that the underground flood-source reservoirs are recharged, making possible subsequent deluges. The high albedo and low thermal inertia of the "snow" relative to the lake ice would allow the "snow" to persist on the highlands much longer than the ice on the lake, thus providing the time needed to incorporate it into the subsurface. The decay of the heat pulse thought to have brought on the cycle-initiating flood progressively lowered the isotherms and moved the $H_2O$ down from the snowpack to the subsurface reservoir. Subsequent pulses of anomalous geothermal activity would initiate a new episode of flooding.

Alternatively, the groundwater seep cycling hypothesis holds that liquid lake water, under its ice cover, drained into a permeable subsurface layer. Concurrent erosion of the highlands/hollows boundary along the southern periphery of the lake by the floodwaters associated with a given flood could have exposed the $0°C$ isotherm, which may be only -1 km from the equilibrium surface interface (e.g., Rossbacher and Ingersoll, 1981). In order for water to be transported toward the flood-source sites, a permeable layer(s) must dip away from the basin so that the underground flood-source reservoirs are downdip from the Chryse lake sites (Figure 4). Schultz et al., (1982) proposed that Chryse basin was formed by a very large impact. Such an impact would have reoriented preexisting horizontal layers so that they would dip away from the impact site for many hundreds of kilometers beyond the basin rim. The dip of the reoriented stratigraphic layers would be steeper near the rim and become increasingly gentle with distance. Subsequent surficial erosion and modification of the circum-Chryse region would have created a regional topographic slope toward the basin while underground layers remained oppositely oriented. A possible indication of an antibasinward dip to the bedding layers composing the circum-Chryse highlands may be the unusual reverse (though very low) gradients of some of the outflow channels (e.g., Simud and Tiu Valles), reported by Lucchitta and Frohlich (1983).

The flow of groundwater can be confined to natural conduits and may travel large distances without much dispersion lateral to the direction of flow (e.g., White, 1969). As long as the water can travel in a permeable zone that lies beneath the $0°C$ isotherm and is sealed against vapor diffusion loss through the material overlying the permeable (phreatic) zone, then this water will move back under the highlands until either the slope of the subsurface layer flats or the water flow is intercepted and ponded by a subsurface obstruction, such as a fault plane. Large faults coincide with several flood-source chaotic terrains and probably played a role in the storage of significant quantities of groundwater. These faults have traces that are circumferential about Chryse and are probably the manifestation of basin tectonics (Schultz et al., 1982), and thus reveal the extent that the subsurface has been disturbed by the formation of Chryse basin. Episodic geothermal heat raised the water from depth up to the surface, where it would burst out to do new flooding, perhaps in the manner suggested by Marras et al. (1977). Simply stated, in this hypothesis, ponded floodwater in Chryse basin drained down an antibasinward dipping subsurface layer, ultimately reaching and recharging the flood-source aquifer.

**Discussion**

We will now discuss the evaluations conducted to determine which of these hypotheses can better explain the apparent recycling of floodwater in the circum-Chryse region. If the airborne precipitation cycling hypothesis is viable, then it may be generally applicable to other portions of Mars and serve as an explanation for other examples of water recycling.

---

**Figure 4.** Idealized and simplified hypothetical cross section of the Chryse region illustrating the concepts described in the groundwater seep hypothesis. The dip of strata underlying the circum-Chryse highlands away from the basin is a structural imprint of the impact that created Chryse basin. The subordinately shown fault is a product of basin ring tectonics. Note that the modern surface of circum-Chryse highlands slopes toward the basin as a consequence of erosion. Considerable vertical exaggeration is shown.
Lake Lifetime and Evaporative Loss Rate

Evaporation Rates

The water vapor flux $E_a$ from the lake's surface to the atmosphere is governed by the water-vapor density gradient within the atmospheric boundary layer (ABL). $E_a = -K_w \frac{\partial \rho}{\partial z}$; $K_w$ is the eddy diffusivity for water vapor. Integrating across the 'interfacial sublayer' to a point $z$ within the 'surface sublayer' (see e.g., Brutsaert [1982] for a description of the various sublayers within the ABL), (1) can be re-expressed in the following form:

$$E_a = \frac{\rho_0 F_0(z, z) - \rho(z) F(z, z)}{F_0(z, z)}$$

where $u^*$ is the 'friction velocity' and

$$F_0(z, z) = \int_{z}^{\infty} K_w dz.$$  

The resistance function $F_z$ embodies the dependence of the mass transfer rate on molecular gas properties and atmospheric stability. $F_z$ can be further broken into two components

$$F_z(z, z) = F_0(z, z) + F_{aDyn}(z, z)$$

corresponding to separate integrations across the interfacial and surface sublayers (first and second terms on the right-hand side, respectively); $L_a$ is the height of the interfacial sublayer.

Within the interfacial sublayer, heat and water vapor transfer involve nonsteady molecular diffusion into Kolmogorov-scale eddies that penetrate from the fully turbulent surface sublayer. Turbulent airflow within this zone can take one of two forms, 'aerodynamically smooth' or 'aerodynamically rough,' depending on the nature of the surface and the viscous length scale $(u^*/v)$, where $v$ is the kinematic viscosity of the $CO_2/H_2O$ gas mixture. For aerodynamically smooth flow, the height of the interfacial sublayer is

$$z_a = \frac{H_{co2}}{u^*}$$

for aerodynamically rough flow, where $z_a$ is known as the "surface roughness length" and $d_s$ is the zero-plane displacement, used to shift the coordinate system to the base of the roughness elements on the surface. Brutsaert [1975] effectively determined the function $F_a(z, z)$ for both flow regimes. In the present notation, he found

$$F_a(z, z) = \frac{1}{C_a^{10}}$$

for aerodynamically smooth flow and

$$F_a(z, z) = \frac{1}{C_a^{10}} \left[ \frac{z}{d_s} \right]$$

for aerodynamically rough flow. $d_s$ is the dimensionless Schmidt number, $K_w$ is the molecular diffusivity for water molecules through a $CO_2$ gas, $z_a = (z^2/d_s)$ is the "roughness Reynolds number," $C_a = 1/\sqrt{14}$, and $C_a = 1/14$. Within the surface sublayer, the eddy diffusivity can be described in terms of the dimensionless water-vapor gradient

$$k(z) = \frac{K_w(z)}{F_0(z, z)}$$

where $k$ is von Karmán's constant and $z$ is the Monin-Obukhov stability parameter. Because heat and water vapor are transferred through the surface sublayer by the same eddy processes, $F(z, z)$ is generally assumed to be equal to the dimensionless temperature gradient $F_w(z, z)$ for this layer. Several parameterizations exist for the $K_w(z)$, and hence for $F(z, z)$. Given $K_w$, the resistance function $K_w(z, z)$ can be determined.

We have evaluated the water vapor flux from a large Martian lake assuming a 7-mbar $CO_2$ atmosphere and temperatures in the range 170 to 240 K. Over this temperature range, the Schmidt number $S_c = 0.48-0.49$ [Glau and Haberle, 1990a] and the kinematic viscosity $v$ varies from $4.0 \times 10^{-6}$ to $9.0 \times 10^{-6}$ m$^2$ s$^{-1}$. The air mass blowing onto the lake is assumed to be perfectly dry (relative humidity $= 0\%$). For the dimensionless gradients $F(z, z)$ and $F_w(z, z)$, we utilize the Businger et al. [1971] parameterization. Because solar radiation is absorbed at depth within the ice, diurnal surface temperature variations are expected to be only a few degrees. Hence, we assume the atmospheric boundary layer is neutrally stable ($z = 0$) above the lake (this assumption would be invalid for a small lake whose dimension is comparable to the thickness of the ABL). The friction velocity was specified using the relation

$$u^* = \left( \frac{D_C}{U(z)} \right)^{1/2}$$

where $U(z)$ is the wind speed and $D_C$ is the momentum resistance function, analogous to $F_0(z, z)$. Since we have assumed neutral atmospheric stability, $F_0$ reduces simply to $F_w$.

$$F_w(z) = \frac{1}{K_w} \left[ \frac{z}{d_s} \right]$$

where the parameter $z_s$ is the momentum roughness length, equal to 0.135 times aerodynamically smooth flow and to the surface roughness length, $z_a$, for aerodynamically rough flow. The surface roughness length $z_a$ for a Martian lake is obviously unknown. We consider two values, 0.03 cm and 1 cm. Measurements over polar snow surfaces, which typically include sastrugi (wind-eroded ridges $< 5$ cm high on snow surfaces), yield roughness lengths between 0.003 and 1 cm [Insley, 1989]. The 0.03-cm value is taken to be representative of these surfaces, while the 1-cm value is used
to simulate a fairly rough surface. For the wind speeds and pressures considered in this paper, airflow within the interfacial sublayer is found to be aerodynamically smooth whenever \( z = 0.03 \) cm and aerodynamically rough when \( z = 1 \) cm. Free convection (Crow and Haberle, 1990b) is not found to occur even under the calmest conditions discussed in this paper.

Calculated water vapor fluxes are shown in Figure 5 as a function of surface temperature \( T_s \) and the wind speed at height of 1000 m above the surface. The solid lines are the results for a smooth surface \( (z = 0.03 \) cm) and the dashed lines are for a relatively rough surface \( (z = 1 \) cm).

### Lake Lifetimes

Once Lake Chryse formed, its lifetime would have depended on its depth and rate of water loss via sublimation from the surface and possibly from groundwater seepage (discussed in a later section). The maximum depth of the lake was limited to \(-2000\) m by the geometry of the Chryse basin, and the most likely initial depth was considerably less than this. As we have seen, the sublimation rate is very sensitive to the surface temperature of the ice, which in turn depends on its albedo.

Using present orbital parameters and solar luminosity, minimum-seasonal, mean-annual, and maximum-seasonal surface temperatures were calculated for ice at the latitude of Chryse (-25°N) by solving the surface energy balance,

\[
\frac{dT_s}{dt} = \frac{R}{\rho c_p} (1 - \alpha) S_0 - \frac{1}{\rho c_p} Q_{\text{cond}} - \frac{1}{\rho c_p} Q_{\text{evap}}
\]

where \( \alpha \) is the Stefan-Boltzman constant and \( \kappa \) is the thermal conductivity of ice; the sensible and latent heat fluxes over the lake would have been small in comparison to the three terms in (12) and hence were found to be negligible in the calculation of \( T_s \).

At equilibrium, the conductive heat flux in the ice immediately beneath the surface \((z = 0)\) is equivalent to the absorbed solar energy,

\[
\frac{dH}{dz} = (1 - \alpha) S_0
\]

where \( \alpha \) is the albedo and \( S_0 \) is the incident solar flux. Equilibrium can be assumed to hold for the mean-annual calculations and is approximately correct at the minimum and maximum points in the seasonal cycle. We have also assumed an atmospheric optical depth of 0.4 and have taken the downwelling infrared flux values, \( IR^2 \), from Haberle and Jakosky (1991).

Although the albedo of the lake ice is unknown, the mostly likely range is 0.3-0.6. The integrated albedo of glacial ice and thick multiyear ice is rarely higher than 0.6 unless covered with a fresh blanket of snow. For multiyear lake ice in Antarctica, McKay et al. (1994) measured albedo values of 0.3-0.4 before the seasonal melt period began. For the 0.3-0.6 albedo range, mean-annual surface temperatures are expected to be in the range of 200-220 K (Table 1). Seasonal temperature variations are expected to be less than ±10 K. Due to the nonlinearity of the water vapor flux with temperature, the use of the seasonal temperature cycle to calculate the annual water vapor flux will lead to an enhancement of 16-23% over that which would be determined using the mean-annual temperature.

Figure 6 shows the expected lifetime of a lake initially 1 km deep in Chryse basin as a function of albedo and wind speed. These results assume water is lost from the lake solely by sublimation and that no additional water is input to the lake during this time. The lifetime of a lake with a different initial depth can be found by simply scaling the values in Figure 6.

### Table 1. Minimum Seasonal, Mean-Annual, and Maximum Seasonal Surface Temperatures (°K) for Ice Using Present Orbital Parameters and Solar Luminosity

<table>
<thead>
<tr>
<th>Albedo</th>
<th>Latitude 25°N</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>210.6</td>
<td>220.6</td>
<td>227.3</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>205.1</td>
<td>214.4</td>
<td>220.7</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>192.7</td>
<td>200.3</td>
<td>205.4</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>185.3</td>
<td>191.9</td>
<td>196.3</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>177.4</td>
<td>182.3</td>
<td>185.8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Albedo</th>
<th>Latitude 0°N</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>215.3</td>
<td>225.4</td>
<td>232.5</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>209.5</td>
<td>218.9</td>
<td>225.6</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>204.0</td>
<td>204.0</td>
<td>205.6</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>198.5</td>
<td>195.1</td>
<td>199.9</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>191.7</td>
<td>184.8</td>
<td>188.6</td>
<td></td>
</tr>
</tbody>
</table>
For the most likely range of albedo (0.3-0.6) and wind speed (<10 m s⁻¹), a 100-m-deep lake is expected to last 10-160 kyr before completely subliming away, while a 1000-m-deep lake would last 96-1610 kyr. We note that if the lake formed early in Martian history when the solar luminosity was perhaps only 70% of present, surface temperatures would have been roughly 10 K less (Table 2). The corresponding water vapor fluxes would have been reduced by a factor of 5, extending the lifetime of the lake by the same factor.

It is interesting to estimate how long liquid water might have existed in such a lake. Once a perennially ice-covered lake achieves equilibrium, its ice thickness is given by

\[ Z_e = \frac{b \ln(T_0/T_f) + c(T_e - T_0) - A \ln(1 - \exp(-b/h))}{F_p - F,} \]

where \( T_e \) is the melting point of water, \( h \) is the extinction pathlength for solar radiation in ice, \( F_p \equiv \rho \nu L_f \) is the latent heat flux due to freezing at the base of the ice cover, \( \rho \) is the ice density, \( \nu \) is the rate of ice formation, \( L_f \) is the latent heat of fusion, \( F_p \) is the geothermal heat flux, and the constants \( b \) and \( c \) are 780 W m⁻² and 0.615 W m⁻² K⁻¹, respectively [McKay et al., 1985].

Once equilibrium is reached, the rate of ice formation at the base of the ice cover equals the mass loss from the surface due to sublimation, \( \nu = E \rho \nu L_f / \sqrt{T_0} \) (15). Figure 6 shows the total amount of water that has been converted to ice at the base of the ice cover.

The Airborne Precipitation Cycling Hypothesis

In this section we consider how moisture from a large body of ice-covered water could provide a source of meteoric water for an accumulation region at high elevation. The system we

<table>
<thead>
<tr>
<th>Albedo</th>
<th>Latitude 25°N</th>
<th></th>
<th></th>
<th></th>
<th>Latitude 0°N</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>198.6</td>
<td>207.1</td>
<td>212.7</td>
<td>202.6</td>
<td>211.1</td>
<td>217.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>194.1</td>
<td>201.8</td>
<td>207.1</td>
<td>197.8</td>
<td>205.6</td>
<td>211.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>189.2</td>
<td>196.2</td>
<td>201.0</td>
<td>192.5</td>
<td>199.7</td>
<td>204.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>184.0</td>
<td>190.1</td>
<td>194.3</td>
<td>186.9</td>
<td>193.2</td>
<td>197.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>178.2</td>
<td>183.3</td>
<td>186.9</td>
<td>180.6</td>
<td>185.9</td>
<td>189.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>171.9</td>
<td>175.7</td>
<td>178.4</td>
<td>175.7</td>
<td>177.6</td>
<td>180.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
are considered is shown schematically in Figure 8. A large body of ice-covered water is assumed to exist at the lower temperature, elevation. Air moving over this water reaches 100% humidity and is assumed to be moistened air then moves upslope and sublimes. The sublimation rate is approximately exponential with time as shown in Table 3. The moistened air then moves upslope by some height \( z = Z \). The moistened air then moves upslope and reheats as it moves up to the accumulation zone. The key controlling variable in this model is the rate of sublimation from the ice-covered field. The key controlling variable in our model is the rate of sublimation from the snow field. The key controlling variable in our model is the rate of sublimation from the snow field.

We now consider the trajectory of a parcel of air beginning at the surface of the lake and moving up to the accumulation zone. The rate of sublimation from the ice-covered surface of the lake is determined solely by sublimation from the lake. The rate of sublimation from the ice-covered surface of the lake is determined solely by sublimation from the lake. The rate of sublimation from the ice-covered surface of the lake is determined solely by sublimation from the lake.

We assume that the air initially contains negligible water and that the flow is sufficient to carry all the water that sublimes from the lake surface into the atmosphere, i.e., the flux of water is determined solely by sublimation from the lake. The moistened air then moves upslope by some height \( z = Z \) and cools by adiabatic expansion. Only when the air parcel reaches the upper layer does the moisture in excess of the saturation at the lower temperature condense onto the surface. This condensation represents the source of \( H_2O \) in the accumulation region. We also consider sublimation in the upslope area as a loss of \( H_2O \) from the accumulation zone.

In this model we assume that winds with an initial relative humidity of 0% move across a frozen lake at altitude \( z = 0 \). Water ice sublimes, and a moist parcel with a relative humidity of 100% forms. It is assumed that this parcel of mass ascends a height \( Z \). With the ascent, \( T_C \), the temperature decreases which causes a fraction \( P \) of the parcel to condense. Note that for the purpose of this evaluation, we assume that all of this fraction of condensate is deposited as snow in the highlands at altitude \( Z \). In addition to the evaporation from the lake, we assume evaporation from the snow field in the highlands at altitude \( Z \). The rate \( R \) of snow accumulation in the highland snow field is a function of (1) the rate of production of the moist parcel of mass, \( z \), \( \Delta \) the amount of loss of this mass due to the ascent of the parcel, and (3) the rate of sublimation from the snow field. The key controlling variable in our model is the rate of sublimation from the ice-covered surface of the lake. The evaporation rate is computed as a function of wind speed and temperature for a 7-mbar atmosphere of \( CO_2 \) using the method of Clow et al. [1988]. In these calculations we have assumed that the planetary boundary layer is isothermal and that the air is completely dry at \( z = 100 \) m. It can be seen from these curves that the sublimation rate is approximately exponential with temperature and thus dominated by the high-temperature period:

\[
E_n = E_0 \exp \left( \frac{\Delta T}{T} \right)
\]

where \( \Delta T = T - T_0 \) for respective wind speeds. We chose \( \Delta T = 6.56 \) K, reflecting a nominal wind speed of 10 m s\(^{-1} \) for our initial calculation.

The global mean-annual temperature is used to characterize the moisture content of the parcel, assuming a simple sinusoidal variation in temperature during a Martian year (temperatures were derived from the model of Kieffer et al. [1977]). There is an enhanced period of evaporation due to peak temperatures greater than the mean-annual temperature.

For simplification, we treat the temperature \( T \) as equal to the mean-annual temperature plus a sinusoidal variation about the mean that has an amplitude of \( \Delta T \). It can be shown that for a sinusoidal variation in temperature for which the amplitude \( \Delta T \) is small, the average sublimation over the period can be approximated by:

\[
\frac{1}{P} \int_0^P E_n dT = \frac{E_0}{P} \int_0^P \left[ \frac{\exp \left( \frac{\Delta T}{T} \right)}{1 + \frac{\exp \left( \frac{\Delta T}{T} \right)}{2}} \right] dT
\]

\[
\Delta T = [1 + \frac{1}{2} \exp \left( \frac{\Delta T}{T} \right)] - 1
\]

where \( P \) is the period, representing 1 year. The term \( [1 + \frac{1}{2} \exp \left( \frac{\Delta T}{T} \right)] \) represents the enhancement over the mean sublimation due to the nonlinear nature of the exponential function.

We now consider the trajectory of a parcel of air beginning at the surface of the lake and moving up to the accumulation zone. At the lake surface the amount of water in the parcel, \( w_p \), is determined by the relative humidity, \( RH_p \) (density of the parcel), \( P_{H_2O} \) times the saturation vapor pressure of water vapor, \( P_{H_2O} \). Over the temperature range considered here (190-200 K), we can approximate \( P_{H_2O} \approx P_{H_2O \, sat} \) where \( P_{H_2O \, sat} \approx 7.215 \) K and \( P_{H_2O} \approx 3.8 \times 10^{-16} \) mbar [Marti and Mauersberger, 1993]. Similarly, the amount of water in the parcel once it has reached the accumulation zone, \( w_p \), is determined by the relative humidity, \( RH \), times the saturation vapor pressure and the temperature there.

The fraction of water condensing on the highlands is defined by \( F = (w_p - w_0) / w_p \), assuming that the relative humidity above the lake and the highlands are the same. Therefore,

\[
F = \frac{RH_p P_{H_2O \, sat}}{P_{H_2O \, sat}} \left[ 1 - \exp \left( \frac{\Delta T}{T} \right) \right] - \frac{RH P_{H_2O \, sat}}{P_{H_2O \, sat}} \left[ 1 - \exp \left( \frac{\Delta T}{T} \right) \right]
\]

where the latent heat, \( \Gamma = \Delta T \Delta \lambda C \), and the change in temperature from the upper elevation to the lower elevation is given by \( \Delta T = T_e - T_p \). To a good approximation, the amount of vapor that condenses during a nominal ascent is not
Table 4. Times for Complete Freeze-Up (Kiloyears)

<table>
<thead>
<tr>
<th>Albedo</th>
<th>Initial Water Depth, m</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td></td>
<td>2.2</td>
<td>6.5</td>
<td>34.7</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>1.0</td>
<td>4.1</td>
<td>16.3</td>
<td>66.0</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>0.67</td>
<td>2.6</td>
<td>10.7</td>
<td>42.9</td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td>0.48</td>
<td>1.9</td>
<td>7.7</td>
<td>36.9</td>
</tr>
</tbody>
</table>

sufficient to alter the moist adiabat very much from the dry adiabat. Thus

\[ T_1 - T_0 = \Delta T = T_s - T_0 = T_s - T_f = T_{scarp} - T_s \]  

(20)

where \( T_1 \) is the temperature of the parcel at the scarp top (i.e., where it snows), \( T_s \) is the temperature at the lake surface, \( T_f = -g/\rho C_p \), and \( \Delta T \) represents the height of the ascent. On Mars, \( g = 3.73 \text{ cm s}^{-2} \) and \( C_p = 8.3 \times 10^7 \text{ cm}^2 \text{ K}^{-1} \text{ mg}^{-1} \), implying that

\[ \Delta T = \frac{-g}{\rho C_p} = 4.45 \times 10^3 \text{ K cm} \text{ kg}^{-1} \].

Therefore, \( \Delta T \) is large.

As previously mentioned, the sublimation from the lake that takes place at mean-annual temperatures, \( E(T) \), is enhanced due to peak temperatures, \( \Delta T_{max} \) greater than mean-annual temperature during the summer time. Thus,

\[ \text{Production} = F \times E(T) \left[ 1 + \frac{1}{3} \left( \frac{\Delta T_{max}}{\alpha} \right) \right] \]  

(21)

Substituting for \( F \) gives

\[ \text{Production} = \left[ 1 + \frac{1}{3} \left( \frac{\Delta T_{max}}{\alpha} \right) \right] \left( \frac{\rho C_p}{\alpha} \right) \left[ \frac{g}{\rho C_p} \right] \]  

(22)

The area loss rate of sublimation from the snow accumulation zone is less than the loss rate from the lake surface because of the decrease in temperature with the increase in altitude. Thus,

\[ \text{Loss} = \frac{E(T)}{a} \left[ 1 + \frac{1}{3} \left( \frac{\Delta T_{max}}{\alpha} \right) \right] \]  

(23)

where \( \Delta T_{max} \) represents the increase in temperature due to peak temperatures at the snowfield during the summer. The net accumulation of snow in the highland snowfield is the difference between the production rate and the loss rate.

We now apply this theoretical framework to a specific case. We assume values that characterize the Chryse basin region of present Mars. For this region, the current annual average temperature is 208 K and the peak summertime temperature is 218 K. The scarp top is 2 km higher than the lake surface. We calculate that the fraction (\( F \)) of \( \text{H}_2\text{O} \) that condenses over the

highlands from the original amount in the 100% RH parcel just above the lake in Chryse is 0.39. We compared the area of the lake to the "equilibrium" area of the highlands snowfield. The snowfield's "equilibrium" area is defined by the area the snow would have to be spread so that its subsequent sublimation rate would exactly match its precipitation rate. The calculated ratio of "equilibrium" highlands snowfield area to Chryse lake area is 3.7, a value which is independent of temperature.

The Groundwater Seep Cycling Hypothesis

Several phenomena are integral to the groundwater seep cycling hypothesis. For this mechanism to work, there must be a permeable layer(s) along which water can travel back toward the flood-source sites. Assuming that antibasinward dipping subsurface layers exist beneath the Chryse highlands region, we evaluate the ability of water, beneath an ice-covered flood lake in Chryse Planitia, to seep into these layers.

To make this evaluation, we constructed a simple model (Figure 9). We began with a lake containing \( 2 \times 10^7 \text{ m}^3 \) liquid water beneath its ice cover. We assumed the liquid portion of the lake is shaped like a circular disk with vertical sides, 200 km in diameter and initially 200 m deep. The lake is allowed to drain through a time-varvbine cross-sectional area \( A_{tr} \) which is less than one-quarter of its circumference (31 \( \times \) \( 10^6 \) m) times the depth of the lake \( d \), which varies with time (Figure 9).

\[ A(t) = \frac{2\pi d^2}{4} \]  

(24)

We assume that the source region for the floodwaters is located 1000 km away from the perimeter of the lake and lies 1 km deeper than the lake's base. The time required to drain the lake completely was calculated for several values of average subsurface permeability. Megaregolith, which is thought to comprise much of the subsurface of Mars, has a permeability estimated by Carr [1979] of approximately 3000 darcies. In comparison, terrestrial basaltic typically have permeabilities ranging from \( 10^{-1} \) to \( 10^2 \) darcies [Davis and DeWiest, 1966].

Table 5. Lake Lifetimes (kiloyears) for \( U = 10 \text{ m s}^{-1} \)

<table>
<thead>
<tr>
<th>Albedo</th>
<th>Initial Water Depth, m</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td></td>
<td>9.5</td>
<td>19.1</td>
<td>38.2</td>
<td>76.4</td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>21.8</td>
<td>43.7</td>
<td>87.4</td>
<td>175</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>34.4</td>
<td>109</td>
<td>218</td>
<td>435</td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td>161</td>
<td>322</td>
<td>643</td>
<td>1290</td>
</tr>
</tbody>
</table>
To determine the total time it takes to drain the lake, we need to calculate the discharge as a function of time \( Q(t) \) through the permeable zone. The integrated discharge is equal to the volume of the lake \( V_{\text{lmax}} \)

\[
\int_0^{V_{\text{lmax}}} Q(t) \, dt = V_{\text{lmax}}
\]

(25)

where

\[
Q(t) = A(t)q(t)
\]

(26)

and

\[
g(t) = -\frac{k}{\mu} \frac{d\Delta H(t)}{L} - h
\]

(27)

The average velocity of the water \( q \) through the permeable zone varies with time because the difference in the hydraulic head decreases as the lake drains. In this equation, which is a form of Darcy's law, \( k \) is the average permeability of the subsurface, \( \rho \) is the water density, \( g \) is the gravitational acceleration for Mars, \( \mu \) is the viscosity of the water, and \( L \) is the total distance the water travels through the permeable zone. We numerically integrated the lake volume to solve for the time it takes to completely drain the lake \( t_{\text{drain}} \) having \( \Delta H(0) = 1.2 \text{ km} \). The change in the hydraulic head as a function of time \( \Delta H(t) \) is given by

\[
\Delta H(t) = h_{\text{source}} - d - h
\]

(28)

where \( h_{\text{source}} \) is the elevation of the lake bed above the source-site deep storage region \( (1 \text{ km}) \) and \( d \) is the lake depth which varies with time. The hydraulic head at the source \( h_s \) is set equal to zero because water is not allowed to pool. If the initial change in hydraulic head is \( 1.2 \text{ km} \) over a distance of \( 1000 \text{ km} \) and if we assume a permeability of \( 1000 \text{ darcies} \), then the lake will completely drain in approximately \( 3300 \text{ years} \). This calculation takes into account the fact that both the lake volume and the wetted surface area of the permeable zone

Figure 8. Schematic diagram illustrating the parameters used for the lake and the uplands condensation region in the airborne hypothesis model.

Figure 9. Illustration of the groundwater seep model. Lake is assumed to be disk-shaped with an initial depth equal to \( 200 \text{ m} \). The cross-sectional area, through which the lake is allowed to drain is equal to one-quarter the circumference of the disk times the depth of the lake at time \( t \). All lake water drains through the lake margin and into the permeable aquifer for a distance equal to \( 1000 \text{ km} \) beyond the lake's edge. An impermeable zone underlies the lake and aquifer. The area subsurface storage at the floodsource site lies \( 1 \text{ km} \) deeper than the base of the lake and is assumed to be much larger than the lake. Therefore, we assume that the hydraulic head at the flood source does not change.
decrease with time, so initially draining is quite rapid. Half of the water will drain into the subsurface in 923 years, and 1/3 of the water will be removed by 1717 years. The time required to drain the lake is inversely proportional to the average subsurface permeability. Therefore, if the average subsurface permeability is actually 100 darcies then it will take an order of magnitude longer to completely drain the lake.

Ultimately, the time constraint for removal of water by seepage is the lifetime of the lake as defined by the rate its ice cover thickens until the lake is frozen solid and by the sublimation rate of H$_2$O from the top of the lake's surface.

Another issue to address is how long does it take for the entire lake volume to reach the source region of the floodwaters. This can easily be calculated by dividing the total distance $L$ the water has to travel by the velocity of the water through the permeable zone $q$ at $t_{final}$ when the lake has fully drained. Because $\Delta t$ within the permeable zone is equal to the elevation of the lake bed at $t_{final}$ we can use equation (27) to calculate $q$ at $t_{final}$. Assuming $k = 10^6$ darcies, the entire volume of water contained in the lake will have reached the source region in 8466 years (Table 6).

We can now calculate the time required for a column of liquid to be vertically transported to the surface, where it is again available for discharge. The transport mechanism is no longer liquid flow, but vertical vapor diffusion from the surface of the deep ponded water to the cold, near-surface regolith. We assume an initially dry regolith column, bounded at the top by the surface, and at the bottom by standing water. The surface temperature is assumed to be the temperature at the annually averaged insolation, calculated from equation (12). We assume that the surface albedo is 0.214, and the lower boundary condition is

$$\frac{dt}{dx} = \frac{Q}{K}$$

where $Q$ is the heat flow and $K$ is the thermal conductivity of the regolith at the bottom of the column. We assume a background heat flow of 0.03 J m$^{-2}$ s$^{-1}$ (Piuke et al., 1982) and a thermal conductivity of 0.6 J m$^{-1}$ K$^{-1}$ [see Clifford, 1993]. As long as the thermal gradient is steeper than 0.02 K m$^{-1}$, the 0°C melting isotherm will lie below the water level (assumed to be at a depth of 3 km). The vapor pressure at the base of the column is the vapor pressure of water at the appropriate temperature. The upper boundary condition is assumed to be the atmospheric H$_2$O abundance, although in practice, the gradient is soon controlled by the precipitation vertically throughout the column. The vertical flux of H$_2$O through the column to a near-surface cold-trap is controlled by

$$q_v = \frac{\rho D_{H_2O} \mu \rho g \Delta T}{P \rho \sigma \Delta T} \frac{dT}{dx}$$

where $D_{H_2O}$ is the effective diffusion coefficient of H$_2$O in CO$_2$, $P_{sat}$ is the saturation vapor pressure of H$_2$O at temperature $T$ of the regolith, $L$ is the latent heat of vaporization, and $\bar{S}$ is a dimensionless empirical factor which has the value of 1.83 ( Jury and Letey, 1979). The term $\rho$ is the density of liquid water, and $R_g$ is the gas constant.

The calculation of the total H$_2$O deposited above the 0°C isotherm is simplified because the regolith is saturated throughout the column. In roughly the upper kilometer, the H$_2$O precipitates as ice. At greater depths, water can precipitate as a liquid if the confining pressure is high enough, and will ultimately drain back down to the water table. From equation (30), we can calculate directly the flux of water passing through the 0°C isotherm at $\sim 3 \times 10^{-2}$ kg m$^{-2}$ s$^{-1}$. Such a flux can transport $\sim 100$ column-meters of ice to the near-surface cryosphere in 10$^9$ years. If a subsequent heat pulse raises the 0°C isotherm into an ice-charged cryosphere, melting can take place, forming and burrowing a charged, confined aquifer. For confined aquifers we invoke perched, near-surface groundwater resting on aquatards, which while permeable to the upward infiltration of H$_2$O vapor over a 10$^9$-year period, are insufficiently impermeable to liquid water so that 'rapid' melting and downslope drainage at the surface depletes the aquifers much more quickly than would vertically downward seepage. We conclude that given the circumstances invoked to initiate a new episode of flooding, ordinary thermal vapor diffusion is adequate to sustain the system.

**Conclusions**

Our analysis of the airborne precipitation cycling hypothesis indicates that this mechanism seems unlikely to effectively operate under current conditions on Mars. Our modeling was idealized to show the most robust case for this hypothesis. We calculate that less than half (~40%) of the H$_2$O carried from Chryse can condense over the highlands. Another result of our analysis is that snowfields with areas larger than 3.7 times the lake area would sublimate faster than they accumulate. It is possible that local topography might cause local "heavy" precipitation within a confined area. The most likely place for this to occur in the Chryse region is directly over the lowland/highlands scarp. Unfortunately, precipitation at the scarp would dump snow back on the southern edge of the lake, leaving very little to travel on to the flood-source sites. The particle size range of snow in the Martian atmosphere varies at least by roughly a factor of 30, from ~64 pm to ~12 pm [Kahn, 1990]. Small variations in particle size result in very different settling rates. Murphy et al. [1990] calculated that it would take ~28 hours for a 10-um particle to fall from...
km above the surface, and ~119 hours for a 1-μm particle to fall from the same height. Parcels of air in which precipitation is taking place traveling as slow as 1 m s⁻¹ would cover a distance of ~400 km in the time it took particles between 1 and 10 μm to fall 1 km to the ground. Assuming that the parcel initially contains all the H₂O that is collected over the lake, that all the condensate forms snow particles that are within 1-10 μm, and that there is no lateral spreading or upward mixing of the parcel, precipitation would take place over an area approximately equal to our model "equilibrium" snowfield. Any change toward more realistic conditions would make retention of snow on the highlands unlikely. Moreover, if a parcel of air with sufficient moisture would simply stay suspended until the air parcel is mixed with upward mixing of the parcel, precipitation would take place over an area approximately equal to our model "equilibrium" snowfield. Parcels of air in which original mass of an ice-covered cylindrical lake (albedo 0.5, 1 km deep, 100-km radius, draining along its rim for one quarter of its circumference into substrata with a permeability of 10⁻¹¹ m s⁻¹) can be modeled to have moved underground before the competing mechanisms of sublimation and freeze-down choked off further water removal. Another encouraging result of our modeling is that floodwater will travel distances equal to the separation between Chryse basin and floodwater-source basins. Certainly of the two hypotheses we formulated and evaluated, does condense over the highlands actually collects on the ground as snow. It is very likely that much, maybe most, of the condensates formed under present Martian conditions would simply stay suspended until the air parcel is mixed with enough dry gas to cause their sublimation before settling to the ground [Kahn, 1990; Michelangelo et al., 1993]. Given all these normally unfavorable conditions, it is difficult to imagine that any more than a few percent of the H₂O leaving a lake in Chryse could collect and remain in the circum-Chryse highlands. Even if several percent H₂O were incorporated into an aquifer, such efficiency would only marginally qualify as a "cycle.

The results of our analysis of the groundwater seep cycling hypothesis are more promising. Optimally, ~1/4 of the original mass of an ice-covered cylindrical lake (albedo 0.5, 1 km deep, 100-km radius, draining along its rim for one quarter of its circumference into substrata with a permeability of 10⁻¹¹ m s⁻¹) can be modeled to have moved underground before the competing mechanisms of sublimation and freeze-down choked off further water removal. Another encouraging result of our modeling is that floodwater will travel distances equal to the separation between Chryse basin and floodwater-source sites in less than 10⁶ years for our "worst case" (k = 10⁻¹¹ darcies) and 10⁴ years for "best case" (k = 10⁻¹⁰ darcies) permeabilities. Like the evaluation of the airborne precipitation cycling hypothesis, our modeling of groundwater seep cycling is idealized to make the most robust cases. Considerations of shallow lakes and relatively impermeable strata will prevent significant amounts of H₂O from returning by this route to its source sites. Also, the hypothesis is absolutely dependent on the reality of the structural situation we have proposed. For the groundwater seep hypothesis to be viable, the water under its ice cover must have extensive access to a ground interface that is 0°C. This is because subfreezing ground has sufficient thermal inertia to seal itself by freezing the H₂O that seeps into its pores. Something on the order of a kilometer of material must be removed to uncover the 0°C isotherm. We propose that access to warmer ground could be accomplished by backwaxing of the highlands/lowlands boundary scarp and gouging of the basin floor in and around where the flood disgorges into Chryse basin. Certainly extensive examples of this type of erosion are observed in association with the Chryse outflow channels [e.g., Baker, 1982; Tanaka and Chapman, 1992].

If the dip of the strata in the circum-Chryse region is as we have hypothesized, and those layers are reasonably permeable and accessible, then perhaps the majority of the water in a Chryse flood lake could have returned to the highlands reservoirs and participated in subsequent flood events. Certainty of the two hypotheses we formulated and evaluated, the groundwater seep cycle seems by far the more viable. Of course, there is the possibility that there has never been a floodwater hydrologic cycle operating in the Chryse region. While observations of landform assemblages seem to imply that such a cycle has existed, they do not absolutely mandate this explanation. Further observations may permit the determination of which mechanism, if either or any, may have operated to recycle the Chryse floodwaters. A principal conclusion of this study is that the supposition of water recycling, while not conclusively demonstrated, is at least reasonable, even under the stringent conditions of the modern Martian environment.

We have identified several tests that hopefully could be conducted with data from forthcoming missions, such as Mars 96 and the Mars Global Surveyor (MGS) spacecraft. First and foremost, for the groundwater seep cycling hypothesis to gain greater acceptance, substantiating evidence for the existence and nature of the putative permeable, antithrusting dipping, subsurface layer must be obtained. The MGS Mars orbiter laser altimeter (MOLA) can provide the topographic information to determine the gradients of the outflow channels and relative elevations of sites where dipping stratigraphic marker horizons might outcrop. The Mars orbiter camera (MOC) and thermal emission spectrometer, both which fly on the MGS, could be used to recognize such outcrops where they might occur at widely scattered locations. High-resolution images of the highlands/lowlands scarp where it borders the Chryse basin acquired from MOC and the Mars 96 camera (e.g., MOLA) could also give a much better determination of the size, shape, and depth of Chryse basin, which bears on the location, areal extent, and depth of the postflood lakes.

Acknowledgments. We wish to express our gratitude to Vic Baker and Tim Parker for their thorough reviews of our manuscript. We would like to thank Bob Haberle, Jim Murphy, and Ray Reynolds for their valuable comments and insight to various issues regarding this study. This research was conducted while J. M. Moore and V. C. Gulick were Associate Research Fellows of the National Research Council. D. R. Baker was supported under NASA grant NAGW-1; G. D. Clow, W. L. Davis, and C. P. McKay were funded under NASA grant RTOP (54-10-80-48); C. B. Baker under NASA grant RTOP (54-10-80-45); and A. P. Zent was supported by NASA grant RTOP 151-01-00-11.

References

Page intentionally left blank
A search for nitrates in Martian meteorites

Monica M. Grady
Department of Mineralogy, The Natural History Museum, London, England, United Kingdom

I. P. Wright and C. T. Pillinger
Department of Earth Sciences, The Open University, Milton Keynes, England, United Kingdom

Abstract. Martian atmospheric nitrogen is highly enriched in \(^{15}\)N; nitrates formed by interaction of the atmosphere with the Martian regolith should therefore also be characterized by an elevated \(^{15}\)N value. A search has been made for nitrates in two Martian meteorites, in order to determine the extent of possible regolith-atmosphere interaction. Shock-produced glass from the Elephant Moraine (EET) A79001 shergottite (E1,149) and a water-soluble extract from Nakhlia were analyzed by Fourier transform infrared (FTIR) spectroscopy and stepped combustion-stable isotope mass spectrometry. FTIR of both meteorites had features at 1375 cm\(^{-1}\) and 1630 cm\(^{-1}\), consistent with nitrates. On account of their low thermal stability, nitrates break down at temperatures below 600°C; in this temperature range, E1,149 yielded \(-1250 \text{ ppb nitrogen with } ^{15}\text{N} - 8\%\). If this nitrogen is from a nitrates, then it cannot be distinguished from terrestrial salts by its isotopic composition. The water-soluble extract from Nakhlia also released nitrogen at low temperatures, approximately 17 ppb with \(^{15}\text{N} - 11\%\). Since Nakhlia is an observed "full," this is unlikely to be a terrestrial weathering product. Nitrates apparently occur in E1,149 and Nakhlia, but in very low abundance, and their origin is unclear. The isotopic composition of the salts, which is within the range of that proposed for Martian magmatic volatiles, is far removed from that of nitrogen in the present-day Martian atmosphere. If the nitrates are Martian in origin, they did not form in recent times from reactions involving atmospheric gases. Rather, the nitrates could be the result of an earlier episode of atmospheric interaction with the regolith, or with implantation of magmatic volatiles introduced during degassing.

Introduction and Summary

The surface of Mars has had an active history: images taken by Mariner 9 and the Viking orbiters testify to episodes of fluvial and aeolian activity, volcanism and cratering. A chronology can be established by relating craters on Mars with events recorded on the lunar surface (e.g., Tanaka, 1986). Timing of surface features in relation to the history of fluvial activity has implications for Martian evolution: liquid water was presumably an important palaeoatmospheric pressure, there is an increased probability of salt deposition and aeolian activity, volcanism and cratering. A chronology can be established by relating craters on Mars with events recorded on the lunar surface (e.g., Tanaka, 1986). On the basis of Fourier transform infrared (FTIR) spectroscopy the petrology of this is some evidence for nitrates in the glass sample. However, FTIR does not distinguish between discrete nitrate minerals, or NO\(_3\) dissolved in the glass (either from trapped Martian atmospheric gases or original magmatic volatiles). E1,149 was therefore also analyzed by stepped combustion; a large portion of the total nitrogen in the sample (1250 ppb) was found to be released at low temperatures, consistent with the breakdown of nitrate minerals. The best estimate of the \(^{15}\text{N} value of this species is approximately 17 ppb nitrogen as nitrates with \(^{15}\text{N} - 11\%\). Assuming that the low-temperature material is from Martian nitrates, the \(^{15}\text{N} data are inconsistent with formation by direct incorporation of present-day Martian atmospheric species (\(N_2\) or NO\(_3\) into the regolith. Any mineral nitrates present in SNC meteorites may therefore have resulted from the redeposition of degassed magmatic species within the Martian crust. The mechanisms of salt formation in Martian meteorites are far from well understood. The carbonate to nitrate ratio of the two meteorites ranges from \(\sim 10\) to 2000. The model of Yang et al. (1977) pre-

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 100, NO. E3, PAGES 5469-5475, MARCH 25, 1995

Copyright 1995 by the American Geophysical Union.

Paper number 94JE02803.
0148-0227/95$05.00

I. P. Wright and C. T. Pillinger
Department of Earth Sciences, The Open University, Milton Keynes, England, United Kingdom
dicts a layer of nitrates 0.3 cm thick over the surface of Mars, from which the presence of carbonate deposits 1 to 600 cm thick might be inferred. This calculation is in agreement with the prediction by Keue (1985) that up to 280 cm of carbonates are stored in the Martian regolith.

Background

Nitrogen, currently present at a level of ~2.7% is the second most abundant constituent of the Martian atmosphere, after CO2. The nitrogen isotopic composition of the present-day Martian atmosphere has been measured directly, by the Viking landers, and indirectly, by analysis of species trapped within pockets of shock-produced glass (lithology C) in the Martian meteorite EET A79001. Nitrogen isotopic composition of the Martian atmosphere has been measured directly, by analysis of species trapped within pockets of shock-produced glass (lithology C) in the Martian meteorite EET A79001. Nitrogen isotopic composition of the present-day Martian atmosphere has been measured directly, by analysis of species trapped within pockets of shock-produced glass (lithology C) in the Martian meteorite EET A79001. Nitrogen isotopic composition of the Martian atmosphere has been measured directly, by analysis of species trapped within pockets of shock-produced glass (lithology C) in the Martian meteorite EET A79001. Nitrogen isotopic composition of the Martian atmosphere has been measured directly, by analysis of species trapped within pockets of shock-produced glass (lithology C) in the Martian meteorite EET A79001.

Both measurements have found the atmosphere to be enriched in $^{15}$N, with $^{15}$N calculated as approximately $+600\%_\text{oo}$ from Viking results (Nye et al., 1976) and about $+320\%_\text{oo}$ from the meteorite (Becker and Pepin, 1984; Wone, 1988); the meteorite data are a lower limit, due to admixture of isotopically lighter indigenous nitrogen. It is widely held that the enrichment in $^{15}$N is a direct consequence of gravitational separation within the atmosphere, coupled with loss of nitrogen atoms from above the exobase (e.g., McElroy et al., 1977). There are several possible mechanisms by which nitrogen might be removed (e.g., Pepin, 1991); however, the present-day level of $^{15}$N enrichment is less than anticipated (Wallin, 1989), as effect which has been variously explained as the result of continued outgassing of nitrogen from the interior acting to dilute the enrichment in the atmosphere (Wallin, 1989), or the existence of an early, dense atmosphere which suppressed the operation of loss processes for the first part of Martian history (e.g., Fox, 1993). Other possibilities include addition of nitrogen from meteoroid or comet impacts, thermal release of regolith volatiles, or breakdown of surficial salts.

Nitrates in the regolith might also provide an important sink for atmospheric nitrogen (e.g., Fanale, 1976). If nitrates do occur on the Martian surface, and their $^{15}$N/$^{14}$N ratio can be determined, an important constraint on the operation of a "nitrogen cycle" on Mars would be provided. Model calculations have predicted that, via the formation of NO, HNO3 and HNO2 in the lower layers of the Martian atmosphere, the regolith might trap nitrite and nitrate anions, leading to the buildup of salts. Integrated over 4.5 x 10^7 years, such a mechanism would contribute the equivalent of a "layer" of nitrates approximately 0.3 cm thick across the whole of the Martian surface (e.g., Wang et al., 1977). This "layer" would be even thicker if the nitrogen content of the atmosphere had been higher in the past.

For terrestrial geochemical studies, it is possible to deduce past environmental processes by analyzing samples of the Earth's atmosphere, in conjunction with precipitated salts and magmatic volatiles. Until material is recovered directly from Mars' surface by a sample return mission, the only way in which an analogous study of the planet might be undertaken is by analyzing the constituents of SNC meteorites, a group of rocks which almost certainly originate from Mars (see McSween [1994] for a review). It is believed that during their history the meteorites resisted sufficiently close to the surface of Mars to have been affected by secondary alteration (e.g., Gooding, 1992). Furthermore, the impact events which removed the meteorites from the Martian surface have, in some cases, caused atmospheric gases to become trapped within the ejecta (e.g., Bogard and Johnson, 1983; Becker and Pepin, 1984; Carr et al., 1985). Wright et al. [1990] deduced a "carbon cycle" for Mars on the basis of the abundance and isotopic composition of carbon-bearing components in SNC meteorites. This paper is an attempt to produce a similar relationship for nitrogen on Mars. An impression of the possible complexities associated with the interactions of nitrogen-bearing components at the Martian surface is shown in Figure 1.

The nitrogen-bearing components that might occur in Martian meteorites include trapped terrestrial atmospheric gas, organic material (i.e., terrestrial contamination), inorganic weathering products (of terrestrial or Martian origin), magmatic nitrogen, and

![Figure 1](image-url)
Martian atmospheric gas trapped in impact glass. Nitrates are unstable, decomposing at low temperatures and readily soluble in water. They might be removed from Martian meteorites during ejection from the planet's surface, passage to Earth or by weathering during the sample's terrestrial history. However, stability may be imparted to nitrates by the formation of mixed salts. The nitrate ion (NO$_3^-$) is in some cases intergrown with the carbonate ion (CO$_3^{2-}$), e.g., KNO$_3$ with calcite, CaCO$_3$ (Field, 1981). Acid sulphate-nitrate hydrates form when nitrogen oxides are distilled onto hydrated sulphates such as MgSO$_4$.6H$_2$O (Florsch et al., 1987). Carbonates and sulphates occur in SNCs [e.g., Carr et al., 1985; Goody et al., 1988; Burgess et al., 1989]; thus nitrates might form stable mixed salts with these species.

Two meteorites were selected for analysis, the Shergotty EET 79001 and Nakhla. Two subfractions from EET 79001 were hand-picked from the bulk meteorite at the Antarctic Meteorite Curatorial Facility, Houston, Texas. The first is El,149, a sample of lithology C (shock-produced glass), in which nitrates might be "protected" (i.e., enrobed in glass), like some of the carbonates. The second fraction, El,239, is a carbonate-rich sample of those derived from the interior of a vug in lithology A [Goody et al., 1988; Wright et al., 1988]. Since El was collected from Antarctica, and has a terrestrial residence age of 12,000 years [Hull and Donohue, 1988], salt formation may have taken place, or the salts modified, on Earth. A non-Antarctic SNC meteorite was therefore also studied in addition. Nakhla is a meteorite retrieved from a temperate location soon after it was observed to fall and which has therefore suffered only minimal terrestrial weathering. Nakhla is also known to contain carbonates, of presumed Martian origin [Carr et al., 1985].

Analytical Methods

The analytical techniques utilized, namely FTIR and stepped combustion-stable isotope mass spectrometry, allow discrete phases to be recognized within a whole-rock sample without physical separation of individual components. A reservoir of whole-rock Nakhla (BM 1911369) was prepared by gently crushing the meteorite, mixing with graphite powder, and placing in a quartz capsule. Aliquots of bulk sample were selected from the reservoir for analysis. In an attempt to concentrate water-soluble salts, such as nitrates, an extract was prepared from Nakhla by suspending - 500 mg of powdered sample in 1 mL of triply distilled, de-ionized water for 20 hours. The resulting solution (which contained fine mineral grains from the meteorite, in addition to any water-soluble salts) was evaporated to dryness, yielding a poorly crystalline residue. The instrument used for FTIR was a Philips PU 9800 infrared spectrophotometer, wavelength range from 225-4400 cm$^{-1}$, continuously purged with CO$_2$-free dry air. Fifty scans were taken for each of the samples and standards. All specimens were finely powdered in a matrix of spectroscopic-grade KBr, dried, then cooled-pressed into a disc ~ 1 cm in diameter. Results from the sample are presented above 50°C. The spectrum is compared here with an absorption band of the CO$_2$-free, dry air sample.

Stalped combustion is the incremental oxidation of nitrogen-bearing components in discrete temperature intervals from room temperature upwards [Jackson et al., 1993]. All gases generated during combustion are converted to N$_2$ gas using a combination of heated CuO and Pt. The nitrogen is then separated cryogenically from other products of combustion (predominantly CO$_2$, SO$_2$, and H$_2$O) prior to admission to a triple-collector noble gas-type mass spectrometer, operated in static mode, for determination of nitrogen isotope composition. Nitrogen yield is gained from calibration of the 28/20 ion beam current. The instrument currently used (known colloquially as FINESSE) can measure the $^{15}$N/$^{14}$N of samples as small as 100 ng of nitrogen to a precision of ±1%. As a by-product of the gas clean-up procedure, carbon yields are also obtained, by measurement of the CO$_2$ gas on a capacitance manometer, to a precision of ±0.1 ng, enabling determination of C/N ratios of the various components.

During stepped combustion, different nitrogen-bearing species exhibit a variety of quantifiable thermal breakdown characteristics. Table 1 summarizes the measured combustion temperature ranges of nitrogen-bearing species anticipated present in Martian meteorites. Data for nitrates were determined by stepped combustion of analytical-grade reagents Mg(NO$_3$)$_2$.6H$_2$O (nitromagnesite) and Ca(NO$_3$)$_2$.4H$_2$O (nitrocalcite) [Grady et al., 1993]. Nitromagnesite decomposed across a single temperature step between 400 and 500°C, but the release of nitrogen from nitrocalcite was more complex, showing a bimodal release with peaks at 200-300°C and 400-500°C.

Results

FTIR

Nitrates and carbonates have strong, characteristic absorptions in the infrared. From the work carried out here, the most prominent feature in all spectra measured for the standard nitrates was a strong singlet peak at approximately 1372 cm$^{-1}$, and a second, nitrate singlet at ~1630 cm$^{-1}$ (Figure 2). Comparison of data from inorganic nitrates with those from carbonates indicated that the two types of salt are resolved using FTIR spectroscopy (the strongest carbonate feature is at ~1425-1450 cm$^{-1}$ [Lowe and Jackson, 1993]).

Analysis of powdered El,149 glass yielded a spectrum with several features (Figure 3) singular at 1375 and 1630 cm$^{-1}$, indicative of nitrates, and a doublet between 1400 and 1450 cm$^{-1}$, which suggests carbonates. FTIR is not able to give an indication of the relative concentrations of the two salt components, and so all that can be concluded is that El,149 appears to contain carbonates and nitrates. The absorptions at 1375 cm$^{-1}$ and 1630 cm$^{-1}$ represent the first evidence that nitrates might occur in Martian meteorites. However, it is not clear from the FTIR data whether the nitrates exist as isolated anions dissolved in the glass, or discrete mineral phases.

The spectra obtained from a powdered whole-rock sample of Nakhla (~2 mg) and the water-extract of the same meteorite (500 mg) are shown in Figure 3. For the whole-rock sample, the FTIR data are dominated by broad features below 1250 cm$^{-1}$.

Table 1: Possible Nitrogen-Bearing Components Released at Low Temperatures From Martian Meteorites

<table>
<thead>
<tr>
<th>Species</th>
<th>Temperature, °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorbed terrestrial atmosphere N$_2$</td>
<td>&lt;100</td>
</tr>
<tr>
<td>Organic materials</td>
<td>200-450</td>
</tr>
<tr>
<td>Nitrate (Ca(NO$_3$)$_2$.4H$_2$O)</td>
<td>200-300, 400-450</td>
</tr>
<tr>
<td>Nitromagnesite (Mg(NO$_3$)$_2$.6H$_2$O)</td>
<td>400-500</td>
</tr>
</tbody>
</table>
Figure 2. FTIR data for two pure nitrate minerals (nitrocalcite and nitromagnesite). Note the characteristic features in both traces at 1375 and 1630 cm⁻¹.

Figure 3. FTIR data for a glass sample of EET A79001 (E1,149), and a whole-rock and water extract from the Nakhla meteorite. The data for E1,149 show evidence for the presence of nitrates and also carbonates. The results for the water extract of Nakhla show a very broad feature in the region where nitrates and carbonates are expected; it is difficult to draw firm conclusions regarding the presence of these minerals in this sample.

Figure 4. Nitrogen released by stepped combustion (room temperature to 700°C) of a single chip (5.842 mg) of E1,149 glass are shown in Figure 4. The yield histogram may be interpreted as indicating either the presence of two components of differing thermal stability or the bimodal release of nitrogen from a single species. The first release of nitrogen (< 1000 ppb) occurs below ~ 320°C. Unfortunately, due to analytical problems, isotopic data for these steps were not all recorded; thus the isotopic composition of the lowest temperature material remains poorly defined, with δ¹⁵N between -35% and +35%. The second release of nitrogen (~ 250 ppb) is between 350°C and 400°C, with δ¹⁵N approximately -14%. If this is from a nitrate, its isotopic composition is not sufficiently distinct from that of Antarctic nitrates [Wada et al., 1981] to assign a specific origin to the material. Above 400°C, nitrogen is released in gradually decreasing quantities with generally increasing δ¹⁵N, up to a maximum of +13.2% at 650°C. C/N ratios range from ~ 30 to 100; interpretation of these values is ambiguous, since at low temperatures the sources of carbon and nitrogen are decoupled: carbon derives from contaminant organic materials of terrestrial biogenic origin, whereas nitrogen can arise from organics, nitrates, or trapped terrestrial atmospheric gases.

Nitrogen yields and δ¹⁵N values for the water-soluble material from Nakhla are shown in Figure 5. The residue comprises approximately 600 ppm of whole-rock Nakhla; hence the nitrogen content of the water-soluble material represents a whole-rock nitrogen concentration of 17 ppb, most of which is liberated between 200°C and 500°C. As in the case of E1,149, the water extract of Nakhla shows a bimodal release of nitrogen at low temperatures, with δ¹⁵N between about -20% and -5%. C/N ratios exhibit the same range as the whole-rock material, and again fail to constrain the likely nature of the nitrogen-bearing species.
nitrogen content. In Nakhla, if the low-temperature nitrogen in the
Figure yields are compared in Figure 6a. For each analysis the highest
temperature. Approximately this did not occur; nitrogen in the water-soluble material is ap-
proximately concentrated in the water extraction. Figure 6a indicates that among the four separate analyses of the two meteorites, all results
been considered: E1,149 [Wright et al., 1988b] and whole-rock Nakhla [Wright et al., 1992]. For these samples, the combination of tem-
perature increments utilized (30°C and 100°C) precludes the ob-
ervation of a low-temperature biomineral release of nitrogen from a
nitate source. Thus, in order to facilitate comparison of data
among the four separate analyses of the two meteorites, all results
have been normalized to a consistent heating program. Nitrogen
yields are summarized in Figure 6a. For each analysis the highest
nitrogen yield and lowest C:N ratio is from the 200-300°C step,
following which there is a general decline in yield with increasing
temperature. E1,239, known to be enriched in minerals with a
low-temperature origin [Crowell et al., 1988], has the highest ni-
trogen content. In Nakhla, if the low-temperature nitrogen in the
whole-rock sample were all from a nitrate source, it should have
been concentrated in the water extraction. Figure 6a indicates that
this did not occur; nitrogen in the water-soluble material is ap-
proximately 3 orders of magnitude less abundant than in the
whole-rock. One possible explanation is that any nitrates in Nakh-
la are present within the carbonate lattice, and hence insoluble in
water. However, this seems unlikely, since carbonates generally
decrepiate on stepped heating between 400°C and 700°C, and so
consisting carbonates would be liberated over the same temperature
range, whereas most of the nitrogen in whole-rock Nakhla is re-
leased below 400°C.

A further constraint on the identity of the low-temperature components of the different fractions of E1 and Nakhla is pro-
vided by the nitrogen isotopic data. E1,149 and E1,239 show
some similarities to each other (Figure 6b): data from E1,239 ind-
dicate that there are at least two nitrogen-bearing components re-
leased below 700°C. The first is liberated below 300°C (85N
about -14‰), comprising some 75% of the low-temperature ni-
trogen. A second, isotopically heavier, component is released
above 300°C, with 85N between -6‰ and -1‰ (which, in itself,
might be two separate nitrogen-bearing entities, as described by
Wright et al. [1988b]). The similarities between E1,239 and
E1,149 lead to the conclusion that if this component is a nitrate,
then distribution of the material within E1 is quite widespread.
Whole-rock Nakhla contains low-temperature nitrogen with
85N in the range -7.6‰ to -1.1‰ (Figure 6c), again indicative of
at least two components. The water-soluble material, on the other
hand, releases nitrogen with an approximately constant isotopic
composition, 85N about -11‰, implying that only one compo-
nent is present. It is possible (indeed, probable) that the whole-
rock sample contains some terrestrial contaminants (e.g., atmos-
pheric nitrogen with 85N about -5‰ and organic materials with
85N about -5‰ to -10‰); these are presumably not transferred to
the water extract. Thus, only indigenous nitrogen, i.e., nitrates
with 85N about -11‰, are present in the water extract. On the
basis of Figure 6a, perhaps as little as 0.1% of the total nitrogen
released below 700°C in Nakhla whole-rock derives from indige-
nous nitrates.

Discussion and Implications
On Earth, three stages have been identified in the evolution of the
nitrogen cycle [Mancinelli and McKay, 1988]. The second
stages occur after biological processes have become estab-
lished, and are therefore probably not relevant to Mars. However,
the first, probiotic stage, might well apply: NO3 species are
formed in the atmosphere by electrical discharge, cosmic ray in-
teractions and by ultraviolet radiation, and subsequently fixed at
the surface by implantation and exchange. Similar processes have
been suggested for the formation of contemporary inorganic ni-
trate deposits in cold regions on Earth, such as Antarctica and
Chile [e.g., Clarridge and Campbell, 1968; Mueller, 1968; Wada
et al., 1991]. In these instances, however, the anions formed by
discharge are washed out of the atmosphere during precipitation,
then deposited by sublimation or evaporation as nitrates, rather
than directly implanted as NO3 species. If nitrates were formed
by similar mechanisms on Mars, then their isotopic composition
should reflect that of the atmosphere from which they originated.
It is apparent that very low levels of nitrate minerals occur in
SNC's (17-1520 ppb nitrogen) with 85N in the range -8 to -14‰,
isotopic compositions that do not reflect the present-day Martian
atmosphere. Assuming that nitrates in E1 and Nakhla are not ter-
restrial contaminants, it is clear that they could not have formed on
the Martian surface in recent times by incorporation of atmos-
pheric species. Alternatively, the nitrates might have formed more
recently by reactions of degassed magmatic volatiles (85N about
20 to -10‰ [Wright et al., 1992]) within the regolith.

Since nitrates are highly soluble salts and Mars has obviously
experienced the effect of mobile water at various stages in its
history, it is perhaps not surprising that nitrates are largely absent
from specimens that came from the near-surface environment. It
is possible that the salts dissolved in liquid water during a period
when fluids percolated through the pore spaces of the Martian
regolith. Any nitrates might therefore have been mobilized and ni-
ter redeposited in a subsurface "hard-pan" layer, or become fra-
zen within permafrost. Although such mechanisms are plausible, they do not account for the presence of other soluble species in SNC meteorites. Halite (NaCl), sylvite (KCl), and the soluble sulphate epsomite (MgSO₄·7H₂O) have been detected in Nakhla [Gooding, 1992]. Herein, halite was the only species identified by X-ray diffraction of the water extract of Nakhla.

Under the current partial pressure of NO calculated for Mars’ atmosphere, nitrates are predicted to be more stable than coexisting carbonate species [Zolotov and Siderov, 1986]. From the data acquired in this study, along with results from Wright et al. [1992] and Wright and Pillinger [1994], the relative proportion of carbonates to nitrates in El and Nakhla are calculated as ~10 and 2200, respectively. The apparent deficit of nitrates in Nakhla might be due to the soluble salts having been washed out of the samples by the action of fluids; in El the same processes may have acted to concentrate the salts. Alternatively, the differences may reflect the admixture of a terrestrial carbonate in the Antarctic sample. Regardless of the difficulties of interpretation, it may be assumed that the carbonate/nitrate ratios are representative of the surface reservoirs in general. If, as has been modeled by Yang et al. [1977], there were a nitrate layer 0.3 cm thick across the surface of Mars, then the corresponding carbonate deposits could be anything from 3 to 600 cm thick. For comparison, Kahn [1985] has suggested up to 140 mbar of CO₂, equivalent to 280 cm of carbonate, stored in the regolith on Mars.

Acknowledgments. The Antarctic Meteorite Working Group is gratefully acknowledged for provision of the fractions from EET A79001. G. Jones and J. Gibson are thanked for assistance with analytical techniques. The paper benefited by reviews from R. H. Becker and A. J. T. Jull. This project was undertaken as part of the Cosmic Mineralogy Programme of the Natural History Museum (M.M.G.). Financial support to I.P.W. and C.T.P. was provided by the Science and Engineering Research Council (now PPARC).

References

Figure 6. Stepped combustion data from room temperature to 700°C for the samples described herein, as well as another subsample of EET A79001 (E1, 239) and Nakhla whole-rock. To facilitate comparisons, the data from each extraction have been calculated to a consistent temperature program. (a) Nitrogen yields for all four samples, (b) δ¹⁵N for the two E1 samples, and (c) δ¹⁵N for the two Nakhla samples.
GRADY ET AL.: NITRATES IN MARTIAN METEORITES 5455


McSween, H.Y., What we have learned about Mars from SNC meteorites, Meteoritics, 29, 757-779, 1994.


M. M. Grady, Department of Mineralogy, The Natural History Museum, Cromwell Road, London, England, United Kingdom. (email: M.Grady@nhm.ac.uk)

C. T. Pillinger and I. P. Wright, Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA, England, United Kingdom. (email: I.P.Wright@open.ac.uk)

(Received April 4, 1994; revised October 10, 1994; accepted October 27, 1994.)
A slightly more massive young Sun as an explanation for warm temperatures on early Mars

Daniel P. Whitmire,1 Laurance R. Doyle,2 Ray T. Reynolds,3 and John J. Matosec

Abstract. The valley network channels on the heavily cratered ancient surface of Mars suggest the presence of liquid water ∼3.8 Gyr ago. However, the implied warm climate is difficult to explain in the context of the standard solar model, even allowing for the maximum CO2 greenhouse heating. In this paper we investigate the astronomical and planetary implications of a nonstandard solar model in which the zero-age, main-sequence Sun had a mass of 1.05 ± 0.03 Me. The excess mass was subsequently lost in a solar wind during the first 1.2 × 10^12 Gyr of the Sun's main sequence phase. The implied mass-loss rate of 4.2 × 10^-12 Me yr^-1, or about 10^12 times that of the current Sun, may be detectable in several nearby young solar type stars.

1. Introduction and Summary

Because of the evidence that early Mars was warmer than can be easily explained in the context of the standard solar model, we have investigated several astronomical models that could in principle increase the solar flux at early Mars. The most plausible of these is one in which the early main sequence Sun was somewhat more massive than today’s Sun. (Other, less promising, astronomical mechanisms for heating early Mars are considered briefly in the appendix.) Because the solar flux at a planet is extremely sensitive to mass (F ∝ M^0.75), the required minimum mass of the young Sun is only 1.03 Me. An upper limit of 1.07 Me is necessary to prevent the loss of water from Earth. We first discuss the constant mass-loss model and the climatic constraints imposed on it. We then consider the nonclimatic astronomical and planetary implications of the model and find that it is consistent with existing planetary and astronomical data. However, a convincing case for the model can be made only if similar mass-loss rates (≥ 2 × 10^-11 Me yr^-1) are found to be a common feature of other young solar-type stars, or by the discovery (or reinterpretation) of supporting meteoritic and lunar implantation data. The recent detection [Güdel et al., 1994] of continuous microwave emission from four nearby G-type dwarfs would, if interpreted in terms of free-free emission from an optically thick wind [Mullan et al., 1992], imply mass-loss rates of ∼ 10^-9 Me yr^-1. This interpretation can be tested by additional measurements at millimeter wavelengths.

2. Background

Assuming the standard solar model, Kasting [1991] has shown that the maximum CO2 greenhouse warming is insufficient to raise the early Martian mean surface temperature to near 273 K, although CO2 can resolve the faint young Sun paradox on Earth [Kasting and Grinspoon, 1991]. Models employing the greenhouse gases methane or ammonia have generally been unsuccessful, as these gases have short photodissociation lifetimes. However, a shielding particulate haze similar to that observed on Titan may mitigate this objection [Chyba and Sagan, 1994]. Some geothermal heating is expected [Spyrey, 1993], but this is not likely to completely explain temperatures required for surface liquid water. Even if the channels were formed by subsurface sapping of groundwater, Martian surface temperatures must have been significantly higher than today to allow groundwater to be mobile near the surface and to recharge the aquifer sources. Controversial evidence of glacial markings would, if substantiated, also indicate high temperatures since precipitation is implied. The issue of warm temperatures on early Mars is also important for the determination of the size of circumstellar habitable zones [Kasting et al., 1995; Whitmire and Reynolds, 1995] and therefore the broader topic of life in the universe.
3. Climatic Constraints on the Mass of the Early Sun

In standard constant-mass stellar models the mass-luminosity relation for stars near 1 solar mass is given by [Jen 1997] as

\[ L \propto M^{4.75} \]

A more recent evolutionary calculation of a mass losing 1.1 M\(_{\odot}\) star gives an effective index of about 4.5 [Boothroyd et al., 1991]. However, observations of G stars in detached binary systems give an index of 4.8 near 1.0 M\(_{\odot}\) [Popper et al., 1986]. (The mass-luminosity index has its maximum value near 1.0 M\(_{\odot}\).) Recognizing that the precise value of the index is likely to depend on composition and age [Breger and McCarthy, 1993], we adopt a value of 4.75 and assume that this is also an adequate approximation when the star’s mass is slowly decreasing.

In the standard model the evolution of the luminosity is given as an analytic function of time up to the present by [Whitmire et al., 1995]:

\[ L(t) = \frac{L_0}{1 - \frac{M(t)}{M_0}} \left( \frac{M(t)}{M_0} \right)^{4.75} \]

where \( L_0 = 0.71(M_0/M_0)^{4.75}L_\odot \) is the luminosity at time \( t = 0 \), \( M_0 \) is the mass at \( t = 0 \) and \( t_0 = 4.6 \) Gyr. We model the solar mass \( M(t) \) as a linear function of time (i.e., constant mass loss) given by

\[ M(t) = M_0 (1 - r(t)) \quad 0 \leq t \leq \tau \]

\[ M(t) = M_0 \quad \tau \leq t \leq t_0 \]

where \( \tau \) is the mass-loss timescale and \( r = (M_0 - M_0)/M_0 \) is the mass-loss fraction. The relation between a planet’s semimajor axis \( a(t) \) and the solar mass is obtained from conservation of angular momentum (or its inverse, the first adiabatic invariant for \( J_3 \) discussed below). The result is \( a(t) \propto M^{-1} \), and therefore the flux at a planet \( F(t) = L(t)/4\pi a^2(t) \) is

\[ F(t) = \frac{F_\odot}{1 - \frac{M(t)}{M_0}} \left( 1 - \frac{M(t)}{M_0} \right)^{4.75} \quad 0 \leq t \leq \tau \]

\[ F(t) = \frac{F_\odot}{1 - \frac{M(t)}{M_0}} \left( 1 - \frac{M(t)}{M_0} \right)^{4.75} \quad \tau \leq t \leq t_0 \]

where \( F_\odot = 0.71(1AU/a(t_0))^2(M_0/M_0)^{4.75} \) is the flux at the planet at \( t = 0 \), located today at \( a(t_0) \), in units of the present solar flux at Earth.

We now impose two independent planetary climate constraints on the flux model. First, we require that the flux at Earth at \( t = 0 \) (when it is a maximum during the first 4.6 Gyr) not be so great that all planetary \( \text{H}_2\text{O} \) would be lost. This would occur through the moist greenhouse effect if the flux exceeded 1.1, in units of the flux at Earth today [Kasting, 1988]. A moist greenhouse occurs when the stratosphere becomes wet and \( \text{H}_2\text{O} \) is lost through UV dissociation and the subsequent loss of the hydrogen to space. In the climate model used to calculate this flux, clouds were not fully taken into account (only their albedo was modeled) and the actual critical flux could be somewhat greater than this, but probably not as high as the runaway greenhouse critical flux of 1.4, for which the oceans evaporate entirely. A critical flux of 1.1 limits the Sun’s zero-age, main-sequence (ZAMS) mass to \( 1.07M_\odot \), independent of the mass-loss timescale \( \tau \). This upper mass limit is fairly insensitive to the precise value of the critical flux, being proportional to \( (\text{critical flux})^{0.76} \).

The second climate constraint is based on the assumption that average annual surface temperatures on Mars should have been at least 273 K at the end of the late heavy bombardment about 3.8 Gyr ago (\( t = 0.8 \) Gyr). The drainage channels and valley networks, which indicate the flow of liquid water, are found primarily in the highly cratered southern highlands and are therefore assumed to date to the end of the period of the late heavy bombardment [Pollack et al., 1993]. This translates to a flux that is at least 13% greater (i.e., \( 1.13 \times 0.32 \)) than predicted by the standard solar model at \( t = 0.8 \) Gyr [Kasting, 1989; Kasting et al., 1993]. From the flux equation \( F(t) \), it can be shown that a 13% increase in flux at \( t = 0.8 \) Gyr requires an initial solar mass of 1.03 M\(_{\odot}\) if we assume that the maximum value of \( r \) is 1.0 Gyr. This timescale limit is based on the apparent absence of any ion implantation evidence of a strong wind more recent than \( \sim 3 \) Gyr ago. Inserting the above upper mass limit of 1.07 M\(_{\odot}\) (determined from Earth’s climate history) into the flux equation (as applied to Mars) yields a minimum value of \( r \) of 1.0 Gyr.

We conclude that a ZAMS solar mass of \( 1.05 \pm 0.02 \) M\(_{\odot}\), corresponding to mass-loss timescales of 1.2-2.3 Gyr, can explain the existence of liquid water on early Mars and avoid the loss of water from Earth. The corresponding mass-loss rates are \( 4 \times 10^{-11} \) M\(_{\odot}\)yr\(^{-1}\), or about 10\(^9\) \times \) that of the current Sun. These results are not very sensitive to uncertainties in the critical cli-
matic fluxes assumed. If mass loss is assumed to be a monotonically decreasing function of time the lower mass limit would be increased but the upper limit would be unaffected. Next we consider other, nonclimactic implications of the model.

4. Nonclimactic Implications of a More Massive Early Sun

4.1. Astronomical

4.1.1. Observed stellar mass loss from main sequence dwarfs. The direct observation of mass-loss rates less than about \(10^{-7} M_{\odot} yr^{-1}\) is difficult in solar-type dwarfs [Brown et al., 1990]. However, because of the presence of a UV continuum from a white dwarf companion, significant mass loss has been detected by Mullan et al. [1989] from the K2V dwarf in the eclipsing detached binary V471 Tauri in the Hyades cluster. The measured mass-loss rate was \(10^{-8} M_{\odot} yr^{-1}\) times that of the present Sun. If this wind has been constant or decreasing during the age of the Hyades, the implied mass loss to date would be \(2 \times 10^{-9} M_{\odot}\). The K star's spin should be tidally synchronized with its orbital period of 0.5 day. This rotational period is in the range of those observed in young solar-type stars [Stauffer and Soderblom, 1991], suggesting that for isolated main sequence stars, significant mass loss is associated with rapid rotation. Observations show that by the age of the Hyades (~0.6 Gyr) this initial very fast rotation has been reduced, though typically still \(\sim 5\times\) that of the Sun today.

Mullan et al. [1992] have reported winds \(\geq 10^{-9} M_{\odot} yr^{-1}\) from several M dwarfs. Evidence of mass loss was based on measurements over a wide range of radio, millimeter, and infrared wavelengths and the fact that free-free emission from an optically thick wind has a well-established characteristic spectrum in which the flux is proportional to \(v^\gamma\), where \(v\) is the frequency. There are several nearby young G stars that may be good candidates for similar observations. The age of HD39587 is determined from its membership in the Ursa Majoris cluster \((\leq 200\, m.y.)\). This star is known to have one or two low-mass companions. The apparently single star HD77205 is also a good candidate. However, although its rapid rotation (period 4.7 days) and CaH and K emission line flux indicate an age of about 700 m.y., the errors on its age determination are rather large. The stars HD153383 (rotation period 3.5 days) and HD206860 \((v \sin i = 10.2\, k m\, s^{-1})\) are also within the potential detection range but again more stringent age constraints are needed.

If significant mass loss is occurring in these young stars, there will be an excess flux given in mJy by [Leitherer and Robert, 1991; Wright and Barlow, 1975; Panagia and Fall, 1975]

\[
S_{\nu} = 2.32 \times 10^{15} (M/2) a^{1/2} (2\mu_\nu)/\mu_m \nu^{(2+3)/2} \rho_m = 4.1 T_2 \rho_2 \nu^2
\]

where \(M\) is the stellar mass-loss rate in \(M_{\odot} yr^{-1}\), \(S_{\nu}\) is the rms ion-core charge of the wind particles, \(\gamma\) is the number of electrons per ion, \(\beta\) is the free-free Gaunt factor, \(v\) is the frequency in Hz, \(n_m\) is the stellar wind velocity in \(k m\, s^{-1}\), \(\mu\) is the mean molecular weight of wind particles, and \(\nu\) is the distance to the star in parsecs (1 parsec = 3.2 light years). For a solar-like stellar wind, \(\nu \approx \beta = 1\). The Gaunt factor is given by \(\beta = 9.77 (1 + 0.13 \log (T_2^{1/2} / 10))\), where \(T_2\) is the stellar wind temperature in degrees Kelvin [Leitherer and Robert, 1991; Abbott et al., 1980].

In the derivation of \(S_{\nu}\), it was assumed that the wind was isothermal, spherically symmetric, and the outflow was stationary with constant velocity. The dependence of the flux on frequency was found not to be very sensitive to changes in spherical symmetry and wind gradients in temperature, velocity, and density [Leitherer and Robert, 1991].

In preliminary work, we have made long integration-time observations of two of the four candidate stars (HD39587 and HD72905) using the National Radio Astronomy Observatory very large array in Socorro, New Mexico. These observations showed that mass loss could be taking place in the HD39587 system but, unfortunately, the flux levels were confused by the unresolved nearby smaller-mass companions, and further observations are required to ensure that the source of the mass loss is the G star. HD72905 shows significant flux at 8.0 GHz, but very little at 4.8 GHz. It appears that the slope of this flux spectrum is too steep to be indicative of mass loss, but further deconvolution is needed. As noted above, the age of this star is relatively uncertain. We have not observed HD115383 or HD208680 long enough to draw any conclusions regarding mass-loss rates less than \(2 \times 10^{-9} M_{\odot} yr^{-1}\).

Recently, Güdel et al. [1994] have reported the discovery of 8.5-GHz microwave emission from four nearby "solar-twin" G dwarfs selected from a set of 15 G dwarfs with \(X\) ray luminosity 100-800 times greater than the quiet Sun. The largest radio flux (0.338 mJy; 13\(\gamma = 10^{-19}\)) and luminosity were found in the young G0V star HD132933 (EK Draconis), which is believed to be a member of the Pleiades cluster \((\sim 70\, m.y.)\) and is often studied as an infant-Sun proxy [Driver and Oman, 1994]. A slightly lower radio luminosity was found for HD 72905 (2 \(\times 10^{-20}\)) and the other two G stars, one of which (GI 755) had no age estimate and the other (HD 225239) was tentatively dated at a surprising \(\sim 10\) Gyr. The emission mechanism is uncertain, since spectra are not yet available. If we interpreted these observations in terms of mass loss, the implied mass-loss rates are \(\lesssim 10^{-8} M_{\odot} yr^{-1}\).

The theoretical arguments (e.g., high corona temperatures) given by Mullan et al. [1992] as motivation for searching for mass loss in M dwarfs also apply to these four solar-like G stars. It appears that the mass-loss model can be tested in these stars by measurements made at millimeter wavelengths. These measurements would test the predicted \(v^{1/3}\) dependence of the flux.
over a large frequency interval from the existing 8.5 GHz to 200 GHz. The predicted 1-mm flux for HD129333 is \( \sim (300/5.6)^{1/2} \times 0.338 \, \text{mJy} \approx 3.6 \, \text{mJy} \). The M star fluxes measured at 1 mm by Mullan et al. were \( \sim 10 \, \text{mJy} \).

4.1.2. Lithium depletion. The element lithium is known to be depleted by two orders of magnitude in the photosphere of the Sun relative to its primordial abundance. Wegmann and Sears [1985] suggested that this depletion could be explained if the Sun was originally somewhat more massive than 1.0 \( \text{M}_\odot \). This would allow the lithium to be convectively mixed into regions with temperatures high enough to destroy it. This idea was independently rediscovered and extended by Boothroyd et al. [1989], who found that an initial mass of 1.04 \( \text{M}_\odot \) would result in the removal of 99% of the initial lithium. Using the latest nuclear reaction rates and opacities, Boothroyd et al. [1991] computed detailed solar models with an initial mass of 1.1-1.2 \( \text{M}_\odot \). They found that this mass resulted in the desired depletion of lithium while at the same time not over depleting the element beryllium. Further indirect support for this explanation of lithium depletion in the Sun has recently been reported by [Gregor, 1995, Soderblom et al., 1990]. Soderblom et al. [1990] found that if mass loss is the explanation for lithium depletion in F and G dwarfs in the Hyades cluster, then the implied mass loss would be \( \approx 0.07 \, \text{M}_\odot \) and the timescale for mass loss would be \( \approx 0.8 \, \text{Gyr} \).

On the other hand, Swenson et al. [1994] find that, although mass loss can be a marginal explanation for the solar lithium problem, it cannot be a significant contributor to the depletion of lithium in Hyades G dwarfs. Rotational mixing associated with magnetic braking on the main sequence has been proposed by [U.Shekher and MacGregor, 1976] as an alternative model to explain the lithium depletion. Other mechanisms have also been proposed [e.g., Svensson et al., 1984; D'Antona and Mazzitelli, 1984] and are further discussed in the work by Boothroyd et al. [1991].

4.1.3. Angular momentum loss. Magnetic braking, which requires an associated stellar wind, is believed to be the explanation for the observed spin down of young solar-type stars after they have arrived on the main sequence [Stauffer, 1995; Soderblom et al., 1990; Biskerker and MacGregor, 1976]. However, in many models, the required change in angular momentum is insensitive to the mass loss [Pauwels et al., 1989], depending instead almost entirely on the magnetic field and stellar wind velocity. An exception is the model of Biskerker et al. [1996, in which the total mass loss was \( \sim 10^{-5} \, \text{M}_\odot \), assuming that the convective envelope was decelerated from the interior of the star. As noted by Grudel et al. [1991], helioseismic observations suggest that the Sun rotates more nearly like a solid body and, consequently, the implied mass loss should be increased by a factor of 20 (= moment of inertia ratio), giving a value roughly consistent with the present model.

Assuming the ZAMS Sun rotated at a rate comparable to typical solar-type stars in the Pleiades cluster, the loss of rotational energy suggests a minimum fractional mass loss of \( \sim \epsilon \left(V_\odot/V_\odot\right)^2 \), where \( V_\odot \) is the ZAMS solar equatorial velocity, \( \sim 200 \, \text{km s}^{-1} \) [Stauffer and Soderblom, 1991]. \( V_\odot, \) is the average early wind velocity at infinity, and \( \epsilon \) is the moment of inertia constant. Unfortunately, \( V_\odot, \) is unknown, but may have been lower than the current average value of \( \sim 400 \, \text{km s}^{-1} \) [Guss and Bacher, 1991]. Observed stellar wind velocities range from tens of \( \text{km s}^{-1} \) for red giants to greater than \( \sim 10^5 \, \text{km s}^{-1} \) for early-type stars.

4.2. Planetary

4.2.1. Planet and planetesimal migration. The effect of slow radial mass loss on planet orbital can be deduced from the adiabatic invariance of the action variables.

\[
J_3 = \frac{2\pi G M_m a}{\kappa}, \quad J_2 = \frac{GM_m \sqrt{\kappa}}{a}, \quad J_1 = \frac{M_m \cos i}{a},
\]

where \( M_m \) is the solar (planet) mass and \( a \) is the semimajor axis. It can be seen that the planetary eccentricities, \( e \), and inclinations, \( i \), are constant during adiabatic mass loss but, as noted earlier, the semimajor axes increase inversely with stellar mass, assuming planetary accretion is negligible.

During the Sun's mass loss phase, any remaining unaccreted planetesimals would migrate at rates different than the planets due to dynamical drag. To model the planetesimal migration process we assume that the early solar wind was radial and isotropic and carried no significant angular momentum. A planetesimal of radius \( r \) would accrete mass from the solar wind at the rate \( \frac{\text{d}M}{\text{d}t}(r^2/4\pi r^2) \). The composition of the solar wind is mostly hydrogen and helium, so almost all of the accreted mass would rapidly volatilize at the same rate and escape from the orbiting planetesimal with negligible thermal speeds; consequently, we take the planetesimal mass and radius to be constant. With these assumptions a straightforward angular momentum analysis results in the differential equation [Whitmire et al., 1991]

\[
\frac{\text{d}a}{\text{d}M} = \left( \frac{2a^3}{M} \frac{r^2}{m} \right)^{1/2}
\]

the solution of which is

\[
a^2 = a_0^2 \left( \frac{M_0}{M} \right)^2 + \frac{M}{4\pi \rho} \left[ 1 - \left( \frac{M_0}{M} \right)^2 \right]
\]

or, when the final mass \( M = M_0 \)

\[
a/a_0 = 1 + \Delta \left( 1 - \frac{1}{\rho M_0^2} \right)
\]

to lowest order in the mass-loss fraction \( \Delta \). In the latter
expression $p$ is the planetesimal density in g cm$^{-3}$, $r$ is the planetesimal radius in units of 10 km, and $a_p$ is the initial semimajor axis in units of AU. In our (elastic-collision/volatilization model), the migration is independent of the solar wind velocity and the functional form of $M(t)$, to first order in $\Delta$. Different assumptions about the angular momentum content of the solar wind and the planetesimal accretion/emission mass rates will lead to different migration formulae. For example, if elastic collisions or sputtering are assumed to be important, the migration formulae will be somewhat different. However, independent of the model, there will, in general, be differential migration between planetesimals and planets.

It is evident from the above migration equation that the drag term is negligible for planets, but not necessarily for planetesimals. The outward migrating planets would perturb any remaining unaccreted planetesimals lying within an original radial band that can be determined by applying the migration equation to both planets and planetesimals. The differential migration between a planet and planetesimal near $a_p$ is then found to be

$$\Delta a \approx \frac{\Delta}{M_{\text{wind}} a_p}$$

Assuming a density of 3.0 (0.5) g cm$^{-3}$ for terrestrial (Jovian) planetesimals, and taking $\Delta \sim 0.1$, the total band cleared by all planets is $\sim 0.2$ AU for planetesimals of radius 10 km. Resonant pumping of planetesimal eccentricities may substantially multiply the total cleared bandwidth. The resulting equilibrium eccentricity in our model would be much greater than that due to drag in the solar nebula [Weidenschilling and Davis, 1985] because of our much smaller drag (or effective density).

If the mass outflow stopped in less than $-200$ Myr, the late heavy bombardment should have continued somewhat beyond $3.8 \pm 0.1$ Gyr ago. The short range of solar flare heavy ions ($<1 \mu m$) and wind-implanted ions ($<1 \mu m$) in meteorites show that the grains were exposed at a time when there was very little intervening material between the Sun and the grains [Walter and Barry, 1991]. This is most consistent with post-T Tauri irradiation. Further, Papazastassou et al. [1974] studied two igneous clasts in Kaskota (the breccia on which Caffee et al. based their argument) and found them to have relatively young Rb-Sr isochron ages ($3.67\pm0.05$ and $3.63\pm0.08$ Gyr). They concluded that implanted solar wind gases and tracks in the breccia must record events during the first 1 Gyr after the formation of the solar system, and not early solar system formation processes [Wood and Pelas, 1991]. Finally, we note that if the early main sequence mass loss was bipolar, as observed in many pre-main sequence T Tauri stars, there may be no meteoritic or lunar implantation record of it.

Appendix: Other Astronomical Mechanisms

Here we comment briefly on several other astronomical mechanisms for heating early Mars in increasing order of plausibility [Whitmire et al., 1996].

A.1. Interstellar Cloud Accretion

Accretion of interstellar matter onto the Sun would increase the solar luminosity due to the release of gravit-
tational potential energy. However, this seems to be an unlikely possibility, since for a $\geq 1\%$ increase in luminosity, the required interstellar cloud hydrogen number density is $\geq 10^6$ cm$^{-3}$ for a relative velocity $\leq 10$ km s$^{-1}$ [McCrea, 1979]. The probability of simultaneously satisfying both of these requirements in 4.6 Gyr is small. Further, for any plausible relative velocity, the resulting transient time through such a dense cloud is likely to be less than the time required to form the valley network channels on Mars.

A.2. Perturbation by a Third Body

The flux at the surface of a planet averaged over an orbital period is

$$F = \frac{T}{2\pi} \frac{1}{a^2 \sqrt{1-e^2}}$$

where $a$ is the semimajor axis and $e$ is the eccentricity. The eccentricity of Mars' orbit can vary from today's value of 0.09 to 0.2 [Laskar, 1994]. The maximum increase in average flux due to this variation is only $\pm 2\%$. The variation in flux between perihelion and aphelion in a single orbit is proportional to $[(1+e)/(1-e)]^2$, which could be as large as 2.2. The degree to which the thermal inertia of a dense early Martian atmosphere and/or oceans would mitigate this large annual increase in solar flux remains to be investigated.

Although the average flux is not sensitive to eccentricity, it is sensitive to the semimajor axis. The required flux enhancement of $\geq 15\%$ (more than the standard solar model value) corresponds to a semimajor axis $\leq 1.62$ AU. If Mars accreted at this distance from the Sun and was subsequently perturbed to its present orbit by another body (or through solar system nonlinear dynamical effects), its present orbit would be expected to overlap its original orbit. 'Today, Mars' perihelion distance (1.38 AU) is consistent with this expectation. Although probably unstable, the occurrence of such a perturbation is not totally ad hoc. The main objection is that the required close encounter would have had to occurred so long after planetary accretion was over.

A.3. Variations in Solar Luminosity

In young solar proxy stars, photometric brightness and magnetic changes are anticorrelated [Balona, 1991; Redick et al., 1990]. Although an increase in photometric brightness does not necessarily imply an increase in bolometric luminosity, it is unlikely that there is no correlation. The near cancellation of the decrease in luminosity due to dark spots by the increase in luminosity due to faculae in the Sun is not representative of younger and more magnetically active solar-type stars [Foukal, 1994]. For these stars it is likely that spots result in net luminosity variations 10-100 times larger than the current solar variation [Redick, 1991]. Young solar-type stars are expected to be heavily spotted, up to several tens-of-percent coverage. In equilibrium, the star's luminosity is fixed by core nuclear reactions, and so the temperature of the radiating surface is higher than it would be without spots. If the spots disappeared on a timescale $\lesssim 10^8$ years the newly exposed area would result in an increase in total luminosity given by $L = L_0 (1+f)$, where $L_0$ is the core luminosity and $f$ is the fraction of the star's surface previously covered by spots [Spruit, 1982]. The luminosity would return to the core value on the convective envelope (thermal) timescale of $\sim 10^8$ years in the case of a G2 dwarf like the Sun. However, this timescale (like that for accretion) is probably too short to explain the Martian valley networks.

A.4. Ecliptic Flux Focusing

A related and more promising spot model is that of ecliptic focusing [Whitmire et al., 1993]. Unlike the current Sun, photometric modeling of young solar-type stars in RS CVn systems, in particular, show a definite tendency for spots to cluster toward northern latitudes [e.g., Redick et al., 1986]. The spotted stars in these binary systems are often young G or K dwarfs. In many cases these stars remain young in the sense that rapid rotation is enforced by spin-orbit synchronization. If spots tend to cluster in the polar regions over timescales of $\sim 10^8$ years, the luminosity is unaffected but the distribution of stellar flux emitted is altered. For young heavily spotted stars the temperature and emitted flux from the spotted regions decreases while that in the equatorial regions increases proportionately (this is not the case for the present Sun). We have investigated a simple model in which both poles are capped down to a stellar latitude $b$. The average enhancement in equatorial flux is then given by

$$\frac{F}{F_0} = \frac{2 \sin \frac{b}{2}}{\sin b}$$

The theoretical maximum enhancement is a factor of $4/\pi$ for 100% coverage (i.e., $b = 0$, corresponding to an infinitesimal equatorial ring). To obtain the early Mars minimum enhancement of 1.13 would require a coverage of $\sim 40\%$. Taking into account limb darkening would reduce the required coverage somewhat. Additional photometric observations of single G dwarfs at the age of the Hyades are needed to see if the amount of coverage at this age is consistent with the focusing model.

Acknowledgments. We thank Dana Backman, Der- moumt Muller and Pat Whitman for valuable suggestions. D.P.W. and J.J.M. acknowledge the Louisiana Educational Quality Support Fund for partial support of this work.

(Received March 9, 1994; revised October 24, 1994; accepted November 23, 1994.)
Diagnostic calculations of the circulation in the Martian atmosphere

Michelle L. Santee
Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena

David Crisp
Earth and Space Sciences Division, Jet Propulsion Laboratory, California Institute of Technology, Pasadena

Abstract. The circulation of the Martian atmosphere during late southern summer is derived from atmospheric temperature and dust distributions retrieved from a subset of the Mariner 9 infrared interferometer spectrometer (IRIS) thermal emission spectra ($L_s = 343^\circ-348^\circ$) (Santee and Crisp, 1993). Zonal-mean zonal winds are calculated by assuming gradient wind balance and zero surface zonal wind. Both hemispheres have intense midlatitude westerly jets with velocities of 80-90 m/s near 50 km; in the southern tropics the winds are easterly with velocities of 40 m/s near 50 km. The net effect of the zonal-mean meridional circulation and large-scale waves can be approximated by the diabatic circulation, which is defined from the atmospheric thermal structure and net radiative heating rates. The radiative transfer model described by Crisp (1990) and Santee (1993) is used to compute solar heating and thermal cooling rates from diurnal averages of the retrieved IRIS temperature and dust distributions. At pressures below 4 mbar, there are large net radiative heating rates (up to 5 K/day) in the equatorial region and large net radiative cooling rates (up to 12 K/day) in the polar regions. These net radiative heating rates are used in a diagnostic stream function model which solves for the meridional and vertical components of the diabatic circulation simultaneously. We find a two-cell circulation, with rising motion over the equator, poleward flow in both hemispheres, sinking motion over both polar regions, and return flow in the lowest atmospheric levels. The maximum poleward velocity is 3 m/s in the tropics at ~55 km altitude, and the maximum vertical velocity is 2.5 cm/s downward over the north pole at ~60 km altitude. If these large transport rates are sustained for an entire season, the Martian atmosphere above the 1-mbar level is overturned in about 38 days. This diabatic circulation is qualitatively similar to the terrestrial diabatic circulation at the comparable season, but is more vigorous.

1. Introduction

Atmospheric transport may be a crucial element in the seasonal cycles of carbon dioxide, dust, and water on Mars, but the extent of its contribution remains uncertain. The only in situ wind measurements were made by the meteorology experiments onboard the Viking landers [Hess et al., 1977]. They acquired detailed temporal records of wind speed and direction, but only for two locations on the planet and only at a fixed elevation (~1.6 m) above the surface. Several authors [e.g., Ryan et al., 1978; Ryan and Henry, 1979; Tillman et al., 1979; Barnes 1980, 1981] have analyzed these data to provide constraints on larger-scale processes and the general circulation.

Recently, Lellouch et al. [1991] have reported the first measurements of winds in the middle atmosphere of Mars (40-70 km), obtained from Doppler shifts of the $J = 2 \rightarrow 1 \, ^{13} CO$ rotational line. Taken during the 1988 opposition, these observations provide extensive coverage of the southern hemisphere near southern summer solstice ($L_s = 279^\circ$). Their results are consistent with the thermal structure of the Martian middle atmosphere derived from infrared heterodyne spectroscopy [Deming et al., 1986; Rothermel et al., 1988]. They are also in general agreement with the modeling results of Barnes [1990, 1991], Pollack et al. [1990], and Haberle et al.
(1993). However, the Lellouch et al. measurements cannot be used to infer wind fields below 40 km.

There have been no direct systematic measurements of altitudes between the surface and approximately 40 km. In the absence of direct wind measurements, exhaustive compilations of cloud structures, including their spatial and temporal variations, have been undertaken for both Mariner 9 and Viking Orbiter images [Loopy et al., 1973; Briggs and Loopy, 1974; Briggs et al., 1977; French et al., 1981; Keesee, 1983]. Although studies of cloud motions and morphology have yielded wind directions, and in some cases, estimates of wind speeds, they suffer from a lack of substantial global coverage. Because regional winds arise from a variety of factors, including the presence of slopes or strong horizontal surface temperature variations (e.g., at the edges of the polar caps or in regions of large albedo or thermal inertia contrasts), these observed winds may not be representative of the mean flow.

Winds in the intermediate region between the surface and the upper portion of the middle atmosphere can be determined indirectly from the observed atmospheric temperature distribution. The thermal structure of the Martian atmosphere has been investigated by remote-sensing instruments on several different spacecraft. The best existing data for studying the vertical temperature distribution on a global scale are thermal emission spectra recorded by the Mariner 9 infrared interferometer spectrometer (IRIS). Using a new technique for the simultaneous retrievals of atmospheric temperatures and airborne dust abundances from IRIS spectra, we derived a global description of the thermal structure and dust loading of the Martian atmosphere for a relatively clear period after a severe dust storm in late southern summer [Santos and Crips, 1983] (hereinafter Paper I).

In Paper I we examined a subset of the IRIS data consisting of approximately 2400 spectra in a 12-day interval extending between $L_s = 343^\circ$ and $348^\circ$. Temperature and dust optical depth profiles were retrieved from each spectrum, and global maps were constructed as functions of latitude $(\lambda = 90^\circ)$, altitude (0–60 km), and Mars local time.

One of the principal conclusions from Paper I is that both daytime and nighttime atmospheric temperatures at altitudes above about 40 km are warmer over the winter (north) polar regions than over the equator or the summer (south) polar regions. These anomalous temperatures are consistent with the thermal structure in the middle atmosphere observed by Doming et al. [1986] and Retherfurd et al. [1988], and they indicate that the atmosphere is not in radiative equilibrium. A similar reversed temperature gradient is present in the terrestrial atmosphere and is known to be associated with the meridional circulation, which produces enough adiabatic heating (or cooling) to overcome the radiative forcing [Loopy, 1964].

In this paper we describe methods for deriving the atmospheric circulation from observed temperatures. Zonal-mean zonal winds are calculated from the observed meridional gradients of the zonally averaged temperature using the gradient thermal wind equation (section 2.1). Both the zonal-mean meridional circulation and large-scale waves contribute to the north-south atmospheric transport. Dunkerton [1978] shows that their net effect can be approximated by the diabatic circulation, i.e., that circulation needed to maintain the observed temperature distribution (warm winter pole, cool tropics) in the presence of the radiative drive. The Transformed Eulerian–Mean (TEM) formulation for the diabatic circulation is discussed in section 2.2. In section 2.3, we present a stream function model which solves for the meridional and vertical components of the diabatic circulation. This model requires values for the net radiative heating rates, which we derive from a radiative transfer model [Crisp, 1990; Santos, 1990] that evaluates the diurnally averaged solar heating and thermal cooling rates for the observed atmospheric temperature and dust distributions.

The results of the zonal wind and diabatic circulation calculations are discussed in section 3. Although similar calculations of the zonal-mean zonal wind have been reported previously [Crowther, 1981; Loopy, 1985; Michalsinski et al., 1987], they have used Mariner 9 meridional cross sections of temperature which were not global in coverage. In addition, although numerical simulations of the Martian atmosphere have been conducted using a full three-dimensional general circulation model [Pollack et al., 1981; 1990; Haberle et al., 1993] and a two-dimensional coupled chemical, radiative, and dynamical model [Moran et al., 1991], our results constitute the first estimation of the diabatic circulation of the Martian atmosphere from actual temperature observations.

2. Diagnostic Calculations of the Atmospheric Circulation

2.1. The Gradient Thermal Winds

In describing the large-scale circulation of the atmosphere, it is common to mathematically distinguish between the zonally symmetric flow and the zonal disturbances (usually called "eddies" or "waves") superimposed on it. Any field variable can be written as the sum of the zonal mean part, denoted by an overbar, and the deviation from the mean, denoted by a prime. Using the zonal wind $u$ for illustration,

$$u(\lambda, \phi, z, t) = \overline{u}(\lambda, \phi, z) + u'(\lambda, \phi, z, t),$$

where $t$ is time, $\lambda$ is longitude, and $\phi$ is latitude. The vertical coordinate $z$ used throughout this paper is not the geometric height but rather the log-pressure altitude: $z = -H \ln(p)$, where $p$ is the atmospheric pressure and the subscript $s$ denotes a value at the surface. $H$ is the mean atmospheric scale height ($H = R T_s / g$) calculated by assuming a constant reference temperature $T_s$ and a mean gas constant $R$, $g$ is the gravitational acceleration (assumed constant).
One definition of the zonal averaging operator is

\[ \bar{u}(\phi, z, t) = \frac{1}{2\pi} \int \lambda \phi, \lambda, t) d\lambda. \]  

(2)

This average is an Eulerian mean, taken at fixed values of the coordinates \( \phi, z, \) and \( t \). Applying the expansion of equation (1) and the zonal averaging operator of equation (2) to the primitive equations of atmospheric motion leads to a reformulated set of equations in which the zonal circulation and the eddy terms have been separated [e.g., Andrews et al., 1987].

The zonal-mean zonal winds can be estimated directly from the observed thermal structure of the atmosphere by combining the meridional momentum equation with the hydrostatic equation to yield

\[ f = \frac{2\pi}{a} \tan \phi \]  

\[ \frac{\partial u}{\partial t} - \frac{R}{aH} \frac{\partial^2 u}{\partial \phi^2} = 0. \]  

(3)

Equation (3), the gradient thermal wind equation, specifies the vertical wind shear due to meridional temperature gradients; \( a \) is the mean radius of the planet, \( f = 2\pi \sin \phi \) is the Coriolis parameter, and \( \Omega \) is the rotation rate. The length of the Martian day is 24.6 hours, the rotation rate \( \Omega = 7.1 \times 10^{-6} \text{s}^{-1} \), and the Coriolis forces are as important on Mars as they are on Earth.

This equation describes a three-way balance between the Coriolis force, the centrifugal force, and the horizontal pressure gradient force. If the centrifugal acceleration term is small compared to the Coriolis term, it can be omitted to yield the geostrophic thermal wind equation:

\[ \frac{\partial u}{\partial t} - \frac{R}{aH} \frac{\partial^2 u}{\partial \phi^2} = 0. \]  

(4)

Equation (4) may not always be valid for Mars. Both equations (3) and (4) highlight the fact that the wind field is strongly coupled to the atmospheric temperature distribution.

To solve equation (3) for the zonal-mean zonal wind \( \bar{u} \), the derivatives with respect to \( \phi \) and \( z \) are approximated by finite differences [e.g., Holton and Williams, 1980]. We divide the \( \phi, z \) space into a grid of \( M \times N \) points, with 10° latitude bins and 30 vertical levels between 6.0 mbar (taken to be the "surface") and 10^{-9} mbar. The grid points are evenly spaced in both the horizontal and vertical directions, and any point on the grid is uniquely identified by the indices \( (m, n) \). We use centered finite difference formulas for the derivatives of the point (\( m, n \)) with respect to \( \phi \). However, because the quantity we are solving for appears in both the \( \partial^2 u/\partial \phi^2 \) and \( \partial u/\partial z \) terms, a forward difference formula is more appropriate for the derivatives with respect to \( z \). Applying the finite difference formulas to equation (3) and grouping like terms together yields a quadratic equation for \( \bar{u}_{m,n} \):

\[ \frac{2a \tan \phi}{a + \phi} \frac{u_{m+1,n} - u_{m,n}}{z_{m+1} - z_{m}} - \frac{2a \tan \phi}{a + \phi} \frac{u_{m,n+1} - u_{m,n}}{z_{n+1} - z_{n}} \bar{u}_{m,n} = 0. \]  

Because the solution for \( \bar{u}_{m,n} \) involves the value of \( \bar{u}_{m+1,n} \), the winds at some reference level must be known. Typically, \( \bar{u} \) is assumed to be zero at the surface. Evidence that surface winds on Mars are generally light (less than 10 m/s) comes from the meteorology experiments onboard the Viking landers [Hsu et al., 1977; Ryan et al., 1978; Ryan and Henry, 1970; Barnes, 1980, 1981] and from constraints imposed by observations of resonant gravity waves in the winter polar region [Briggs and Leovy, 1974].

There are circumstances under which surface winds are not negligible. Viking lander observations revealed that surface winds rose steadily with the onset of global dust storms, with gusts reaching 30 m/s [Ryan and Henry, 1979]. Numerical simulations of global dust storms predict that (up to a point) higher dust opacities are accompanied by stronger surface winds [Roveri et al., 1982]. Other modeling studies [Roveri et al., 1979] indicate that the magnitude of the surface wind is substantially enhanced (up to 30 m/s) when the condensation rate onto, or the sublimation rate from, the seasonal polar caps is at a maximum. Because this data subset represents a relatively clear period after a global dust storm has subsided (Paper I), and because the polar cap at this season is relatively stable (section 2.2), we assume negligible surface winds for the purposes of calculating the vertical wind shear.

2.2. The Diabatic Circulation

The evolution in the understanding of wave-mean flow interactions in the terrestrial atmosphere has been discussed in depth by Dunkerton [1972], Dunkerton [1980], Hess [1980], Matsumoto [1980], Mahlman et al. [1984], and others. These concepts were reviewed and applied to Mars by Santee [1983], and will only be briefly summarized here. The splitting of the flow into eddy fluxes and a mean meridional circulation is an artificial distinction imposed by the Eulerian-mean approach. In actuality, these two processes are highly interrelated. One of the earliest investigations to use a numerical general circulation model to study quantitatively the relative importance of the mean meridional circulation and the eddies in transporting trace substances in the terrestrial stratosphere was that of Hunt and Manabe [1968]. Their work demonstrated that eddy fluxes due to planetary waves in the stratosphere are almost counteracted by advection due to the mean meridional circulation. This near cancellation of the two transport processes is a consequence of the fact that stratospheric planetary waves approximately satisfy the "noninteraction" theorem. This theorem, which was established by several authors and generalized by Andrews and McIntyre [1970, 1978], states that waves which are linear, steady (constant amplitude), and conservative (not being internally forced or dissipated) induce a mean meridional cell which acts to identically
canceled the eddy fluxes of the waves themselves. Under these conditions, the waves cause no net transport to the basic flow. Because the eddy heat and momentum fluxes tend to force a compensating mean meridional circulation, the Eulerian-mean framework is generally the small residual of two large terms of opposite sign.

The near cancellation of the eddy and zonal-mean components suggests that the conventional Eulerian mean is not well suited for defining the transport circulation. An alternative is the Lagrangian mean. Instead of taking a simple zonal average along a latitude circle, the Lagrangian mean involves averaging with respect to a curved material tube consisting of a set of fluid parcels [e.g., Hsu, 1980; Matsuno, 1980]. In the absence of waves, the material tube coincides with a latitude circle; under the influence of waves, it is distorted into a wavy shape. The motion of the center of mass of the tube defines the Lagrangian-mean motion. The Lagrangian–mean and Eulerian–mean circulations are quite distinct in general and may, in fact, be of opposite sign. However, net mass transport is achieved entirely by the Lagrangian-mean flow [Dunkerton, 1978].

Although conceptually Lagrangian means are more appropriate for transport calculations, in practice the direct application of the Lagrangian–mean formalism to atmospheric data encounters serious technical difficulties. Eulerian quantities are much easier to calculate. Eulerian quantities are much easier to calculate.

An improvement over the traditional Eulerian formulation was suggested by Andrews and McIntyre [1976]. They introduced a "residual" mean meridional circulation \( \left( \bar{v}^*, \bar{w}^* \right) \) defined through the transformation

\[
\bar{v}^* = \bar{v} - \frac{1}{a \cos \phi} \left( \frac{\rho \bar{w} \bar{v}^*}{\frac{\partial \phi}{\partial \bar{z}}} \right)
\]

\[
\bar{w}^* = \bar{w} - \frac{1}{a \cos \phi} \left( \frac{\rho \bar{v} \bar{w}^*}{\frac{\partial \phi}{\partial \bar{z}}} \right)
\]

where \( \bar{v} \) and \( \bar{w} \) are the zonal-mean velocity components in the meridional and vertical directions, respectively, \( \rho \) is the atmospheric density, the \( \bar{v} \bar{w} \) term represents an eddy heat flux, and \( \phi \) is the specific heat at constant pressure, which for a pure CO2 atmosphere can be obtained from an expression of the form [Touloukian and Makita, 1970]

\[
\phi(T) = 443 \times 10^{-3} + 1.688 \times 10^{-7} T^2 + 3.470 \times 10^{-11} T^3.
\]

The term \( \phi(T) \) is the adiabatic lapse rate.

After using the definitions in equation (6) to substitute in for \( \bar{v} \) and \( \bar{w} \) in the Eulerian–mean version of the primitive equations and performing many algebraic manipulations, a set of TEM equations is obtained. For completeness, the full set of TEM equations is given here:

\[
\frac{\partial \bar{u}}{\partial \bar{t}} = \frac{\bar{u}^* \cos \phi}{a \cos \phi} \left( \bar{u} \cos \phi \right) - f \bar{v}^* + \bar{w} \frac{\partial \bar{u}}{\partial \bar{z}} =
\]

\[
\bar{v} + \frac{1}{\rho a \cos \phi} \nabla \cdot \bar{F} = \nabla \cdot \bar{F} (8a)
\]

\[
\frac{\partial \bar{v}}{\partial \bar{t}} + \frac{\bar{v}^* \cos \phi}{a \cos \phi} \left( \bar{v} \cos \phi \right) + \frac{1}{\rho a \cos \phi} \left( \bar{w} \bar{v}^* \cos \phi \right) = \frac{\partial \bar{v}}{\partial \bar{z}} = \nabla \cdot \bar{F} = \frac{\bar{g}}{\bar{a}^2} \frac{\partial H}{\partial \bar{z}} = \frac{\partial \bar{p}}{\partial \bar{z}}
\]

\[
\frac{\partial \bar{w}}{\partial \bar{t}} + \frac{\bar{w}^* \cos \phi}{a \cos \phi} \left( \bar{w} \cos \phi \right) + \frac{1}{\rho a \cos \phi} \left( \bar{u} \bar{w}^* \cos \phi \right) = \frac{\partial \bar{w}}{\partial \bar{z}} = \frac{\partial \bar{w}}{\partial \bar{z}} = \frac{\bar{g}}{\bar{a}^2} \frac{\partial H}{\partial \bar{z}} = \frac{\partial \bar{p}}{\partial \bar{z}}
\]

The first two equations express momentum balance in the zonal and meridional directions, respectively. In equation (8a), X represents the zonal component of friction or other unspecified nonconservative forces. The vector \( \bar{F} \) is called the Eliassen–Palm (EP) flux. The divergence of the EP flux is a direct measure of the degree of interaction between eddies and the mean flow at each latitude and height, and is therefore of prime importance in determining the total wave forcing of the zonal–mean state [Edmon et al., 1980]. Assuming steady state conditions and neglecting the drag term, equation (8a) reduces to a balance between the force per unit mass exerted on \( \bar{u} \) by the mean meridional circulation and the force per unit mass produced by the waves. Thus a nonzero EP flux divergence implies the existence of a nonzero residual mean circulation. In equation (8b), \( \bar{G} \) represents all the terms that disrupt the gradient wind balance between \( \bar{u} \) and the geopotential \( \bar{\Phi} \). In most meteorological applications, \( \bar{G} \) is small [Andrews et al., 1987]. The third equation, which is written in terms of the geopotential, expresses hydrostatic balance in the vertical.

Equation (8d), the continuity equation, describes the conservation of mass. The source and sink terms, represented by \( \partial \bar{u}/\partial \bar{t} \), are formally included in this equation to account for the condensation and sublimation of a significant fraction of the CO2 atmosphere at high latitudes over the course of a year. Approximately 25% of the atmospheric mass is cycled into and out of the seasonal polar caps [Hess et al., 1979], but these terms are most important during the fall, when the cap is growing, and during the spring, when it is receding. A corresponding large-amplitude seasonal oscillation in the daily–average atmospheric pressure was observed at both Viking lander sites [Hess et al., 1979]. There was a deep pressure minimum near \( L_s = 350^\circ \), signaling maximum CO2 accumulation onto the south polar cap at the end of southern winter. A secondary, much shallower minimum occurred near \( L_s = 346^\circ \), signaling maximum CO2 accumulation onto the north polar cap at the end of northern winter. The daily–average pressure at both sites remained relatively constant from \( L_s = 346^\circ \) to 350, suggesting that the north polar cap was stable during this period (at least during the Viking year). It is therefore probable that the north polar cap was neither
growing nor retreating in any significant way during the time this data subset was recorded. \(L_s = 343^\circ - 348^\circ\).

Furthermore, based on a systematic review of wind direction information obtained from a variety of indicators, Kärn [1983] found that recession of the polar cap in the north produces much smaller mass fluxes than in the south, and hence has very little influence on wind directions during early northern spring. He concluded that during the late winter and early spring seasons in the southern hemisphere, diabatic heat transport is determined by the temperature field, not cap flows. In addition, Pollock et al. [1999] used a three-dimensional general circulation model of the Mars atmosphere to conduct numerical simulations of polar processes for different seasons and found very small net diabatic heating. Simulations at \(L_s = 342^\circ\) (with a visible dust optical depth \(\tau = 1\) ) indicated that the importance of heat transport by the condensation flow (mass flow toward or away from polar regions where \(CO_2\) is condensing or subliming) is relatively small at this season, except at the very center of the cap. Consequently, we neglect the \(\partial \mathbf{v} / \partial t\) term for these calculations.

Finally, equation (8) is the thermodynamic energy equation. This equation relates the time rate of change of the temperature following the fluid motion to the net diabatic heating \(Q\). The term "diabatic" refers to a change in temperature involving the actual addition or removal of heat (as opposed to adiabatic heating). Diabatic processes include absorption of solar radiation, absorption and emission of infrared radiation, and latent heat release. Also, mixing of different air masses within a convective layer leads to diabatic heating through conduction, friction, and diffusion [Holten and Baader, 1978; Wallace and Hobbs, 1977]; however, for most meteorological applications these contributions are neglected. We also neglect the latent heat component because temperatures in our data subset rarely approach the condensation temperature. Therefore, the diabatic heating reduces to the net radiative heating.

As discussed above, in the Eulerian-mean framework the net transport is the small residual between nearly equal, but opposite, eddy and mean circulation terms. In the transformed mean formalism, the problem has been formulated directly in terms of the residual. One of the principal advantages to the TEM equations is that \(\mathbf{V}^*\) and \(\mathbf{w}^*\) represent the velocity components whose adiabatic heating and/or cooling is not balanced by eddy heat fluxes, and they are therefore the part of the mean circulation which is relevant for mass transport. A second advantage is that the residual mean circulation \((\mathbf{V}^*, \mathbf{w}^*)\) provides a description of atmospheric motions that is close to the actual time-averaged behavior of fluid parcels.

Daveberton [1978] argued that under certain approximations the residual (hence mass transport) circulation can be inferred solely from the Eulerian-mean, net diabatic circulation. He showed that the "diabatic circulation." The main approximation involves the omission of the "wave heating" term on the right-hand side of equation (8). The generalized noninteraction theorem can be used to prove that for steady, linear, conservative disturbances, the EP flux divergence and the wave heating both vanish [Andrews and McIntyre, 1976]. Andrews et al. [1987] state that the wave-heating term is negligible for quasi-geostrophic motions (in fact, it drops out of the quasi-geostrophic form of the TEM equations) and for steady, conservative gravity waves. Nichman and Leung [1986] estimated this term for the terrestrial midlatitude atmosphere from satellite data and found it to be much smaller than the other terms in the thermodynamic energy equation. Similar arguments have been made by other researchers [e.g., Holton and Wehrbein, 1980; Crisp, 1983; Solomon et al., 1988; Rosenfeld et al., 1987; Gille et al., 1987; Shue, 1988]. With these precedents, and in the absence of information about the atmosphere of Mars, we will also neglect the wave heating.

In actuality, because the observed climatological temperature field is far from the radiative equilibrium structure, the EP flux divergence must be nonzero and the waves must be transient, nonlinear, or nonconservative. Thus, the diabatic circulation approximation introduces an essential inconsistency: the wave heating is negligible, but the EP flux divergence is not. That is, eddy transience and dissipation are neglected in the thermodynamic energy equation, but not in the momentum equation. This inconsistency can be justified because the wave heating is most likely much smaller than the radiative heating, whereas the EP flux divergence is most likely much larger than the drag. It is believed that the "breaking" of vertically propagating gravity waves provides the majority of the forcing of a nonzero EP flux divergence in the terrestrial mesosphere [Holton, 1983; Andrews et al., 1987]. Barnes [1990, 1991] has suggested that breaking topographically forced gravity waves could explain the large deviations from radiative equilibrium that have been observed [Deming et al., 1986; Rothermel et al., 1988] in the middle atmosphere of Mars.

The terms in equation (8) associated with the meridional velocity \((\mathbf{V}^*, \mathbf{w}^*)\) and seasonal change \((\partial \mathbf{v} / \partial x\) are also customarily ignored [e.g., Dunkerton, 1978; Rosenfeld et al., 1987; Gille et al., 1987]. This leads to a very simplified balance equation:

\[
\mathbf{V}^* \left( \frac{\partial \mathbf{v}}{\partial z} + \frac{\partial \mathbf{w}}{\partial x} \right) = Q \tag{9}
\]

Equation (9) states that the net vertical transport can be obtained from the diabatic heating and the static stability alone. Equation (8) is often combined with the continuity equation (equation (4)) and used as an iterative procedure to estimate the residual meridional circulation \((\mathbf{V}^*, \mathbf{w}^*)\). Although this strategy provides an adequate first approximation to the diabatic circulation, it is predicated on an inherent discrepancy, and it depends strongly on a proper normalization of the net radiative heating rate to ensure that global mass balance constraints are satisfied (i.e., nonzero net heating rates indicate net mass transport into or out of the system; see, e.g., the review by Shue [1988]). Therefore, we chose not to use a procedure based on equation (9).
2.3. The Stream Function Model

In this paper we adopt an approach that solves for $\Psi$ and $\Psi'$ simultaneously. The meridional advection term is retained in the thermodynamic energy equation, but the $\partial T/\partial t$ term is of little consequence due to the limited seasonal extent (only 12 days) of the data subset analyzed in Paper I. Neglecting the $\partial T/\partial t$ and wave-heating terms, equation (8e) reduces to

$$
\frac{\partial \Psi}{\partial \phi} + \frac{\partial \Psi'}{\partial \phi} \left( \frac{\partial T}{\partial z} + \frac{\partial \Psi}{\partial z} \right) = \mathcal{Q}.
$$

This equation is coupled to the continuity equation through the definition of a zonal-mean stream function, $\Psi$. The form of $\Psi$ was adopted from Shia et al. [1989]:

$$
\Psi' = \frac{1}{\cos \phi} \frac{\partial \Psi}{\partial \phi} \left( \frac{\partial T}{\partial z} + \frac{\partial \Psi}{\partial z} \right)
$$

$$
\Psi = \frac{1}{\cos \phi} \frac{\partial \Psi}{\partial \phi} \left( \frac{\partial T}{\partial z} + \frac{\partial \Psi}{\partial z} \right)\sin \phi.
$$

Substituting these expressions into equation (10) yields

$$
\frac{1}{\cos \phi} \left( \frac{\partial \Psi}{\partial \phi} \frac{\partial \Psi}{\partial \phi} \right) \frac{\partial T}{\partial z} + \frac{1}{\cos \phi} \left( \frac{\partial \Psi}{\partial \phi} \frac{\partial \Psi}{\partial \phi} \right) \frac{\partial \Psi}{\partial \phi} \cos \phi = \mathcal{Q}.
$$

This mathematical "trick" changes the character of the equation and allows it to be integrated numerically. An additional benefit is that in this form of the equation, the diabatic heating appears only as a derivative with respect to latitude, eliminating the need for a heating rate normalization.

Our approach for solving equation (13) is described in Appendix A. The solution depends on the boundary conditions that are applied. We assume that $\Psi = 0$ on the boundaries of the domain, which are the south pole, the north pole, the surface, and the top of the atmosphere:

$$
\psi_{n,n} = \psi_{n-1,n} = \psi_{n,n-1} = \psi_{n+1,n} = 0.
$$

This requires that there is no mass flow through the boundaries, and assumes nonacceleration conditions. These assumptions may introduce some error near the surface, where eddy heat fluxes can be large. With these boundary conditions, we only need to solve for interior grid points, resulting in a reduction in the total dimensions of the grid such that $M = M - 2$, $N = N - 2$.

The stream function model was validated by computing the diabatic circulation of the terrestrial atmosphere for four representative months (January, April, July, October). The meridional temperature cross sections for these calculations were obtained from an 8-year average of National Meteorological Center (NMC) data (1979-1986, inclusive), and the ozone profiles were obtained from a 7-year average of stratospheric aerosol and gas experiment (SAGE I) data [M. Gerstell, personal communication, 1992]. The circulation above the 20-km level deduced from the NMC temperatures using the stream function model was compared to a number of other recent diagnostic calculations. These studies, which employ a variety of radiative transfer schemes and input temperature and ozone data, include Hitchman and Leovy [1986], Solomon et al. [1986], Rosenfield et al. [1987], Gille et al. [1987], Shae et al. [1989], and Yang et al. [1990]. Despite the differences in both data and models, all these studies produce generally similar fields, and our results compare favorably to them in overall structure and actual velocities.

3. Results and Discussion

We use the diagnostic methods described in section 2 to define two aspects of Martian atmospheric dynamics: the zonal winds, and the net meridional circulation. These calculations require knowledge of the atmospheric temperature structure. The retrieved 0200 LT and 1400 LT meridional cross sections of atmospheric temperature presented in Paper I were averaged to produce a diurnal-mean temperature distribution, which is shown in Figure 1a. In section 3.1 the zonal-mean zonal winds are discussed. Although the principal results are obtained from the gradient thermal wind equation, winds obtained from the geostrophic thermal wind equation are also included to facilitate comparisons with previous studies.

Solar heating rates and thermal cooling rates are calculated from the temperature (Figure 1a) and dust op-
Santee and Crisp: Circulation in the Martian Atmosphere

Figure 1a. Diurnal-mean temperatures obtained by averaging the 1400 LT and 0200 LT temperature maps retrieved from the Mariner 9 IRIS spectra in Paper I. For consistency with the results of Paper I, the vertical coordinate in this and all subsequent figures is the atmospheric pressure $p$. The approximate altitude $z$ corresponding to a given pressure level is also included in these figures (see section 2.1). This correlation of altitude with pressure is based on the following values: $p_f = 6.0 \text{ mbar}$, $R = 191.0 \text{ J K}^{-1} \text{ kg}^{-1}$, $g = 3.74 \text{ m/s}^2$, and $T_0 = 196 \text{ K}$ (the globally and diurnally averaged value of the atmospheric temperature at the surface), resulting in a mean scale height $H = 10 \text{ km}$.

Figure 1b. Dust optical depth distribution obtained from the 1400 LT opacity map retrieved from Mariner 9 IRIS spectra in Paper I.

Finally, in section 3.3 we present the vertical and meridional components of the diabatic circulation diagnosed from the net heating rates using the stream function model, along with the EP flux divergence calculated from the zonal momentum equation. In addition, the diabatic circulation is compared to results from previous modeling simulations, and some implications of the strength of the circulation are discussed.

3.1. Zonal–Mean Zonal Winds

Zonal winds computed from the gradient thermal wind equation (equation (3)) assuming zero wind velocity at the surface are shown in Figure 2a. One of the characteristics of the thermal wind equation is that the zonal wind increases with altitude in the presence of equator-to-pole temperature gradients. Because the horizontal temperature gradients are steeper in the northern (winter) midlatitudes (see Figure 1a), the eastward (westerly) winds aloft are slightly more intense over the northern regions, although both hemispheres have strong midlatitude eastward jets (with velocities of 80–90 m/s near 50 km). Winds in the equatorial regions, where the Coriolis parameter approaches zero, cannot be determined from gradient wind balance and are omitted from this figure. In the southern tropics the winds are westward (easterly), but there is only a small corresponding region of easterlies at the top of the domain in the northern hemisphere.

A second calculation was performed with the same temperatures but using the simpler geostrophic thermal wind equation (equation (4)). The overall morphology of the geostrophic thermal wind field (shown in Figure 2b) is consistent with that of the gradient thermal wind field (Figure 2a), but the geostrophic upper level westerly jets are up to 30 m/s stronger. A similar overestimation of the zonal–mean westerlies has been found in calculations of the terrestrial zonal winds. In a comparison of gradient and geostrophic mean zonal winds, Marks [1989] notes that their spatial structures are very

Figure 2a. Zonal–mean zonal winds, in m/s, calculated from the gradient thermal wind equation. Positive values represent eastward winds; negative contours are dashed, the zero contour is thicker, and the contour interval is nonuniform.
similar but that the maximum values of the positive gradient winds are up to 20% less than the corresponding geostrophic values, and that the differences between them are greatest in the regions of strong winds. This arises from the fact that the centrifugal acceleration term, absent in the geostrophic wind equation, is proportional to the zonal wind speed.

Several different calculations of the zonal winds based on Mariner 9 data have been reported previously, but none of these are global in coverage. Briggs and Leovy [1974] find evidence for resonant gravity waves in the middle latitudes in Mariner 9 television images taken during northern winter. Using gravity-wave theory, along with temperatures in the 45°-50°N latitude range obtained from inversion of IRIS spectra recorded on one orbit near \( L_S = 330° \) and a wavelength for the lee waves of 30 km, Briggs and Leovy estimate the surface zonal wind to be 10 m/s. From the geostrophic approximation they calculate a profile of the zonal wind up to 30-km altitude. Their results show the wind speed increasing up the column to a maximum value at the top of the domain of nearly 120 m/s.

Pollack et al. [1981] present geostrophic zonal winds computed from Mariner 9 IRIS data as one means of validating the NASA/Ames Mars general circulation model (GCM). These cross sections encompass the area between 50°S and 70°N, from the surface to the 0.3-mbar level, and include spectra spanning \( L_S = 43°-84° \). This season is roughly midway between the spring equinox and the summer solstice in the northern hemisphere. The zonal wind field is characterized by a strong eastward jet (with velocities exceeding 120 m/s) between 1.0 mbar and 0.5 mbar at 50°S latitude, much weaker westerly winds (with velocities of 20-30 m/s) at that altitude between 60°N and 80°N latitude, and eastward winds of unspecified magnitude in the region between the equator and 40°N. The occurrence of eastward winds in the northern tropics, rather than the southern tropics as in Figure 2, and the much weaker eastward winds in the northern high latitudes, are consequences of the fact that Pollack et al. [1981] analyzed IRIS spectra representing a different season from those of Paper I (late southern summer).

Geostrophic zonal winds calculated from Mariner 9 IRIS spectra at the same season as those of Paper I are reported by Conrath [1981]. Conrath constructs a northern hemisphere mean meridional temperature cross section by averaging temperature profiles retrieved from spectra recorded during \( L_S = 330°-350° \). He finds an eastward jet of 100 m/s at about 0.2 mbar and 60°N. Michelangeli et al. [1987] use the Conrath [1981] temperature distribution, but assume gradient wind balance to compute the zonal winds. They present a cross section of the zonal-mean gradient wind between 15°N-85°N and 5.0-0.2 mbar which exhibits the same features as Figure 2a. The ageostrophic correction to the zonal-mean zonal wind reduces the maximum eastward velocity from 160 m/s to 100 m/s, which is in better agreement with the results displayed in Figure 2a. Hoberle et al. [1993] use the NASA/Ames Mars GCM to simulate the response of the Martian atmospheric circulation to variations in season and dust loading. Comparisons with these GCM results provide insight into the behavior depicted in Figure 2. Because the interval represented by our data subset \( (L_S = 343°-348°) \) falls between solstice and equinox, we will examine the GCM results at both of these seasons. GCM experiments for northern winter solstice \( (L_S = 335°-340°) \) under relatively clear conditions reveal zonal winds which are easterly in the southern hemisphere (with peak velocities of 60 m/s) and westerly in the northern hemisphere (with peak velocities of 120 m/s). The intense westerly jet is centered near 80°N-60°N and 35 to 40-km altitude, whereas the eastward jet in the southern hemisphere is much broader, covering the entire tropical region at 40-km altitude. The GCM zonal wind field for a period just past northern spring equinox \( (L_S = 0°-22°) \) is very symmetric with respect to the equator. Westerlies appear in both hemispheres, but these westerly jets (peak velocities of 80 m/s at 50°S and 40 m/s at 60°N) are weaker than those present during solstice conditions due to weaker latitudinal temperature gradients. Basteries have formed in the tropical region of both hemispheres above 30-km altitude and near the surface, although not in the intermediate region of the atmosphere. Thus, Figure 2 represents a transitional period from a season in which the winds are easterly throughout the southern hemisphere and westerly throughout the northern hemisphere, to a season in which the winds are westerly at the higher latitudes of both hemispheres and easterly in the tropics.

Zonally averaged zonal wind and temperature fields from the NASA/Ames Mars GCM are presented by Barnes et al. [1983] for relatively low-dust conditions \( (r < 0.3) \) in late northern winter \( (L_S = 332°-348°) \). Because this experiment approximates well the conditions of our data subset in terms of both season and dust loading, it provides an opportunity to compare winds inferred from data and those simulated by the GCM.

---

**Figure 2b. Zonal-mean zonal winds, in m/s, calculated from the geostrophic thermal wind equation. Positive values represent eastward winds, negative contours are dashed, the zero contour is thicker, and the contour interval is nonuniform.**
Although westerlies appear in both hemispheres in the simulation, there is a much more pronounced interhemispheric difference in the strength of the jets than is seen in Figure 2. The peak velocity of the GCM northern hemisphere jet (centered near 50°N-60°N and 35 to 40-km altitude) exceeds 100 m/s, whereas the peak velocity of the GCM southern hemisphere jet (centered near 60°S and 40-km altitude) is only about 30 m/s.

This difference in peak intensity of the westerly jets is reduced to a factor of two when GCM fields for a more limited seasonal range (Ls = 343°-346°) are examined (J. Barnes, personal communication, 1994), but in either case the GCM late southern summer southern hemisphere jet is considerably weaker than that obtained from analysis of Mariner 9 data. Sensitivity tests indicate that a major portion of this difference is attributable to the fact that GCM temperatures at high southern latitudes, above the near-surface region, are roughly 10 K warmer than those in Figure 1a. One possible explanation for warmer simulated temperatures in this region is that the GCM horizontal dust distribution is uniform (except for topography effects), whereas the retrieved horizontal dust distribution falls off rapidly poleward of 40°S (see Figure 1b). Larger airborne dust abundances at high southern latitudes in the model would produce increased solar heating and warmer temperatures in this region, leading to weaker latitudinal temperature gradients and less intense westerlies.

3.2. Net Radiative Heating Rates

Net radiative heating rates for relatively low-dust conditions (r(μm) ~ 0.2) are shown in Figure 3. At pressures below 4.0 mbar, there are large net heating rates (up to 5 K/d) in the equatorial region and large net cooling rates (up to 12 K/d) in the polar regions. The heating and cooling rates are approximately symmetric about the equator, although the polar cooling is much weaker in the southern hemisphere (where the sun is shining). Because no large temperature changes occur during the 12 days represented by this subset of IRIS data (see Paper I), these results imply the existence of a nonradiative mechanism which acts to cool the tropics (by up to 5 K/d) and warm the polar regions (by up to 12 K/d) in order to maintain the observed temperature structure in the presence of this radiative forcing. For example, upwelling over the equator would produce expansional cooling to offset the net radiative heating, while downwelling over the poles would produce compressional heating to offset the net radiative cooling.

3.3. The Diabatic Circulation

3.3.1. Meridional and vertical velocities. The dynamical heat transport in the meridional plane that balances the net radiative heating is called the diabatic circulation. In the absence of wave transience or dissipation, the diabatic circulation represents the true net motion of air parcels. The diagnostic stream function model described in section 2.3 is used to simultaneously solve for the meridional and vertical components of the diabatic circulation. This model uses the diurnal-mean temperatures of Figure 1a and the net radiative heating rates of Figure 3. The resulting meridional and vertical velocities are depicted in Figures 4a and 4b, respectively. The mass-weighted stream function can be defined [Reiter, 1975]

\[ \psi_m = 2 \pi \rho_s c^2 \theta \phi. \]  

The density of the atmosphere at the surface is \( \rho_s = \rho_f H = 1.6 \times 10^{-2} \text{ kg/m}^3 \). Equation (15) was used to construct the streamlines of mass flow, which are shown in Figure 4c.

Figures 4a-4c depict a two-cell diabatic circulation, with rising motion in the equatorial region and sinking motion over both poles. Figures 4a and 4c show transport from the equator toward the poles through most of the atmosphere, with return flow in the low-
Comparison of Figure 3 and Figure 4b shows that the vertical velocity structure strongly reflects the diabatic heating distribution. The strength of the meridional circulation is directly proportional to the strength of the radiative drive because this circulation provides the necessary adiabatic heating and/or cooling compensation. The diabatic circulation presented in Figures 4a-4c is qualitatively similar to the terrestrial diabatic circulation at the comparable season, but it is more vigorous. Comparisons with the terrestrial diabatic circulation diagnosed for March or April by Solomon et al. [1986], Rosenfield et al. [1987], Gille et al. [1987], Shine [1989] and Yang et al. [1990] reveal that the late northern winter/late southern summer Martian circulation is roughly five times stronger. A similar factor is obtained from GCM results by Haberle et al. [1993].

The results of Figure 4 can be compared to those of Moreau et al. [1991], which were derived from a two-dimensional coupled chemical, radiative, and dynamical model of the dust-free Martian atmosphere. Moreau et al. obtain the meridional distribution of temperature at $L_{SO} = 551^\circ$ by explicitly including radiative transfer and dynamical heat transport in the solution of the thermo-dynamic energy equation. Their transport model describes a transformed Eulerian mean meridional circulation which is forced by gravity wave dissipation, but neglects the effects of atmospheric tides. Perhaps the most serious shortcoming of their model is that it completely neglects the radiative impact of airborne dust, which dominates the solar heating even during relatively "clear" periods like that examined here. Their radiative transfer model includes other limitations as well. For example, they use an empirical parameterization for the CO$_2$ absorption at near-infrared wavelengths which does not include Voigt effects, and consequently is not valid for the low-pressure Martian atmosphere. Their thermal cooling rate model also neglects Voigt effects, and above 40-km altitude employs a Newtonian cooling formulation. These simplifications result in modeled temperatures which are about 20 K cooler than those of Figure 1a throughout most of the atmosphere. They present a plot of the stream function, the general character of which agrees well with Figure 4c. Although they do not include plots of the velocity fields, they quote values for the maximum meridional velocity of 1.6 m/s at 50 km and 45°N latitude, and the maximum vertical velocity downward over the polar regions of 1.3 cm/s. The diabatic circulation computed from IRIS observations (Figure 4) is stronger than the circulation derived by Moreau et al. [1991]. We find a maximum meridional velocity of 3 m/s at about 55 km at 20°N latitude, and a maximum downward vertical velocity of 2.5 cm/s at 60 km over the north pole.

The Martian atmosphere contains substantial quantities of dust even during non-dust-storm conditions, and we account for the consequent enhancement of the radiative forcing in our diagnostic models. Because the thermal structure the atmosphere would have in the absence of dust is not an observable quantity, our diagnostic model cannot be used to infer the dust-free cir-
calculation. However, model simulations [Haberre et al., 1982, 1993] indicate that the introduction of dust into the atmosphere causes a significant strengthening of the circulation. The stream function model results can also be compared to the Mars GCM results of Haberre et al. [1993] (see the discussion of this model in section 3.1). Haberre et al. present plots of the mass stream function and meridional and vertical velocities for the northern spring equinox ($L_s = 90^\circ - 23^\circ$). This circulation consists of two cells of fairly comparable strength whose common rising branch is centered just south of the equator. The peak mass flux associated with these cells ($10^{19}$ kg/s) is in good agreement with the maximum contour of Figure 4c. The overall pattern of the vertical velocity conforms well to that of Figure 4b, although the GCM-generated field exhibits substantially more structure and smaller values than the stream function model results. The GCM meridional velocities are also smaller than results from the stream function model, and they tend to be negative over a larger range of latitudes.

3.3.2. EP flux divergence. We can now calculate the EP flux divergence from equation (8a) by assuming steady state conditions and neglecting drag. Figure 5 shows the quantity $(p\cos \phi)^{-1} \nabla \cdot \mathbf{F}$, which is a mean zonal force per unit mass, scaled to units of $\text{ms}^{-1}/\text{d}$. At most levels of the atmosphere in both hemispheres, the EP flux divergence is negative. In these regions, the waves exert a negative force (i.e., a drag) on the zonal mean flow. This forcing is consistent with an induced residual circulation having a poleward meridional velocity at high altitudes and a downward vertical velocity over the poles. Examination of the magnitudes of the individual terms on the left-hand side of equation (8a) reveals that the force produced by the mean meridional circulation is dominated by the Coriolis torque $(F_C)$, although the contributions from the other terms are not negligible. For this reason, the $\nabla \cdot \mathbf{F}$ and $\mathbf{\nabla} \mathbf{v}$ fields are strongly correlated. The magnitude of the EP flux divergence is also correlated with the values of $\nabla \cdot \mathbf{v}$ (cf. Figure 4b and Figure 5).

The EP flux divergence identifies regions where the noninteraction theorem is being violated. As discussed in section 2.2, in regions where the $\nabla \cdot \mathbf{F}$ values are large, the waves must be strongly transient, nonconservative, or nonlinear. In Figure 5, the largest (negative) values of $\nabla \cdot \mathbf{F}$ occur at high latitudes in the winter hemisphere at altitudes above about 40 km ($=0.1$ mbar). This coincides with the area distinguished by temperatures far in excess of the radiative equilibrium values [see Paper 1; also Deming et al. [1980] and Rothermel et al. [1988]). The results of Figure 5 are consistent with those of Barnes [1990], who uses a simplified dynamical model of the zonal mean flow to investigate the effects of breaking topographically forced internal gravity waves on the circulation in the middle Martian atmosphere. His model results show that breaking gravity waves induce little high-latitude warming at levels below $\sim 49$ km; however, they have a significant impact in the 40 to 80-km region, generating a strong mean meridional circulation that acts to produce very warm winter polar temperatures.

3.3.3. Overturning time. There is no consensus on the nomenclature for the different altitude regions of the Martian atmosphere. For the sake of analogy with terrestrial conventions, in this section we will call the region of the atmosphere with $p > 1$ mbar the "stratosphere" (although this designation is sometimes given to the region from the ground to 45 km [Zurek, 1992]), and the region with $p < 1$ mbar the "stratosphere" (although it could also be called the mesosphere [Zurek, 1992]). Thus, the net meridional circulation described by Figures 4a-4c provides a mechanism for exchange between "tropospheric" and "stratospheric" air masses. The tropical branch of the circulation carries significant amounts of tropospheric air into the stratosphere, which, for reasons of mass continuity, must return to the troposphere at high latitudes. If advection by the net meridional circulation is the most important process involved in troposphere-stratosphere exchange, then the characteristic advective exchange time, or residence time, for parcels in the stratosphere is defined [Reiter, 1975; Chamberlain and Hunten, 1987]:

$$t_A^1 = \frac{1}{\dot{M}} \frac{dM}{dt}$$

where $\dot{M}$ is the total stratospheric air mass. We estimate the mass of the Martian "stratosphere" (the atmosphere above 1 mbar, or approximately 20 km) to be $\dot{M} = 4 \times 10^{18}$ kg. The $dM/dt$ term in equation (16) is estimated from the mass-weighted stream function $\Psi_m$. The physical meaning of $\Psi_m$ can be illuminated by substituting equation (15) into equations (11a) and (11b) and rearranging:

$$\Psi_m(\phi, t) = -2\pi p_a \cos \phi \int_0^t \int_{-\infty}^{\infty} \mathbf{v} \cdot \mathbf{F} \, d\lambda \, dt$$

Figure 5. Eliassen-Palm flux divergence ($10^{-1} \nabla \cdot \mathbf{F}$) in units of $\text{ms}^{-1}/\text{d}$. Negative contours are dashed, the zero contour is thicker, and the contour interval is nonuniform.
The first expression represents the vertically integrated mass flux in the region from 0 to $z$ across latitude $\phi$, and the second expression represents the horizontally integrated mass flux in the region from $-\xi$ to $z$ across a pressure surface, denoted by the log-pressure coordinate $\xi$. The calculation of the characteristic timescale for troposphere-stratosphere exchange requires a value for the integrated mass flux across the 1-mb pressure surface in the region of equatorial upwelling where the diabatic vertical velocity $w$ is positive. Therefore, from Figures 4b and 4c and equation (17b), the term $dM/dt$ can be approximated by \( \Psi_m(20^0N, 1\text{ mbar}) - \Psi_m(40^0S, 1\text{ mbar}) \approx 1.2 \times 10^4 \text{ kg/s}. \) This value of $dM/dt$ leads to a residence time $t_R \approx 3.2 \times 10^6 \text{ s} \approx 38 \text{ days}$. 

Because fluid parcels are constrained to travel around closed streamlines, to satisfy mass continuity, the horizontal and vertical transport times must be equal. If these large transport rates are sustained for an entire season, the Martian atmosphere above the 1-mb level is ventilated in about 38 days. This value of the exchange time is roughly comparable to modeling results obtained by Barnes [1986]. Barnes computes vertical velocities of several centimeters per second, and concludes that the exchange time between the lower portion of the middle atmosphere (10-40 km) and the upper portion of the middle atmosphere (40-100 km) is 5-10 days. In sharp contrast, Shin et al. [1989] show that it takes about 1.8 years for the terrestrial stratosphere to be completely overturned.

3.3.4. Eddy diffusion coefficient. The simplest approach to modeling meridional and vertical transport in the Martian atmosphere is to parameterize these processes with eddy diffusion coefficients. Then the characteristic vertical diffusion time, the dynamical time associated with vertical mixing, is given by [Chamberlain and Hunten, 1987]

\[
t_{\text{R}} = \frac{H^2}{K_{zz}},
\]

where $K_{zz}$ is the vertical eddy diffusion coefficient. $K_{zz}$ has been estimated by several authors using different methods. Comrie [1975] deduces the dust optical depth during the dissipation phase of the 1971 dust storm from Mariner 9 IRIS temperature profiles and a simplified atmospheric heating model. To explain the comparable exponential decay rates of the dust opacity at two different atmospheric levels, he requires a vertical eddy diffusion coefficient $K_{zz} \geq 10^2 \text{ m}^2/\text{s}$ between the surface and 50 km. Zurek [1976] estimates the eddy diffusion coefficient for the turbulence generated by the diurnal atmospheric tide in the middle Martian atmosphere (30-40 km) to be $K_{zz} \sim 5 \times 10^2 \text{ m}^2/\text{s}$. Kong and McElroy [1977] investigate a variety of photochemical models and also conclude that a mixing coefficient of about $K_{zz} \sim 5 \times 10^2 \text{ m}^2/\text{s}$ at 40 km is required to match the observed abundances of various atmospheric constituents. Toon et al. [1977] use a multiple scattering radiative transfer code to find the best fit synthetic spectra to Mariner 9 IRIS observations. They conclude that the size distribution of the dust particles did not change appreciably during the dissipation of the dust storm, and they invoke vertical mixing on the order of $10^3 \text{ m}^2/\text{s}$ to explain the constancy of the distribution. Anderson and Leovy [1978] examine Mariner 9 television reflectance profiles which crossed the limb of Mars. From an analysis of the decay of dust scale heights they determine the vertical eddy mixing coefficient to be $K_{zz} \sim 10^3 \text{ m}^2/\text{s}$ in the 40 to 60-km height range.

Using a value of $K_{zz} = 10^2 \text{ m}^2/\text{s}$ in equation (18) yields $t_R = 10^6 \text{ s} \sim 3.2 \text{ days}$. This estimate of the transport timescale due to diffusion is an order of magnitude less than the estimate of the advective timescale. If correct, this would imply that vertical transport in the Martian atmosphere is primarily diffusive, rather than advective, in nature. In contrast, based on an upper limit of $K_{zz} = 1 \text{ m}^2/\text{s}$ for the stratosphere of the Earth from Shin et al. [1986], the characteristic diffusion time for the terrestrial stratosphere is 1.5 years, comparable to the timescale due to advection, and both processes are important for tracer transport.

In actuality, the temperature and net heating rate fields calculated from the Mariner 9 spectra provide strong evidence for the importance of advection in the Martian atmosphere because they require the upgradient meridional transport of the heat rather than advection. The determinations of the eddy diffusion coefficient outlined above parameterize all contributions to the vertical transport. The derived values of the eddy diffusion coefficient therefore include the effects of large-scale organized motions, such as the mean meridional circulation, and are consequently much larger than they would be if they represented only the small-scale eddy mixing processes. Although these large values of $K_{zz}$ may be valid for the dustiest conditions or in limited regions, e.g., within the boundary layer, they are not appropriate for application on a global scale.

Recent photochemical models do not appear to require as strong vertical mixing in the lowest region of the atmosphere as the earlier calculations indicated [Nair et al., 1994, and references therein]. Moreau et al. [1991] calculate a meridional cross section of $K_{zz}$ for $L = 63^1$ using a two-dimensional model which accounts for both advection due to the residual mean meridional circulation and eddy mixing associated with absorption of waves. As discussed in section 3.3.1, several simplifications in their radiative transfer model and wave parameterization scheme lead to a residual circulation that is weaker than that of Figure 4. They find that a value of $K_{zz} = 30 \text{ m}^2/\text{s}$ is representative of the majority of the atmosphere below 50-km altitude. This value of $K_{zz}$ yields $t_R = 3.8 \times 10^6 \text{ s} \sim 38 \text{ days}$, in agreement with the timescale for vertical advection calculated above. Using dynamical arguments, Phumb and Mahman [1987] maintain that advective and dif-
fusive transport effects have comparable magnitudes in the terrestrial atmosphere (globally, but not necessarily locally). Although the information currently available is insufficient to prove conclusively whether either advective or diffusive processes dominate in the Martian atmosphere, the assumption that their transport timescales are equivalent leads to a vertical eddy diffusion coefficient two orders of magnitude smaller than the commonly used value.

3.3.5. Transport issues. The influence of the meridional circulation on a chemical species can be evaluated by comparing the dynamical time constant to the photochemical time constant, which is calculated for each species from its rate of loss. One of the most radiatively important trace gases in the Martian atmosphere is ozone. O₃ molecules are temporarily removed from the atmosphere by photolysis, but its overall abundance is not affected because of the rapid recombination of O and O₂. The four main reactions leading to the destruction of odd oxygen (O + O₂) in the Martian atmosphere are:

\[
\begin{align*}
\text{NO}_2 + O & \rightarrow O_3 + OH \\
O + OH & \rightarrow O_3 + H \\
O_3 + H & \rightarrow O_3 + OH \\
\text{NO}_2 + O & \rightarrow \text{NO} + O_2.
\end{align*}
\]  

The chemical lifetime for odd oxygen is then given by:

\[
\tau_{O^+} = \frac{|[O_3]|}{k_1[O][\text{NO}_2] + k_2[O][OH] + k_3[\text{O}_3][H] + k_4[\text{NO}_2][O]},
\]  

where the square brackets signify a concentration in molecules/cm³, and k₁–k₄ are the rate constants for reactions (10a)–(10d), respectively. These reactions are evaluated for midlatitude conditions using the Caltech-JPL one-dimensional photochemical model of the Martian atmosphere [You et al., 1988; Nair et al., 1994]. Nair et al. [1994] find that good agreement between modeled and observed CO, O₂, and O₃ can be achieved by varying the rate constants for a few key reactions within their published recommended uncertainties. All model results presented here are obtained using the modified rate constants. Two different water vapor profiles are considered: a low-water case with an integrated column abundance of 8.0 pr μm, and a high-water case with an integrated column abundance of 8.0 pr μm. The water vapor is assumed to be well mixed up to the altitude where condensation first occurs, above which it roughly follows the saturation curve. Results from the two water vapor cases are sufficiently similar for our purposes that only the high-water values are included here. Further details of the photochemical model initialization and constraints can be found in Nair et al. [1994]. Model results at 40°-km altitude (H. Nair, personal communication, 1994) indicate that the chemical time constant for odd oxygen is about \(\tau_{O^+} = 4200 \text{ s} = 1.2 \text{ hours}\). (For comparison, the photochemical lifetime of ozone based solely on photolysis is 625 s.) Because the photochemical time constant is much shorter than the dynamical time constant, ozone is in a state of photochemical equilibrium. Although the circulation may affect the ozone distribution indirectly, through the effects of temperature or by altering the concentration of water vapor (to which ozone is strongly coupled), it does not have a direct impact.

There are atmospheric constituents, e.g., CO, whose distributions are controlled by dynamics. The only significant loss mechanism for CO in the Martian atmosphere is

\[
\text{CO} + \text{OH} \rightarrow \text{CO}_2 + \text{H}.
\]  

The chemical lifetime for CO is \(\tau_{\text{CO}} = 1.5 \times 10^7 \text{ s} \approx 175 \text{ days}\) (H. Nair, personal communication, 1994). CO should therefore be sufficiently long-lived to be thoroughly mixed by atmospheric motions. This conclusion is supported by the work of Clancy et al. [1990], who show that ground-based microwave spectra of CO in the atmosphere of Mars can be fit well by assuming a CO mixing ratio which is constant in altitude and latitude. However, photochemical model results [Nair et al., 1994] indicate that the CO mixing ratio actually increases slightly with altitude for a \(k_{\mu_\tau}\) profile varying exponentially from 10 to 1000 m²/s between the surface and 40 km.

A similar comparison can be undertaken for airborne dust, although in this case the only removal process is gravitational settling. Following Conrath [1970], the terminal velocity \(w_T\) of the dust particles is given by the Stokes–Cunningham relation for spherical particles for the case where the molecular mean free path \(\lambda\) is large compared to the particle diameter \(d\):

\[
w_T = \left[ \frac{6 \pi \eta d}{m} \right] \left[ 1 + \frac{2}{d} \left( \frac{\lambda}{\mu} \right) \right].
\]  

The following values are assumed: dust particle density \(m = 3 \times 10^3 \text{ kg/m}^3\); molecular viscosity \(\eta = 1.5 \times 10^{-5} \text{ kg/m/s}\); and reference pressure \(P_r = 25 \text{ mbar}\) at which \(\lambda = 2.2 \times 10^{-4} \text{ m}\). Once the fall velocity has been calculated for given values of \(d\) and \(\mu\), the characteristic settling time is found by

\[
t_s = \frac{H}{w_T}.
\]  

The characteristic settling time for a 1-μm-diameter particle at 1 mbar is \(t_s \approx 20 \text{ days}\). The fact that the settling time is comparable to the timescale for vertical advection implies that the dust is well mixed. Any 1-μm dust particles entrained by local dust storms at 1 mbar (~20 km), which is the upper limit on the local dust cloud height inferred by Peterfreund and Kieffer [1979], have a residence time sufficiently long to allow for horizontal transport. Assuming the dust remains suspended for 25 days and 1 m/s is a representative meridional velocity, the dust particles would be trans-
ported 2100 km, or approximately 40% of the distance between the equator and the pole. In contrast, the characteristic settling time for a 5-μm-diameter particle varies from 4 days at 1 mbar, to 10-20 days at 5 mbar (~2 km). The ascending motion in the non-perturbed atmosphere, after the dissipation of the dust storm, is not intense enough to support dust particles of this size over large vertical extent.

4. Summary

We presented diagnostic methods for defining the circulation of the atmosphere from a knowledge of its thermal structure. We used diurnal averages of meridional cross sections of atmospheric temperature and dust optical depth retrieved from Mariner 9 IRIS thermal emission spectra for a 12-day interval in late southern summer ($L_2 = 243°-368°$). The retrieval methods and the resulting global maps of temperature and dust optical depth are described by Santer and Crisp [1993].

Zonal-mean zonal winds were derived from the meridional gradients of the observed temperatures assuming gradient wind balance and zero surface zonal wind. Outside of the tropics in both hemispheres, the zonal winds are westerly, with intense midlatitude jets near 50- km altitude. In the southern tropics the winds are easterly at all altitudes, but there is only a small corresponding region of easterlies near 60 km in the northern tropics.

Solar heating and thermal cooling rates were computed from the diurnally averaged IRIS temperature and dust distributions using the radiative transfer model described by Crisp [1990] and in the appendix of Santer [1993]. At pressures below 4 mbar, there are large net radiative heating rates (up to 5 K/d) in the equatorial region and large net radiative cooling rates (up to 12 K/d) in the polar regions.

The diabatic meridional circulation was calculated from the net radiative heating rates using a diagnostic stream function model which solves for the vertical and meridional velocities simultaneously. At this season, the circulation consists of two cells, with rising motion over the equatorial region, poleward flow in both hemispheres, sinking motion over both polar regions, and return flow in the lowest atmospheric levels. The maximum meridional velocity is 3 m/s at about 55 km at 20°N latitude. The vertical velocity field is fairly symmetric; although the winds in the winter hemisphere are stronger, with a maximum downward vertical velocity of 2.5 cm/s at 60 km over the north pole. If these large transport rates are sustained for an entire season, the Martian atmosphere above the 1 mbar level is ventilated in about 38 days. The assumption that the advective and diffusive transport timescales should be comparable leads to a vertical eddy diffusion coefficient of about 30 m$^2$/s, two orders of magnitude smaller than the commonly used value.

The EP flux divergence was calculated from the zonal momentum equation. This quantity identifies regions where the noninteraction theorem is being violated. At most levels of the atmosphere in both hemispheres, the EP flux divergence is negative, indicating that the waves exert a drag on the zonal-mean flow. The largest values occur at high latitudes in the winter hemisphere at altitudes above 40 km, which coincides with the region of anomalously warm temperatures described by Santer and Crisp [1993].

This paper presents the first global estimates of the meridional circulation in the Martian atmosphere diagnosed from actual temperature observations. This diabatic circulation is qualitatively similar to that of Earth's atmosphere at the corresponding season, but is more vigorous. It is also more vigorous than the residual mean meridional circulation derived previously from a two-dimensional tracer transport model of the Martian atmosphere [Moreau et al., 1991]. This diabatic meridional circulation places constraints on the atmospheric transport of dust in late southern summer, during non-dust-storm conditions. A comparison of characteristic settling and meridional transport times reveals that 1-μm dust particles can be transported over global scales, whereas 5-μm particles do not remain suspended in the atmosphere for periods sufficient for long-range transport.

Although this study has yielded insight into the magnitude and direction of atmospheric transport processes on Mars, it does not provide a complete picture of the circulation patterns. We have investigated one 12-day interval in late southern summer during a relatively clear period after a severe dust storm. It is uncertain how well the diagnosed circulation describes that of other seasons or other years; the proximity to the equinox of the study period suggests that the derived velocities may be smaller than average over the course of this particular Mars year, but the fact that the dust optical depth is still slightly above the background value suggests that the derived velocities may be larger than those for a more typical year at the same season. Even if all available spectra had been processed, the entire Mariner 9 data set encompasses only two Martian seasons, and for much of that time the IRIS data are too sparse to differentiate diurnal and longitudinal variations for the construction of reliable zonal means. A more comprehensive examination of the meridional circulation over an annual cycle must await new observations.

Appendix A: Solution to the Stream Function Equation

The solution to equation (13) is described here in some detail. To simplify the notation, $A$, $B$, $C$, $D$, and $E$ are introduced for the coefficients of the $\psi$ terms, and $F$ is used for the right-hand side of equation (13):

$$A\psi + B\frac{\partial\psi}{\partial z} + C\frac{\partial^2\psi}{\partial y^2} + D\frac{\partial^2\psi}{\partial y^2} + E\frac{\partial^2\psi}{\partial z^2} = F$$ \hspace{1cm} (A1)

The simplest approach to solving equation (A1) is to use finite differences to approximate the derivatives on a grid of points in $\psi$ and $z$. Centered finite difference
formulas are applied to both the $\psi$ terms and their coefficients in equation (A1). The coefficients in their expanded finite difference forms are written as follows:

\[
A_{mn} = \frac{1}{aH \cos \phi_m} \left[ \frac{T_{m+1,n} - 2T_{m,n} + T_{m-1,n}}{\Delta \phi} \right] + \tan \phi_m \left( \frac{T_{m+1,n} - T_{m-1,n}}{\phi_{m+1} - \phi_{m-1}} \right)
\]

\[
B_{mn} = -H A_{mn}
\]

\[
C_{mn} = \frac{1}{a \cos \phi_m} \left[ \frac{T_{m+1,n} - T_{m,n}}{H} \right] \left( \frac{1}{\phi_{m+1} - \phi_m} \right) + \frac{1}{\Delta \phi} \left( \frac{1}{\phi_{m+1} - \phi_m} \right) \left( \frac{1}{\phi_{m+1} - \phi_m} \right)
\]

\[
D_{mn} = \frac{1}{a \cos \phi_m} \left[ \frac{T_{m+1,n} - T_{m,n}}{H} \right] \left( \frac{1}{\phi_{m+1} - \phi_m} \right) + \frac{1}{\Delta \phi} \left( \frac{1}{\phi_{m+1} - \phi_m} \right)
\]

The introduction of a second set of coefficients will again simplify the notation:

\[
A_{mn} \psi_{m-1,n} + B_{mn} \psi_{m+1,n} + C_{mn} \psi_{m+1,n+1} = F_{mn}
\]

Equation (A5) represents a system of simultaneous equations for the $\psi$ that can be compactly expressed using matrix notation:

\[
A_n \tilde{\psi}_n + B_n \tilde{\psi}_n + C_n \tilde{\psi}_{n+1} = \tilde{F}_n
\]

where $\tilde{\psi}_n$ and $\tilde{F}_n$ are $M \times 1$ column vectors composed of the individual $\psi$ and $f$:

\[
\psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_M \end{bmatrix}, \quad \tilde{F}_n = \begin{bmatrix} f_1 \\ \vdots \\ f_M \end{bmatrix}
\]

and $A_n$, $B_n$, and $C_n$ are $M \times M$ tridiagonal matrices composed of the coefficients of the $\psi$ terms:

\[
A_n = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & 0 & 0 \\ a_{2,1} & a_{2,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{M-1,1} & b_{M-1,1} \\ 0 & 0 & \cdots & b_{M-1,1} & c_{M-1,1} \end{bmatrix}
\]

\[
B_n = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & 0 & 0 \\ c_{2,1} & c_{2,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{M-1,1} & d_{M-1,1} \\ 0 & 0 & \cdots & d_{M-1,1} & e_{M-1,1} \end{bmatrix}
\]

\[
C_n = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & 0 & 0 \\ e_{2,1} & e_{2,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & e_{M-1,1} & f_{M-1,1} \\ 0 & 0 & \cdots & f_{M-1,1} & g_{M-1,1} \end{bmatrix}
\]
5480 SANTTEE AND CRISP: CIRCULATION IN THE MARTIAN ATMOSPHERE

Of course, equations (A6)-(A8) hold for all \( N \) grid levels, and are therefore also more compactly expressed in matrix form:

\[
\begin{pmatrix}
   h_{1,n} & i_{1,n} & 0 & \cdots & \cdots & 0 \\
   \varphi_{2,n} & h_{2,n} & i_{2,n} & 0 & \cdots & \cdots \\
   0 & \varphi_{3,n} & h_{3,n} & i_{3,n} & 0 & \cdots \\
   \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\
   \vdots & \vdots & \vdots & \cdots & \ddots & \ddots & \ddots \\
   0 & 0 & 0 & \cdots & \varphi_{M-1,n} & h_{M-1,n} & i_{M-1,n} \\
   0 & 0 & 0 & \cdots & \varphi_{M,n} & h_{M,n} & i_{M,n} \\
\end{pmatrix}
\]

\[ (A8c) \]

The supermatrix on the left-hand side is an \( N \times N \) tridiagonal matrix whose elements are themselves tridiagonal matrices. Because this is an ill-conditioned matrix, special precautions must be taken in inverting it to solve for the \( \psi \).

Y. Yung (personal communication, 1991) provided a routine for solving this set of equations; its structure is similar to that of programs designed to invert tridiagonal matrices with scalar elements [Press et al., 1989]. Once the tridiagonal solver has yielded values for the \( \psi \), the definition of the stream function is used to calculate the residual circulation:

\[
\begin{align*}
\mathbf{\Phi}_{m,n} &= \frac{1}{\cos \phi_m} \left[ \psi_{m,n} - \psi_{m,n-1} - \psi_{m+1,n-1} + \psi_{m+1,n} \right] \\
\mathbf{\Phi}_{m,n} &= \frac{1}{\cos \phi_m} \left[ \psi_{m+1,n} - \psi_{m,n-1} - \psi_{m+1,n-1} + \psi_{m,n} \right]
\end{align*}
\]

\[ (A10) \]

Appendix B: Optical Properties of Dust in the Martian Atmosphere

Mariner 9 IRIS observations provide the best existing constraints on the wavelength dependence of the dust optical properties at thermal infrared wavelengths, but no quantitative information about the dust optical properties at solar wavelengths. Subsequent spacecraft observations of the dusty Martian atmosphere have yielded additional insights into these properties, but only in a relatively narrow range of visible wavelengths [Pollack et al., 1977; Pollack, 1982; Drossart et al., 1991; Clancy and Lee, 1991]. Much less is known about the dust properties at other wavelengths. The minimum spectral information needed to assess the impact of dust on the net radiative forcing of the Martian atmosphere is the ratio of the broadband solar and thermal optical depths, but even this ratio is uncertain by about a factor of two [cf. Zurek, 1982; M Marvin, 1986; Toon et al., 1977; Clancy et al., this issue].

For this study, we combine data from a variety of sources to construct a more complete wavelength-dependent description of the dust optical properties. At wavelengths longer than 5 \( \mu \)m, we assume that palagonite is an acceptable spectral analog for Mars dust, and we adopt the refractive index measurements of Roush et al. [1991]. These data are used in a Mie scattering model [Crisp, 1986] to derive the wavelength-dependent extinction efficiencies, single scattering albedos, and phase functions for the particle size distributions published by Toon et al. [1977], Drossart et al. [1991], and Clancy and Lee [1991].

This approach is not possible at shorter wavelengths, where refractive index data for palagonite are not available. We therefore derive these quantities from observations of the albedo of the dusty Martian atmosphere during dust-storm conditions, including ground-based 0.3 to 1.0-\( \mu \)m spectrophotometric measurements [McCord et al., 1977] and Viking Orbiter infrared thermal mapper (IRTM) visible channel measurements [Christensen et al., 1988]. At wavelengths where no dust-storm observations are available, we assume that the albedo of the dusty Martian atmosphere is similar to that of a relatively bright region on the Martian surface [McCord et al., 1977]. The dusty-atmosphere albedo spectra compiled from these data are shown in Figure A1.

Figure A1. Albedos of a dusty Martian atmosphere used to derive dust optical properties. The nominal albedo (solid line) are near 0.3 at solar wavelengths, consistent with Viking IRTM observations. The other curves are scaled by constant factors to yield solar albedos near 0.375 (dashed line), 0.30 (dash-dot line), and 0.3 (dash-triple-dot line).
To derive the wavelength-dependent optical constants \((n, k)\) from these measurements, we developed a simple retrieval algorithm consisting of the Mie scattering model, a multiple scattering model [Stamnes et al., 1988; Bell and Crisp, 1993], and an iteration scheme based on Newton's method. This approach produces physically reasonable single scattering optical properties for each of the dust size distributions listed above. The wavelength-dependent extinction efficiency \(Q_{ext}\), single scattering albedo \(\omega_s\), and particle scattering asymmetry parameter \(g\) are shown in Figures A2-A4.

The absorption coefficients corresponding to the imaginary index values for the Toon et al. [1977] size distribution provide the best agreement with available laboratory measurements of terrestrial palagonites [Clark et al., 1990]. Surface radiance calculations using the optical properties derived for this size distribution also produce the best fit to the radiation field measured by the Viking Lander cameras [Pollack et al., 1977]. Although these methods cannot yield a truly unique description of the Mars dust, they should be adequate for estimating the dust solar forcing, since they are consistent with the available observational constraints both at the top of the Martian atmosphere and at the surface.

These dust optical properties and the atmospheric temperatures and dust distributions derived from Mariner 9 IRIS observations (Figure 1) are used in the radiative transfer model [Crisp, 1990; Santee, 1993] to generate a two-dimensional (latitude/altitude) description of the solar heating rates, the thermal cooling rates, and the net radiative heating rates for late southern summer on Mars. The IRIS observations for this season indicate 9-\(\mu\)m dust optical depths between 0.2 and 0.3 above 6 mbar (Figure 1b). Because the optical depths are specified at 9 \(\mu\)m in our calculations, the thermal cooling rates are relatively insensitive to the size distribution. However, the solar heating rates are strongly dependent on the size distribution, because this property determines the ratio of the solar and thermal optical depths. For the optical constants and size distributions described above, 9-\(\mu\)m optical depths near 0.2 correspond to solar optical depths between 0.35 and 0.6. Size distributions consisting primarily of smaller particles [e.g., Clancy and Lee, 1991; Drossart et al., 1991] produce the largest solar/thermal ratios, and the largest solar heating rates. Size distributions with larger particles [e.g., Toon et al., 1977] produce much smaller solar/thermal ratios and solar heating rates for the observed 9-\(\mu\)m dust amounts.

No large variations in the globally averaged temperatures are observed during the late southern summer period studied here. The globally averaged solar heating rates and thermal cooling rates should therefore be roughly equal. For the smaller size distributions, we calculate globally averaged solar heating rates that exceed the globally averaged thermal cooling rates by almost...
a factor of two. Only simulations that use dust optical properties derived for the highest visible albedos and largest particles roughly satisfy global balance. We therefore adopt the optical properties derived for dust-atmosphere visible albedos near 0.4 and the Toon et al. [1977] size distribution as our nominal values.

Acknowledgments. We would like to thank Y. Yang for providing the tridiagonal solver and H. Nair for discussions concerning the CIT/JPL photochemical model results. We are especially grateful to R.-L. Shia for suggesting how to solve the stream function equation. Insightful and constructive reviews by J. Barnes and C. Leovy were greatly appreciated. M.L.S. gratefully acknowledges support from the NASA Graduate Student Researchers Program. Some of this work was funded by grants from the NASA Planetary Atmospheres Program to the Jet Propulsion Laboratory of Caltech, Division of Geological and Planetary Sciences, California Institute of Technology, contribution 5383.


Western boundary currents in the Martian atmosphere: Numerical simulations and observational evidence

M. M. Joshi, S. R. Lewis, P. L. Read and D. C. Catling
Atmospheric, Oceanic and Planetary Physics, Department of Physics, Oxford University, Oxford, England

Abstract. Western boundary currents (hereafter WBCs) are an intensification of meridional fluid flow along the eastward facing flank of a boundary. They occur in the Earth's oceans, the best known instance being the Gulf Stream, and in the terrestrial atmosphere, an example of which is the East African Jet. We have investigated WBCs in numerical simulations of the Martian atmosphere, where they occur in the presence of large longitudinal topographical gradients, combined with the $\beta$ effect. We suggest that WBCs have already been simulated by other Martian atmospheric models, although not identified as such. The intensity of these currents is dependent on basic model parameters, notably surface drag. We show that for physically reasonable values of drag, frictional forces dominate the WBCs' behavior. The requirement of zero net cross-equatorial time-mean mass flow produces a "return flow" in simulations with relatively low surface drag, which acts in opposition to the Tharsis WBC. We also show that the intense flow associated with WBCs advects potential vorticity across the equator, although the presence of strong radiative damping inhibits low-level inertial instabilities that might result from this. Slope winds are found to have a profound effect on WBC structure, especially where they have a component parallel to the jet. In these cases, slope winds can cancel out or reinforce WBCs, depending on their direction. Enhanced low-level winds associated with WBCs may be a factor that influences the locations of dust storm generation. The high vertical shear associated with WBCs may also play a role in the maintenance and subsequent decay of these dust storms. Some observational evidence may exist for WBCs in the form of the alignment of bright depositional dust streaks, wave clouds, and in wind measurements taken during the parachute descent phase of Viking Lander 1 through the planetary boundary layer.

Introduction

One of the most prominent phenomena observed in the oceans is that of western intensification. This occurs close to western boundaries of oceans where the flow reaches far higher speeds than in their interior. An example of this is the Gulf Stream, which is a narrow and intense current off the eastern coast of North America. This feature, and others like it, have been explained in terms of western boundary currents (WBCs), which are caused by a combination of the presence of the meridional boundary and the variation of the strength of the Coriolis force with latitude [Gill, 1980]. WBCs can also exist in the terrestrial troposphere, although in this case the "western boundary" is not truly a boundary, but a meridionally aligned mountain range which reaches heights of the order of 3-4 km. The intense flow that occurs on the eastern flank of the East African mountain chain has been interpreted as a WBC [Anderson, 1976]. This jet significantly affects the South-East Asia monsoon circulation by controlling the cross-equatorial flow between its high- and low-pressure systems. Indeed, recent simulations of this system suggest that without the effect of topography the monsoon circulation would not take place (M. J. Rodwell and B. J. Hoskins, A model of the Asian summer monsoon, II, Cross-equatorial flow and PV behavior, submitted to Journal of Atmospheric Science, 1994; hereinafter referred to as submitted manuscript). Prompted by these studies, and the fact that equatorial topography on Mars is of very high amplitude (O(1 scale height or 10 km)), and has large longitudinal gradients associated with it, we looked for WBCs in numerical simulations of the Martian atmosphere. As on Earth, the totally averaged circulation of Mars is thought to be dominated at most times of the year by a cross-equatorial Hadley cell [Lindzen and Hou, 1988], whose lower branch, according to Martian global circulation models (GCMs), is confined in the vertical be-
large longitudinal topographical gradients occur. These low latitudes, especially in places where the cross-equatorial flow associated with the Hadley cell should be strongly affected by the presence of this topography at low latitudes, especially in places where large topographical gradients occur. One such location is the eastern flank of the Tharsis Plateau, a large volcanic outcrop that straddles the equator and lies between 160°W and 90°W. This rises about 10 km above the global mean geoid [Wu, 1980].

In an earlier paper [Joshi et al., 1994] (henceforth paper 1), we first reported some preliminary model results indicating the likely existence of WBCs in the Martian atmosphere. It was noted that the nature of the flows was affected by surface drag, and that WBCs may influence the locations of dust storms. Some possible observational evidence was also provided. In this paper, more detailed results on the structure of WBCs given differencing model parameters are discussed. Some calculations have been carried out to determine whether frictional or inertial forces control the simulated WBCs, and these results are then compared with the scale analyses performed in paper 1. The advection of potential vorticity by the jet is also examined, since some studies of WBCs in the terrestrial atmosphere suggest that cross-equatorial transport of this quantity can cause inertial instabilities in subtropical latitudes (M. J. Rodwell and B. J. Hoskins, submitted manuscript, 1994). Results are then presented from an enhanced model with more realistic representations of forcing and dissipation processes, and similar analyses performed. The influence of WBCs on the generation, maintenance, and decay of dust storms is discussed, and a more quantitative description of possible observational evidence for the existence of WBCs from Viking lander entry data is presented. Finally, because of the sparsity of data available for Mars, we suggest some future observational methods to test our theoretical predictions of WBCs.

Basic Theory of WBCs and Scale Analysis

Barotropic Representations

WBCs can be simulated in a two-dimensional barotropic fluid using the shallow water equations to describe a fluid confined by two meridional boundaries and forced, e.g., by a surface wind stress which varies sinusoidally with latitude in one hemisphere, representing the transition from tropical easterlies to midlatitude westerlies [Gill, 1980]. The circulation spins up to a steady state everywhere except at the western boundary, where in the absence of frictional forces, the current becomes stronger with time but narrower, eventually reaching an infinitesimal width. In reality, other physical processes not represented above, such as frictional and/or nonlinear inertial terms, become significant before this happens, causing the WBC to reach a finite magnitude in the steady state. Where nonlinear inertial terms dominate, the width of an inertial WBC will tend to \( (u/\beta)^{1/2} \), where \( u \) is a typical wind speed and \( \beta \) is the variation of Coriolis force with latitude, while the width of a frictionally controlled WBC will depend on the friction parameterization itself. In the case of linear Rayleigh friction, represented schematically as \( \partial u/\partial t = -u/v \), where \( v \) is the timescale over which the friction acts, the WBC width tends to \( W \approx 1/\sqrt{v} \) [Gill, 1980].

This scaling argument can be generalized to show how the width of a WBC changes using a more complicated frictional parameterization. We consider here a quadratic drag parameterization of the form

\[
(\partial u/\partial t)_{drag} = -c_2(\sigma u^2)/H, \quad (1)
\]

where \( c_2 \) is the nondimensional drag coefficient, \( H \) is the scale height, and \( \sigma \) is a terrain-following vertical coordinate \( p/p_0 \), where \( p_0 \) is surface pressure, as commonly used in GCM studies for the Earth and Mars. In this case, we evaluate the derivative shown above and approximate it across the interface between the interior flow and the ground by finite differences

\[
(\partial u/\partial t)_{drag} = -(c_2u\Delta u)/(H\Delta \sigma). \quad (2)
\]

Assuming that \( u \) is zero at the ground, \( \Delta u \approx |u_0| - |u_0| \) where \( |u_0| = 0 \). Hence the frictional acceleration can be approximated by

\[
(\partial u/\partial t)_{drag} = -2c_2u_0\Delta u/(H\Delta \sigma), \quad (3)
\]

where \( \Delta \sigma \) in reality is the distance in \( \sigma \) between the ground and the lowest model level. This leads naturally to a frictional drag timescale \( \tau_f = H\Delta \sigma/(2c_2u_0) \), suggesting that the width of a frictionally controlled boundary current will tend to

\[
W \approx 1/\sqrt{\tau_f} = 2c_2u_0/(GH\Delta \sigma). \quad (4)
\]

Baroclinic Representations

The barotropic representation discussed above describes the formation of an idealized WBC adjacent to a western boundary of effectively infinite height (so no flow takes place over it). In an atmosphere, however, toponography is never sufficiently high to completely block the flow at all levels, and WBCs will consequently only extend over a limited range in altitude. The proper description of atmospheric WBCs therefore requires a more complex three-dimensional (baroclinic) model to capture their vertical and horizontal structure. Despite this, the considerations outlined above should also apply in determining the horizontal width of WBCs close to the topographical features which control them in relation to the parameterization of friction and/or inertial terms. However, the vertical structure and extent of a WBC remains to be determined.

If the horizontal scale derived above may be taken to be comparable to the relevant radius of deformation (either \( ND_f/\beta \) at midlatitudes or \( ND/\beta \)) near the
equation), would lead naturally to a vertical scale 
\[ D \approx \frac{fW/N}{\alpha} \] at midlatitudes or 
\[ D \approx \frac{\beta W^2}{N} \] in the tropics. Thus we might expect inertially controlled WBCs to penetrate over a characteristic depth in the atmosphere, and to adopt a particular aspect ratio which depends on the ratio \( N/\alpha \) or \( N/\beta \). Substituting values for \( W, N, \) and \( \beta \) typical of the simulations of Martian WBCs discussed below (\( W = 1000 \text{ km}, N = 10^{-2} \text{ s}^{-1} \), \( \beta = 4 \times 10^{-11} \text{ m}^{-1}\text{s}^{-1} \)) leads to estimates for their vertical scale of around 2-4 km, comparable to the amplitude of the surface topography itself.

**Numerical Models**

The Simplest Three-Dimensional Model

The simple general circulation model (henceforth SGCM), is a fully nonlinear, multilevel primitive equation model, based on the model originally developed by Hoskins and Simmons [1975]. Model fields are represented in the horizontal by spherical harmonics, and a finite difference formulation is used in the vertical. The convention adopted to describe the model resolution is as follows: \( N_{x/0} \) denotes a run having model fields truncated at total horizontal wavenumber \( n \), with \( k \) levels in the vertical (\( T_{21} \) resolution is approximately equivalent to a grid-point resolution of \( 6^\circ \times 6^\circ \), while \( T_{10} \) corresponds to \( 12^\circ \times 12^\circ \) and \( T_{42} \) to \( 3^\circ \times 3^\circ \)).

The coordinate system is \((\lambda, \phi, \sigma)\) where \( \lambda \) is longitude, \( \phi \) is latitude, \( \sigma \) is as defined previously. This vertical coordinate has the advantage that it follows the terrain, thus providing an accurate representation of orographic forcing. The topography data set used is the U.S. Geological Survey digital terrain model (DTM) [Wu, 1976], smoothed to the relevant resolution by a low-pass averaging process, and filtered in spectral space by the method of Hoskins [1986] in order to remove structure at resolution scale.

The only forcing and dissipation processes present are (1) Newtonian relaxation towards a prescribed zonal mean state, (2) hyperdiffusion of temperature and momentum, and (3) Rayleigh friction, which are represented by the four terms on the right-hand side of the following thermodynamic and momentum equations:

\[
\frac{dT}{dt} = \frac{T - T^*}{\tau_v(\sigma)} + k\nu T \tag{6}
\]

\[
\frac{d[U, V]}{dt} - f [V, U] + \nabla \Phi = \left[ \frac{[U, V]}{T(\sigma)} + k\nu T [U, V] \right] \tag{6}
\]

where \( \tau_v \) varies with \( \sigma \), and \( T^* \) represents an approximation to a radiative equilibrium state. It is important to note that the Newtonian relaxation acts on \( T \) fields interpolated onto pressure surfaces. The interpolation therefore isolates the purely mechanical effect of topography, and does not model the thermal effect.

One important consequence of this is that the SGCM does not represent sloping winds [Sasai and Stuh, 1993].

SGCM runs were carried out with 20 levels in the vertical domain reaching from 1 km above the ground to 80-km altitude, with a spacing of approximately 1.5 km near the ground, rising to around 6 km at the model top. Parameter \( \gamma \) was set to 3 sols (a sol is a Martian day) up to 40-km altitude, then linearly decreased with altitude above that level to reach 0.5 sols at 80 km. This also acts as a "sponge layer", damping upward propagating eddies produced by the model, and attenuating their reflection from the top of the model. Rayleigh friction was set to vary with height, from 20 sols at a height of 40 km linearly to 1 sol at the model top, and effectively acted at model levels above 50 km as a simplified parameterization of gravity wave drag. It also functioned as a parameterization of surface drag at the lowest model level. The sensitivity tests detailed below were carried out by changing \( \gamma \) in the lowest model level, while keeping it constant everywhere else.

The SGCM has two limitations when trying to represent the subtropical Hadley circulation. First, it has no convective adjustment scheme to simulate vertical mixing stimulated by surface heating. Because such a scheme will tend locally to enhance vertical motion via the thermodynamic equation, this leads to a Hadley cell far smaller in magnitude in the SGCM than that produced by more realistic GCMs [e.g., Haberle et al., 1993b]. A consequence of this is smaller cross-equatorial velocities and hence weaker WBCs. Second, more realistic GCM simulations suggest that the Hadley cell in the southern summer is typically a factor of 2 stronger than its northern summer counterpart [Haberle et al., 1993b].

The first problem can be resolved by running a more complex model that includes a parameterization of vertical energy transfer by convection (see below). The second, however, can only be addressed by a full GCM that explicitly represents the seasonal variation of isolation, which is beyond the scope of the present work.

The Intermediate Three-Dimensional Model

The intermediate general circulation model (henceforth ICGM) is structured in a very similar way to the SGCM as described above, except that some of the main physical processes are handled in a more realistic manner. Friction is parameterized by a quadratic drag law in the lowest model level only. The momentum tendency in this level due to friction is shown in equation (2). The temperature tendency in the lowest model level is given by

\[
\frac{dT}{dt} \text{resulting from surface heat} = \frac{\beta}{\rho \, \gamma} \left( \frac{\mathbf{H}_s}{\mathbf{U}_s} \right) \tag{7}
\]

where \( \mathbf{H}_s \) is the sensible heat, which in stable conditions is given by

\[
H_s = \rho \, c_w |u| \Delta \theta \tag{8}
\]

and unstable conditions by

\[
H_s = \rho \, c_w \left( |u| + A \frac{\Delta \theta^{1/2}}{\beta} \right) \Delta \theta \tag{9}
\]
where $\rho_s$ is the surface density, $c_d$ is the drag coefficient for heat transfer and is equal to $c_d/S$, and $A$ is a stability parameter set to 500 m s$^{-1}$. The surface potential temperature $\theta$ is specified as varying with latitude only. These processes mix momentum and heat in the vertical far more efficiently than linear processes such as Ekman or Rayleigh friction, and so their effect on WBCs may be substantial. A convective adjustment scheme is included which adjusts the vertical model temperature profile from a convectively unstable state to a stable one (determined from observations), thus transferring heat upwards through the model. This scheme does not mix momentum in the vertical. For a fuller description of the IGCM, see Collins [1993].

Infrared cooling in the atmosphere is represented by a uniform cooling of 10 K sol$^{-1}$, which is an approximation to cooling rates produced in GCMs [Hourdin, 1992]. The IGCM was run with 10 levels in the vertical equispaced in $\sigma$, such that the top of the model was at $\sigma = 0.05 \approx 30$ km.

The IGCM, therefore, while being only slightly more computationally expensive than the SGCM, is different from a full Martian GCM in two main respects: it has a very simple representation of surface heating and radiative cooling, and does not represent the meridional source-sink flow caused by condensation and sublimation of CO$_2$ at the poles. This model can therefore serve as a useful intermediate step between the SGCM above and a full GCM.

It should be noted at this point that the SGCM only models the mechanical forcing effects of topography because equilibrium temperature in the lowest model level varies with elevation. In the IGCM, however, surface temperature does not vary with elevation. The IGCM will therefore simulate both the thermal as well as the mechanical forcing effects of topography. The latter case is the more realistic one for Mars [Webster, 1977].

A consequence of this is that the IGCM simulates upslope winds, which exist in balance with upward motion above elevated regions. These winds occur in the same places as WBCs, have the same horizontal scales, as well as similar magnitudes. The along-slope component of these winds is small at the equator as the strength of the Coriolis force there is small [Savijärvi and Siili, 1993]. However, as is shown later, the presence of significant meridional topographic slopes leads to significant interactions between these winds and WBCs.

Experiments and Mean States

In all of the simulations described here, both the SGCM and IGCM were typically run for 60 sols. The 15-sol time-mean fields were averaged in time from sols 45-60, representing the statistically stationary state reached by the model simulation.

Most SGCM experiments were performed at $T21L20$ resolution, though some cases described below were done at $T10$ and $T42$ resolution. All these cases employed temperature relaxation states representative of either of the two solstice seasons. The southern winter temperature state is shown in Figure 1, along with the resulting time-mean state. The former is an approximation to a radiative-convective equilibrium state produced by Haberle et al., [1982], and the latter is the resultant time-mean temperature state. The relaxation state is warmer at the pole than GCM simulations of the same season [Haberle et al., 1993b]. However, this offset makes no difference to the results presented here.

The main differences between the two are due to the advection of temperature at high altitudes into the win-

![Figure 1](image1.png)

Figure 1. (Top) $T_{eq}$ (K), an approximation to the radiative equilibrium temperature, as used for the SGCM southern winter solstice runs and (Bottom) resulting 15 sols time mean temperature $\bar{T}$. 

---

**References**

- Collins, [1993]
- Hourdin, [1992]
- Haberle et al., [1982]
- Haberle et al., [1993b]
ter midlatitudes, and are largely consistent with other models [Haberle et al., 1993b].

In the southern winter case, \( \tau \) at the surface was varied between 0.2 and 10 sols, while in the northern winter case it took the values 1 sol or 10 sols. A few SGCM runs were performed either at T10L20 with \( \tau = 10 \) sols or at T42L20 with 0.2 sols \( \leq \tau \leq 10 \) sols to test the effects of varying the horizontal resolution.

The IGCM experiments described here were performed (1) for southern winter, with surface friction drag coefficient \( c_d \) having the values either 0.001 and 0.005, and (2) for one northern winter run, with \( c_d = 0.001 \). These values are consistent with values of \( c_d \) used in GCM studies of Mars for frost-free surfaces, such as 0.001 [Haberle et al., 1993b] and 0.002 [Hovius et al., 1993].

The time-mean temperature state produced by the IGCM in southern winter is shown in Figure 2. This resolves the sharp latitudinal temperature gradient at 50°S better than the SGCM (see Figure 1) because in the IGCM, the steep surface meridional temperature gradient is explicitly specified.

**SGCM Results**

**An Idealized Equatorial WBC**

To illustrate the formation of WBCs more clearly prior to investigating more realistic simulations, an "idealized" run was performed in which the only topographic feature consisted of a uniformly shaped mountain. This had the shape

\[
M(\lambda, \phi) = A_0 \exp\left[-(\lambda - \lambda_0)^2 + (\phi - \phi_0)^2/W_1/W_2^2\right] \tag{10}
\]

where the peak amplitude \( A_0 = 7 \) km, and the widths in longitude and latitude are \( W_1 = 30° \) and \( W_2 = 30° \), respectively (representing an idealization of the Tharsis plateau). Figure 3 shows 15 sol time-mean velocity vectors at the end of a 60 sol run, at the lowest model level (\( \sigma = 0.9 \approx 1 \) km above the surface), with topography contours overlaid, \( \tau = 1 \) sol in all cases.

In the absence of topography, control integrations with no surface orographic features show no longitudinal variation of \( \zeta \). In the idealized mountain case, however, the cross-equatorial flow is confined to the eastern flank of the Gaussian mountain at a longitude of 50°W.

Since this is the only significant area of cross-equatorial flow at the lowest model level, we deduce that on Mars this intense flow is the response of the Hadley cell to the presence of the large horizontal topographical gradient at 50°W.

This phenomenon is not a mesoscale along-slope wind类似 of the type investigated, e.g., by Savijärvi and Sillén [1993], as it is present only on the eastern side of the idealized Tharsis. Moreover, as shown in paper 1, the variation of jet width with varying surface drag is consistent with simple theories of WBCs. We conclude therefore that the area of intense meridional flow at 50°W can be interpreted as a WBC.

The response of the Hadley cell to more realistic orography, while being broadly of the same form, is expected to have more complex structure than the above idealized case. Results are now presented for the full Martian topography, mainly at T21 resolution.

**Subsidiary Simulations With Full Topography**

Low-level winds from the two southern winter runs are shown in Figure 4, which shows horizontal velocity vectors at \( \sigma = 0.9 \), superimposed on contours of topography, for two different values of \( \tau \). As expected, winds are stronger in the low-drag case than in the high-drag case, by about a factor of 2. However, in both cases, the positions of the WBCs produced are similar. One WBC lies on the eastern flank of the Tharsis Plateau, in Chryse Planitia at a longitude of about 50°W, in association with the large longitudinal gradient of topography there. The other lies on the eastern slopes of Syrtis Major, around longitude 80°E, although the ratio of its intensity to that of the Tharsis WBC is smaller in the northern winter case [see below]. The lower branch of the Hadley cell has therefore been funnelled into two relatively narrow channels, each about 15-30° wide. This funneling is expected to have a profound effect on the cross-equatorial transport of heat, momentum, and constituents.

WBCs exist in the same places during northern winter simulations, as shown in Figure 5, although they are weaker at their cores than their southern winter counterparts. This is despite a variation of no more than 20% in the strengths of the Hadley cells in the four "real" topography simulations discussed so far. This might be due to the fact that in northern winter, fluid

![Figure 2. Southern winter 15 sols time-mean temperature \( T \) (K) in the IGCM. Note that the vertical scale is different from that shown in Figure 1 due to the different vertical model domains.](image-url)
Figure 3. Time-mean velocity vectors at the lowest model level ($\sigma = 0.9$, equivalent to about 1 km above the surface) for two runs at T21 resolution with $\tau = 1$ sol. The top diagram is from a simulation with no topography and exhibits no latitudinal variation of $v$ at the equator. The bottom diagram is from a simulation with an "idealized" Tharsis and clearly shows a WBC at 50°W. The topography contours are at 1 km intervals. The dotted contours are an artifact of the spectral transform method.

crossing the equator has to flow slightly uphill (see contours on Figure 4) and will therefore be relatively reduced in intensity. This result is repeated in the IGCM (see below).

The jet structure can be more clearly seen by examining longitude-height sections of time-mean meridional wind $v$ at the equator. Two examples of these are shown in Figure 6, for southern winter simulations performed with $\tau = 1$ sol and 10 sols. The Tharsis WBC in the $\tau = 10$ sol case is wider than the $\tau = 1$ sol case, consistent with the barotropic theory of inertial WBCs (see text). The wider WBC also has the deeper structure, consistent with the simple vertical scaling arguments presented above. A comparison of the Syrtis WBC structure between the $\tau = 1$ sol and $\tau = 10$ sols cases also yields the same result.

Cross-Equatorial Mass Flow

The strength of the Hadley circulation, as measured by the maximum value of the time-mean mass stream function, in the presence of varying topography and surface drag, is summarized in Table 1. In the case with no topography, peak transport is diminished when drag is reduced, consistent with the arguments outlined in paper 1. The opposite happens in the presence of topography, where the increased strength of the WBCs counteracts the lessening drag, leading to greater mass transport. The strength of the Hadley cell is principally determined by the large-scale thermal structure, and so is relatively unaffected by WBCs. However, the largest mass transport of all four cases occurs with $\tau = 10$ sols, in the presence of topography.

In the $\tau = 1$ sol case, the core of the Tharsis WBC lies about 2 km below the Syrtis WBC, while in the $\tau = 10$ sols case, it lies at about the same level. This, allied with the fact that the Tharsis WBC is the more intense of the two, suggests that more of the cross-equatorial mass transport in the Hadley cell takes place in the Tharsis WBC, although the proportion of mass carried by each will again be dependent on surface drag.
Figure 4. Time-mean velocity vectors at the lowest model level for two southern winter simulations with T21 topography. The top diagram is from a simulation with $T=1$ sol and the bottom from one with $T=10$ sols. The contours show orography contours at 1-km intervals.

The total mass flow below 10 km (i.e., the lower branch only) was analyzed in order to determine the fraction typically concentrated into WBCs. In both northern winter and southern winter simulations, the results were similar, and are shown in Table 2. In the $T=1$ sol case, the two jets transport 65% of the total mass across about 80° of longitude. In the $T=10$ sol case, when the effects of the return flow are included, about 70% of the total mass of fluid is transported across these longitudes. This shows that WBCs do indeed transport the bulk of the fluid carried in the lower branch of the Hadley cell.

One curious result, only observed in the lowest drag ($T=10$ sol) simulations, is a "return flow". This is an area of high winds directed away from the WBC and back toward the equator that can be observed at 20°W in Figure 4. One explanation for this phenomenon is that there is such a large amount of mass transported across the equator by the intense WBCs in the $T=10$ sols case, that it cannot be returned across the equator solely by the upper branch of the SGCM Hadley circulation. A consequence of this is the development of a low-level return flow, which can be thought of as an overshoot of the Tharsis WBC, and is depicted in Figure 7. This shows the time-averaged cross-equatorial mass transport in the southern winter $T=10$ sols run, integrated over height from the lowest model level to the model top. The major areas of northward mass transport coincide with WBCs identified earlier (cf. Figure 4; lower panel), and the southward return flow is evident again at 20°W. As there is no condensation flow in the SGCM, the time-mean globally integrated value of cross-equatorial mass transport must be zero. This criterion can apparently only be satisfied by the existence of the return flow.

Potential Vorticity

Topography is important at middle latitudes in enabling net meridional flow to satisfy the requirement of geostrophy. However, in the tropics, this geostrophic constraint is of lesser importance, and the primary constraint on fluid crossing the equator arises from the necessity to conserve potential vorticity (PV). In the absence of diabatic and frictional effects, fluid which crosses the equator will inevitably find itself having PV of opposite sign to $f$. Such a situation satisfies a nec-
essay (though not sufficient) condition for inertial instability, which would result in significant modification of the flow. In practice, however, friction and/or diabatic effects are not negligible and tend to balance the effects of cross-equatorial PV advection. Such effects are therefore crucial in enabling the Hadley circulation to cross the equator.

WBCs will tend to advect PV across the equator, which could lead to the possible development of inertial instability. Denoting Ertel’s PV by \( P \), it is defined as

\[
P = \zeta \cdot \nabla \theta \rho
\]  

(11)

where \( \zeta \) is absolute vorticity, \( \theta \) is potential temperature, and \( \rho \) is density. A necessary (but not sufficient) condition for inertial instability is that \( fP < 0 \). In a region of strong cross-equatorial flow, such a condition might well be caused by the advection of positive (negative) \( P \) into the southern (northern) hemispheres. \( P \) was analysed at an isentropic level coinciding with the jet to verify this hypothesis. The 232 K potential temperature surface was chosen for this analysis.

It was found that areas of negative \( fP \) only existed in the simulations with \( \tau = 10 \) sols. Since these values of drag are probably not physically realistic for Mars, it is concluded that inertial instability will not result from the cross-equatorial flow associated with WBCs. This contrasts with similar work done for the East African Jet by M. J. Rodwell and B. J. Hoskins (submitted manuscript, 1994), who suggested that inertial instability might result from high winds associated with the jet. It is suggested that the strong Newtonian relaxation in the SGCM does not allow anomalies of \( fP \) to build up, as they might on Earth.

In order to fully understand the behavior of \( P \) in the Tharsis WBC, a \( P \) budget analysis, such as that carried out by M. J. Rodwell and B. J. Hoskins (submitted manuscript, 1994) for the Somali jet and monsoon circulation on Earth, should ideally be performed. This would indicate the relative importance of terms such as advection, diabatic heating, and friction. However, work such as this is beyond the scope of this paper.

One point to note when trying to analyze isentropic \( P \) in the Martian tropics is that, in southern winter, \( \theta \) surfaces tend to slope downward toward the north at the equator as \( T_e < 0 \). This slope is in the same direction as the north-south topography gradient present at this
location (see Figure 4), implying that the surface has an (approximately) constant potential temperature. This is not the case in northern winter, where \( \theta \) contours intersect the ground, as can be seen by comparing the top and bottom panels of Figure 8. Therefore, in the northern winter PV21 case, there is relatively more advection of \( P \) across \( \theta \) surfaces, rather than along \( \theta \) surfaces, as is the case in southern winter, and indeed typically in the East African Jet as well.

Sensitivity to Model Parameters

It was shown in paper 1 that the variation of jet width was consistent with the barotropic theories outlined above. Frictional forces dominated in the Tharsis WBC for values of \( r < 1 \) sol. We now explicitly find the ratio of frictional to inertial forces by calculating the term

\[
R = (u/\tau) / (u \cdot V_u)
\]

in the two jets. For \( r = 1 \) sol, \( R \approx 2-3 \) in the jet core, while for \( r = 10 \) sols, \( R \approx 0.2 \). These results hold for both southern and northern winter simulations. It therefore appears that WBCs are frictionally controlled for all but the most inviscid cases.

As shown in paper 1, the effect of higher horizontal resolution (T42) on the structure of the jet was minimal. This result is perhaps surprising as the minimum longitudinal width of the WBC in any of the simulations was about 15°, which might not be expected to be fully resolved at T21. However, the core of the jet had the same strength in both cases, indicating that T21 res-
A longitude-height plot of time-mean meridional velocity $v$ in the IGCM, with $c_d = 0.001$, is shown in Figure 11. In the northern winter case, meridional winds in the Tharsis WBC are $O(20 \text{ m s}^{-1})$, and are stronger than in southern winter, where they only reach $O(15 \text{ m s}^{-1})$. This is largely as expected due to the stronger northern winter Hadley cell. In both cases, $v$ is not at its most intense in the lowest model level, as it is in the SGCM, indicating the presence of much stronger drag at the lowest model level of the IGCM than in the corresponding level of the SGCM.

There is a distinct seasonal asymmetry in Hadley circulation strength and structure. The peak meridional stream function obtained in the IGCM is $6 \times 10^9 \text{ kg s}^{-1}$ in a simulation of northern winter solstice, with $c_d = 0.001$, as compared to a peak of $4 \times 10^9 \text{ kg s}^{-1}$ in the corresponding southern winter solstice simulation. The northern winter Hadley cell is therefore about 1.5 times as strong as its southern winter counterpart.

Full GCMs show that the intensity of the Hadley circulation during northern winter is about 2 times greater than during southern winter. The subsolar insolation is 43% greater during northern winter than in southern winter, and so cannot fully account for this asymmetry. Zurek et al. [1992] suggested the influence of other factors such as orography and wave activity.

Our IGCM simulations show that even with no variation in insolation due to orbital eccentricity, i.e., a...
JOSH ET AL.: MARTIAN WESTERN BOUNDARY CURRENTS

Time mean \( v \) at Equator (Interval 5.0 ms\(^{-1}\))

\[ T_{10} = 10 \text{ sols} \]

Figure 9. A longitude-height plot showing \( v \) at the equator in a (top) T10 simulation and (bottom) T42. The equatorial topography, resolved to the appropriate resolution, has been masked out in the same way as Figure 6.

The zonal-mean temperature state reflected between hemispheres for each solstice, the intensity of the Hadley cell is a factor of 1.5 greater during northern winter solstice. We therefore suggest that a major influence on increased Hadley circulation intensity during northern winter is that its rising branch is in an area of elevated terrain relative to the northern hemisphere. The relatively low surface pressure associated with this leads to an enhanced Hadley circulation at this time.

The northern winter Tharsis WBC is only about 1.2 times as strong as its southern winter counterpart. This can be partly explained as before in terms of fluid flowing uphill (southwards) and therefore flowing with relatively lower intensity than otherwise might be expected. The northern winter Syrtis WBC is a factor of 4 stronger than its southern winter counterpart. This large asymmetry can be explained in terms of slope winds (see next subsection).

The ratio of frictional to inertial forces in the jets (\( R \); see equation (12)), is greater than unity for all simulations, indicating that frictional forces dominate WBCs in this model and are therefore likely to do so on Mars.

Influence of Slope Winds

As shown earlier, the IGCM will represent upslope winds as a consequence of the lack of any surface temperature variation with elevation. On the eastern side of the Tharsis Plateau, although there is some meridional topographical gradient present, the slope is primarily directed from the west downward to the east. Upslope winds will therefore flow E to W. This causes a veering of the jet toward the west, but no significant change in cross-equatorial flow.

On the eastern side of Syrtis (in Isidis), however, the slope at the equator is actually downward toward the northeast (see Figure 4). A component of the upslope winds therefore flows across the equator from north to south. It therefore appears that in northern winter, upslope winds reinforce the Syrtis WBC, and cancel it out in southern winter. Precisely the opposite result might be expected for downslope winds, which reach their peak strength just before dawn [Sagan and Solomon, 1993], although a model which represents diurnal variation is needed to test this hypothesis.

WBCs and Dust

WBCs may play an important role in the generation of global and great dust storms, since dust raising is critically dependent on surface stress, which is a function of wind speed [Greeley et al., 1993] and this, in turn, is...
Figure 10. Time-mean velocity vectors at the lowest model level for two IGCM simulations with T21 topography. The top diagram represents northern winter solstice, and the bottom southern winter. The contours show orography contours at 1-km intervals.

the result of the superposition of winds caused by different forcing mechanisms [Leovy et al., 1973]. Time-mean winds, when added to winds resulting from transient processes such as tides, can cause locally high winds at the surface during the day, when the turbulent boundary layer couples these winds to winds at the surface [Hubert et al., 1994a].

Large eastward-facing slopes in low latitudes are locations where global dust storms have been seen to originate, especially in Hellas (60°E, 40°S), Chryse Planitia (40°W, 0°N) and to a lesser extent Litis (60°E, 10°N) [Martin and Zurek, 1993]. The latter two are coincident with the locations of WBCs in the SGCM and IGCM, and in addition, areas of high surface wind stresses produced by GCMs also correspond to locations of WBCs [see Greeley et al., 1993, Figure 2e].

WBCs will have an effect on all phases of global dust storms. As shown above, they can help to cause dust raising. Second, there is a positive feedback mechanism present: the more dust that is raised, the more intense the Hadley cell will be. This suggests that, since WBCs are the response of the Hadley cell in the presence of topography, WBCs will be even more intense at this time. They will therefore provide even higher wind stresses than in relatively dust-free times and will consequently help dust storm maintenance. The third phase is that of dust storm decay, which is thought to occur when large amounts of airborne dust decouple the surface winds from winds above the planetary boundary layer (PBL) due to the increased static stability [Leovy et al., 1973]. However, the criterion for boundary layer stability is that the Richardson number $\text{R}i = N^2/\text{u}_\alpha^2$ be less than a critical value $\text{R}i_c \approx 0.2$. This decay mechanism will therefore be opposed, and possibly even nullified, by the intense “dusty state” low-level winds acting to enhance $\text{u}_\alpha$, and therefore lower $\text{R}i$. Model results support this conclusion (J. Murphy, personal communication, 1994). Areas of high near-surface vertical shear, such as those found in WBCs, may therefore locally arrest the decay of global dust storms.

Historically, more dust activity has been observed during southern summer not only because this is the season of maximum solar insolation, but also because the southern hemisphere is most easily seen from Earth
when the two planets are closest to each other. However, some dust activity has been observed on Mars during southern winter, especially on the eastern and southern flanks of the Tharsis plateau, e.g., in early 1984 [Martin and Zurek, 1993], which is consistent with the presence of intense WBCs in that season, and also with the SGCM results shown above.

**Observations of WBCs**

**Current Observational Evidence for WBCs**

There are three types of observational evidence: evidence of low-level direction, such as dust streaks or lee waves; evidence of high wind strength, such as dust storm activity as presented in the previous section; and evidence that sheds light on both of these physical aspects, such as wave clouds. We now present observations which, while not conclusively proving the existence of WBCs, are entirely consistent with their presence.

On Mars, bright depositional streaks observed in subtropical latitudes are believed to form in late southern summer and autumn [Green et al., 1992], and are thought to be representative of global-scale winds at these times [Green et al., 1992]. Bright streaks are associated with times of high static stability, when dust in the atmosphere is deposited on the ground in the lee of small-scale topographical features such as craters. The directions of the streaks then show the direction of the surface wind. GCM simulations show that during late southern summer, a cross-equatorial Hadley cell is present, although it is lower in intensity than during southern summer solstice [Haberle et al., 1993b]. This implies that the Tharsis WBC will decay in strength as southern summer progresses. The orientation of bright streaks, when averaged over 5° x 5° cells, infer N-S wind flow and are especially coherent at 40-50°W, which is coincident with the longitude and direction of the Tharsis WBC [Kahn et al., 1992].

A possible signature of WBCs can be found by comparing SGCM and IGCM results with those obtained from a one-dimensional (1-D) Martian boundary layer model [Haberle et al., 1993a], as well as with winds obtained during the parachute descent of Viking Landers 1 (henceforth VL1) [Seiff, 1993]. When the 1-D model winds were compared with values of $v$ and $w$ obtained during VL1 entry, the modeled values were 6 times less than the observed ones and rotated clockwise with increasing height, which was opposite to the sense in which the VL1 winds rotated. The magnitude and rotation of the observed winds could be accounted for by significant advection of cold fluid from the southeast [Haberle et al., 1993a]. One mechanism that would
explain this anomalous advection is the Tharsis WBC, which could lead to large winds at the VLI site. This can be seen in Figure 4, at 48°W, 22°N (the coordinates of the VLI site). There is significant meridional advection caused by the WBC which does not occur anywhere else at this latitude. Moreover, the decrease in meridional wind strength with altitude in the lowest 4 km of the SGCM and IGCM (see Figures 6 and 11) would cause an anticyclonic turning of the wind with height.

VL1 winds can be compared with time-mean values obtained from the SGCM during southern winter with \( r \) set to 10 sols and 1 sol, and with the IGCM, with \( c_w \) set to 0.001. For reference, a "control" simulation having no topography, with low drag (\( r = 10 \) sols), is also diagnosed, to show the effect of the Hadley cell on a "flat" Mars. While it is not expected that the SGCM can properly simulate winds in the Martian boundary layer, the effect of the WBCs on winds at the VLI site can be gauged qualitatively. Figure 12 shows a comparison of the 1-D PBL model, inferred VLI winds at 1.45 km above the surface at the landing site, and model winds at the 6.0-mbar level, which corresponds to 1.45 km above the VLI site.

The control run (arrow c) does not correctly simulate the direction or magnitude of the observed winds (arrow a). The runs with topography do show significant meridional winds, consistent with (arrow a), although their zonal components are wrong. This might be explained in the low drag case (arrow a) by the fact that the return flow causes the westerly turning of the WBC at this location, which might not happen in reality. Case d has wind vectors in the right direction, but a factor of 2 too small. The IGCM case (case f) exhibits significant meridional winds, albeit weaker than those observed. To conclude, the presence of topography is necessary to simulate observed meridional winds at VLI, which is consistent with the large values of \( v \) being produced by the Tharsis WBC.

Future Observational Tests

WBCs may be indirectly observed in the future by measuring the near-surface temperature signal associated with advection of cold fluid across the equator by the WBCs. The SGCM produces a strong thermal signal, as shown in paper 1, although its magnitude may be significantly different from that actually observed in reality because the SGCM has such a simplified parametrization of surface drag. A problem with this method is that not only is the vertical scale of the signal small (approximately 3 km), but it is also very near the surface, and so would be difficult to measure by conventional remote sounding instruments. However, as mentioned in paper 1, it might be possible for a network of surface weather stations to measure this thermal signal, as well as other WBC signatures such as wind, if some stations were deployed along WBC longitudes.

Wave clouds are produced by shear instabilities near the surface. They take the form of periodic cloud cylinders, with the axes of the cylinders being aligned perpendicular to the mean wind. These clouds have been observed around the Tharsis Plateau [Kahn, 1984], although their directions are almost randomly distributed. Given more observations of these phenomena, more supporting evidence for WBCs might be accumulated.

Visual evidence in the form of dust storms and condensate clouds will shed light on low-level wind speed and direction, and so might help to provide evidence for the presence of WBCs.

Balloons have the unique capability to determine wind directly in both the boundary layer and the free atmosphere. They can also provide coverage of other meteorological variables such as temperature and pressure. Tethered balloons, which rise and sink daily, would be particularly useful for determining vertical profiles through the boundary layer at WBC locations.

Conclusions

This paper has shown that WBCs are the manifestation of the lower branch of the Martian cross-equatorial Hadley cell in the presence of the large equatorial topography on Mars, and the latitudinal gradient of planetary vorticity. They have been simulated previously in numerical models, but have not been investigated in detail until now. The most robust WBC produced lies on the eastern flank of the Tharsis Plateau, due to the large longitudinal topographical gradients there. Most of the mass transported across the equator by the
lower branch of the Hadley cell is within this WBC. The peak strength of the Hadley cell is increased by approximately 10% by the presence of intense WBCs in the low-drag case, despite the fact that WBCs only have a relatively localized effect on the atmosphere’s thermal structure.

The character of the WBCs produced was found to be dependent on the strength of the surface drag applied, ranging from being weak flows with a large longitudinal extent (\(\approx 40^\circ\)) in the high drag cases to narrow and intense flows (\(\approx 10^\circ\)) in simulations with low drag. Moreover, their width changed in a manner roughly consistent with simple barotropic theory, reaching a minimum at about \(\tau = 1\) sol, and then increasing as \(\tau^{1/2}\). SGCM simulations with low drag resulted in a “return flow” consisting of fluid flow across the equator in the opposite sense to the WBC. It is thought that this phenomenon is due to the constraint of the model having to satisfy the continuity equation requirement of no net time-mean cross-equatorial flow. For values of surface drag thought to be “reasonable” for Mars, WBCs were shown to be controlled by frictional forces. This implies that the \(\tau = 10\) sols simulation probably has unrealistically low surface drag, so that although the return flow is an interesting phenomenon, it is unlikely to occur in reality.

There is significant advection of Etel’s PV across the equator, which when combined with the presence of high winds in the low-drag simulation, suggests that low-level inertial instabilities should not be present in the equatorial regions. This is probably a consequence of the strong diabatic heating present on Mars. If these sorts of instabilities do occur on Mars, they might be important in the context of dust storm generation, as they would lead to higher than average near-surface transient winds. It is suggested that future work might consider explicitly calculating the PV sources and sinks present, as done for the Earth by M. J. Rodwell and B. J. Hoskins (submitted manuscript, 1994).

The differences between model resolution at T21 and T42 were found to be small. This was a surprising result, since the minimum width of WBCs produced in the models was 15°, equivalent to about 2.5 grid spacing at T21, or 5 at T42. As shown in paper 1, T21 resolution was sufficient to resolve WBC cores. However, T10 simulations carried out in this paper show that T10 is not sufficient to resolve the core of the Tharsis WBC, suggesting that T21 resolution is about the minimum model resolution needed to simulate WBCs.

The IGCM, with its more realistic treatment of surface fluxes of heat and momentum, produces similar WBCs to the SGCM. However, in this case it was found that surface drag dominated in the lowest model level. This result is a consequence of having a representation of surface drag that mixes momentum in the vertical very efficiently.

Uplift winds are found to drastically affect the character of the Syrtis WBC, by reinforcing it during northern winter and cancelling it during southern winter. Downside winds might be expected to have the opposite effect. The Tharsis WBC is more robust, not having its intensity significantly affected by slope winds because they flow east to west in this case.

The PBL varies quite dramatically over the course of one sol. Modeling studies show that the PBL can reach heights of 5 km during late afternoon (Mather et al., 1993a; Saslavsky and Stull, 1993) and observationally the PBL depth for VL1 entry during late afternoon has been estimated as 6.5 km above the ground (Steff, 1977). During the night, however, it “collapses”, decoupling the winds in the atmosphere from the surface. This implies that momentum transfer through the PBL at this time will be relatively inefficient. Hence, the structure of WBCs might be expected to change over the course of a day. However, the variability is not easy to predict. First, the high winds associated with WBCs might act to destabilize the PBL (see below). Second, as WBCs are set up by propagating Rossby waves over timescales \(\sim 1/\omega\), it is not clear whether they can react to dynamical changes having timescales less than a sol. Again, the only way to test this hypothesis is to analyze data from GCMs.

The top of the PBL is defined as the height at which the Richardson number, \(Ri\), first exceeds the critical Richardson number \(Ri_c\), as vigorous heat and momentum exchange in the vertical due to sub grid-scale processes such as turbulent mixing are only assumed to occur when \(Ri < 1\). Near-surface winds may act to stop the decay of global Hadley cells, which would lead to stronger WBCs and surface summer westerlies, thereby counteracting the stabilizing effect of dust on the PBL.

Some evidence for WBCs may already exist, in the form of meridionally aligned depositional dust streaks as well as in VL1 entry data, and a means of measuring temperature fluxes caused by WBCs, although difficult, has been suggested. It is finally concluded that although we have demonstrated that it is extremely likely that WBCs of some sort exist in the Martian atmosphere, our knowledge of their general character, intensity, and effect on the circulation remains incomplete, and will do so for at least future measurements of the near-surface Martian circulation are made.
Acknowledgments. The authors would like to thank M. Collins, R. Haberle, J. Murphy, the anonymous reviewers of paper 1 as well as the reviewers of this paper for their helpful comments. This work formed part of a D.Phil. project in the Department of Physics, Oxford University. M. M. Joshi was supported by a scholarship from the Gamot Great Committee of the UK Meteorological Office. S. R. Lewis and D. C. Catling were supported by the UK Science and Engineering Research Council.

References


D. C. Catling, M. M. Joshi, S. R. Lewis, and P. L. Read, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, England (e-mail: catling@oxpln3.atm.ox.ac.uk; joshi@atm.ox.ac.uk; lewis@atm.ox.ac.uk; read@gordon.atm.ox.ac.uk)

(Received March 24, 1994; revised September 27, 1994; accepted October 14, 1994.)
The sensitivity of the Martian surface pressure and atmospheric mass budget to various parameters: A comparison between numerical simulations and Viking observations

Frédéric Hourdin, François Forget, and Olivier Talagrand
Laboratoire de Météorologie Dynamique du Centre National de la Recherche Scientifique, Ecole Normale Supérieure, Paris, France

Abstract. The sensitivity of the Martian atmospheric circulation to a number of poorly known or strongly varying parameters (surface roughness length, atmospheric optical depth, CO₂ ice albedo, and thermal emissivity) is investigated through experiments performed with the Martian version of the atmospheric general circulation model of Laboratoire de Météorologie Dynamique, with a rather coarse horizontal resolution (a grid with 32 points in longitude and 24 points in latitude). The results are evaluated primarily on the basis of comparisons with the surface pressure records of the Viking mission. To that end, the records are decomposed into long-period seasonal variations due to mass exchange with the polar caps and latitudinal redistribution of mass, and short-period variations due to transient longitudinally propagating waves. The sensitivity experiments include a 5-year control simulation and shorter simulations (a little longer than 1 year) performed with "perturbed" parameter values. The main conclusions are that (1) a change of horizontal resolution (twice as many points in each direction) mostly affects the transient waves, (2) surface roughness lengths have a significant impact on the near-surface wind and, as a matter of consequence, on the latitudinal redistribution of mass, (3) atmospheric dust optical depth has a significant impact on radiative balance and dynamics, and (4) CO₂ ice albedo and thermal emissivity strongly influence mass exchange between the atmosphere and the polar caps. In view of this last conclusion, an automatic procedure is implemented through which the albedo and emissivity of each of the two polar caps are determined, together with the total (i.e., including the caps) atmospheric CO₂ content, in such a way as to get the closest fit of the model to the Viking pressure measurements.

1. Introduction

The meteorology of the planet Mars is probably, after that of Earth, the one which has been most studied and is best understood. This is due, first, to the large number of observations acquired in the seventies on the occasion of several spacecraft missions (the Soviet Mars missions and the American Mariner and Viking missions). In particular, the two Viking landers have recorded near-surface temperature, pressure, and winds over several Martian years. Such records are unique for extraterrestrial planets. The Viking measurements, especially the pressure measurements, contain very instructive information on the global atmospheric circulation. The large-amplitude seasonal oscillations of the pressure are due to the variations of the atmospheric mass (which result from condensation-sublimation of a substantial fraction of the atmospheric carbon dioxide in the polar caps), but also to internal latitudinal mass redistribution. The more rapid oscillations of the surface pressure, with periods of 2 to 5 sols (sols are Martian solar days), are signatures of the transient planetary waves which are present, at least in the northern hemisphere, during autumn and winter. Our relatively good knowledge of the Martian meteorology is also due to the results obtained with atmospheric general circulation models (GCMs). The usefulness of GCMs in this particular instance is certainly due in great part to the similarity of the atmospheric circulations of Earth and Mars. As early as the late sixties, the then recently developed terrestrial GCM of the University of California, Los Angeles, was successfully adapted to Mars [Leovy and Mintz, 1969]. The same model, on which various improvements were continuously performed, was later used both for the preparation of the Viking mission and for many studies of the Martian meteorology performed at NASA/Ames Research Cen-
ter by J. B. Pollack and collaborators [Pollack et al., 1981, 1990; Esposito et al., 1993; Barnes et al., 1990].

More recently, a new version of the terrestrial GCM of Laboratoire de Météorologie Dynamique (LMD) has been adapted to Mars [Hourdin, 1992; Hourdin et al., 1993]. The LMD GCM was the first self-consistent model to simulate the Martian atmospheric circulation over more than a year. Despite the somewhat arbitrary use of an a priori set of model parameters, some of which are poorly known from observation, the model was able to reproduce rather accurately many observed features of the Martian atmosphere, such as the thermal structure derived from Mariner 9 temperature retrievals, and the long- and short-period oscillations of the surface pressure observed by the Viking landers. Moreover, these simulations have shown that the large seasonal pressure variations were not only due, as usually thought until then, to mass exchange between the atmosphere and the polar caps, but were also strongly influenced by internal mass redistribution associated to the atmospheric circulation [Tolgyessy et al., 1991, Hourdin et al., 1993; referred to as paper 1 hereinafter, and Pollack et al., 1993].

The present paper, which is a continuation of paper 1, is mainly devoted to sensitivity studies of the LMD model to some very uncertain or strongly varying parameters (surface roughness length, atmospheric optical depth, CO2 ice albedo, and thermal emissivity) with emphasis on their impact on the atmospheric mass budget and on the meteorological contribution to the annual pressure cycle. Similar studies have already been performed by several authors: Patie and Ingersoll [1985], Wood and Paige [1992], and Paige and Wood [1992] have concentrated on the mass budget with a model of the energetics of the polar caps. Pollack et al. [1993] performed a more complete study, since they included in their energy budget model the effect of dynamics (heat transport and meteorological contribution to the pressure variations) using the results of off-line simulations performed with the NASA Ames GCM on a series of selected short periods. The present work is new in that the sensitivity experiments described here have been performed with a self-consistent GCM allowing for complex feedback between dynamics and diabatic processes.

We try here to answer the following questions: (1) What is the uncertainty in the model resulting from the uncertainties in model parameters or parameterizations? (2) Can the model parameters be constrained from the available observations (and especially from the Viking surface pressure measurements)? (3) What information can be derived from the GCM on the polar processes which dominate the mass cycle?

To that end, we propose simple diagnostics for the Viking pressure measurements, well adapted to the validation of GCMs and to the study of the annual pressure cycle. Those diagnostics are the subject of section 2. We then give in section 3 a brief description of the atmospheric model (already described in paper 1) and present in section 4 a 5-year simulation used afterward as control for the sensitivity experiments. In section 5 we test the uncertainties related to the most uncertain model parameters and, in section 6, we study the impact of the variations of the dust amount in the atmosphere. We take advantage of the high sensitivity of the atmospheric mass budget to the CO2 ice radiative properties (visible albedo and thermal emissivity) to derive an automatic best fit procedure to the Viking pressure. This procedure and the best fit results are presented in section 7.

2. Analysis of the Viking Surface Pressure Measurements

The records of the surface pressure, acquired over almost 2 Martian years by the Viking 2 lander, and over more than 3 years by the Viking 1 lander [Tholins and Geyser, 1987], have been widely used for studies of the Martian transient planetary waves [Thiébaux, 1988; Barnes, 1980, 1981], for comparisons with the results of seasonal energy balance models [Pollack et al., 1993; Paige and Ingersoll, 1985; Wood and Paige, 1992; Paige and Wood, 1992], and, in paper 1, for direct comparisons with GCM simulations. We present here simple diagnostics, well suited to GCM validations, based on a decomposition of the observed variations between a seasonal component (which is itself decomposed into a total atmospheric mass contribution and a meteorological contribution) and a shorter term transient component. In the present paper, which does not discuss diurnal cycle and the associated thermal tides (known to be very strong on Mars), we use directly the diurnally averaged pressure data.

Seasonal Variations

The large-amplitude seasonal variations of the surface pressure can easily be isolated, as proposed by Tholins et al. [1992], by retaining the first harmonics of the annual cycle. Following the same procedure, we retain only the observations performed after the 1977-B dust storm (clear-sky conditions), but we use solar longitude $L_0$ instead of real time $t$ as the temporal variable of our Fourier analysis. $L_0$ is a more appropriate coordinate for the seasonal cycle, since both the inclination of the planet, $\sin f = \sin L_0 \sin \delta_0$ (where $\delta_0 = 24^\circ$ is the Martian obliquity), and the distance to the sun, $r = p/(1 + c \cos L_0) \sim p(1 - e \cos L_0)$ (where $p = 1.51$ AU and $e = 0.093$), are almost first harmonics of a year in terms of solar longitude (and not in terms of real time).

The values of the amplitude and phase lag of the first eight harmonics are given in Table 1 for the two Viking landers. The quality of the fit (measured from the root mean square of the difference between the data and the synthetic spectra) reaches a minimum value for $N \geq 5$, the residual value corresponding to the mean amplitude of the transient oscillations (about 0.04 mbar for Viking 1 and 0.1 mbar for Viking 2). Somewhat arbitrarily, the eight-harmonic fit (compared to Viking data...
Table 1. Amplitude (\(P'(k)\)) and Phase Lag (\(L_2^{(k)}\), Solar Longitude of theMaximum Pressure) of the First Eight Harmonics of the Seasonal Pressure Variations Observed by the Viking Landers (Clear-Sky Conditions)

<table>
<thead>
<tr>
<th>Site</th>
<th>Mode</th>
<th>(P'(k)), mbar</th>
<th>(L_2^{(k)}), deg</th>
<th>Mean Quadratic Error, mbar</th>
</tr>
</thead>
<tbody>
<tr>
<td>VL1</td>
<td>0</td>
<td>7.9623</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6.6897</td>
<td>310.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.4853</td>
<td>66.76</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0154</td>
<td>39.87</td>
<td>0.0532</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0135</td>
<td>68.53</td>
<td>0.0369</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0150</td>
<td>33.30</td>
<td>0.0383</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.0023</td>
<td>12.62</td>
<td>0.0378</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0028</td>
<td>19.85</td>
<td>0.0376</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.0022</td>
<td>13.11</td>
<td>0.0376</td>
</tr>
<tr>
<td>VL2</td>
<td>0</td>
<td>6.6669</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6.0320</td>
<td>308.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0051</td>
<td>66.62</td>
<td>0.1139</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0422</td>
<td>25.48</td>
<td>0.1130</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0391</td>
<td>71.05</td>
<td>0.1094</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0355</td>
<td>65.98</td>
<td>0.1086</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.0113</td>
<td>46.37</td>
<td>0.1081</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0122</td>
<td>28.39</td>
<td>0.1081</td>
</tr>
</tbody>
</table>

\(c\) is the mean quadratic difference between the Viking individual data and the synthetic values reconstructed from the first \(N\) harmonics:

\[
p^{(N)}(s) = \sum_{k=0}^{N} p^{(k)} \cos \left(k \times (L_2 - L_2^{(k)})\right)
\]

in Figure 1) is retained as a reference seasonal pressure cycle in the present study and used for direct comparison with simulated pressures. However, for both landing sites, the first two harmonics explain the main part of the seasonal variations. Note that the two-harmonic fit is much better when solar longitude is used instead of real time as a spectral coordinate. For the time coordinate, for instance, the amplitude of the third harmonic at Viking 1 is of the order of 0.1 mbar, and the root mean square of the difference between the two-harmonic fit and the data is of the order of 0.075 mbar.

As illustrated in Figure 2 and Table 1, harmonic 2 really corresponds to the seasonal cycle with two minima during the two winters, lagging the solstices by about 66°. The minimum of insolation on one hemisphere at winter solstice corresponds to the maximum condensation rate, the maximum mass of the corresponding cap occurring much later. Harmonic 1 reflects the asymmetry between both winters: mainly forced by orbital eccentricity, it can also be affected by hemispheric asymmetries in orography, in the dust content, and in cloud coverage over caps or amplified by albedo feedbacks on the polar caps. The maximum pressure of the first harmonic also lags the maximum planetary insolation, at perihelion, by about 60°.

**Various Contributions to the Local Seasonal Pressure Variations**

As discussed above, the seasonal pressure variations are due in part to the variations of the total atmospheric mass \(M_{\text{atm}}\) resulting from the mass exchange between the polar caps and the atmosphere. It is convenient to introduce the planetary mean of the surface pressure \(p_{\text{atm}} = g M_{\text{atm}} / (4\pi a^2)\), where \(a\) is the planetary radius and \(g\) the gravity, and to define in the same way an equivalent surface pressure for the northern and southern polar caps, \(p_{\text{VL1}}\) and \(p_{\text{VL2}}\) respectively, and a total...
pressure $p_{pt,t} = p_{atm} + p_{~} + p_{s}$, which is independent of time.

However, as first pointed out by Talagrand et al. [1991], and studied in greater details in paper 1 and by Pollack et al. [1993], the seasonal evolution of the surface pressure at a particular point of the planet can differ significantly, under the effect of latitudinal mass redistribution, from the evolution of the mean surface pressure $p_{atm}$. It is convenient to introduce here the ratio $\alpha$ of the local pressure to the mean atmospheric surface pressure. This ratio will be called the "meteorological factor" hereafter. As shown in paper 1, the horizontal variations of $\alpha$ can be accurately decomposed into the sum of a geographic and a dynamical component. The geographic component (accounting for the horizontal pressure variations along slopes) varies in time because of the strong seasonal variations of temperature. The dynamical component is due to the geostrophic balance between the surface pressure and the near-surface horizontal winds. It is particularly responsible for strong latitudinal pressure variations in balance with the strong Martian zonal winds. Those zonal winds are produced themselves by latitudinal angular momentum redistribution by the condensation flow and thermally driven atmospheric circulation.

The absolute meteorological contribution cannot be deduced from Viking observations. However, the relative meteorological factor between the two Viking sites, defined as

$$\alpha_{VL2/1} = \frac{p_{VL2} - p_{VL1}}{p_{VL2} + p_{VL1}}$$

(1)

can be used for model validation. Figure 3 shows the relative meteorological factor as computed from the eight-harmonic fits to the Viking 1 and Viking 2 data. The maximum of the curve near northern winter solstice can be explained by the very cold temperatures which strongly enhance the pressure in the lowest Viking 2 site. This effect is partly compensated by the presence of the strong winter jet which should correspond by geostrophic balance to a northward decreasing pressure.

As explained in paper 1, the disappearance of the near-surface westerlies during the great 1977-B dust storm ($L_s \sim 280$) was responsible for the sudden increase of the surface pressure recorded by the Viking 2 lander. It is noteworthy that models must be used in addition to observations of the surface pressure in order to deduce the evolution of the atmospheric mass which requires the knowledge of the absolute meteorological factor. The relative factor $\alpha_{VL2/1}$ can only help for the model validation.

Transient Eddies

Another important feature of the observed pressure variations is the presence of short-period fluctuations linked to transient planetary waves. Without going into a complete study of these transient waves, which would require the use of more sophisticated analysis techniques, we can easily determine the seasonal evolution of the amplitude of the transient oscillations by retaining the running quadratic mean, $\sigma_N$, over $N$ con-
secutive data, of the difference between the pressure observation for a given sol and the corresponding value computed from the eight-harmonic reference pressure cycle. In order that the seasonal amplitude of the transient waves, but nothing else, be retained in the running mean, the value of $N$ must be significantly larger than the period of the transient waves, and significantly smaller than the length of seasons. As already stated, the analysis is based on diurnally averaged data and thus the transient eddies do not account for the thermal tides.

In Figure 4 we show the time evolution of $\sigma_{T2}$ (thick solid curve) and $\sigma_{T0}$ (dotted curve), which appear to be very similar. For diagnostics, we will retain the eight-harmonic fit to $\sigma_{T0}$ (thin solid line). Transient eddies occur from northern autumn equinox to early northern spring. The amplitude of the perturbations at the Viking 2 site is about twice that at Viking 1 and shows a two-peak structure with a relative minimum at northern winter solstice.

3. Model

General Description

The LMD Martian GCM has been described in paper 1. The dynamical part is based on a finite difference formulation of the "primitive equations" of meteorology developed for the terrestrial GCM of LMD (Sedov and Laval, 1984). We use a more recent formulation (developed by R. Sadourny and P. Le Van and described by F. Hourdin and O. Talagrand, Superposition of planetary atmospheres: A numerical study, submitted to Journal of Atmospheric Science, 1994) which allows us to change automatically the distribution of grid points in longitude and/or latitude. The formulation exactly conserves (1) the total atmospheric mass (in absence of CO$_2$ condensation), (2) the potential energy and its square for adiabatic motions, (3) the potential enstrophy for barotropic flows, and (4) the total angular momentum for axisymmetric flows.

The physical part has already been presented in detail in paper 1: (1) the radiative transfer is computed using a code adapted by Hourdin [1992] from the code currently used at the European Center for Medium-Range Weather Forecasts (ECMWF) [Fouquart and Bonnel, 1980; Morcrette et al., 1986]; it includes absorption and emission by carbon dioxide and dust in the thermal infrared and absorption and scattering by dust in the visible; (2) a simple formulation is used for the vertical turbulent mixing in the whole atmosphere; (3) vertically unstable profiles are prevented by applying a vertical convective adjustment which conserves energy; momentum is also mixed in the unstable layer, the intensity of the mixing depending on the intensity of the vertical instability; (4) the temperature of the surface is computed using an 11-level model of thermal conduction in the soil which correctly simulates the response for oscillatory forcing with periods going from a tenth of day up to a few years; (5) condensation and sublimation of carbon dioxide are computed in a simple energy and mass conserving manner. If somewhere in the atmosphere, the temperature falls below the condensation temperature $T_{CO_2}$, condensation occurs in an amount appropriate to restore $T_{CO_2}$ by latent heat release. All the condensed CO$_2$ is assumed to precipitate instantly to the ground, the surface pressure being consequently modified. At the surface, the temperature of the frost is kept at the condensation value either by condensing atmospheric CO$_2$ or by sublimating CO$_2$ ice. The approximation for the pressure-vapor curve used in paper 1 is replaced by a more accurate relationship based on the Clausius-Clapeyron relation for perfect gas. Assuming that the latent heat of sublimation $L$ is independent of temperature, the vapor pressure curve reduces to

$$
\ln (p/p_0) = \frac{L}{R} \left( \frac{1}{T_0} - \frac{1}{T_{CO_2}} \right)
$$

(where $R$ is the gas constant and $T_0$ is the condensation temperature corresponding to the pressure chosen as a reference, here $p_0 = 1$ mbar). $L = 5.9 \times 10^5$ J kg$^{-1}$ and $T_0 = 136.3$ K are fixed in the range of experimental values [e.g., James et al., 1992]. The change of the pressure-vapor curve was found to have a minor effect on the atmospheric mass budget.

The only significant change with respect to paper 1 is linked to the diurnal cycle. It was intentionally turned off in the previous study. Here, the radiative forcing is computed about 40 times per day. It then becomes necessary to introduce the dependency of the surface drag coefficient (a constant in paper 1) with both the magnitude of wind and vertical stability above the surface which are known to vary significantly between night and day. In the new version, we use the formulation developed by Louis [1979] for the terrestrial planetary boundary layer, in the form which has been implemented at ECMWF.

The drag coefficient is given by

$$
C = \left( \frac{k}{\ln(\frac{\zeta}{z})} \right)^2 \frac{1}{f} \frac{\zeta}{\bar{v}} \frac{z}{\bar{u}}
$$

(3)

where $\zeta$ is the roughness length, $z$ is the height of the middle of the first atmospheric layer, $k = 0.4$ is the von Karman constant, and

$$
R_c = \frac{2 \pi (\theta - T_s)}{\theta |V|}
$$

(4)

is the Richardson number, where $\theta = T(p_u/p_v)$ and $|V|$ are the potential temperature and wind velocity in the first atmospheric layer and $T_s$ the surface temperature. Two different functions $f$ are used for momentum ($f_m$) and potential temperature ($f_s$). In either case, a different formulation is used, depending on whether the atmosphere is stable ($R_c > 0$),

$$
f_m = \frac{1}{1 + 2 \lambda R_c \sqrt{1 + \lambda R_c}}
$$

(5)
or unstable \((R_e < 0)\),

\[
\begin{align*}
\mathcal{F}_n &= 1 - \frac{3 \sqrt{C}}{1 + 3C \left[ \sqrt{\frac{\ln (x/a)}{\ln (1 + \frac{x}{a})}} \right]} \\
\mathcal{F}_n &= 1 - \frac{3 \sqrt{C}}{1 + 3C \left[ \sqrt{\frac{\ln (x/a)}{\ln (1 + \frac{x}{a})}} \right]} \\
\mathcal{F}_n &= 1 - \frac{3 \sqrt{C}}{1 + 3C \left[ \sqrt{\frac{\ln (x/a)}{\ln (1 + \frac{x}{a})}} \right]}
\end{align*}
\]

As in the original terrestrial version, we use \(C = 1.47\), \(d = 5\).

Grid

As in paper 1, we use two different horizontal resolutions in the present study: a low horizontal resolution based on a grid with 24 points in latitude and 32 points in longitude and a high resolution with 48 points in latitude and 64 points in longitude. In the vertical, the atmosphere is discretized using 15 sigma levels \(\sigma\) is the pressure normalized by its surface value \(p_0/p\), the distribution of which is given in paper 1. The middle of the first layer is located at about 40 m above the surface and the 15th at about 60 km (the pressure is zero at the top of this layer). As pointed out in paper 1, even a coarse horizontal resolution (the low resolution) is sufficient to reproduce, at least qualitatively, the main features of the available observations of the Martian atmospheric circulation. The low resolution was generally retained for sensitivity studies.

Surface Conditions

In addition to the model parameters, three maps are introduced for the surface conditions: geography, visible albedo, and thermal inertia. These maps correspond to the consortium data set, completed by more recent estimation of the surface thermal inertia and albedo in the polar regions [Keegan et al., 1991; Paige and Keegan, 1991], which were provided to us by J. B. Pollack.

4. The Control Simulation

Description

In the present section, we describe a first simulation, which is afterward used as a reference for the sensitivity study. In this "control simulation," the dust visible optical depth is a constant over time and space. Its value is set to 0.2, which corresponds to typical clear-sky conditions [Martin, 1986]. Vertically, the dust mixing ratio is almost constant from the surface up to the 0.3-mbar level and then rapidly decreases, as proposed by Pollack et al. [1990]. The mixing length for vertical turbulent mixing is fixed to 35 m as in the LMD terrestrial GCM. The surface roughness is set to 1 cm everywhere, a value typical of near Viking lander conditions [Sutton et al., 1978]. The emissivity of the bare surface (without ice) is fixed to 0.98, as suggested by Santee and Crisp [1993]. The CO\(_2\)-ice thermal emissivity \(e\) and visible albedo \(A\) are set to 0.7 and 0.5, respectively (Pollack et al. [1993] have found that those values give reasonable fits to the Viking pressure observations). The dust optical properties are fixed, following Pollack [1982], with single-scattering albedo and asymmetry factor in the visible of 0.85 and 0.79, respectively, and a ratio between the absorption efficiency in the thermal infrared \(Q_{\text{abs}}(\text{IR})\) and extinction efficiency in the visible \(Q_{\text{ext}}(\text{Vis})\) of 0.2.

As mentioned in paper 1, the interaction between explicitly resolved and unresolved motions is parameterized using an iterated Laplacian acting on the potential temperature and wind components. The operator depends upon one time constant, \(\tau_{\text{diss}}\), which corresponds to the time of dissipation of the smallest resolved scales. The Laplacian is iterated in order to be more selective in the smaller horizontal scales not acting directly on the large-scale dynamics. Here, the number of iterations \(\tau_{\text{diss}}\) is fixed to 2, and \(\tau_{\text{diss}}\) is set to 20,000 s. The dissipation is of the order of 0.2 s for zonal wave numbers \(n \sim 12\), of a few tens of s for \(n \sim 6\), and of more than a few tens of s for \(n \lesssim 5\).

The initial state was taken from an old simulation and, after a phase of stabilization (shorter than one Martian year) due to the modification of model parameters, we simulated 5 consecutive Martian years in order to estimate the interannual variability of the model results, which is of course, crucial for sensitivity studies. For comparison with Viking observations, the simulated surface pressure is interpolated at the two Viking sites from the four closest grid points. The Viking 1 point is somewhat shifted in longitude (from 47.9W to 42.5W) in order to match the currently assumed latitude of the landing site, 1.5 km below the reference altitude of Mars. For Viking 2 (which is generally assumed to be 1 km below Viking 1), since none of the four closest grid points is below -2 km, the surface pressure is extrapolated downward based on the hydrostatic relationship using the near surface simulated air temperature.

Viking 1 Pressure

The eight-harmonic fit to the simulated pressure at Viking 1 is reported in Figure 5 (thin solid curves). The interannual variability of the simulated seasonal pressure variations appears to be very weak. The amplitude of the seasonal variations is comparable to Viking 1 observations (dashed curve). This amplitude was some-
Figure 5. Eight harmonic Ls fits to the seasonal evolution of the surface pressure at Viking 1. The thin solid curves, almost superimposed, correspond to the five consecutive years of the control simulation. The thick dashed curve (shown in Figure 1) corresponds to Viking 1 observations.

Figure 6. Time evolution of the mean atmospheric surface pressure \( \bar{p}_{\text{atm}} \) (upper panel) and that of the equivalent surface pressure of the northern \( p_N \) and southern cap \( p_S \) for the first year of the control simulation. The sum of the three components \( p_{\text{tot}} \) (not shown) does not evolve along seasons and is equal to 6.854 mbar for the control experiment.

what overestimated in the simulations of paper 1. The difference is mainly due to the change of the albedo and emissivity of the pole ice, whereas the introduction of the diurnal cycle (which was tested independently) appears to have a minor effect on the seasonal pressure variations. In the present simulation, the minima and maxima occur approximately at the right seasons. However, a significant discrepancy exists between observations and model, the second pressure maximum being too high in the simulation. The amplitude of the second harmonic is also too large if compared to that of the first harmonic (compare Table 2 for the simulated pressure with Table 1) resulting from an insufficient asymmetry between the two hemispheres.

Planetary Mean of the Surface Pressure

The planetary mean of the surface pressure, \( \bar{p}_{\text{atm}} \), is shown in Figure 6 for the first year of the control simulation, as well as the equivalent pressure of the CO2 trapped in each cap, \( p_N \) and \( p_S \). First, it must be noticed that there is no permanent polar cap in our numerical experiments (this is a common point of most numerical studies of the atmospheric mass budget on Mars). Since the maximum mass of a given cap coincides with a bare pole in the other hemisphere, each pressure minimum really coincides with the maximum mass of the winter cap. The pressure maxima correspond to periods where the two caps are present but very small. The interannual variability of those three curves is very weak and would not be visible on the graph. Thus the small variability in the Viking 1 pressure (Figure 5) is caused almost exclusively by changes in the internal mass redistribution (meteorological component).

Relative Pressure at the Two Sites and Interannual Variability

Figure 7 shows the simulated relative pressure difference between the two landing sites for the five years. The amplitude is in good agreement with Viking observations with a maximum near northern winter solstice and a minimum at southern winter solstice, but high-frequency variations are weaker in the simulation. The simulation shows year-to-year variability but smaller than the discrepancy from Viking data. These interannual variations are due to modifications in the global latitudinal pressure variations. We will concen-
The Viking 2 pressure is about 0.55% lower in the second year. This is directly related to a change in the zonal wind shown in Figure 8. The upper panel shows the mean zonal wind for the first two years for $L_s = 295$ to 305. We recognize (1) the strong eastward jet in the northern middle and high latitude in geostrophic balance with the strong latitudinal gradient at the edge of the polar cap; (2) the near-surface eastward jet in the southern tropics created, as the Indian monsoon winds on Earth, by transport of angular momentum across the equator in the lower branch of the Hadley cell; and (3) rather weak winds in the rest of the southern hemisphere. The difference between year 1 and year 2 (lower panel) can be interpreted as an equatorward displacement of the winter jet.

The wind in the third layer of the GCM ($p/p_o = 0.966$ and $z \sim 300$ m) is shown for the two years in the upper panel of Figure 9 and the latitudinal pressure variation corresponding to geostrophic balance with this near surface zonal wind is reported in the middle panel of the same figure. The relative difference between year 1 and year 2 of this dynamical component is the solid curve of...
Figure 10. Amplitude of the transient pressure variations at Viking sites. The thin solid curve (shown in Figure 4) is the observation, and the thick curves correspond to the five successive years of the control simulation.

The lower panel. The corresponding pressure changes at Viking latitudes (22N for Viking 1 and 48N for Viking 2) are compatible with the change in the pressure simulated at Viking sites, which means that these pressure changes are almost totally explained by the modification of the mean zonal winds. Note also the good agreement between the actually simulated pressure modification and that reconstructed from the orographical and dynamical contributions.

Transients

Finally, the seasonal evolution of the amplitude of the simulated transient eddies is compared to Viking data in Figure 10. The agreement with Viking data is rather good in view of the very coarse horizontal resolution of this simulation. This is probably the consequence of the importance of transient waves with small wave numbers, which can be well simulated, even with such a coarse resolution. Once again, there is an interannual variability, but less than the discrepancy with observations. The model does not simulate the decrease of the amplitude of the transient eddies near northern winter solstice, and the decrease of this amplitude at the end of the northern winter is somewhat anticipated.

5. Sensitivity to Model Parameters for Clear-Sky Conditions

We present a series of numerical experiments performed by changing some model parameters, starting at $L_2 = 90^\circ$ (from a state of the control simulation) just before the formation of the southern polar cap in order to minimize the memory of the climatic system.

In all the simulations the northern cap disappears near $L_3 = 90$, after which the seasonal mass cycle adjusts in the sense that the total atmospheric mass is almost unchanged after one more year of simulation. Each sensitivity experiment thus includes an adjusting period going from $L_3 = 330^\circ$ to $L_3 = 98^\circ$ (date of the Viking 1 landing) followed by a 1-year simulation used for comparisons with the control simulation and Viking observations.

In this first set of experiments, the dust optical depth is kept to its "clear-sky" value: $r = 0.2$.

Horizontal Resolution

A simulation was performed at high horizontal resolution (64 longitude points and 48 latitude points) using exactly the same set of parameters as for the control simulation except for the near-surface horizontal diffusion time constant, which was set to $\tau_{\text{diff}} = 12,000$ s. This corresponds to a horizontal dissipation about 8 times weaker than that of the control simulation at a given spatial scale.

The seasonal variations of the surface pressure are compared to the control experiment and to the observations in Figure 11. In terms of the Viking 1 seasonal pressure variations, a simple quadratic error analysis would suggest that the agreement with observations is worse with the high resolution. However, it must be kept in mind that the total $\text{CO}_2$ mass is an adjustable parameter which directly influences the mean value of the simulated pressure. In fact, the results are somewhat better for the high resolution, since the differ-
ence between the two pressure maxima is somewhat enhanced. The change does not come from the atmospheric mass budget, which is almost unmodified (see the lower curves in Figure 11), but from the meteorological contribution $\alpha_{V1}$.

The representation of the differential pressure between Viking 1 and 2 is also closer to observations as shown in Figure 12, with a better amplitude and stronger rapid oscillations (Figure 7) than for the low resolution. However, the improvement of the amplitude must be taken rather carefully: it can be the consequence of changes in the representation of the winds but also of the modification of the representation of the geology at Viking sites.

For transient eddies (Figure 13), the main improvement, with respect to the control simulation (see Figure 10), is their longer persistence during late winter. Note that the main discrepancy with Viking observations is an overestimation of the amplitude of the transients near northern winter solstice.

In all the sensitivity experiments performed herein, we use the low horizontal resolution (in order to be able to perform a large number of experiments at a reasonable numerical cost), although a finer horizontal resolution would have been preferable to correctly represent the atmospheric dynamics.

Surface Roughness

Sutton et al. [1978] deduced values of 0.1-1 cm for the surface roughness length, based on the camera images of the surface around the Viking landers. These values, which were derived for local studies of the planetary boundary layer, are not representative of a mean Martian surface and may also not be appropriate for GCM's. It must be noted that even for terrestrial GCM's, a good determination of the roughness length is an open question. At the same time, this surface roughness strongly influences the near-surface wind and, in turn, the dynamical component of the horizontal pressure variations. Some recent experiments performed with the LMD terrestrial climate GCM, have shown, for instance, that a decrease by a factor of 2 of the surface drag coefficient leads to latitudinal pressure gradients 25% steeper in the southern midlatitudes (30-60S), where because of the absence of a significant orography, the horizontal pressure variations are completely dominated by the geostrophic balance with near-surface winds (dynamical component).

Four simulations were conducted for $z_0 = 0.1$ mm, 1 mm, 10 cm and 1 m (the control value was 1 cm). The simulated atmospheric pressure pattern is almost identical for the five experiments (including control), but there is significant effect on the relative pressure difference between Viking 2 and 1 (Figure 14): for the period $L_s = 300$ to 307, for instance, $\alpha_{V2/1} = 9.7\%$ for $z_0 = 1$ mm and $\alpha_{V2/1} = 10.7\%$ for $z_0 = 10$ cm (diamonds in Figure 14). As for the interannual variations of the control experiment, the local pressure changes at Viking sites are representative of the modification of the zonally averaged surface pressure (a 0.5% decrease at 22N and a 0.6% increase at 48N, lower panel of Figure 15) and are a direct consequence of the reduction of the intensity of the near-surface zonal wind (upper panel). Note that the change does not affect significantly (in comparison to the interannual variations) the global zonal wind distribution, but selectively, the near-surface winds. It is also noticeable that such large

**Figure 12.** The $\alpha_{V2/1}$ computed from observations and from the results of the high-resolution simulation (64 longitude points and 48 latitude points).

**Figure 13.** Amplitude of the transient pressure variations at Viking sites computed from observations and from the results of the high-resolution simulation (64 longitude points and 48 latitude points).
changes in the surface drag do not affect significantly the amplitude of the transient pressure variations (not shown).

Caps Albedo and Emissivity

Four experiments were performed by changing the ice visible albedo $A$ or thermal emissivity $\varepsilon$ simultaneously on the two caps. The four experiments correspond to $A = 0.4$, $A = 0.6$, $\varepsilon = 0.6$, and $\varepsilon = 0.8$. Albedo and emissivity have a direct impact on the energy budget on the caps and therefore strongly affect the mass budget. This is illustrated in Figure 16 and Figure 17, which show the time evolution of the equivalent pressure of the two caps. As expected, a reduction of the ice albedo (Figure 16) increases the absorption of the solar radiation, and reduces the mass of the polar cap. Emissivity acts in the same direction (Figure 17): less thermal emission by the cap results in smaller polar caps. However, the effect of the change in albedo and that of emissivity are not completely equivalent: a change in the ice emissivity has a rather constant effect during both the formation and recession phase, whereas albedo changes have a much weaker effect during the cap formation which mainly occurs in the polar night.

The dynamical component, $\alpha_{p2/1}$, and the transient pressure variations are not significantly affected by the modifications of the ice properties.

Other Uncertain Parameters

The choice of the parameters which have been varied involve very different physical processes and may give, all together, an overview of the sensitivity of the simulated atmospheric circulation to the values of the model parameters: the changes of horizontal resolution and dissipation time constants directly influence...
HOURDIN ET AL.: MARTIAN SURFACE PRESSURE AND ATMOSPHERIC MASS

Figure 17. Time evolution of the equivalent pressure of both polar caps for various values of the ice emissivity: $\epsilon = 0.6$, 0.7 (control), and 0.8.

Figure 18. Time evolution of the equivalent polar cap pressures $p_N$ and $p_S$ for the various dust sensitivity experiments ($\tau = 0.2, 0.5, 1$, and 2.5).

6. Dusty Conditions

Another series of sensitivity experiments was performed for different values of the dust opacity. Dust optical depths show a very large spatial and temporal variability on Mars, large enough to significantly affect the atmospheric energy and mass budget [e.g., Pollack et al., 1990, 1993]. Unfortunately, there is no complete climatology of the Martian dust optical depth, at least for a year without great dust storms, which could be used to constrain numerical simulations (the best temporal coverage corresponds to the Viking infrared thermal mapper (ITRM) data analyzed by Martin and Richardson [1993]). In this section, we first present simple sensitivity experiments performed with various values of the global dust optical depth: $\tau = 0.2, 0.5, 1$, and 2.5. A stronger horizontal dissipation in the upper atmosphere was needed for model stability when increasing the dust optical depth, probably due to shorter radiative time constants. For $\tau = 0.2, 0.5$, and 1, we use the same time constant as for the control experiment in the lower atmosphere ($\tau_{\text{diss}} = 20,000$ s), but $\tau_{\text{diss}} = 4000$ s above 0.15 mbar. For $\tau = 2.5$, we set $\tau_{\text{diss}} = 15,000$ s in the lower atmosphere and 2000 s in the upper atmosphere. The $\tau = 0.2$ experiment (same value as for the control experiment) was performed in order to make the three first experiments ($\tau = 0.2, 0.5$, and 1) identical except for the dust optical depth, and to check the impact of changing the horizontal dissipation parameters for the $\tau = 0.2$ case.

Effect on the Mass Cycle

The first effect of an increase of the dust opacity in our model is a global increase of the amount of carbon dioxide trapped in the caps, as shown in Figure 18, somewhat in contradiction with the results obtained by Pollack et al. [1990], which showed a small decrease of the condensation rate at $L_\phi = 379$ for increasing dust optical depths. In fact, the reduction of the atmospheric mass, in our simulations, is mainly due to a reduction of the sublimation rate during cap recession, especially near spring equinox, the formation phase being much less affected. The extension of the caps is also modified (Figure 19): for larger values of $\tau$, the caps are smaller during their formation but larger at the end of the recession.
Figure 19. Time evolution of the mean latitude of the edge of the polar caps for the dust sensitivity experiments ($\tau = 0.2$, 0.5, 1, and 2.5).

The sensitivity of the climate to the dust optical depths involves radiative as well as dynamical mechanisms. Here, we center our analyses on the $\tau = 0.2$ and $\tau = 1$ experiments (which are representative of the seasonal variability of dust for years without global dust storm) and on a period just before northern solstice ($L_s = 253 - 259$) when the northern cap is forming while the southern cap is still subliming.

Modifications of the Radiative Budget

The solar radiation (noted SW hereafter) reaching the surface is much smaller in the $\tau = 1$ than in the $\tau = 0.2$ experiment, by about 16% in midlatitude and 40% (40 W m$^{-2}$) on the southern cap (upper panel in Figure 20; note that the insolation is null from the northern pole down to 67°N during that period). The total planetary albedo is also somewhat reduced in the $\tau = 1$ experiment (see SW, second panel of Figure 20), the increase of the atmospheric albedo, resulting from dust scattering, being smaller than the reduction of the visible light reflected by the surface, especially on the highly reflecting polar ice. As a result, for $\tau = 1$, more visible radiation is absorbed by the global system (atmosphere plus solid planet), and much more by the atmosphere alone, especially in the summer hemisphere, which increases the latitudinal thermal contrasts.

The surface temperature is almost identical in the two simulations, as can be seen from the surface thermal emission ($e_T$, Figure 20), which is only significantly affected near the edge of the southern cap due to a larger ice cover in the 70-80°S region at that particular season (also responsible for the larger planetary albedo for $\tau = 1$ in this latitude range). The atmospheric thermal radiation absorbed by the surface ($e_T LW^s$, Figure 20) is larger in the $\tau = 1$ experiment, as expected from the larger atmospheric emissivity. This flux is also affected by the modifications of the atmospheric temperature shown in Figure 21. The atmosphere is globally warmer in the $\tau = 1$ experiment, as a consequence of the larger direct absorption of solar radiation, but colder by about 5 K over the forming cap, up to the 3-mbar level. On the north pole itself, the temperature is not affected and corresponds in both cases to a condensing atmosphere.
Figure 21. Zonally averaged temperature simulated just before northern winter solstice ($L_s = 253-259$) for the $\tau = 0.2$ (upper panel) and the $\tau = 1$ (middle panel) experiment as well as the difference between both (lower panel).

Energy and Mass Budget in the Polar Regions

In fact, there is a great difference between the processes controlling the formation and the recession of the polar caps. During recession, the mass exchange is controlled by the surface heat balance (compare upper and lower panels). Note that the atmospheric condensation extends to much lower latitudes for the $\tau = 0.2$ than for the $\tau = 1$ simulation (not shown).

On the northern pole surprisingly, although the air temperature decreases with altitude, the thermal emission to space, $LW_\text{top}$, is (1) somewhat larger than the surface thermal emission for both simulations and (2) larger in the $\tau = 1$ than in the $\tau = 0.2$ experiment. Both results are the consequence of the low thermal inertia of the icy surface, which minimizes the role of the emission emitted by the surface and enhances the role of the thermal emission by the atmosphere (in particular through its reflection on the surface) as can be checked with a simple analytical one-layer radiative transfer model.

Note that for an ice emissivity $\epsilon = 1$, Pollack et al. [1990] found a decrease of the total condensation rate at this season for increasing dust amounts.

Finally in our simulations, the net effect, in terms of atmospheric radiative budget at the top of the atmosphere, of an increase of $\tau$ from 0.2 to 1 are (Figure 22, upper panel): (1) an increase of the hemispheric asymmetry with a larger gain of energy in the summer hemisphere and smaller loss in the winter hemisphere (including the polar region); and (2) a decrease of the strong energetic gain on the subliming cap (except at the edge of the cap due to the different cap extensions of the two simulations).

Energy and Mass Budget in the Polar Regions

In fact, there is a great difference between the processes controlling the formation and the recession of the polar caps. During recession, the mass exchange is controlled by the surface heat balance (compare upper and lower panels). Note that the atmospheric condensation extends to much lower latitudes for the $\tau = 0.2$ than for the $\tau = 1$ simulation (not shown).

On the northern pole surprisingly, although the air temperature decreases with altitude, the thermal emission to space, $LW_\text{top}$, is (1) somewhat larger than the surface thermal emission for both simulations and (2) larger in the $\tau = 1$ than in the $\tau = 0.2$ experiment. Both results are the consequence of the low thermal inertia of the icy surface, which minimizes the role of the emission emitted by the surface and enhances the role of the thermal emission by the atmosphere (in particular through its reflection on the surface) as can be checked with a simple analytical one-layer radiative transfer model.

Note that for an ice emissivity $\epsilon = 1$, Pollack et al. [1990] found a decrease of the total condensation rate at this season for increasing dust amounts.

Finally in our simulations, the net effect, in terms of atmospheric radiative budget at the top of the atmosphere, of an increase of $\tau$ from 0.2 to 1 are (Figure 22, upper panel): (1) an increase of the hemispheric asymmetry with a larger gain of energy in the summer hemisphere and smaller loss in the winter hemisphere (including the polar region); and (2) a decrease of the strong energetic gain on the subliming cap (except at the edge of the cap due to the different cap extensions of the two simulations).
lower panel of Figure 22) which is itself very close to
the radiative balance at the top of the atmosphere, due
to the weak role of the thermally driven circulation,
shown in the upper panel of Figure 23. To the con-
try, during winter, condensation occurs both in the
atmosphere and on the surface, and since the radiative
fluxes are much smaller, the latitudinal heat advection
plays a significant role in the model sensitivity.
What we call thermally driven circulation hereafter is
the total circulation minus the condensation flow (sec-
ond panel of Figure 23). As proposed by Pollack et al.
[1990], the condensation flow is defined at a given lati-
titude as the zonally and vertically averaged meridional
mass flux. The condensation flow accounts for the lati-
tudinal mass redistribution and includes, in fact, in ad-
dition to condensation, the meteorological latitudinal
mass redistribution. On the subliming cap, the conden-
sation flow is the principal component of the latitudinal
energy transport and just corresponds to the transport
toward lower latitudes of the energy the atmosphere has
gained by increasing its mass.
The cap recession is thus essentially controlled by the
radiative balance, giving a simple explanation for the
major difference between the $\tau = 0.2$ and $\tau = 1$ simu-
lations. In late winter, when the latitudinal extension
of the winter cap is maximum and the sublimation rate
is already high in the midlatitudes, the smaller incident
flux on the surface plus smaller extent of the cap in the
$\tau = 1$ simulation explains the much slower cap reces-
sion at this season. Since the recession is slower, the
cap finally becomes larger than in the $\tau = 0.2$ case,
after which the smaller radiative flux in the $\tau = 1$ ex-
periment is partially compensated by the larger surface
of sublimation and leads to almost identical recession
rates, in terms of mass, in late spring.
During the cap formation, the slightly larger atmo-
spheric thermal emission to space in the $\tau = 1$ exper-
iment tends to reduce the total (atmosphere plus sur-
fave) condensation. Heat advection by overturning cells
(zonal meridional circulation minus condensation flow)
is much stronger in midlatitudes than for the $\tau = 0.2$
simulation, while the transient eddies, which transport
energy from middle to high latitudes in the winter hemi-
sphere (as expected from baroclinic eddies), are strongly reduced.

The increased heating by the mean meridional circulation is mainly sensitive near the edge of the cap (50-70N) responsible for a weaker condensation in that region for $r = 1$ (lower panel of Figure 22) and for the smaller cap extension during the formation phase (Figure 19). To the contrary, the reduction of the poleward energy transport by transient eddies has a major impact in very high latitudes (70-85N), where it is responsible for a larger condensation rate in the $r = 1$ experiment. As a consequence, the mass of the cap is more concentrated near the pole in the $r = 1$ experiment.

Seasonal Pressure Variations

The change in the condensation-sublimation affects directly the mean atmospheric pressure $p_{\text{atm}}$ (lower curves in Figure 24). For increasing dust opacities, the Viking 1 pressures (upper curves) are closer to observations, in that the second pressure maximum is more reduced than the first one. This is due in part to the modification of the meteorological component $\alpha_{\text{VL1}}$ shown in Figure 25. It must be noticed however, that $\alpha_{\text{VL1}}$ is mainly sensitive to the dust optical depth near northern summer solstice, which is in reality, the period of minimum atmospheric dust content [Martin, 1986].

The effect on the differential pressure between Viking 2 and 1 (Figure 26) is very different, in that it is large only during northern autumn and winter. It corresponds to a strong increase of the Viking 2 pressure for increasing dust opacities. As for the control and roughness sensitivity experiments, the modification of $\alpha_{\text{VL2}}$ (by about 2% for $L_0 = 295 - 300$) is representative of the modification of the zonally averaged surface pressure (see Figure 27): the main effect is a strong pressure increase in the northern middle and high latitudes (by about 6% at 60N), corresponding to a weakening of the circumpolar depression due to the reduction of the near-surface eastward winds in the 30-60N region. In the present case, modifications of the near-surface winds

Figure 25. The $\alpha_{\text{VL1}} = p_{\text{VL1}}/p_{\text{atm}}$ for the dust sensitivity experiments ($r = 0.2, 0.5, 1,$ and 2.5).

Figure 26. Relative pressure difference $\alpha_{\text{VL2}}$ between the two Viking sites for the dust sensitivity experiments ($r = 0.2, 0.5, 1,$ and 2.5).

Figure 27. (Upper panel) Mean zonal wind (m s$^{-1}$) in the third atmospheric layer ($p/p_{\text{atm}} = 0.966$ and $z \sim 300$ m) for the $r = 0.2$ and $r = 1$ experiments for the period $L_0 = 295 - 300$. (Lower panel) Various contributions to the pressure modification between $r = 0.2$ and $r = 1$ (percent): dynamical and orographical contributions and their combined effect (computed as the sum of the latitudinal derivative of the two contributions) as well as the actual pressure change.
from an increase of the dust visible optical depth from Figure 28. Zonal wind (m r structure are the signature eddies Transient Eddies We also performed a sensitivity experiment varying n instructed in the range of the variability of the control simulations for increasing optical depths, coinciding with the decrease of the latitudinal energy transport mentioned previously. More remarkable is the fact that the amplitude for the transient eddies was much less reduced during the first global dust storm 1977-A, Ls = 280 – 300, than during the 1977-B dust storm, Ls = 280 – 300. Note also that the variance of the transient pressure variations in the r = 0.2 experiment for the period Ls = 280 – 300 (about 0.03-0.04 mbar) is of the order of that observed by Viking 2 during the 1977-B dust storm. The explanation for this seasonal behavior would require much more sophisticated diagnostics and is beyond the scope of the present paper, but it is worthwhile noticing that Martian GCM’s seem to be well suited for further studies of the Martian transients (see Barnes [1980, 1981], and Barnes et al. [1993] for more complete studies).

Time-Varying Dust Opacity

We also performed a sensitivity experiment varying the dust opacity as a simple cosine function of solar longitude (r = 0.5 + 0.3 cos(Ls – Ls0)), with Ls0 = 280, which qualitatively matches the seasonal evolution deduced for the Viking years, except for the period of dust storms [Pollack, 1982; Martin, 1989].

This experiment is compared to the previous experiments for constant optical depths, in terms of the two first harmonics of the pressure simulated at Viking 1 (Table 3). The improvement of the simulations with increasing optical depth (described previously) is very clear in the two-harmonic decomposition: as r increases, the ratio between the first and second harmonic increases. The experiment with variable optical depth (VAR1) does not show any significant improvement, even when compared to the r = 0.2 experiment.

Uncertainties Arising From Dust Optical Properties

Some of the results presented above are rather sensitive to the numerical values adopted for the dust optical properties. In particular, we use in the present paper a ratio QIR/QVIS (Vis) of 0.2 taken from Pollack [1982], whereas the radiative code of the Ames GCM, which accounts for scattering in the thermal infrared,

Table 3. Phase Lag (Solar Longitude of the Maximum Pressure) and Amplitude of the First Harmonics of the Pressure Seasonal Cycle as Simulated at Viking 1 for the Various Dust Sensitivity Experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>g(1)</th>
<th>g(2)</th>
<th>L(1)</th>
<th>L(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r = 0.2</td>
<td>9.84</td>
<td>0.58</td>
<td>320.7</td>
<td>0.423</td>
</tr>
<tr>
<td>r = 0.5</td>
<td>7.90</td>
<td>0.68</td>
<td>316.6</td>
<td>0.602</td>
</tr>
<tr>
<td>r = 1</td>
<td>7.84</td>
<td>0.68</td>
<td>313.3</td>
<td>0.573</td>
</tr>
<tr>
<td>r = 2.5</td>
<td>7.78</td>
<td>0.64</td>
<td>311.4</td>
<td>0.586</td>
</tr>
<tr>
<td>VAR1</td>
<td>7.96</td>
<td>0.66</td>
<td>314.2</td>
<td>0.621</td>
</tr>
<tr>
<td>VAR2</td>
<td>7.94</td>
<td>0.60</td>
<td>314.7</td>
<td>0.624</td>
</tr>
</tbody>
</table>
The impact of this choice was tested numerically by rerunning the \( r = 0.2 \) and \( r = 1 \) experiments for \( Q_{\text{abs}}(\text{IR})/Q_{\text{ext}}(\text{Vis}) = 0.4 \) and \( Q_{\text{abs}}(\text{IR})/Q_{\text{ext}}(\text{Vis}) = 1 \). In both cases, contrary to the \( Q_{\text{abs}}(\text{IR})/Q_{\text{ext}}(\text{Vis}) = 0.2 \) case, the surface temperature is globally warmer in the \( r = 1 \) experiment, the enhancement of the greenhouse effect being larger than the diminution of the amount of solar energy reaching the surface. Note also that for increasing optical depths, the total condensation on the forming cap increases for \( r = 1 \), which provides an additional explanation for the difference between the present results and that published by Pollack et al. [1990].

The simulation with time-varying dust opacity was also rerun for \( Q_{\text{abs}}(\text{IR})/Q_{\text{ext}}(\text{Vis}) = 0.4 \), which did not produce any significant improvement in terms of comparison with Viking seasonal pressure variations (experiment VAR2 in Table 3).

7. Best Fit Simulations

Method

The method we propose in order to fit Viking pressure data is based on the decomposition of the local surface pressure presented above: for Viking 1, for instance,

\[
P_{\text{VL1}} = \rho_{\text{VL1}} \times \text{Patm} = \rho_{\text{VL1}} \times (P_{\text{tot}} - P_{\text{s}} - P_{\text{N}})
\]

(in that section, we always refer to the eight-harmonic fit instead of the pressure variations themselves). Hopefully, the time evolution of \( \rho_{\text{VL1}} \) which cannot be deduced from observation, is much less sensitive to the model parameters than \( \rho_{\text{VL1}} t_{\text{fit}} \). For all the low-resolution simulations with \( r = 0.2 \), this factor does not vary by more than 3%. Even the experiments with varying dust opacities are within this range, the effect of dust opacity on the value of \( \rho_{\text{VL1}} t_{\text{fit}} \) being maximum near northern summer solstice (Figure 25) when the atmosphere is clearer. Finally, for our best fit simulation, \( \rho_{\text{VL1}} t_{\text{fit}} \) is just taken from the simulation used as a basis for the best fit procedure.

We then take advantage of the high sensitivity of the atmospheric pressure to the ice albedo \( A \) and emissivity \( e \) to fit the Viking 1 pressure variations by allowing, for the two caps, independent values of both parameters, \( (A_N, t_{\text{N}}) \) and \( (A_S, t_{\text{S}}) \) for the northern and southern cap, respectively. We also allow variations of the total amount (ice plus gas) of carbon dioxide \( \text{Ptot} \). Assuming that the CO\(_2\) trapped in one particular cap is to a first order independent of the total CO\(_2\) amount \( \text{Ptot} \) and of the ice properties of the other cap, the mean surface pressure \( \text{Patm} \) can be written as

\[
\text{Patm} = \rho_{\text{tot}} (A_N, t_{\text{N}}) - \rho_{\text{tot}} (A_S, t_{\text{S}})
\]

The change \( \delta\text{Patm} \) of the mean atmospheric pressure, resulting from small changes of these five parameters, can be written formally as

\[
\delta\text{Patm} = \delta\text{Ptot} - \delta A_N \frac{\partial \text{Patm}}{\partial A_N} \delta A_N
\]

\[
= \delta\text{Ptot} - \delta A_N \frac{\partial \text{Patm}}{\partial A_N} \delta A_N - \delta E_N \frac{\partial \text{Patm}}{\partial E_N} \delta E_N
\]

The sensitivity functions, \( \delta\text{Ptot}/\delta A_N, \delta\text{Ptot}/\delta E_N, \delta\text{Ptot}/\delta A_S, \) and \( \delta\text{Ptot}/\delta E_S \), computed from the results of the sensitivity experiments performed with \( A = 0.4-0.6 \) and \( e = 0.6-0.8 \) are shown in Figure 30. The functions themselves (thin curves) show rapid oscillations. This may be partly explained by the fact that the ice cover is not fractional: either a mesh is completely ice, or it is bare. If, due to a change in the ice parameters, the icing of a mesh occurs later or sooner, it makes a discontinuous change in the sensitivity. We retained smoothed functions (thick curves) for the best fit procedure.

The differences between the response to emissivity and albedo changes mentioned previously appears very clearly in Figure 30: at the beginning of the cap formation, an increase of the ice emissivity increases the mass of the cap about three times more than an equivalent increase of the ice albedo, whereas albedo and emissivity have an equivalent effect in the late recession period. This is especially important for our present purpose since, by acting independently on the four ice parameters, we act on functions which are strongly uncorrelated, with four maxima rather regularly distributed within a year.

![Figure 30. Sensitivity of the equivalent pressure of a given polar cap to a change of its albedo or emissivity, computed using the results of the sensitivity experiments presented previously. For instance, \( \delta\text{Ptot}/\delta A_S \) is computed as \( \delta\text{Ptot}/\delta A_S \) = \( \text{Ptot} (A_S = 0.6) - \text{Ptot} (A_S = 0.4) \) / 0.2. The thin lines correspond to the direct computation, and the thick (dotted or dashed) curves correspond to the smoothed functions used for the best fit procedure.](image-url)
Finally, the best fit values of the parameters can be obtained by minimizing the following cost function measuring the difference between the simulated and observed pressures:

$$J (p_{001}, A_N, \epsilon_N, A_S, \epsilon_S) = $$

$$\sum_{\text{days}} \left\{ P_{\text{V1}} - \alpha_{\text{V1}} (p_{001} - F_N - P_S - \delta A_N \frac{\partial P_N}{\partial A_N} - \delta \epsilon_N \frac{\partial \epsilon_N}{\partial A_N} - \delta A_S \frac{\partial P_S}{\partial A_S} - \delta \epsilon_S \frac{\partial \epsilon_S}{\partial A_S}) \right\}^2 \tag{12}$$

(where the sum is done with one value per day on one Martian year). The cost function $J$ can be easily minimized by cancelling of its derivative which is equivalent to inverting a five-equation linear system. However, such a direct minimization produces extreme values of the emissivity and albedo. In fact, good fits could also be obtained by imposing one of the ice parameters. We thus introduce a supplementary constraint: we look for the fit which requires the minimum artificial asymmetry between the two hemispheres by adding to the cost function $J$ a second term

$$\xi \left( (A_S - A_N)^2 + (\epsilon_S - \epsilon_N)^2 \right) \tag{13}$$

accounting for hemispheric asymmetries, where $\xi$ represents the relative importance given to this asymmetry with respect to the departure from Viking 1 data (in fact, $\xi$ has been normalized in order to have a transition region around $\xi = 1$). For very high values of $\xi$, we obtain the best fit with the same ice properties for both caps. For $\xi = 0$, we obtain the best unconstrained fit.

Once we have obtained, for a given value of $\xi$, a set of best fit parameters $p_{001}, A_N, \epsilon_N, A_S$ and $\epsilon_S$, it is possible to compute a synthetic pressure curve at Viking 1 as

$$P = \alpha_{\text{V1}} (p_{001} - F_N - P_S - \delta A_N \frac{\partial P_N}{\partial A_N} - \delta \epsilon_N \frac{\partial \epsilon_N}{\partial A_N} - \delta A_S \frac{\partial P_S}{\partial A_S} - \delta \epsilon_S \frac{\partial \epsilon_S}{\partial A_S}) \tag{14}$$

**Best Fit Results**

We first present results of the best fit procedure applied to the control simulation. Figure 31 shows, for values of $\xi$ varying from $10^{-3}$ to $10^{6}$, the values of the best fit albedos and emissivities as well as the mean quadratic error between the synthetic and observed Viking 1 pressure. For $\xi < 1$, the best fit values are very different for the two caps, especially for albedo with $A_S = 0.3$ and $A_N = 0.65$. On the contrary, for large $\xi$, we find values of the ice parameters rather close the control values, with $\epsilon_S = \epsilon_N = 0.64$ and $A_S = A_N = 0.5$. Figure 32 shows the synthetic pressure at Viking 1 as obtained for $\xi = 0$, $\xi = 23$, and $\xi = 10^6$. Of course, the best agreement with Viking data (thin solid curve) is obtained for $\xi = 0$, but the agreement is still satisfactory for $\xi = 23$. The corresponding values of the ice parameters are given in Table 4. The difference $A_N - A_S$ is twice smaller for $\xi = 23$ than for $\xi = 0$. As expected also, it is mainly the first harmonic of the seasonal cycle which is sensitive to the value of $\xi$.

---

**Figure 31.** Best fit values of the ice albedo and emissivity computed for various values of the constraint coefficient $\xi$ and the corresponding mean quadratic difference between the synthetic and observed Viking 1 pressure.

**Figure 32.** Three synthetic best fits for three different values of the constraint parameter $\xi$ ($10^6$, 23, and 0). The pressure at Viking 1 is also shown (solid curve). All the curves correspond to eight-harmonic fits.
The best fit procedure was validated, a posteriori, by performing a new simulation analogous to the control simulation except that the ice parameters and total atmospheric mass were changed to the best fit values of the $\xi = 23$ case. The mean quadratic difference between the synthetic (deduced from the best fit algorithm) and simulated seasonal pressure evolutions is 0.035 mbar, which is surprisingly good in view of the simplicity of the best fit procedure.

The same approach was applied to the high-resolution simulation. Since the annual mass cycle is very close in the control and high-resolution simulations, we used the sensitivity functions computed from the low-resolution experiments. After some tests, we chose the $\xi = 10$ case ($p_{\text{tot}} = 6.152$ mbar, $A_N = 0.57$, $\epsilon_N = 0.53$, $A_S = 0.41$, and $\epsilon_S = 0.62$), which gives a mean quadratic error of 0.042 mbar, to perform a new high resolution simulation. The simulated pressure at the Viking sites are presented in Figure 33.

Note that when we do not allow differences between the northern and southern ice properties, the best fit algorithm gives a mean quadratic error of 0.064 mbar only (much lower than for the low resolution) for $p_{\text{tot}} = 6.527$ mbar, $A = 0.51$, and $\epsilon = 0.58$.

The best fit procedure was also applied for other dust scenarios, some of which (like the $r = 1$ experiment) were closer to the Viking observations in terms of the two first harmonics of the seasonal cycle than for the control experiment. But at the same time, with more aerosols, the atmospheric mass budget is less sensitive to albedo or emissivity changes. As a consequence, all the best fits with high or time-varying dust opacities led to more extreme values of the ice emissivity and albedo.

8. Concluding Remarks

Best Fit Parameters and Polar Processes

We have derived an algorithm which provides best fits to the Viking pressure seasonal cycle by varying independently the albedos and emissivities of the two polar caps. Thanks to the automatic procedure we use, we obtain values of the mean quadratic difference with Viking observations $r = 0.04-0.05$ mbar when north-south asymmetries are allowed and $r = 0.1$ mbar when identical values are imposed for the two caps. For comparison, the best fit published by Pollack et al. [1993] corresponds to $r = 0.13$.

As Wood and Paige [1992] and Pollack et al. [1993] results, our study confirms the necessity for low values of the polar cap emissivity (of the order of 0.7). It also suggests the necessity for an asymmetry between the southern and the northern polar caps with a lower emissivity and larger albedo in the north. Note that, except for the $r = 1$ simulation, our best fit values of the cap albedos lie within the range of observations [see Pollack et al., 1993, Table 4].

As already discussed, for instance, by Pollack et al. [1990], the necessity for low emissivity can be interpreted either in terms of real CO$_2$ ice properties or in terms of physical processes not accounted for in the model. Those low emissivity values are, at first approximation, compatible with the low values of the thermal emission recorded by the IRTM instrument aboard Viking orbiter, which have been recently analyzed extensively by F. Forget et al. [Low brightness temperatures of Martian Polar caps: CO$_2$ clouds or low emissivity? manuscript in preparation, 1995].

If not due to a real low emissivity of the icy surface, this low thermal emission could be due for instance to opaque CO$_2$ clouds colder than the surface [Pollack et al., 1990, 1993]. An alternative or additional explanation could arise from our rather poor knowledge of the polar orography. In fact, on the ice caps, there is a direct link, through the Clausius-Clapeyron law (2) and hydrostatic balance, between surface temperature and altitude variations:

$$\delta z = \frac{L}{g} \frac{dT}{T} \sim -1.6 \times 10^3 \frac{dT}{T}.$$ (15)

An underestimate by 1.6 km of the elevation of the

---

**Table 4. Best Fit Parameters for $\xi = 0$, $\xi = 23$ and $\xi = 10^4$, and Mean Quadratic Error Based on the Low-Resolution Control Simulation**

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$p_{\text{tot}}$ (mbar)</th>
<th>$A_N$</th>
<th>$\epsilon_N$</th>
<th>$A_S$</th>
<th>$\epsilon_S$</th>
<th>Error, mbar</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6.58</td>
<td>0.54</td>
<td>0.47</td>
<td>0.40</td>
<td>0.70</td>
<td>0.033</td>
</tr>
<tr>
<td>23</td>
<td>6.60</td>
<td>0.58</td>
<td>0.53</td>
<td>0.41</td>
<td>0.66</td>
<td>0.051</td>
</tr>
<tr>
<td>10$^4$</td>
<td>6.67</td>
<td>0.50</td>
<td>0.64</td>
<td>0.50</td>
<td>0.64</td>
<td>0.094</td>
</tr>
</tbody>
</table>

---

**Figure 33. High-resolution best fit simulation ($p_{\text{tot}} = 6.152$ mbar, $A_N = 0.57$, $\epsilon_N = 0.53$, $A_S = 0.41$, and $\epsilon_S = 0.62$) of the pressure at the Viking sites and the observations smoothed by retaining the eight first harmonics of the seasonal cycle.**
surface orography would correspond to an overestimation of the surface temperature in the model by about 1.4 K (the cap temperature being of the order of 140 K), which should be compensated, in order to reproduce good global condensation rates, by a decrease of the surface emissivity by 4%.

Improvements in our understanding of the polar atmospheric energy and mass budget on the polar caps will probably require more sophisticated parameterizations of the surface properties and atmospheric condensation, including the representation of CO₂-ice clouds, a careful comparison of the numerical results with the observations of the thermal emission by the Viking IRTM instruments as well as a better knowledge of the actual orography in the polar regions.

Interannual Variability and Dust

One intriguing result concerning the seasonal pressure variations recorded by the Viking landers is the very weak interannual variability. In particular, the pressure cycle was not significantly affected by the two global dust storms of the first Viking year, the temporary increase of the surface pressure at Viking 2 during the second dust storm being mainly due to modifications of the dynamical component [Pollack et al., 1989; paper 1].

For a given set of model parameters, the simulated interannual variability of the seasonal mass cycle is also very weak (see Figure 5) and compatible with observations, but, on the other hand, this cycle depends strongly on the atmospheric dust content, in a complex manner involving modifications of the radiative budget and atmospheric dynamics.

The link with observation is premature at this stage and would require a more realistic representation of the temporal and spatial variations of the atmospheric dust content. This could be achieved either by introducing in the model a climatology of the atmospheric dust content, or by modeling directly the atmospheric dust transport into various components, already proposed in paper 1, has proved to be a powerful diagnostic tool for GCM simulations. The surface pressure appears to be especially sensitive to modifications of the dynamical component arising from modifications of the mean zonal winds following changes of the atmospheric angular momentum budget. This dynamical component is particularly sensitive to surface roughness and atmospheric dust content. Note that landing probes at 30° and 60° latitude in each hemisphere (the latitudes which are most affected by the dynamical component) plus one at the equator would probably be optimum to constrain the dynamical component and, in turn, some model parameters and the annual mass cycle.

Figure 34. Planetary averaged surface pressure $P_{\text{avg}} = P_{\text{VL1}} \times \alpha_{\text{VL1}}$, where $P_{\text{VL1}}$ is the eight-harmonic fit to the seasonal pressure variations observed by Viking 1 and $\alpha_{\text{VL1}}$ is taken from the high-resolution sensitivity experiment (dashed curve) and high-resolution best fit simulation (thick solid curve). The two thick solid curves have been computed by adding ±1.5% to the meteorological factor of the high-resolution best fit simulation.

One output of the present study is also the determination of the time evolution of the atmospheric mass from the Viking observations. Figure 34 shows the seasonal evolution of the planetary averaged surface pressure reconstructed from the Viking 1 pressure observations scaled by the meteorological factor $\alpha_{\text{VL1}}$, taken from the high-resolution sensitivity experiment (thick dashed curve) and best fit experiment (thick solid curve). The two thin curves, computed by adding ±1.5% to the meteorological factor of the best fit simulation, correspond to the maximum uncertainty estimated from the maximum variations of the meteorological factor in all the sensitivity experiments using a constant $\tau = 0.2$ or a time-varying dust optical depth.

Note also that the agreement of the simulated and observed relative pressure between the two Viking landers can be obtained only for a rather short range of values for the altitude difference between the two sites. The dashed curve in Figure 12 was computed assuming that Viking 2 was 980 m below Viking 1. A change by 50 m of this altitude difference would have shifted $\alpha_{\text{VL1}}$ by about 0.5%, resulting in a rather bad fit. This gives an estimation of the altitude difference: $dz = -980 \pm 50$ m.

Transient Eddies

The amplitude of the transient eddies in the clear-sky high-resolution simulation is in good agreement with Viking observations, except near northern winter sol-
stes, when it is strongly overestimated. The sensitivity experiments suggest that the increase of the dust optical depth during winter is responsible for the reduction of the transient activity in that season. In good agreement with Viking observations, this reduction is found to be stronger near northern winter solstice. Note also that the amplitude of the transient waves is much smaller in the southern hemisphere (not shown), as also found with the NASA/Ames GCM (Barnes et al., 1992).

Implications for Spatial Exploration

We have mentioned above how GCM's can be used for the preparation of spacecraft missions, for instance, with the conjunction with spacecraft observations. This approach also suggests that the increase of the dust optical depth is much stronger near northern winter solstice. Note also that the amplitude of the transient waves is much smaller in the southern hemisphere (not shown), as also found with the NASA/Ames GCM (Barnes et al., 1992).

It also appears clearly from the present study that one priority of the future Martian missions must be the determination of the Martian orography. An accurate knowledge of the orography is of prime importance both (1) for estimation of the energy budget over the caps, and (2) for fully efficient use of local pressure measurements such as the Viking measurements.

This study underlines more generally the usefulness of using self-consistent models such as GCM's in conjunction with spacecraft observations. This approach could be made much more systematic by using the "data assimilation techniques" developed for the purpose of operational weather forecasting. For that goal, we are currently developing at LMD the adjoint map of our Martian GCM, which we intend to use for four-dimensional variational assimilation (e.g., Talagrand and Courtier, 1987; Talagrand and Courtier, 1991) of the data of future Martian missions.

Acknowledgments. The computational means were provided by Institut du Developpement et des Ressources en Informatique Scientifique of Centre National de la Recherche Scientifique. The Viking pressure measurements were produced by Tillman and O'Keefe (1987) and Tillman (1988) and provided to us by Steven Lee. They are available in digital form through the Planetary Data System. For information on ordering the data set(s) or related documentation, contact the PDS Planetary Atmospheres Discipline Node (303-492-5348; e-mail: Isrsyrite.asgs.colorado.edu). SPAAN: ORION:(J.LEE); Steven Lee, Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80309 on the PDS Central Node (361-305-6026; e-mail: SPAAN:JPLPDS-PDS.OPERATOR). SPAAN: Jet Propulsion Laboratory, MS 525-354L, 4800 Oak Grove Drive, Pasadena, CA 91109. We are also grateful to James Murphy and Robert Haberle for their useful comments on the manuscript.

References


HOURDIN ET AL.: MARTIAN SURFACE PRESSURE AND ATMOSPHERIC MASS


Tillman, J. E., and W. Guest, Atmospheric pressure, point by point values: VL 1 sols 1-2245 and VL 2 sols 1-1050, technical report, National Space Science Data Center, Greenbelt, MD, 1987.


Tillman, J. E., and W. Guest, Atmospheric pressure, point by point values: VL 1 sols 1-2245 and VL 2 sols 1-1050, technical report, National Space Science Data Center, Greenbelt, MD, 1987.


Page intentionally left blank
Post-Phobos model for the altitude and size distribution of dust in the low Martian atmosphere

Eric Chassefière
Service d'Astronomie, Centre National de la Recherche Scientifique, Versailles-Le-Buisson, France

P. Drossart
Département de Recherche Spatiale, Observatoire de Paris-Meudon, Meudon, France

O. Korabiev
Space Research Institute, Moscow, Russia

Abstract. Four experiments flown on board Phobos 2 provided information on the characteristics of the dust particles suspended in the Martian atmosphere: Auguste (UV-visible-IR spectrometer working in solar occultation geometry), ISM (IR spectrometer measuring the light of the Sun reflected by the planet), Termoskan (scanning radiometer mapping the planetary thermal radiation), KRFM (UV-visible multichannel spectrometer providing limb-to-limb profiles). These experiments, which sounded equatorial regions (20°S-20°N) near the northern spring equinox (Ls=0-20°), are shown to yield a reasonably consistent picture of the dust distribution over the whole altitude range from the ground level, or just above, outside the boundary layer, up to ~25 km. The vertical profiles of particle volume mixing ratio and effective (projected area-weighted) radius deduced from Auguste measurements, performed in the 15-25 km altitude range, are extrapolated down to the ground by using a simple, physical parameterization of the altitude dependence of dust mixing ratio and radius. This parameterization, which must be understood as describing the vertical distribution of dust in the zonal average and on the mesoscale in latitude, assumes that gravitational settling and eddy diffusion are the only two processes driving vertical dust transport. The vertically averaged effective radius and optical depth of dust particles, as well as vertical profiles of related quantities, are obtained. Optical depth at 1.0-μm wavelength is found to be 0.2 on average, with a typical variation of ±0.1 with time and space. This result is similar to that obtained from ISM spectra analysis. It is also consistent with the Termoskan and KRFM measurements, which yield near-infrared optical depths of 0.12-0.26 and 0.12-0.24, respectively. The particle number density near the surface, as derived from extrapolation of solar occultation profiles, is in the range 1-3 cm⁻³, in good agreement with Termoskan results (1-2 cm⁻³). The scale height of the dust volume mixing ratio just above the surface is ~8-9 km on average, that is, of the same order as the background atmospheric scale height. The vertically averaged effective radius of dust particles is found to lie in the range 1.7±0.2 μm, possibly ~2 μm in the case of a large effective variance of 0.4. The most likely ISM value is 1.2 μm, with a rather large uncertainty of ±0.4 μm, mainly due to the fact that the spectral dependence of the Minnaert coefficient is not well known. Because ISM data used in the present work were obtained on the Tharsis plateau, at a mean altitude of ~7 km, the ISM radius must be compared to the Auguste vertically averaged radius for ~7 km, that is, ~1.5±0.2 μm. Auguste and ISM radii are therefore consistent at the 1-σ level. Three typical vertical profiles of the dust particle radius and number density, obtained by averaging all solar occultation profiles, including their extrapolated parts below ~15 km, are proposed as reference models, for three selected values of the effective variance of the particle size distribution (0.10, 0.25, and 0.40).

Introduction

It has been known since the Mariner 9 and Viking missions that a persistent dust opacity, varying from a few tenths to nearly 5 during global dust storms at visible wavelengths, characterizes the Martian atmosphere. The dust cycle is expected to play an important role in determining the climate of Mars through, notably, its influence on the atmospheric thermal structure, the ground albedo, and the interhemispheric transport of water and carbon dioxide ice. It was recently pointed out by Murphy et al. [1993] that the temporal evolution of the dust opacity during global dust storms cannot be well reproduced by numerical simulations. One of the most critical aspects is that the size distribution of particles suspended in the atmosphere is not preserved over space and time, with a rapid depletion of large particles through gravitational settling. The question of the size distribution of the suspended dust particles, including
dependence with altitude, is therefore of crucial importance for a better understanding of the Martian dust cycle. First results on the composition and size of the dust particles were obtained from Mariner 9 UV and IR spectrometer data (Conrath, 1975; Pang and Ajello, 1977; Toem et al., 1977). Information was also obtained with TV experiments (Anderson and Leovy, 1974). Later observations of aerosols from Viking Orbiter and Lander cameras, as discussed respectively by Clancy and Lee (1991) and Pollack et al. (1979), allowed the picture previously obtained to be completed and refined. The effective (projected area-weighted) radius of the dust particles, as seen from the Martian surface, is ~2.5 μm, although a dust particle size 5-10 times smaller cannot be ruled out (Clancy and Lee, 1991).

Within the large measurement uncertainties and the interdependent sensitivities on composition, particle size, and optical depth, the measurements suggest that the optical parameters of the particles (single scattering albedo, asymmetry factor) do not vary with place and time. The size distribution of dust is constant at subsolar latitudes. A compilation of all results obtained from spacecraft observations up to the Viking mission is given by Murphy et al. (1993).

More recently, the Martian aerosols were observed by different instruments onboard the Phobos 2 spacecraft. The vertical profiles of the mixing ratio and size of the dust particles in the 15-25 km altitude range were deduced from solar occultation measurements performed by the Auguste instrument [Korablev et al., 1993; Chassefière et al., 1992]. These measurements were made in a narrow band of latitude in equatorial regions (0-20°N) near northern spring equinox (Ls=0-20°). The effective radius of particles was found to vary from 0.8 μm at 25 km up to 1.6 μm at 15 km. Their number density is ~0.3 cm⁻³ in the same altitude range. The effective variance of the distribution is ~0.25. The vertical optical depth of dust at 1.9-μm wavelength is ~0.2, that is, about the minimum value observed from Viking landers, suggesting a somewhat clearer year even at this season of minimum dust loading. A small effective radius of ~1.25 μm was inferred from the analysis of the dust component observed by the ISM infrared spectrometer [Drossan et al., 1991]. Spectrophotometry of Mars by the KRFM instrument yielded a small value of the dust optical depth, ~0.1-0.2 [Moroz et al., 1993]. Finally, the thermal infrared emission of the limb of Mars, as measured by the scanning radiometer Termoskan, suggested a small visible optical depth of ~0.13, with a number density of dust particles at the ground level of 1-2 cm⁻³ [Moroz et al., 1995]. All pre-cited Phobos measurements were made in equatorial regions (20°S-20°N).

The aim of this paper is to reconcile these various estimates and thereby obtain a consistent view of Phobos results. Whereas solar occultation measurements provided for the first time the vertical structure of the dust in the 15-25 km altitude range, infrared imaging allowed the dust in the low atmosphere to be characterized, typically for the first atmospheric scale height above the ground. It is tempting to merge these two kinds of measurements in order to infer the complete profile of the dust from the ground up to ~25 km altitude. Additional information provided by other instruments may be used to check the global consistency of the results. The rather marked similarity between the nine vertical profiles of dust extinction coefficient obtained by solar occultation at various longitudes shows that atmospheric conditions do not change much in space and time at these low latitudes during this early northern spring period. This result is consistent with the pre-Phobos picture, although the dust amount indicated by Phobos observations is quite low, as compared to the higher opacities observed by Viking during the same Mars season. It suggests, together with the small optical depth of dust, a typical, stable equatorial atmosphere. Other arguments in this sense are developed by Rosenqvist and Chassefière (this issue) from a more general analysis of Phobos results. In this context, it seems interesting, and useful for modelers of the dust cycle, to examine the different results obtained by the Phobos mission with the main objective to propose a single synthetic vertical profile of the dust (size, number density).

Solar occultation profiles are extrapolated from the 15-25 km altitude range down to the ground, that is, the zero reference altitude (altitude of the reference ellipsoid), by using a specific parametrization of the altitude dependence of the volume of particles per unit atmospheric volume, which will be called volume mixing ratio in the following, and the particle effective (projected-area weighted) radius. Vertically averaged quantities obtained in this way are compared to those observed by the infrared spectrometer and other instruments in order to check the validity of the extrapolation procedure. The parametrization used for extrapolation is based on the simplified steady state model used by Chassefière et al. [1992]. Gravitational settling is assumed to be balanced by eddy diffusion, parameterized by an empirical eddy diffusion coefficient K. The validity of the steady state model will be reassured in the concluding section. In particular, the fact that no marked variation of the dust column density with altitude is measured on the flank of the volcanoes by ISM suggests additional mechanisms at low atmospheric levels [Drossan et al., 1991].

The Data

The main characteristic of the instrument is listed in Table 1.

Table 1. Instrument Characteristics

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Wavelength Range, μm</th>
<th>Wavelengths Used for Dust Retrieval, μm</th>
<th>Observation Mode</th>
<th>Altitude Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auguste</td>
<td>0.21-0.33, 0.76/0.94, 1.93/1.7</td>
<td>1.9, 3.7</td>
<td>solar occultation</td>
<td>15-25 km at limb</td>
</tr>
<tr>
<td>ISM</td>
<td>0.7-3.2</td>
<td>1-2</td>
<td>surface mapping</td>
<td>column</td>
</tr>
<tr>
<td>KRFM</td>
<td>0.3-0.6</td>
<td>1-2</td>
<td>limb-to-limb scan</td>
<td>column</td>
</tr>
<tr>
<td>Termoskan</td>
<td>7-13</td>
<td>1-2</td>
<td>limb scan</td>
<td>column</td>
</tr>
</tbody>
</table>

These values are scale height.
Solar Occultation Data

The infrared spectrometer on board the Phobos spacecraft has been designed for solar occultation measurements of the Martian atmosphere. It is a part of a more complete instrument analyzing the spectrum of the solar radiation in different wavelength ranges. The principle of the Auguste experiment is to measure from the Phobos orbit the spectrum of the Sun modified by atmospheric extinction during sunset. The infrared spectrometer provides two spectra of 20 and 16 elements in the ranges 5292-5372 cm⁻¹ (2.95 μm) and 2797-2740 cm⁻¹ (3.7 μm) with a resolution of 1200. A detailed description of the instrument is given by Blamont et al. [1989] and Krasnopolsky et al. [1989]. Because of the large apparent diameter of the Sun modified by atmospheric extinction during sunset, the instrument analyzing the spectrum of the solar radiation in the Martian atmosphere. It is a part of a more complete project for solar occultation observations, which provide information on the vertical profile of the dust volume mixing ratio and particle radius, and of the atmosphere. The altitude of the line of sight resulting from uncertainties on CO₂ absorption and atmospheric profile. The coordinates (longitude, latitude) of the tangent point are shown in Figure 1.

Infrared Spectrometer Data

The observations taken by the infrared spectrometer ISM, on board the Phobos spacecraft, are described by Bibring et al. [1989] and Combes et al. [1991]. This spectrometer is operating close to opposition, with phase angles typically ranging from 5° to 10°. Therefore, the contribution of dust opacity is seen in a different viewing geometry as compared to the Auguste experiment. However, the main difficulty in these observations is to extract the dust component from the contribution of the surface. A procedure has been developed by Drossart et al. [1991] to extract the backscattered dust contribution from an average reflectance spectrum, obtained in the region of Pavonis Mons (outside the high-altitude ranges close to the volcano), recorded on March 14, 1989, with a phase angle of 6°. In this paper, we use the dust spectrum retrieved in Figure 4 of this Drossart et al., in contrast to using an improved calibration procedure for ISM. (S. Ehrig, personal communication, 1994). This spectrum is recovered under the assumption of small optical depth for suspended dust, and uniform surface properties along the sequence of observations. Both hypotheses are consistently verified along the retrieval procedure. The reflectance of the ISM spectra can be shown to verify

\[ R = a_q \cdot K \cdot H(n-d) \]

where \( H \) is the vertically averaged phase function of the aerosols, \( a_q \) the vertically averaged single scattering.

Table 2. Times and Altitude Ranges of the Nine Solar Occultation Measurements of the Dust (All for Sunset Observations)

<table>
<thead>
<tr>
<th>Occultation</th>
<th>Date in 1989</th>
<th>Occultation Time, UT</th>
<th>A_q, deg</th>
<th>A_q, max, km</th>
<th>A_q, min, km</th>
<th>ΔZ, km</th>
</tr>
</thead>
<tbody>
<tr>
<td>23212158</td>
<td>Feb. 2</td>
<td>1636:57</td>
<td>1.9</td>
<td>24</td>
<td>15</td>
<td>+3, -2</td>
</tr>
<tr>
<td>2360639</td>
<td>March 7</td>
<td>0115:28</td>
<td>8.9</td>
<td>26</td>
<td>15</td>
<td>+4, -2</td>
</tr>
<tr>
<td>2362241</td>
<td>March 7</td>
<td>1717:14</td>
<td>9.2</td>
<td>25</td>
<td>20</td>
<td>+6, -3</td>
</tr>
<tr>
<td>2421504</td>
<td>March 13</td>
<td>0940:01</td>
<td>12.0</td>
<td>27</td>
<td>16</td>
<td>+2, -2</td>
</tr>
<tr>
<td>2440712</td>
<td>March 15</td>
<td>0148:22</td>
<td>12.8</td>
<td>26</td>
<td>15</td>
<td>+2, -3</td>
</tr>
<tr>
<td>2441514</td>
<td>March 15</td>
<td>0949:58</td>
<td>12.9</td>
<td>24</td>
<td>13</td>
<td>+2, -2</td>
</tr>
<tr>
<td>2450717</td>
<td>March 16</td>
<td>0153:19</td>
<td>13.3</td>
<td>23</td>
<td>16</td>
<td>+2, -2</td>
</tr>
<tr>
<td>2510743</td>
<td>March 22</td>
<td>0219:08</td>
<td>16.1</td>
<td>28</td>
<td>17</td>
<td>+4, -2</td>
</tr>
<tr>
<td>2560312</td>
<td>March 26</td>
<td>2147:43</td>
<td>18.4</td>
<td>24</td>
<td>17</td>
<td>+2,-3</td>
</tr>
</tbody>
</table>
measurements of other Phobos instruments recorded at the characteristics, and can be checked against the characteristics is a decreasing slope. As will be shown in following, this property can be related to the aerosol opacity roughly estimate that a multiplicative factor of 1.2 must be applied to transform the visible opacity into an infrared value (at 1.9 pm), calculated in the way already described, is 0.12-0.24. The dust scale height was estimated to be of the order of the gaseous atmosphere scale height, and the number density of dust particles near the surface was found to be in the range 1-2 cm⁻³.

Other Experiments
A series of five limb-to-limb photometric profiles was obtained by the KRFM instrument at low latitudes in the southern hemisphere [Morez et al., 1993]. The traces of the scanning line of sight are shown in Figure 1 for each one of the five observations. All profiles were obtained in eight selected spectral bands in the 315-550 nm spectral range. The spectral and spatial resolutions of the spectrometer are ≈10 (λ/Δλ) and ≈30 km, respectively. Thanks to an appropriate modeling of the brightness of the Martian disk, it was possible to determine several important characteristics of dust and cloud particles. In particular, the visible optical depth of the dust component on the morning limb in equatorial regions was found to lie in the range 0.1-0.2 at the time of the measurements (Lp=5'-16'). Because opacities derived from solar occultation measurements are obtained at 1.9-μm wavelength, it is necessary to scale the KRFM value. By using the optical index given in the appendix, one can roughly estimate that a multiplicative factor of 1.2 must be applied to transform the visible opacity into an infrared opacity [e.g., Hansen and Travis, 1974]. The resulting KRFM estimate at 1.9 μm is ≈0.12-0.24.

The thermal radiation of Mars in the 7-13 μm spectral range has been measured by the Termoskan instrument [Morez et al., 1994]. In particular, information was obtained near the limb of the planet where the atmosphere contributes significantly to the planetary thermal radiation. Termoskan is a scanning radiometer able to measure the spatial distribution of brightness with a high spatial resolution (~2 km). Eight adjacent, regularly spaced (~20 km intervals) cross-limb profiles of brightness were extracted from one image obtained on March 26, 1989, at ~20° S latitude (Figure 1). A visible optical depth of 0.13 (+0.09, -0.03) was inferred for the dust component. The corresponding infrared value (at 1.9 μm), calculated in the way already described, is 0.16 (+0.10, -0.04). The dust scale height was estimated to be in the range 1-2 cm⁻³.

Extrapolation of Solar Occultation Profiles

Treatment of Auguste Data
The corrections applied by Korablev et al. [1993] to the Auguste data take into account the nonlinearity of the detector and the change of the signal due to both the solar limb darkening and the partial loss of the solar disk when the pointing is disturbed. These corrections generally improve the aspect of the vertical profiles of the extinction coefficients \( \Delta \Sigma'(z) \) and \( \Delta \Sigma'(z) \) at 1.9 μm and 3.7 μm, respectively. Nevertheless, they induce small oscillations of the ratio \( \Delta \Sigma'(z)/\Delta \Sigma'(z) \), and therefore of the inferred radius of the particles. Because these oscillations occur on a vertical scale smaller than the projected size of the field of view (~4 km), they may be unambiguously attributed to an instrumental bias in the corrective algorithms. As shown hereafter, the vertical profile of the radius \( r(z) \) is extrapolated from ~15 km down to the ground by parameterizing the vertical variation of its "scale height" \( h(r) \approx 4 \log (r/h)^{-1} \), and even small oscillations of \( r(z) \) imply well-marked variations of \( h(r) \). It is emphasized that the radius obtained by comparing the uncorrected extinction coefficients is near the radius obtained by using the corrected values but does not present the small oscillations resulting from the instrumental bias. On the other hand, the accuracy of the
extinction coefficient profiles is improved when using Korablev’s corrections. For this reason, these corrections are used only for retrieving the extinction coefficient profiles, and therefore the volume mixing ratio of particles, but not for the ratio. Corrected and uncorrected profiles at 1.9 µm and 3.7 µm are shown in Figure 2 for occultation 2212158 (February 20). Although the general shape of the profiles is not changed by instrumental corrections, the irregularities of the uncorrected profiles, which cannot be from atmospheric origin due to the low vertical resolution (~4 km), are removed by the corrective algorithm.

The Method

The objective of the present section is to explain how the vertical profiles of particle radius and dust volume mixing ratio are extrapolated down to the ground in order to derive the ranges of the possible values of the effective radius and the optical depth, as seen in backscattering viewing geometry by the ISM instrument. As previously explained, nine vertical profiles of the extinction coefficient Σ(z) and Σ’(z) at 1.9 and 3.7 µm are obtained in an average altitude range of 15-25 km (see Table 2). Let us denote by a and b the effective radius and effective variance of the dust size distribution. The shape of the size distribution of particles is assumed to be [Hansen and Travis, 1974]

\[
db(a,b) = \frac{1.6b}{a} \frac{r}{e} e^{-\frac{r}{a}}
\]

The ratios \( \frac{\Sigma'(z)}{\Sigma(z)} \) and \( \sigma^2/\sigma^2 \), where \( \sigma^2 \) and \( \sigma^2 \) are the extinction cross sections calculated using the same optical index as Pollack et al. (1979), which depend on \( a, h, \) and a roughness parameter \( \rho \) (see appendix), are denoted by \( \sigma(a,b) \) and \( \sigma(a,h,p) \), respectively. For each couple \( (b,p) \), the value \( n_b(p) \) of index \( n \) such as \( \sigma(a,b,p) = \sigma(a,b) \) may be derived at any altitude \( z \). The number density of the particles may be deduced:

\[
n_b(p)(z) = \frac{\Sigma'(z)}{\Sigma(z)} n_b(p)(3)
\]

Let us denote by \( v_b(p) \) the average volume of one particle, obtained by averaging \( 4/3\pi r^3 \) over the size distribution given by (2). The volume mixing ratio of dust may be deduced:

\[
V(z) = \frac{\int n(z) \sigma(z) dz}{\int n(z) \sigma(z) dz}
\]

Figure 2. The extinction coefficients at 1.9 µm (thick lines) and 3.7 µm (thin lines) are plotted as a function of the altitude for occultation 2212158 (February 20). Although the general shape of the profiles is not changed by instrumental corrections, the irregularities of the uncorrected profiles, which cannot be from atmospheric origin due to the low vertical resolution (~4 km), are removed by the corrective algorithm. In this way, the vertical optical depth \( \tau \) at 1.9 µm and the vertically averaged effective radius \( A \) may be calculated:

\[
\tau = \int n(z) \sigma(z) dz
\]

and

\[
A = \frac{\int n(z) \sigma(z) dz}{\int n(z) \sigma(z) dz}
\]

where \( r(z) \) is the particle projected area, \( \pi r^2 \), averaged over the distribution given by (2). Practically, the integration may be stopped at the highest altitude where some extinction is detected, that is, \( z_{max} \approx 25 \) km (see Table 2). At this altitude, the extinction coefficient is lower than \( 10^{-5} \) km\(^{-1}\) and the residual optical depth \( \tau(z_{max}) \) is not in excess of \( 10^{-5} \times 10 \), that is, 0.01, using an atmospheric scale height as the characteristic vertical scale. This value is 1 order of magnitude below the uncertainty which will be finally derived. In the same way, the contribution of dust particles located above \( z_{max} \) to the vertically averaged effective radius \( A \) is weak. The values of \( \tau \) and \( A \) given by (5) and (6) depend on the assumed values of the effective variance \( \sigma^2 \) and roughness factor \( \rho \), and the associated derived value of the effective radius \( a \).

Extrapolation of the Dust Volume Mixing Ratio

The scale height of the volume mixing ratio \( V(z) \) of dust particles may be written [see Chassefière et al., 1992]

\[
H' = \frac{R}{1 + V_{sed}/(V_{edd})}
\]

where \( H \) is the scale height of the gas, CO2 atmosphere, \( V_{sed} \) the settling velocity of dust particles, and \( V_{edd} \) their eddy diffusion velocity defined as the ratio \( KH \), where \( K \) is the eddy diffusion coefficient. Values of \( K \) found in this way are of the order of \( 10^{8} \) cm\(^2\) s\(^{-1}\) and physical implications of this result have been extensively discussed in previous papers [Korablev et al., 1991; Chassefière et al., 1992]. Equation (7) is a direct consequence of the assumed balance between the eddy diffusion and settling fluxes of the dust. \( V_{sed} \) varies as the product of the particle radius by the inverse of the atmospheric density in relevant conditions.

By neglecting in first approximation the vertical variation of the radius, which is slow compared to the exponential
variation of the atmospheric density, the settling velocity $V_{set}$ may be assumed to vary as the inverse of the atmospheric density. Equation (7) may be written -

$$H'(z) = \frac{H}{1 + \sigma \exp(\Delta H)}$$

(8)

where $c$ is assumed to be constant with altitude. In this simplified parameterization, it is assumed that the vertical variation of the ratio $V_{set}/V_{edd}$ is dominated by the exponential component of the settling velocity and any rapid variation of $K$ over the first 25 km above the ground is ruled out. This assumption is a priori justified by the fact that $K$ seems to remain about constant between 15 km and 65 km [Rosenqvist and Chassefière, this issue] as derived from solar occultation data obtained at sunset. In addition, the average error bar on the total volume mixing ratio of dust deduced from the present study is about a factor 2.

The principle of the extrapolation is to fit the parameterized profile (8) to the observed profile of the volume mixing ratio scale height in the region where this profile is observed (15-25 km). To this end, the observed $H'(z)$ profile is averaged over the first 4 km above the lowest sounded altitude $z_{min}$, and its standard deviation $\sigma H'$ around $<H'>$ is calculated at the corresponding average altitude $<z>$ $(=z_{min} + 2$ km). The choice of a recovering zone of 4-km thickness is motivated by two independent arguments. First, the vertical resolution of the IR spectrometer is of the order of 4 km, and no major information is lost when averaging over 4 km altitude. Second, a slight change of the slope of the dust volume mixing ratio profile above 20 km altitude was noted by Korablev et al. (1993), and it seems better to remove points above 20 km altitude to connect theoretical and observed profiles. By adjusting the parameterized value of the dust scale height (8) to the mean ($<H'>$) and extreme ($<H'>\pm 2\sigma H'$) values at altitude $<z>$, it is possible to define a mean and two extreme values of $c$ at the 1-sigma level, therefore of the $H'(z)$ profile. The extrapolation of $V(z)$ below $z_{min}$ is performed by using the different possible profiles of $H'(z)$.

The dust scale height $H'$ obtained over the first atmospheric scale height above the ground is given in Table 3 for the nine available occultations. It is in the range 7.5-9.5 km.

Extrapolation of the Dust Effective Radius

The following differential equation has been obtained by Chassefière et al. [1992] for the effective radius $a$: $\frac{d \ln(a)}{dz} = -\frac{V_{set}}{V_{edd}}$ (9)

where $b$ is the effective variance of the dust size distribution. It has been analytically shown that $b$ is expected to remain constant with altitude [Chassefière et al., 1992]. These results have been obtained by solving the balance equation (eddy diffusion/settling) for each radius range independently from each other, which is justified by the fact that particles are coupled to the fluid atmosphere over characteristic times much smaller than the coagulation time, and by deriving the dependence of the size distribution with altitude. The scale

Figure 3a. The vertical profile of the effective radius scale height is shown in the right upper corner for occultation 2212158. The five extrapolated profiles (dashed lines) are obtained by adjusting the parameterized profile to the observed scale height at altitude $<H'>$ and to scale heights shifted by one half ($<H'> \pm \sigma H'$) standard deviation. The vertical profile of the effective radius, including extrapolated profiles, is shown in the left upper corner. The corresponding observed and extrapolated profiles for the dust volume mixing ratio and its scale height are plotted respectively in the left and right lower corners of the figure, respectively.
height $h$ of the effective radius $a(t)$, defined as the height over which $a(t)$ decreases by a factor $\exp(1)$, is therefore

$$h = \frac{N}{\rho V_{dust}}$$

and varies nearly as the atmospheric density:

$$h = h_0 \exp (-\alpha H)$$

As for deriving the vertical profile from the dust volume mixing ratio, and for the same reasons, $K$ is assumed to not vary significantly between the ground and 25-km altitude. As shown here before, the vertical variation of $h(a)$ over the altitude range where dust is observed agrees approximately with the law given by (11) in seven cases out of nine, with a definition decrease of $h(a)$ between 15-km and 25-km. This variation must be taken into account when extrapolating $h(a)$ down to the ground. The only free parameter is the radius scale height $h_0$ at the ground level. As in the case for the dust volume mixing ratio, the average value of $h(a)$ over a thickness of 4 km above $a_{min}$ and the associated standard deviation are calculated, allowing a mean and two extreme values of $h_0$ at the 1-σ level to be defined. The effective radius $a$ is extrapolated by using the different possible extrapolated profiles of $h$. Occultation 22121-28 (Figure 3a) is one of the two previously mentioned occultations for which the scale height of the radius above 20-km altitude does not follow well the parameterized law. The case of occultation 2360639 (Figure 3b), for which the scale height of the radius remains inside the envelope over the whole observed altitude range, is more representative of the general case. A direct examination of the radius profile in Figure 3b shows a well-marked curvature between 15-km and 25-km altitude, and a linear extrapolation would be clearly less realistic than our parameterized approach.

**Treatment of ISM Data**

To model the ISM spectra, the backscattering component of the dust has been simulated, in the single scattering approximation. According to (1), the backscattered dust component is completely defined from the vertically averaged $\mu_0 N$ and from the total optical depth $\tau_0$. These quantities are calculated using a Mie calculation, modified to take into account the scattering by irregular particles [Droz, 1990], with the introduction of a roughness factor which takes into account the irregularities of the scatterers. The parameters of the model are therefore a size distribution (we use the same form as (2) used for the solar occultation data), with an effective radius $a$ and a variance $b$. The optical constants are taken from Pollack et al. [1973], interpolated to each wavelength. In the calculations, a roughness factor of 0.067 was applied. It can be shown that this factor influences mostly the short-wavelength part of the spectrum, and is not predominant in the 1-2 μm range. The simulation is also not very sensitive to the variance $b$, which has been taken equal to 0.25 in the calculations, according to the solar occultation data analysis. Given an effective radius, the optical depth $\tau_0$ is measured from the comparison with ISM dust spectrum (Figure 4). It can be seen that the largest radius corresponds to the faster dust spectrum. We conclude from this analysis that an effective radius between 0.8 and 1.6 μm provides a good fit to the dust spectrum, with larger radii giving too flat spectra, and smaller radii too small backscattering at 2 μm. The even broader range of radius on the small side is consistent with error bars due to the surface Mie parameter. Larger radii give a poorer fit, and the size distribution of Foss et al. [1977] cannot account for the dust spectral shape between 1 and 2 μm. Nevertheless, it must be emphasized that we cannot exclude from ISM spectra bimodal distributions, with a large particle component ($a > 2$ μm) giving a flat dust backscattering component in addition to the decreasing slope due to the scattering by smaller particles.

**Results**

The possible values of the vertically averaged effective radius $a(t)$ and optical depth $\tau(t)$ at 1.9 μm (5) are shown in Figure 5 for the nine occultations. The effective variance $b$ is supposed to be 0.25, which is the most likely value derived from previous analyses ($b=0.25$15 [Chassefiere et al., 1992]; $b=0.25$15 [Keravel et al., 1993]), and the roughness factor $\rho$ is 0.067. Due to the uncertainty on the observed dust volume mixing ratio and effective radius in the altitude range where the theoretical profiles are constrained.
by the observed profiles, there is not a single value of the couple \((A, \theta)\) but an extended range which is represented by a box (solid line). In order to include the uncertainty on the altitude (see Table 2), the same calculations are made for the two extreme altitudes at 1 km, and the corresponding boxes are indicated with dashed lines. Except in the case of occultation 2362241, for which the estimates of optical depth and radius are inaccurate, there is a rather good agreement among the different observations. The large uncertainties for occultation 2362241 are linked to the fact that the observation is restricted to a small altitude range of 20-25 km, due to instrumental problems (see Table 2). As previously mentioned, the scale height of the dust volume mixing ratio tends to decrease above 20 km and, due to the fact that the parameterized profile is constrained by the observed profile in the range 20-25 km instead of 15-20 km in the other cases, the optical depth and radius in the lower atmosphere could be overestimated in this particular case.

Uncertainties resulting, on the one hand, from the extrapolation process and, on the other hand, from the not well-known altitude are combined (by adding variances, not standard deviations) for these results in Figure 6, with error bars for both IR optical depth \(\tau\) and effective radius \(r_{\text{eff}}\). Corresponding values of \(\tau\) and \(r_{\text{eff}}\) are given in Table 3. No correlation between optical depth and radius is exhibited. In addition, these two quantities are not correlated to the position (longitude, latitude, see Table 3) of the observed region on Mars. Apart from occultations 2212158 and 2362241, the optical depth at 1.9 \(\mu\)m is concentrated around a modal value of 0.2 and the effective radius of particles is in the range 1.5-1.8 \(\mu\)m. Histograms of optical depth and effective radius are also shown in Figure 6. The Gaussian

### Table 3. Characteristics of the Dust as Inferred From Phobos 2 Measurements

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Long. (\phi), deg</th>
<th>Lat. (\theta), deg</th>
<th>(\tau_{1.9}), (\mu)m</th>
<th>(r_{\text{eff}}, \mu)m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auguste</td>
<td>1.9 60.9 1.6</td>
<td>0.44</td>
<td>7.5</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.20,-0.09)</td>
<td>(x2.7;1.9)</td>
<td>(+0.1)</td>
</tr>
<tr>
<td></td>
<td>8.9 52.6 8.7</td>
<td>0.21</td>
<td>9</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.09,-0.03)</td>
<td>(x2.6;1.8)</td>
<td>(+0.1)</td>
</tr>
<tr>
<td></td>
<td>9.2 286.6 9.0</td>
<td>0.60</td>
<td>8</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.09,-0.22)</td>
<td>(x1.8;1.3)</td>
<td>(+0.4)</td>
</tr>
<tr>
<td></td>
<td>12.0 117.4 12.0</td>
<td>0.23</td>
<td>9</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.06,-0.04)</td>
<td>(x1.5)</td>
<td>(+0.1)</td>
</tr>
<tr>
<td></td>
<td>12.8 343.3 12.8</td>
<td>0.22</td>
<td>9</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.06,-0.05)</td>
<td>(x2.8;2.0)</td>
<td>(+0.1)</td>
</tr>
<tr>
<td></td>
<td>12.9 100.5 13.0</td>
<td>0.18</td>
<td>9</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.04,-0.03)</td>
<td>(x2.3;1.6)</td>
<td>(+0.2)</td>
</tr>
<tr>
<td></td>
<td>13.3 334.8 13.2</td>
<td>0.22</td>
<td>8.5</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.08,-0.05)</td>
<td>(x2.9;2.1)</td>
<td>(+0.1)</td>
</tr>
<tr>
<td></td>
<td>16.1 282.9 16.1</td>
<td>0.16</td>
<td>9.5</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.06,-0.02)</td>
<td>(x2.4;1.7)</td>
<td>(+0.1)</td>
</tr>
<tr>
<td></td>
<td>18.4 168.3 18.5</td>
<td>0.14</td>
<td>8.5</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.06,-0.04)</td>
<td>(x2.6;1.7)</td>
<td>(+0.2)</td>
</tr>
<tr>
<td>ISM</td>
<td>12.4 95/160 2.8/2.7</td>
<td>0.220.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0.10,-0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temsamkan</td>
<td>18.0 168/160 -22</td>
<td>0.16</td>
<td>+11</td>
<td>+1.2</td>
</tr>
<tr>
<td>KRPM</td>
<td>8.4 168/12 3/7</td>
<td>(+0.10,-0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.3 128/28 -7/12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.8 242/77 -8/13</td>
<td>0.12-0.24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>12.3 228/70 -8/13</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.6 176/2 -11/19</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4. Atmospheric reflectances retrieved from ISM observations (circles) on Pavonis Mons (March 14, 1989) compared with simulated radiances according to the procedure described in the text. Error bars correspond to the uncertainties in the (unknown) Minnaert surface coefficient. The solid lines correspond to three calculations with an optical depth of 0.2, with $a = 0.8$ (bottom), 1.2 (middle) and 1.6 $\mu$m (top).

The contributions of the nine occultations are cumulated in unit ranges of 0.2 $\mu$m for the radius and 0.1 for the optical depth. The contribution of occultation 2302241 is spread over wide intervals of radius and optical depth due to the large uncertainties. Well-marked maxima are obtained around 1.7 $\mu$m (±0.3 $\mu$m) for the radius and 0.2 (±0.1) for the optical depth. The peak of the effective radius is narrower for a small effective variance ($b=0.10$, Figure 7) and wider for a large effective variance ($b=0.40$, Figure 7), but the range is not changed. The assumption of spherical particles ($p=0$) does not change the picture, but an increase in $p$ ($p=0.2$) yields a small decrease of the modal effective radius (Figure 7). The nine vertical profiles of dust particle effective radius, extinction coefficient, volume mixing ratio, and number density have been averaged, and corresponding mean profiles are shown in Figures 8, 9, 10 and 11. Both average instrumental and geographical error bars at the 1-$\sigma$ level are represented by couples of extreme profiles. The instrumental

\[ b=0.25, \ p=0.067 \]

Figure 5. Domains of the possible values of the vertically averaged effective radius $A$ and vertical optical depth $\tau$ at 1.9 $\mu$m for the nine IR occultations. The three domains indicated for each occultation correspond to the most likely (solid line) and extreme (dashed lines) values of the lowest sounded altitude.
error bar is, for one given occultation, the error resulting from the different uncertainties linked to the measurement (instrumental noise, uncertain altitude, error in the retrieval of size and number density). The geographical error bar is the empirical standard deviation obtained from a statistical treatment of the results of the nine occultations at each altitude level. It is an indication of the natural variability of profiles with time and space. The geographical error bar is generally larger than the instrumental one in these data, except for the dust volume mixing ratio (Figure 10). The mixing ratio of the dust suspended in the atmosphere seems therefore to be almost constant with space and time, as expected if transport dominates. The number density, radius, and optical depth of particles are more variable. It should be noted that the volume mixing ratio of the dust is obtained without any assumption on the radius, whereas number density of particles depends on both mixing ratio and radius. The retrieval of number density, radius, and optical depth is therefore expected to be more model dependent.

Nevertheless, we think that the geographical variability of dust size, and derived quantities at constant volume mixing ratio, is real.

The particle number density averaged over the first atmospheric scale height is given for the nine available occultations in Table 3. The major source of uncertainty in this case is the effective variance $b$ of the particle size distribution. This may be seen in Figure 12, where the vertical profiles of effective radius and number density, averaged over the nine occultations, are plotted as a function of altitude for various values of $b$ and $p$. The radius and number density profiles corresponding to the mean ($b=0.25$) and extreme ($b=0.10, 0.40$) cases are given in Table 4 for a plausible roughness parameter of 0.067. It should be noted that a simple exponential extrapolation of optical depths measured by occultation would lead to a larger value of the total optical depth ($≈0.25$ rather than 0.2), although the vertical profile of number density would remain rather similar to those proposed in Table 4.
Figure 8. The mean vertical profile of the dust effective radius is shown with a thick solid line. Extreme profiles at the 1-sigma level corresponding to the averaged instrumental error bar and geographical variability are plotted with dashed and thin solid lines, respectively.

The dust optical depth derived from the present study (0.1-0.3, Figure 6) is in good agreement with KRFM and Termoskan ranges, 0.12-0.26 and 0.12-0.20 (Table 3), respectively. It is similar to the ISM value (Table 3). A detailed examination of Table 3 and Figure 7 shows that there is a certain variability of optical depth with space and time. This is proven by Figure 9, where it can be seen that the instrumental error bar is a factor of 2 smaller than the geographical one. The dust scale height, and more significantly the number density of dust particles near the surface, as deduced from Termoskan measurements, are in good agreement with Auguste values (Table 3). Seven of the nine values of number density obtained from solar occultation measurements are in agreement with Auguste values (Table 3). One of the most interesting points is the relatively small value of the effective radius deduced from both Auguste (1.7±0.2 μm) and ISM (1.2±0.4 μm), as compared to the generally accepted pre-Phobos value (=2.5 μm [Toon et al., 1977; Pollack et al., 1979]). Clancy and Lee [1991] have argued that the Viking data are consistent with a dust particle size 5-10 times smaller than previously admitted, but there is no confirmation of such a small radius in Phobos data. It is emphasized that the variability in space and time of the effective radius, as deduced from solar occultation measurements, is small. There is one exception (occultation 2362241, Tables 2, 3) but, as already mentioned, this particular case is less reliable than the others. The particle effective radius, derived from Auguste, is never smaller than 1.5 μm. It is nevertheless emphasized that the ISM observation selected in the present work was performed on the Tharsis plateau, westward of Pavonis Mons, and that the footprint of the field of view covered a wide altitude range from <0 km up to ~10 km, with a mean altitude of the sounded region of ~7 km. The averaged value of 1.7 μm, obtained at zero reference altitude, must be reduced by about 0.2 μm, which put Auguste and ISM data in good agreement at the 1-sigma level. Cross validation of data is therefore satisfactory, and the extrapolation used to extend solar occultation vertical profiles from >15 km, which is the lowest sounded altitude in occultation, down to zero reference altitude may therefore be considered as realistic.

Figure 9. Same as Figure 8 for the dust extinction coefficient.

Figure 10. Same as Figure 8 for the dust volume mixing ratio.

Figure 11. Same as Figure 8 for the dust number density.
Discussion

The approach which was used to extrapolate vertical profiles of dust volume mixing ratio and particle radius from the range where they are measured (15-25 km down to the ground is based on the assumption that vertical transport may be treated in terms of eddy diffusion. Prior to any assessment of the validity of this concept, it is necessary to estimate the relevant temporal and spatial scales. The timescale for eddy mixing is equal to ~10 days by assuming a value of the eddy diffusion coefficient of $10^6$ cm$^2$ s$^{-1}$ (Korablev et al., 1993). The typical time for a dust particle of radius $\sim 2$ $\mu$m to fall one atmospheric scale height at the 2mbar level is also ~10 days (Kahn et al., 1992). These times are much smaller than the mean congelation time ($\sim 10^7$ s) and particles may be supposed to not interact. On the other hand, they are definitely greater than 1 sol (1 Martian solar day) and the vertical structure of dust is not expected to follow possible diurnal variations of atmospheric conditions, at least outside the boundary layer (2-4 km). Spatial scales are obtained by multiplying timescales by the wind velocities in the relevant altitude range (0-25 km). By assuming typical values of the zonal and meridional wind velocities of $\sim 10^3$ m s$^{-1}$ and $\sim 1$ m s$^{-1}$, respectively, zonal and meridional scales are $\sim 10^5$ km and $\sim 10^3$ km, respectively. The coefficient $K$ used in the steady state model must therefore be considered as a zonally averaged empirical quantity including both vertical mixing and meridional advective transport on the mesoscale.

Although it is not appropriate in a rigorous way, this kind of assumption is currently used in terrestrial photochemical models. Because there is no dramatic change of the observed dust profile and inferred quantities for varying time and longitude of observation, the use of a spatially averaged value of $K$ is justified.

It is emphasized that the steady state model is not used to fit the data, but only to parametrize the vertical variation law of some dust characteristics. In particular, the possibility that gravitational settling of particles should be considered by upwelling, with typical ascending velocities of some mm s$^{-1}$ in equatorial regions, does not invalidate the extrapolation laws used in the present work (8) and (11).

Although it might introduce a small error due to the variation of the vertical wind velocity over the relevant altitude range (2-25 km), the possible presence of upwelling is implicitly taken into account when fitting data to the model in the sounded altitude range. In a more general way, the range of extrapolated profiles obtained for each observed profile, as shown in Figure 3, does not seem to be very dependent on possible large-scale disturbing effects. Moreover, the validity of the extrapolation method is supported by the lack of any slope breaking around 15-km altitude on profiles presented in Figures 8-12.

There are several ways dust particles may be lifted from the surface into the atmosphere. The recent model of Haberle et al. (1993 b) shows a strong diurnal variation of the height of the boundary layer. During the morning and the beginning of the afternoon, the surface is heated by sunlight. The overlying atmosphere is warmed by ground thermal radiation, with a regular increase of the height of the convective layer from 100 m up to $\sim 4$ km at 16 hours, followed by a rapid collapse due to the fact that the ground temperature falls below the atmospheric temperature. It is thought that direct heating of the ground might give rise to strong upward transport of dust by convection, with subsequent formation of sporadic dust-devils or convective vortices (Kahn et al., 1992). These kinds of structures have been observed, with dust columns of 1 km in width reaching 6 km in altitude, near the top of the boundary layer. It was also argued that, in addition to direct solar heating, regional winds might generate the initial shear stress required to lift surface dust. Although no upwind dust cloud larger than a few kilometers across has been observed, the role of local storms in redistributing the dust could be of prime importance. The fact that no marked variation of the dust column density with altitude was measured from Phobos 2 on the flank of the volcanoes (Droz et al., 1991) is perhaps the consequence of such a phenomenon. It was argued by the authors that strong slope winds, induced by the solar heating of the slopes, should be responsible for transport of dust and result in the complete redistribution of the suspended dust over the whole area of volcanoes. In this way, the layer of dust would follow the relief instead of being confined in low regions.
Table 4. Three Reference Models for the Vertical Distribution of the Number Density and Effective Radius of Dust Particles

<table>
<thead>
<tr>
<th>Altitude, km</th>
<th>( n ), cm(^{-3} )</th>
<th>( r_{\text{eff}} ), ( \mu m )</th>
<th>( n ), cm(^{-3} )</th>
<th>( r_{\text{eff}} ), ( \mu m )</th>
<th>( n ), cm(^{-3} )</th>
<th>( r_{\text{eff}} ), ( \mu m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.26</td>
<td>1.90</td>
<td>1.31</td>
<td>1.76</td>
<td>5.56</td>
<td>2.05</td>
</tr>
<tr>
<td>1</td>
<td>2.06</td>
<td>1.88</td>
<td>1.18</td>
<td>1.75</td>
<td>5.09</td>
<td>2.03</td>
</tr>
<tr>
<td>2</td>
<td>1.88</td>
<td>1.87</td>
<td>1.07</td>
<td>1.74</td>
<td>4.67</td>
<td>2.01</td>
</tr>
<tr>
<td>3</td>
<td>1.71</td>
<td>1.85</td>
<td>0.97</td>
<td>1.73</td>
<td>4.28</td>
<td>1.99</td>
</tr>
<tr>
<td>4</td>
<td>1.56</td>
<td>1.84</td>
<td>0.87</td>
<td>1.72</td>
<td>3.93</td>
<td>1.97</td>
</tr>
<tr>
<td>5</td>
<td>1.43</td>
<td>1.82</td>
<td>0.79</td>
<td>1.71</td>
<td>3.62</td>
<td>1.94</td>
</tr>
<tr>
<td>6</td>
<td>1.30</td>
<td>1.80</td>
<td>0.72</td>
<td>1.70</td>
<td>3.34</td>
<td>1.91</td>
</tr>
<tr>
<td>7</td>
<td>1.19</td>
<td>1.78</td>
<td>0.65</td>
<td>1.68</td>
<td>3.08</td>
<td>1.88</td>
</tr>
<tr>
<td>8</td>
<td>1.09</td>
<td>1.75</td>
<td>0.59</td>
<td>1.67</td>
<td>2.85</td>
<td>1.85</td>
</tr>
<tr>
<td>9</td>
<td>1.00</td>
<td>1.73</td>
<td>0.53</td>
<td>1.65</td>
<td>2.65</td>
<td>1.81</td>
</tr>
<tr>
<td>10</td>
<td>0.92</td>
<td>1.70</td>
<td>0.48</td>
<td>1.64</td>
<td>2.47</td>
<td>1.78</td>
</tr>
<tr>
<td>11</td>
<td>0.85</td>
<td>1.67</td>
<td>0.43</td>
<td>1.62</td>
<td>2.31</td>
<td>1.74</td>
</tr>
<tr>
<td>12</td>
<td>0.78</td>
<td>1.64</td>
<td>0.39</td>
<td>1.60</td>
<td>2.16</td>
<td>1.69</td>
</tr>
<tr>
<td>13</td>
<td>0.72</td>
<td>1.61</td>
<td>0.36</td>
<td>1.58</td>
<td>2.04</td>
<td>1.65</td>
</tr>
<tr>
<td>14</td>
<td>0.68</td>
<td>1.57</td>
<td>0.33</td>
<td>1.55</td>
<td>1.94</td>
<td>1.60</td>
</tr>
<tr>
<td>15</td>
<td>0.64</td>
<td>1.54</td>
<td>0.30</td>
<td>1.53</td>
<td>1.86</td>
<td>1.55</td>
</tr>
<tr>
<td>16</td>
<td>0.60</td>
<td>1.49</td>
<td>0.27</td>
<td>1.50</td>
<td>1.79</td>
<td>1.50</td>
</tr>
<tr>
<td>17</td>
<td>0.56</td>
<td>1.45</td>
<td>0.25</td>
<td>1.47</td>
<td>1.72</td>
<td>1.44</td>
</tr>
<tr>
<td>18</td>
<td>0.52</td>
<td>1.41</td>
<td>0.23</td>
<td>1.44</td>
<td>1.65</td>
<td>1.38</td>
</tr>
<tr>
<td>19</td>
<td>0.48</td>
<td>1.36</td>
<td>0.20</td>
<td>1.41</td>
<td>1.56</td>
<td>1.33</td>
</tr>
<tr>
<td>20</td>
<td>0.44</td>
<td>1.32</td>
<td>0.18</td>
<td>1.38</td>
<td>1.47</td>
<td>1.27</td>
</tr>
<tr>
<td>21</td>
<td>0.40</td>
<td>1.26</td>
<td>0.16</td>
<td>1.34</td>
<td>1.40</td>
<td>1.20</td>
</tr>
<tr>
<td>22</td>
<td>0.38</td>
<td>1.20</td>
<td>0.14</td>
<td>1.29</td>
<td>1.32</td>
<td>1.14</td>
</tr>
<tr>
<td>23</td>
<td>0.35</td>
<td>1.15</td>
<td>0.13</td>
<td>1.25</td>
<td>1.25</td>
<td>1.08</td>
</tr>
<tr>
<td>24</td>
<td>0.33</td>
<td>1.09</td>
<td>0.12</td>
<td>1.20</td>
<td>1.20</td>
<td>1.02</td>
</tr>
</tbody>
</table>

and no vertical variation of the optical depth should be seen on a small scale or even mesoscale.

If this were the case on a global scale, a correlation between the optical depths derived from solar occultation measurements, which are calculated with respect to the common zero reference altitude (altitude of the reference ellipsoid), and the actual altitude of the sounded region would be expected. It is definitely not the case. Small optical depths of \( 4.15 \pm 0.20 \) are obtained indifferently above high regions, like the flanks of Aracraeus Mons (occultation 2441514), and low regions like Amazonis Planitia (occultation 2510743). Any redistribution of dust by convection in the low atmosphere does not propagate to higher atmospheric levels, where solar occultation measurements were made (15-25 km). Since the timescale for vertical mixing is \( \approx 10 \) days, horizontal homogenization must occur over a typical scale of \( 10^5 \) km (by assuming a near-surface wind velocity of \( 1 \) m s\(^{-1} \)), which suggests that convective redistribution of dust must be regional rather than global. It must be noted that dust optical depths at 1.9-\( \mu m \) wavelength, as shown in Table 3, must be reduced by \( \approx 30\% \) on average when stopping the integration of the extinction coefficient at the ground level, instead of the zero reference altitude. Profiles presented in Figures 8-12 must be considered as averages over the low-latitude reference geoid. They should be cut at the ground altitude in any modeling using them as input data. In summary, the Phobos observational results are self-consistent. However,
uncertainties in dust properties remain large, particularly near the surface.

The sporadic presence of expanding low dust clouds may result in an excess of large particles in the first half scale height above the ground on the mesoscale. Nevertheless, particles whose radius is larger than \( \sim 10 \mu \) m and whose settling time between \( \lesssim 5 \) km altitude and the ground is lower than \( \sim 10 \) (typical period of one diurnal cycle of the boundary layer) cannot be efficiently levitated. Typical mean profiles presented in Figures 8-12 might be locally modified below \( \lesssim 5 \) km altitude, with a possible additional component of large particles, whose radius should be in the range 2-10 \( \mu \) m. There is no clear evidence for such a phenomenon in the Phobos data.

The vertically averaged effective radius derived from Phobos data peaks around a modal value of \( \sim 1.7 \mu \) m (Figures 6, 7, Table 3), which is smaller than the value derived from Viking lander measurements [Pollack et al., 1979], \( \sim 2.3 \mu \) m. The Phobos value of the radius at the ground level is \( \sim 1.9 \mu \) m (Figure 8). This becomes nearly \( 2.1 \mu \) m if a large effective variance of 0.4 is assumed (Figure 12). It must be noted that the Viking value is consistent at 1-, considering geographical variability, with Phobos values. This is mainly due to occultations 2362241, which provides a larger variance of the radius (2.3 \( \mu \) m). Although this particular occultation, which is responsible for the high-radius wing (2 to 3 \( \mu \) m) of the histograms shown in Figures 6 and 7, is less reliable than others, for reasons already explained, no strong disagreement appears between Viking and Phobos data. A factor of 2. The roughness parameter \( p \) [Drossart, 1987] might be modified for modeling purposes. There is nevertheless some uncertainty due to the fact that the variance of the size distribution of particles is not accurately known.

Appendix : Calculation of Scattering Extinction Cross Sections for Irregular Silicate Particles

The physical parameters used to calculate the scattering extinction cross sections for a given size distribution are as follows.

1. The real and imaginary optical index [Pollack et al., 1975] at 1.9 \( \mu \) m: \( N = 1.51 + 0.0026 i \), and at 3.7 \( \mu \) m: \( N = 1.48 - 0.0040 i \).
2. The roughness parameter \( p \) [Drozdz, 1990], which parameterizes the scale of irregularities, as compared to the size of the scatterer.
3. The two parameters of the size distribution, according to the analytical form used in (2): the effective (projected area-weighted) radius of particles and the effective variance of the particle size distribution.

A tabulation is obtained for the scattering cross section for a set of parameters in the following ranges: 0.1-1 \( \mu \) m, 0.1-10 \( \mu \) m, 0.4-0.7. The roughness parameter has been estimated from comparison to both ISM spectra and Viking dust properties [Pollack et al., 1979]. A most likely value of 0.067 is adopted for calculations. Practically, the extinction cross section may be shown to be nearly insensitive to \( p \) for effective radius larger than \( \sim 1 \mu \) m. The value of \( p \) has therefore very little effect on the results of this study, implying that the higher differences are due to the larger effective radius of dust particles found to lie in the range 1-2 \( \mu \) m. The sensitivity of results to \( p \) is tested.

Acknowledgments. We thank B. Zurek and J. Schaeffer for their very constructive referee reports. We are indebted to N. Ray for her help in the preparation of the camera-ready manuscript.

References


Blamont, J.E., E. Chaussier, J.P. Goutail, B. Mige, M. Nassas,


Blamont, J.E., E. Chaussier, J.P. Goutail, B. Mige, M. Nassas,


Haberle, R.M., J.E. Blamont, J.P. Goutail, B. Mige, M. Nassas,


E. Chassefière, Service d’Astronomie du CNRS, BP 3, 91371 Vincennes-la-Bouisse Cedex, France. (e-mail: eric.chassefier@cnes.jussieu.fr)
P. Desvaux, DESSPA, Observatoire de Paris-Meudon, 92195 Meudon, France.
O. Korablev, Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow, GEP-7, Russia.
(Received February 22, 1994; revised December 22, 1994; accepted December 27, 1994.)
Page intentionally left blank
A reexamination of the relationship between eddy mixing and \( \text{O}_2 \) in the Martian middle atmosphere

Jan Rosencwaig
Département de Recherche Spatiale, Observatoire de Paris-Meudon, Meudon, France

E. Chassefière
Service d’Astronomie du Centre National de la Recherche Scientifique, Verrières-le-Buisson, France

Abstract. A value of the eddy diffusion coefficient \( K \) of \( (1.5 \times 10^6 \) cm\(^2\) s\(^{-1}\) in the middle atmosphere of Mars was obtained from Phobos 2 solar occultation measurements of dust, ozone, and clouds at low latitude. The aim of the present study is to complete this picture by using a steady state photochemical one-dimensional model. The main regulation mechanism of \( \text{O}_2 \) is the reaction of \( \text{O} \) with itself, whose rate depends on the value of \( K \) in the middle atmosphere. By comparing calculated and observed values of the \( \text{O}_2 \) abundance, an upper limit of \( (2 \times 10^7 \) cm\(^2\) s\(^{-1}\) on \( K \) is inferred. By including an additional constraint provided by \( H_2 \) balance, a lower limit of \( (8 \times 10^6 \) cm\(^2\) s\(^{-1}\) may be placed. It results from the present analysis that the most realistic value of \( K \) to be used in works resorting to one-dimensional modeling (long-term evolution, escape, surface/atmosphere exchanges) is \( (10^7 \) cm\(^2\) s\(^{-1}\) rather than \( (10^6 \) cm\(^2\) s\(^{-1}\). The difference between theoretical and observational values might be due to the regular occurrence of global dust storms, whose effect should be to increase the yearly average value of \( K \). The present study suggests less than 3 precipitable micrometer (p-\( \mu \)m) of the yearly averaged water vapor column, unless \( H_2O \) is confined in a layer near the ground. Although the first possibility seems more probable, the second hypothesis cannot be ruled out. It could reflect a continuous supply of \( H_2O \) from the regolith to the atmosphere on a seasonal scale. The loss to production ratio of \( C_OH \), which has a lifetime of \( >5 \) years, is shown to depart from unity by no more than \( +10\% \) over a wide range of atmospheric conditions. The stability of the Martian atmosphere is therefore realized in the classical frame of homogeneous chemistry.

Introduction

Recent observations of dust, ozone, and clouds in the middle atmosphere of Mars from the Phobos 2 spacecraft by the solar occultation technique have stimulated a modelling effort, partly focused on the estimate of the eddy diffusion coefficient in different altitude ranges from \( +15 \) km up to nearly \( +45 \) km. The specific interest of Phobos solar occultation measurements is that they cover a wide range of longitudes, although limited to a rather narrow latitude band \((0°-20°) \) and a small interval of solar aereentric longitude \((L_s=0°-20°) \) at the time of northern spring equinox. The small variability of the dust distribution and inferred parameters with space and time shows that atmospheric conditions in the low and middle equatorial atmosphere of Mars are zonally homogeneous. These data provide constraints on eddy mixing strength and some other atmospheric parameters (humidity level, temperature) in three different altitude ranges, i.e., \(+15-30 \) km [Korablev et al., 1993], \(+37-55 \) km [Blamont and Chassefière, 1993], and \(+45-60 \) km [Chassefière et al., 1992].

The vertical structure and size distribution of the dust component between \( 15 \) km and \( 30 \) km altitude were estimated by Korablev et al. [1993]. By using a one-dimensional model involving gravitational settling and eddy diffusion as vertical transport processes, the value of the eddy diffusion coefficient \( K \) below \( 30 \) km altitude was found to be \( (10^6 \) cm\(^2\) s\(^{-1}\) within a factor of 2. Ozone was observed in the 37 to \( 55 \) km altitude range [Blamont and Chassefière, 1993]. By using a simple one-dimensional photochemical model, a good fit to data was obtained for \( K = 2 \times 10^6 \) cm\(^2\) s\(^{-1}\) within a factor of 2. The fit was improved by assuming a steep increase of \( K \) for increasing altitude around \( 40-45 \) km, from \( 3 \times 10^6 \) cm\(^2\) s\(^{-1}\) below \( 40 \) km up to \( 10^7 \) cm\(^2\) s\(^{-1}\) above \( 45 \) km. The analysis and modelling of cloud layers observed in the \( 45 \) to \( 65 \) km altitude range [Chassefière et al., 1992] suggested a most likely value of \( K \) of \( 3 \times 10^6 \) cm\(^2\) s\(^{-1}\) within 1 order of magnitude. Apart from one of these clouds, for which a complete confidence interval was obtained \((K=4 \times 10^6 \) to \( 10^7 \) cm\(^2\) s\(^{-1}\)) only upper limits on \( K \) were derived. A synthetic view of all previous results is given in Figure 1. The large scatter on \( K \) found in the last case presumably reflects an increased variability at higher atmospheric levels. It is emphasized that a common moderate value of the eddy diffusion coefficient \( K = 1.5 \times 10^6 \) cm\(^2\) s\(^{-1}\) accounts for all observed phenomena.

The fact that no reliable information exists in the 60 to 100 km range was already mentioned by Krasnopolsky [1993a]. A lower limit on \( K \) of \( 2 \times 10^7 \) cm\(^2\) s\(^{-1}\) at 50 to 70 km altitude was deduced from the analysis of a high cloud observed in the Viking orbiter images [Kahn, 1990; Chassefière et al., 1992]. This cloud was detected at
Figure 1. Values of the eddy diffusion coefficient as inferred from Phobos 2 observations (dust: box with dashed lines; ozone: box with solid lines; clouds: boxes with dotted lines) and Viking upper atmosphere measurements (V2, triangles, V1, circles). A mean vertical profile of $K$ consistent with observations is shown with a thick solid line. An alternate profile, showing a steep increase around 40-45 km, is plotted with a thin solid line.

midlatitudes in early northern winter ($\mu=139^\circ$). It is emphasized that $K$ is expected to be highly variable in this altitude range due to the necessarily sporadic character of turbulence generated by internal gravity waves, whose breaking level is extremely sensitive to atmospheric thermal profile and wave characteristics [Theodore et al., 1993]. Above 100 km, the vertical profile of $K$ has been derived from $N_2$, $Ar$, CO, and O$_3$ measurements by the two Viking entry probes [Kier and McElroy, 1977].

Steady state models used for the different analyses previously mentioned were supposed to describe the behavior of the atmosphere on short timescales, typically 1 day to 1 week. Moreover, despite the fact that phenomena cannot be considered as local but rather acting on mesoscale, the vertical profiles of background quantities (T, H$_2$O, CO, O$_2$, $H_2$) were fixed in an external way, either by using simultaneous measurements (H$_2$O), or from global-scale Earth-based observations for long-lifetime species (CO, O$_2$). In other words, the abundances of background species were not equilibrated by using the model on a global scale, due to the fact that relevant temporal and spatial scales are much smaller than 1 year and 1 planetary radius, respectively.

As explained by Blamont and Chassefière [1993], the vertical structure of O$_3$ in the middle atmosphere crucially depends on the altitude profile of the eddy diffusion coefficient $K$. The value of $K$ in the middle atmosphere similarly influences the O$_3$ mixing ratio on the long-term. Indeed, processes involved in the short-term balance of O$_3$ and long-term balance of O$_2$ are basically the same: the main two reactions yielding O$_2$ ($O+O+CO_2$; $O+HO_2$) deplete O$_3$. The balance of O$_2$ is therefore expected to be sensitive to $K$.

In Figure 1, the profiles of K and O$_2$ as derived from the photochemical model are shown. The model used here is described in detail by Blamont and Chassefière [1993]. The chemical rates are those used by Moreau et al. [1994]. The reactions are listed in Table 1. All other reactions used in more complete models may be proven to be of very little importance, with rates lower by at least 3 orders of magnitude than the rates of the corresponding main reactions. Condensation is taken into account for H$_2$O and H$_2$O$_2$. The condensation law for H$_2$O$_2$ is the one given by Krasnopolsky [1986]. Dust opacity, as derived from Phobos 2 observations, is shown in Table 1.
2 solar occultation data [Chaseffre et al., 1992], is included in the calculation of UV fluxes. The thermal profile is basically taken from Seiff [1982] for the lower atmosphere. At higher levels, we have also considered the infrared thermal mapper (IRTM) data [Martin et al., 1979]. The pressure at the surface has been taken from Hess et al. [1988] with a value of 7 mbar at 0 km ground altitude for $L_p=30-35^\circ$. Several tests corresponding to different Martian seasons have been performed to estimate the influence of temperature and pressure profiles on the $O_2$ equilibrium abundance. The results show that the influence of pressure and temperature is quantitatively weak as compared to that of $K$ and water vapor.

The thermal profile corresponding to an approximate average vertical distribution. Thus, in the following, we fix the surface pressure at a value of 7 mbar and use a unique standard thermal profile corresponding to an approximate average on the year, the season, and the day: $T=205 K$ at 0 km altitude, 175 K at 25 km, and 145 K at 50 km. It is emphasized that this profile fits well the one proposed by Krasnopolsky et al. [1991] from solar occultation measurements of water vapor.

In the present work, the total hydrogen escape flux $\Phi(H)$ is parameterized in the following way:

$$\Phi(H) = \Phi(H)_{eq} \frac{p(H_2)}{p(H_2)_{eq}}$$

where $\Phi(H)_{eq}$ is the present value of the hydrogen escape flux, $p(H_2)$ is the equilibrium value of the $H_2$ mixing ratio, as given by the model, and $p(H_2)_{eq}$ is the present value of the $H_2$ mixing ratio in this study. $\Phi(H)_{eq}$ is fixed at 8.3x10$^{-5}$ [Kong and McElroy, 1977]. The present values of the total escape flux of hydrogen $\Phi(H)$ is assumed to be 2x10$^{8}$ cm$^{-2}$s$^{-1}$ (see Table 1 of Krasnopolsky [1993b]). The oxygen escape flux $\Phi(O_2)$ is assumed to be twice as small as $\Phi(H)$ [Lin and Donahue, 1976].

**Photochemical Regulation of $O_2$**

Production and loss reactions of $O_2$ are summarized in Figure 2. The production of $O_2$ by $R_{12}$ and $R_{15}$, which are of very minor importance, are omitted for simplicity. The main $O_2$ production process is photodissociation of ozone ($R_3$), which represents 40-75% of the total production. Since reaction of $O$ with $O_2$ ($R_6$), which is the only one yielding $O_3$, is the major loss process for molecular oxygen (45-75%), $O_3$ balance ensures more than half of the $O_2$ balance. In the same way, the balance between $R_7$ and $R_8$ is roughly maintained through $H_2O_2$ balance, which ensures another third of the $O_2$ balance. The ratio of the loss to the production rate (LPR) of $O_2$ may be formally written:

$$LPR = \frac{P(O_2)}{P(O_2)_{eq}}$$

Neglecting terms with weights smaller than 1%, one obtains

$$LPR = 1 + \sum_{j=1}^{6} T_j$$

where the analytical expressions and the orders of magnitude of terms $T_j$ are given in Figure 2. Positive and negative sums (respectively $T_1+T_2+T_3+T_4$ and $T_5+T_6$) are in the range between 2% and 20% of the total LPR. The loss to production ratio of $O_2$ varies from 0.9 to 1.05 as the eddy diffusion coefficient $K$ varies from 10$^9$ to 10$^{10}$ cm$^2$ s$^{-1}$. This means that the balance between $O_2$ production and loss rates is realized within 10% through the balances of short-lived $O_3$ and $H_2O_2$. A first step of the $O_2$ equilibrium is reached in less than 1 hour through the balance of $O_3$. In a second step, the balance of $H_2O_2$ is reached in about 1 hour, ensuring a rough equilibrium between the production and loss rates of molecular oxygen. Fine adjustment is realized only when the sum of the six terms in (3) becomes equal to zero, over a time-scale of the order of the $O_2$ lifetime ($\sim 10^4$ years), or even more in the case of escape ($\sim 10^5$ years).

A brief examination of the regulation terms is necessary at this step of the work. Figure 3 shows the reaction rates of the 22 photochemical reactions involved in the model as a function of altitude (Figure 3a) and the associated vertical profiles of the number densities of the constituents for standard conditions (Figure 3b). The effect of the condensation of $H_2O_2$ may be seen on profiles at altitudes between 30 and 40 km. This figure shows that the atmosphere can be roughly divided in four layers, which are [0-20 km], [20-40 km], [40-60 km], and altitudes above 60 km. The dominant balance reactions of $O_3$, $O$, $H_2O_2$, and $H_2O$ (cold hydrogen family) in these four layers are given in Table 2.

The term of photodissociation ($R_1$), whose value is weak ($\sim 3-5$%), is a damping term, since it is directly proportional to $O_2$. Any relative variation of $O_2$ will be translated in a similar relative variation of the $O_2$ photodissociation rate.
Figure 3a. Reaction rates as a function of altitude for the 22 photochemical reactions listed in Table 1 for standard conditions: \( K = 10^7 \text{ cm}^2 \text{ s}^{-1} \), \([\text{H}_2\text{O}]/[\text{CO}_2] = 5.6 \times 10^{-3} \), \( c(\text{water vapor column amount}) = 11 \text{ pp} \mu \text{m} \), \( h(\text{H}_2\text{O} \text{ mixing ratio scale height}) = 3 \text{ km} \).

Figure 3b. Vertical profiles of the number densities of the constituents.
because the penetrating solar ultraviolet flux is very little affected by a small variation of \( \text{O}_2 \), which forms \( \text{O}_3 \) when interacting with \( \text{O}_2 \) in the presence of \( \text{CO}_2 \), with a strong decrease of \( \text{H}^* \) which may be linked to a relatively strong loss of \( \text{H}^* \) at high altitude. This simple consideration suggests, as confirmed by calculations, that the total amount of \( \text{H}^* \) and \( \text{O}_2 \) must globally increase. At this step, the equilibrium of \( \text{H}^* \) atoms over the whole atmosphere can be approximately written as follows, by keeping only major terms:

\[
\frac{\mathrm{d}[\text{CO}][\text{OH}]}{\mathrm{dt}} = \frac{[\text{O}_2][\text{H}][\text{CO}_2]}{\text{dt}}
\]

(\ref{eq:4})

The left side of the equation is nearly invariable, as may be deduced from the study by \cite{Chassefiere1991}. Consequently, the product \([\text{H}][\text{O}_2] \) is expected to be roughly constant for varying \( \text{H}_2\text{O} \). Thus, as water vapor amount increases, \( \text{O}_2 \) decreases (see flow chart in Table 2). There is a clear anticorrelation between the equilibrium value of \( \text{O}_2 \) and the water vapor concentration.

**Case of Molecular Hydrogen**

The only process leading to the formation of \( \text{H}_2 \) is the reaction of \( \text{H} \) with \( \text{H}_2\text{O} \) (R1). The main loss processes are dissociation by \( \text{O}(\text{D}) (\text{R}_6) \) and thermal escape of hydrogen atoms from the upper atmosphere. Any variation of \( \text{H}_2 \) therefore yields a variation of the same sign of the \( \text{H}_2 \) loss rate, with a slow damping over a timescale of the order of the \( \text{H}_2 \) lifetime \( (=5 \times 10^4 \text{ years}) \). The amplitude of this variation depends on the relative weights of escape and chemical \( (\text{R}_3) \) terms, which are of the same order of magnitude. It is emphasized that the photochemical behavior of \( \text{H}_2 \) differs from that of \( \text{CO} \) and \( \text{O}_2 \). \( \text{H}_2 \) is strongly regulated by both dissociation by \( \text{O}(\text{D}) \) and thermal escape, and its equilibrium level is well defined by the balance between production and loss, contrary to the cases of \( \text{CO} \) and \( \text{O}_2 \), which are only

<table>
<thead>
<tr>
<th>( \text{Response of O}_2 ) to a variation of ( K )</th>
<th>( \text{Response of O}_2 ) to a variation of ( \text{H}_2\text{O} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>( \text{H}_2\text{O} )</td>
</tr>
<tr>
<td>( \text{O} ) (\text{O} downward transport)</td>
<td>( \text{I} ) (\text{I} upward transport)</td>
</tr>
<tr>
<td>( \text{O}_2 ) (25 km)</td>
<td>( \text{H}_2\text{O} ) (25 km)</td>
</tr>
</tbody>
</table>

Figure 4. Response of the equilibrium abundance of \( \text{O}_2 \) to variations of \( K \) and \( \text{H}_2\text{O} \) concentration.
Table 2. Main Balance Reactions of Short-Lived Constituents as a Function of Altitude

<table>
<thead>
<tr>
<th>Constituent</th>
<th>0-20 km</th>
<th>20-40 km</th>
<th>40-60 km</th>
<th>&gt;60 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2k&lt;sub&gt;4&lt;/sub&gt; + k&lt;sub&gt;5&lt;/sub&gt;</td>
<td>k&lt;sub&gt;7&lt;/sub&gt; + k&lt;sub&gt;10&lt;/sub&gt;</td>
<td>k&lt;sub&gt;7&lt;/sub&gt; + k&lt;sub&gt;10&lt;/sub&gt;</td>
<td>k&lt;sub&gt;7&lt;/sub&gt; + k&lt;sub&gt;5&lt;/sub&gt;</td>
</tr>
<tr>
<td>O&lt;sub&gt;2&lt;/sub&gt;</td>
<td>J&lt;sub&gt;1&lt;/sub&gt; + J&lt;sub&gt;2&lt;/sub&gt; → k&lt;sub&gt;7&lt;/sub&gt;</td>
<td>J&lt;sub&gt;1&lt;/sub&gt; + k&lt;sub&gt;7&lt;/sub&gt;</td>
<td>J&lt;sub&gt;1&lt;/sub&gt; + k&lt;sub&gt;7&lt;/sub&gt;</td>
<td>J&lt;sub&gt;1&lt;/sub&gt; + k&lt;sub&gt;5&lt;/sub&gt;</td>
</tr>
<tr>
<td>H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>k&lt;sub&gt;4&lt;/sub&gt; + k&lt;sub&gt;7&lt;/sub&gt; + k&lt;sub&gt;14&lt;/sub&gt;</td>
<td>k&lt;sub&gt;4&lt;/sub&gt; + k&lt;sub&gt;7&lt;/sub&gt;</td>
<td>k&lt;sub&gt;4&lt;/sub&gt; + k&lt;sub&gt;7&lt;/sub&gt;</td>
<td>k&lt;sub&gt;4&lt;/sub&gt; + k&lt;sub&gt;14&lt;/sub&gt;</td>
</tr>
<tr>
<td>H&lt;sup&gt;+&lt;/sup&gt;</td>
<td>J&lt;sub&gt;3&lt;/sub&gt; + k&lt;sub&gt;2&lt;/sub&gt; + k&lt;sub&gt;14&lt;/sub&gt;</td>
<td>J&lt;sub&gt;3&lt;/sub&gt; + k&lt;sub&gt;14&lt;/sub&gt;</td>
<td>J&lt;sub&gt;3&lt;/sub&gt; + k&lt;sub&gt;14&lt;/sub&gt;</td>
<td>J&lt;sub&gt;3&lt;/sub&gt; + k&lt;sub&gt;14&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Density terms are systematically omitted in these systems of algebraic equations for convenience.

weakly regulated at their equilibrium values. Because there is a well-established anticorrelation between O<sub>2</sub> and H<sub>2</sub> in the Martian atmosphere, the equilibrium value of H<sub>2</sub> is expected to depend on K and H<sub>2</sub>O in the same way as O<sub>2</sub>. The H<sub>2</sub> equilibrium abundance is expected to be very sensitive to the vertical distribution of water vapor. Indeed, the formation of odd hydrogen species like H and H<sub>2</sub>O, which react together to form H<sub>2</sub>, is principally due to the photodissociation of H<sub>2</sub>O.

Results

The observed mixing ratio of molecular oxygen is assumed to be 1.3 ± 0.3 x 10<sup>-3</sup> for the present study [Barker, 1972; Carleton and Traub, 1972; Trauger and Lune, 1983]. Anderson [1974] proposed an H<sub>2</sub> mixing ratio of 2 x 10<sup>-5</sup>, consistent with the observed H density profile in the high atmosphere. A reanalysis of these data by Kong and McElroy [1977] yields a value twice as small of 8.3 ± 10<sup>-5</sup>. Values obtained by Shimazaki [1989], as presented in his Table 7, are in the range 7 x 10<sup>-5</sup> - 1.5 x 10<sup>-5</sup>. We therefore adopt a range of 7 x 10<sup>-5</sup> - 2 x 10<sup>-5</sup>, which accounts well for the major part of the existing estimates.

The CO mixing ratio is fixed at its observed value (8 x 10<sup>-5</sup>). The equilibrium mixing ratios of O<sub>2</sub> and H<sub>2</sub> are plotted as a function of K in Figure 5. Two cases are considered.
Infrared measurements of reflected sunlight performed from Phobos 2 show that the ground in equatorial regions is too small (Figure 5). By the way, we must study, uniformly mixed or not, are systematically obtained by decreasing the scale height h of the water vapor absolute abundance, as derived from the analysis of lines of constant water vapor column density (in pr-pm) are also plotted.

Figure 6. Permitted domain of the near surface water vapor mixing ratio p_w and 1/h (h is the scale height of the water vapor mixing ratio), as derived from the analysis of O_2 for K=5x10^7 cm^2 s^-1. The gray area corresponds to expected values of the O_2 mixing ratio (1.3x10^5 cm^-2). The white area corresponds to an O_2 mixing ratio larger than 1.5x10^5 cm^-2 and the black area to an O_2 mixing ratio lower than 10^5 cm^-2. Lines of constant water vapor column density (in pr-pm) are also plotted.
should be no consistent field of solutions. As previously mentioned, values larger than $2 \times 10^7 \text{ cm}^2 \text{ s}^{-1}$ are excluded, since, in this case, no solution exists for $O_2$. The widest range of solutions is obtained for $K = 5 \times 10^6 \text{ cm}^2 \text{ s}^{-1}$. By fixing $K$ below 15 km to a nominal value of $1.5 \times 10^6 \text{ cm}^2 \text{ s}^{-1}$, it is also possible to analyze the effect of a change of $K$ restricted to the middle atmosphere. The widest range of solutions is obtained for a slightly larger value of $K (=7 \times 10^6 \text{ cm}^2 \text{ s}^{-1})$. There is no more upper limit in this case. The relation between $h$ and $c$ is not significantly modified.

Results are summarized in Table 3.

Concerning CO, whose mixing ratio is fixed at its observed value ($8 \times 10^{-6}$), it is emphasized that the loss to production ratio is in the range $0.03-1.17$ in the whole domain of $K$ and $c$ considered in the previous study. The upper value of 1.17 becomes 1.09 by excluding very low values of the eddy mixing coefficient ($2 \times 10^5 \text{ cm}^2 \text{ s}^{-1}$). The balance of CO, whose lifetime is 511 years, is realized with an accuracy better than 10%. This behavior is expected from a theoretical analysis of processes regulating CO [Chassefière, 1991]. Isolines corresponding to different values of the relative difference between the loss to production ratio of CO and unity are plotted in the plane $[K,c]$ (Figure 10). The range of $K$ for which the balance of CO is realized with an accuracy better than 5% is $1 \times 10^6-5 \times 10^6 \text{ cm}^2 \text{ s}^{-1}$ with a most likely value of $10^7 \text{ cm}^2 \text{ s}^{-1}$ (isoline at 1%). Although no strong constraint may be derived on $K$ or $c$ from the necessity to account for an accurate CO balance, a consistent picture of

Figure 7. Integrated photodissociation rate of $H_2O$ as a function of the water vapor column for different scale heights of the water vapor mixing ratio in the 3—10 km range.

Figure 8. Same as Figure 6 for $H_2$. The gray area corresponds to expected values of the $H_2$ mixing ratio ($7 \times 10^{-6}$ to $2 \times 10^{-5}$). The white area corresponds to an $H_2$ mixing ratio lower than $7 \times 10^{-6}$ and the dark area to an $H_2$ mixing ratio larger than $2 \times 10^{-5}$.
balances governing long-lifetime constituents (CO, O₂, H₂) may be found in one-dimensional photochemical models by assuming a small average water vapor column of \( \rho_{w} = 3 \) pp-my. By the way, it must be noted that present results differ in a significant way from those obtained by Arreya and Gu [1994] which suggested that heterogeneous chemistry is needed for the stability of the Martian atmosphere. However, the presence of heterogeneous chemistry such as adsorption of HOx on aerosols [Anbar et al., 1993], or sticking of molecules on suspended dust particles [Rosenqvist et al., 1992], could play a significant role for the variability of the horizontal and vertical distributions of some atmospheric constituents.

**Discussion**

The eddy diffusion coefficient and its variability are summarized in Table 4. The mean observational value is somewhat smaller than the one inferred from the model. A possible reason should be the regular occurrence of global dust storms. Suspended dust particles reach heights in excess of 40 km [Leovy et al., 1972; Haberle et al., 1982] in a few days. Then the storm covers much of the planet for weeks. The mean vertical velocity of particles is thus of the order of 10-20 cm s⁻¹ during storms as compared to 1-2 cm s⁻¹ during still conditions. The peatmosphere conditions are typically restored after several months. Consequently, the regular occurrence of global and local dust storms should increase the yearly diffusion by a factor of at least five as compared to \( K \) for clear conditions only, such as during the Phobos observations.

A direct consequence of the present work is that the best value of \( K \) to be used in all works resorting to one-dimensional modeling (long-term evolution, escape, surface/atmosphere exchanges) is 10⁷ cm² s⁻¹ rather than 1.5\times10⁶ cm² s⁻¹.

Another interesting consequence is the possibility that water vapor should be confined in the first half atmospheric scale height above the ground. There are actually two possibilities: either the globally averaged water vapor column is around \( \rho_{w} = 3 \) pp-my, that is, definitely lower than values derived from Viking and Phobos measurements (\( \rho_{w} > 10 \) pp-my), or it is larger than 3 pp-my. However, in the latter case, the water vapor in excess must be confined in the first half scale height. As shown in the present study, water vapor molecules present in the low atmosphere (\( z < 5 \) km) are not efficiently dissociated by solar photons. The water vapor content required for global photochemical balance is therefore \( \rho_{w} = 3 \) pp-my as well. A larger mixing ratio below this level, in the boundary layer, could play a significant role for the variability of the horizontal and vertical distributions of some atmospheric constituents.

### Table 3. Permitted Ranges of the H₂O Mixing Ratio Scale Height \( h \) as a function of \( K(z) \)

<table>
<thead>
<tr>
<th>( K ), cm² s⁻¹</th>
<th>H₂O Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-35km</td>
<td>35-80km</td>
</tr>
<tr>
<td>( 4\times10^{5} ) ( \rho_{w} )</td>
<td>( 1\times10^{7} ) ( \rho_{w} )</td>
</tr>
<tr>
<td>( 1.5\times10^{6} ) ( \rho_{w} )</td>
<td>( 3\times10^{6} ) ( \rho_{w} )</td>
</tr>
<tr>
<td>( 1.5\times10^{6} ) ( \rho_{w} )</td>
<td>( 7\times10^{5} ) ( \rho_{w} )</td>
</tr>
<tr>
<td>( 1.5\times10^{6} ) ( \rho_{w} )</td>
<td>( 1\times10^{7} ) ( \rho_{w} )</td>
</tr>
<tr>
<td>( 1.5\times10^{6} ) ( \rho_{w} )</td>
<td>( 5\times10^{5} ) ( \rho_{w} )</td>
</tr>
</tbody>
</table>

\( h < 50 \) km \( h < 8.5 \) km \( h < 8.5 \) km

\( h < 95 \) km \( h < 8.0 \) km

\( h < 85 \) km

\( h < 7.5 \) km
maintained over one half Martian year, a total amount of
\(~500\) pr-\(\mu\)m must therefore be transferred from the regolith to
the atmosphere in northern hemisphere during spring and
summer time. By assuming, with Zent et al. [1993], a typical
concentration of regolith \(2\) kg m\(^{-2}\), it is equivalent to
the \(H_2O\) quantity present in the first \(25\) cm below the
surface. The corresponding seasonal flux is \(\approx 1.5 \times 10^{13}\) kg m\(^{-2}\)
\(\times 1\) day, that is, a value of the same order as diurnal fluxes
calculated by Zent et al. [1993] in their Figure 7. Removal of
water molecules by meridional diffusion or transport of water
ice clouds toward polar regions must occur on a global scale
in order to balance the \(H_2O\) input flux.

Acknowledgments. This research was supported by the Programme
Nacional de Planèstica de l'Institut Nacional des Sciences de l'Univers.
We are indebted to N. Rey for her help in the preparation of the
camera-ready manuscript.

References

Aber, A.D., M.T. Lee, H.A. Nair, and Y.L. Yong, Adsorption of
\(CO_2\) on aerosol surfaces: Implications for the atmosphere of Mars,

Andersson, D.E., Mariner 6, 7, and 9 ultraviolet spectrometer
experiment: Analysis of hydrogen Lyman alpha data. J. Geophys.

Atreya, S.K., and Z.O. Go, Stability of the Martian atmosphere: Is
heterogeneous catalysis essential to Mars, J. Geophys. Res., 99, 13,133-

Barker, E.S., Detection of molecular oxygen in the Martian

Bennet, J.C., and E. Chassefière, First detection of ozone in the
middle atmosphere of Mars from solar occultation measurements,

Carleton, N.B., and W.A. Traub, Detection of molecular oxygen on

Chassefière, P., Photochemical regulation of \(CO\) on Mars. Geophys.

Chassefière, E., J.E. Blemont, V.A. Krasnopolsky, O.V. Kobalev,
S.K. Ayrey, and R.A. West, Vertical structure and size distributions of
Mars aerosols from solar occultation measurements, Icarus, 97, 36-69,

Chassefière, E., P. Dessart, and O. Kobalev, Post-Phobos model for
the altitude and size distribution of dust in the low Martian atmosphere,

Clancy, R.T., A.W. Grossman, and D.O. Mulleran, Mapping of Mars
water vapor with the Very Large Array, Icarus, 100, 48-59, 1992.

Davies, D.W., The vertical distribution of Mars water vapor, J.

Dowell, S.L., Vertical distribution of \(CO_2\) in the atmosphere of Mars,

Dowell, S.L., J.A. Ryan, J.E. Tillman, R.M. Heays, and C.B. Leovy,
The annual cycle of pressure on Mars measured by Viking landers 1 and

Fontes, D.M., Aeronomy of the Lower Atmosphere of Mars, Rev.

Grossman, R.T., and E.S. Barker, Comparison of ground-based and
Viking orbiter measurements of Martian water vapor: Variability of the

Gulkis, S., Detection of molecular oxygen on Mars, Science, 177,

Habeb, C.B. Loosy, and J.B. Pollack, Some aspects of global dust storms
on the atmospheric circulation of Mars, Icarus, 50, 322-
337, 1982.

Hess, S.L., The vertical distribution of water vapor in the atmosphere

Hess, S.L., J.A. Ryan, J.E. Tillman, R.M. Heays, and C.B. Leovy,
The annual cycle of pressure on Mars measured by Viking landers 1 and

Hess, S.L., J.A. Ryan, J.E. Tillman, R.M. Heays, and C.B. Leovy,
The annual cycle of pressure on Mars measured by Viking landers 1 and

Hunten, D.M., The vertical distribution of Mars water vapor: Variability of

Johansen, B.M., and R.M. Habeb, The seasonal behavior of water on


E. Chassefière, Service d'Aeronomie du CNRS, BP 3, 91371, Versailles-le-Bouquet Cedex, France.
Page intentionally left blank
A case for ancient evaporite basins on Mars

Randall D. Forsythe
Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte

James R. Zimbelman
Center for Earth and Planetary Studies, Natural Air and Space Museum, Smithsonian Institution, Washington, DC

Abstract. Observations indicate that a Martian analog to the Earth's salt pans and saline lakes of arid regions may have existed in crater-basins during Mars' early (Noachian) epoch. Terraced and channelized crater-basin point to ponding of surface water as well as possible prolonged and evolving base levels. In addition, supportive (evaporite basin) analogs are offered for three other morphologic features of Martian crater-basins. An evaporite basin model for crater-basins on Mars has major implications for the mechanical, chemical, and even biological processes that potentially have operated in Mars' past, and represent a spectrum of potential mineral resources.

Introduction

There is irrefutable evidence for the presence of water ice on Mars. While the surface pressure of the Martian atmosphere is only 0.6% that of Earth, it contains a precipitable 12 μm (in a global layer) of water vapor [Farmer et al., 1977]. The polar deposits today are estimated to include 2 x 10^5 km^3 of water ice [Carr, 1986], and variable lines of evidence suggest the presence of ice-rich frozen ground at middle to high latitudes [Fanale, 1976; Squyres, 1989]. In addition to the possibility of near-surface ground ice today, there is a growing body of work that supports an active role for water in the shaping of Martian landforms during its past epochs [Squyres, 1989]. However, since present atmospheric conditions preclude the surface stability of either liquid water at all latitudes or ice at low latitudes, there have emerged two general points of view in attempts to explain the origin of fluvial and other water-related landforms.

One set of interpretations attempts to minimize the need for dramatically different climatic conditions than those found today. The ancient valley networks, documented to exist within the highlands of Mars' southern hemisphere, are seen as most likely related to snowmelt from subsurface groundwater reservoirs. However, there is no evidence to suggest that the erosion in the highlands greatly postdates the formation of the cratered terrain during the Noachian epoch on Mars [Tomark, 1986]. The major outflow channels fringing the boundary between the southern hemisphere's highlands and the lowlands of the northern hemisphere can be related to catastrophic flooding [Baker and Milone, 1974; Baker, 1982] of water possibly released over a prolonged period of time from permafrost-confined groundwater systems [Carr, 1979], and surface water lakes, if they existed, were potentially coated with thick layers of ice [Nedell et al., 1987]. Common to these models is the existence of an early Martian geologic epoch with ground water sources (of uncertain origin) that over time have become inactive in low-latitude areas.

From a contrasting perspective, other interpretations require past atmospheric conditions to vary dramatically from those of today, and potentially include the stability of water and ice to equatorial latitudes [Masursky et al., 1971; Pollack et al., 1977; Baker et al., 1991; Parker et al., 1993]. These views which generally beg warmer and wetter climatic models explicitly (or implicitly) invoke mechanisms for recharge directly to the Martian surface, and imply a complex hydrologic cycle once operated on Mars. Ancient valley networks, particularly the degraded ones, are interpreted as runoff channels requiring maintenance of high water tables [Baker and Partridge, 1979; Bruckneridge, 1990], the outflow channels are visualized as potentially debouching into ancient seas or oceans [Guilek and Baker, 1990] and an active hydrologic cycle is seen to potentially persist well beyond Mars' earliest (Noachian) epoch.

Here new observations and arguments are presented, based on hydrologic comparisons of Martian crater-basins to terrestrial drainage basins in arid and semiariad areas, and used to advance further the hypothesis for a complex hydrologic cycle in Mars' past that may have included open hydraulic conditions between surface water and groundwater and mechanisms for recharging these reservoirs.

A Case for Base Level and Groundwater-Controlled Drainage on Mars

In the absence of recharge to an unconfined groundwater system, the upper surface of groundwater (i.e., the water table) relaxes to a global minimum equipotential surface, which on Earth is approximated as mean sea level [Freeze and Cherry, 1979]. However, the hydrologic processes of precipitation, infiltration, and base flow permit on Earth a dynamic equilibrium between surface topography and the underlying water table. The equilibrium is established by the manner in which the minimum potential of hydraulic head (sea level) is also the base level, below which fluvial erosion cannot occur [Blissom, 1978], and inversely, the manner in which local streams, which are responsible for eroding and shaping topography, limit the rise of the water table during recharge periods. On a more regional
level, the same boundary conditions are applicable to internally drained depressions [Bloom, 1978], wherein a local base level for fluvial erosion is represented by the corresponding minimum of the water table within the basin. Partially because of the role of groundwater in determining base levels, streams undergo systematic changes in their channel morphologies from headwater ephemeral regimes (generally without significant floodplains) to base level perennial regimes (with lower gradients and floodplains) [Leopold et al., 1964]. In the latter, periods of climatic and watershed stability (i.e., tectonic quiescence) permit streams to be "graded" or in a state of quasi-equilibrium [Dana, 1902; Mckin, 1948; Strahler, 1952], and it is common during these periods of stability for processes responsible for the lateral migration of channels to be more efficient than the processes of downcutting. The result is the formation of floodplains. Floodplains formed in this fashion, after being subjected to a period of downcutting, are left as terraces or benches along the margins of the valley. Alternative (i.e., nonterrestrial) mechanisms for the formation of fluvial valley terraces on Mars remain possible, perhaps (hypothetically speaking) related in some fashion to changing discharges over an ice-rich permafrost. However, if the ancient valley networks of Mars' southern highlands were produced by fluvial erosion occurring under Earth-like conditions, then similar constraints and conditions can be postulated. One condition of such an analog is that Martian drainage systems would also represent an equilibrium between surface and groundwater systems. The important point being an Earth-like hydraulic coupling between ground water and surface water systems in Mars' past is a plausible hypothesis if geomorphic evidence exists for fluvial drainage under steady base level conditions. Below we describe one ancient Mars drainage basin for which there is evidence in support of fluvial and lacustrine erosion under base level conditions that was sufficiently long-lived in nature to produce what are here interpreted as fluvial and lacustrine terraces.

The middle latitude to equatorial region of the southern highlands has been one of the primary areas from which evidence has emerged for Noachian fluvial erosion [e.g., Baker and Panridge, 1986; Pieri, 1980] and extensive resurfacing [Jones, 1974; Chapman and Jones, 1977; Craddock and Maxwell, 1990, 1993]. Within the Memnonia Northwest quadrangle a segment of a north trending early drainage system, located approximately 200 km south of the highlands/lowlnds boundary at 174.6°W, 14.6°S, has been preserved that contains both fluvial and lacustrine features that are interpreted here on geomorphic grounds to support the existence of base level-controlled fluvial

Figure 1a. A terraced crater-basin south of Mars' dichotomy boundary with both an inflowing (lower right) and outflowing (top center) channel. Note the trace of the main terrace up the inflow channel, and the lower level of the outflow channel. Scale bar is 10 km. (Viking Orbiter frame 438512, centered at 174.6°W, 14.6°S).
Localities A-D mark areas near meanders with differential terrace development (see text) and lacustrine systems. The main fluvial channel can be seen to drain into, and away from, a crater basin. Moderate resolution (63 m/pixel) Viking images of this basin (Figure 1a) shows it to have a polygonal ground plan with an average diameter of approximately 35 km, and a crater rim that, from shadow length measurements, is estimated to rise 0.7 to 1 km above the floor. A prominent terrace or bench, at a height of 200-220 m above the basin floor, is traceable around the entire basin, and extends up into the inflowing channel (Figures 1a and 1b). In addition, an upper smaller bench is traceable just above the main terrace, and two additional, but poorly preserved, erosional benches may exist along the lower slopes (Figure 1c).

The origins of these terraces are of paramount importance to the interpretation that this channel and basin represent a base level-controlled drainage system akin to terrestrial watersheds. There are two questions related to potential ambiguities to be addressed: (1) Are, in fact, the benches of fluvial and lacustrine origin such that their elevations can be interpreted as former channel flood plain levels, and lake stands, respectively; and (2) does the existence and the levels of the postulated lake benches reflect stands of water controlled by a regional groundwater table? These two questions are addressed separately below.

Erosional benches or terraces can only be produced by two sets of alternative mechanisms. Either the terraces represent levels of lithologic differences within the Martian crust that have, in turn, controlled incision by whatever means (e.g., gully, mass wasting, or fluvial), or they have been formed by a water-related mechanism [Bloom, 1978]. Fluvial terraces within river valleys can be either degradational (strath terrace) or aggradational (fill terrace) but, regardless, represent uplifted or drowned flood plains [Leopold et al., 1964].

If the terraces seen in Figure 1 are due to nonfluvial or nonlacustrine processes, then these processes should also operate in surrounding areas free from possible fluvial or lacustrine influences. Craddock and Maxwell [1993] have demonstrated that pervasive resurfacing has occurred within a marginal (equatorial) province of the highlands, affecting the Martian crust to a depth of 290 to 2300 m. Yet, despite the evidence for extensive erosion, terraces in the Memnonia NW quadrangle (Figure 1) are restricted exclusively to fluvial valleys and degraded crater basins.

A fluvial or lacustrine origin for the terraces seen in Figure 1 is also supported by three specific attributes of the terraces, which would not be easily explained under other mechanisms. First, the main terrace, which can be traced into, and around, the crater basin, ends abruptly at the outlet, despite the continuing relief along the walls of the valley (Figure 1d, location A). A fluvial or lacustrine origin for the terraces explains this asymmetry, since graded fluvial systems are expected to develop preferentially upstream from a base level. Also note that the outlet channel, which is deeply incised at the outlet, opens up to a broad valley floor (i.e., flood plain) at point B (Figure 1d); again a relationship consistent with the existence of local base level influences. Second, if a river terrace represents a former flood plain, then its width should vary in a consistent manner with its precursory flood plain. The latter would broaden as the channel approaches the local base level. Figure 1b shows a closeup of the inflowing channel. In this image one can see that both the channel floor and terrace levels become systematically narrower upstream. Third, the relationships of the valley terrace to channel meanders suggests that they were formed early on in the development of the valley. At locality B (Figure 1b) one can see that the terrace
Locality A is the uppermost terrace; locality west margins of the basin (Viking Orbiter frame 438312). terrace, and depression, it may does not exist on the outside margin of the mound. Similarly at the next meander, Locality C (Figure 1b), the terrace is narrower than it is upstream or downstream. These relationships are expected among river terraces which are incised by later river downcutting. Thus, the terraces described within this system have a number of specific attributes which support a fluvial and lacustrine origin.

The second concern is the possibility that, rather than having an erosional base level defined by a local groundwater depression, it may have been controlled by the existence of a hypothetically impermeable permafrost layer, as envisaged by Cuffey [1993]. This possibility cannot be ruled out with existing data. However, there are three reasons that support the former hypothesis as being the more plausible alternative in the Memnonia region. There is little doubt that a permafrost zone has existed on Mars, given the abundance and distribution of presumed ice-related geomorphic features [e.g., Squyres and Carr, 1986; Squyres, 1989] and the estimates of current past stability of ground ice [Leighton and Murray, 1966; Farmer and Domn, 1979; Feinle et al., 1986; Pollack et al., 1987b, 1990; Mollen and Jakosky, 1993]. However, the hypothesis that a thick permafrost zone could have perched Noachian drainage systems such as the one described above is potentially problematic on three grounds.

First, if permafrost did indeed exist within the southern highlands during the creation of the valley networks, then one should be able to identify features characteristic of geomorphic processes operating in regions with ground ice. As noted above, Mars preserves many geomorphic features in its current polar and midlatitude regions that are interpreted to be related to ground ice (e.g., "terrane softening" of Squyres and Carr [1986]). However, ice-free alternatives do exist for some of these geomorphic features [Zimbelman et al., 1989]. The preservation of fluvial features throughout much of the highlands indicates that a sufficient record of past surface processes is available. As pointed out by Craddock and Maxwell [1993], the only potential evidence for ice-related processes in the equatorial area are postcrater degradation features (interpreted to be glacial in origin) suggested to be present in some of the more elevated regions by Baker et al. [1991] and Kargel et al. [1992]. Rampart craters have been cited as possible indicators of subsurface volatiles on Mars, and these craters are present at all latitudes (see Squyres et al. [1992] for a detailed discussion). The precise role of volatiles in the rampart crater formation process remains ambiguous [Squyres et al., 1992], but the equatorial rampart craters have been interpreted to indicate that water or ice is present in the subsurface, although at a much greater depth than at midlatitudes [Kazmin et al., 1988].

Second, while the permafrost zone is commonly described as impermeable, no where on Earth has it been demonstrated that it operates as a pure aquifuge [Price and Cherry, 1979; Williams 1970, Williams and Smith, 1989]. While the hydrogeologic literature for permafrost has presented the permafrost zone as impermeable when considering water budgets on timescales pertinent to our understanding of water resources (e.g., Williams, 1970), estimates in the laboratory, as well as from field studies, indicate that it is best characterized as an aquitard, or aquiclude, of very low hydraulic conductivity on geologic timescales [Williams and Smith, 1989]. The 45 - 50° K difference in the mean annual surface temperatures between the coldest permafrost regions on Earth (-5°C to 0°C) and the warmest (equatorial) latitudes on Mars (approximately -55°C) suggest that the hydraulic conductivities of permafrost on Mars under ice-rich conditions would be somewhat lower than that observed for terrestrial conditions. However, the in situ large-scale transmissivities of hydrogeologic units are influenced by both diffuse and nondiffuse flow. Laboratory measurements of the hydraulic conductivity of ice-rich permafrost speak to only the former component.

Third, when surface water persists over permafrost it will have a dramatic affect on the underlying permafrost, producing an extensive zone of thawing both at the top and base of the permafrost zone [Lauckebreech, 1971]. While thermal modeling of these effects for Martian conditions is beyond the scope of this
FORSYTHE AND ZIMBELMAN: ANCIENT EVAPORITE BASINS ON MARS

Figure 1d. Closeup of the outflowing channel where no terrace has been developed (location A), and the incised channel opens up downstream to a floodplain (location B) (Viking Orbiter frames 43881.12 and -13).

Paper, it should be noted that even within the continuous permafrost regime of Alaska (with permafrost thicknesses of several hundred meters), permafrost is either very deep or absent under lakes and large rivers [Yoshikawa et al., 1955]. Along the latter there are also frequent, but sporadic, areas of groundwater discharge that produce characteristic "spurs" or icings [Williams, 1970] that may reflect the development of nondiffuse flow conditions within the underlying or surrounding permafrost. The colder mean annual surface temperatures on Mars again could have a mitigating effect on comparisons drawn here to terrestrial permafrost regions. Thus, while the case for a permafrost-controlled base level within the drainage basin described above cannot be unequivocally ruled out, it faces a significant number of challenges to proceed with the case for an interacting groundwater and surface water drainage system in the Memnonia region. Here we prefer to leave the valley network drainage systems in direct contact with a Martian regolith, for which porosity estimates at the surface range from 20 to 50% [Clifford, 1993], and permeability estimates range from 10 to 103 darcies [Carr, 1979; MacKinnon and Tanaka, 1989; Squyres et al., 1992]. Under these conditions a hydraulic connection should exist between surface and ground water systems. In summary,
our preferred explanation for the terraces along the drainage system is that they represent erosional base level changes over time associated with a warmer, wetter, early climate, begging a terrestrial watershed analog for such a Martian drainage system.

**Morphologic Features of Crater Basins Consistent With Terrestrial Evaporite Basins**

The basin described above is not an isolated occurrence; many other basins exist in the southern highlands near the equatorial boundary that may also have been part of comparable surface drainage systems. Some of these have both inflow and outflow channels (e.g., that described above); others have inflow and no outflow channel (e.g., Memnonia, figures 8 and 11) and still others have terraces with neither obvious stream inflow or outflow (Figure 2). In the latter case, water inputs could be purely by base flow, and losses by evaporation and/or sublimation. These conditions are very common in the Atacama desert (discussed below). Given that one accepts the arguments presented above for a fluvial or lacustrine interpretation of these terraces, a steady base level must have operated for some significant geologic period for such terraces to form. To maintain these base levels, in the face of either periodic or steady inputs, requires commensurate outputs, either through baseflow or evaporation (or sublimation), to balance the water budget (i.e., conserve mass). But as argued above, maintenance of a base level gives the probable hydrogeologic characteristics of the highlands' regolith is problematic without that base level also being the water table. Given that the terraces exist in topographic depressions, groundwater could have provided seepage to these basins, leaving evaporative or sublimative losses as the most likely means of balancing the water exchanges (implied by the geomorphic features). Losses to the atmosphere by evaporation (or perhaps sublimation) within closed basins necessarily leads to saline conditions [Puskar, 1986; Storzer and Ericksen, 1974].

Thus, many of these closed basins may be akin to the salt flats of the western United States in that their deposits could record a history of progressive desiccation from saline lake to salt pan conditions in time [Puskar, 1985]. An additional consideration that may influence lake or groundwater salinities under Martian conditions is the exclusion of solutes through freezing. In addition to terraced crater-basins, all basins of the highlands with depths beneath a local base level would, like terrestrial basins of arid and semiarid areas, accumulate admixtures of wind, stream sediment, colluvium, and evaporites. On Earth these basins are strongly zoned, with evaporite floors, clastic, and marginal facies dominated by salt-cemented clastic units [Puskar, 1981; Storzer and Ericksen, 1974]. Sharp contrasts in albedo are commonly found at the basin margins, where the albedos of salts or wetted salts stand in sharp contrast to that of the clastic materials composing the fringing alluvial or colluvial fans, badhas, and pediments (Figure 3). Several basins on Mars exhibit a narrow curvilinear marginal fringe zonation involving linear zones with higher and/or lower relative albedos than that observed in upland or central basin floor areas (Figure 4). This suggests the presence of a very narrow basin margin deposit with unique spectral characteristics (e.g., mineralogy; see Table 1). Their analogs on Earth (e.g., Death Valley, Mojave Desert, Salar de Atacama, Atacama Desert) typically mark the location of groundwater seepage and/or pedoturbation. A further distinctive feature of Salars (salt pans) in the Chilean parts of the Atacama Desert are marginal salt fingers or channels, developed seaward of marginal seeps or springs toward the prevailing wind direction. Winds passing across the marginal zone, where fresh salt is often found, become charged with salt-laden aerosols and etch linear trails in the prevailing wind direction across the salars (Figure 3 and 5). One Martian crater-basin located at 167.15° longitude, -10.26° latitude, in the Meridiani northwest quadrangle exhibits long and narrow high albedo tracks extending into the basin floor deposits (Figure 5). The features appear morphologically distinct from the array of Martian wind features previously described within craters [Ward, 1979; Pyeskwad, 1981], and have aspect ratios much greater than typical Martian slide features. These linear tracks are good candidates for salt fingers.

A Martian analogy to terrestrial salars or sebkhas would predict that crater-basins within the Martian highlands that sit at levels near the regional base level would have filled with admixtures of evaporite, clastic, alluvial, and colluvial materials. Two Martian inselbergs formed of partially exhumed crater-basin deposits exist which have appropriate characteristics for an evaporite facies deposit: "White Rock" [Ward, 1979; Williams and Zimbelman, 1994], and a similar deposit in the Bucurial crater [Christensen, 1983]. Both of these Martian inselbergs have relatively high albedos (see Figure 6), but not as high or white as some of the filtered or processed images might lead one to conclude [Davies, 1979]. It is, however, fallacious to suggest that because modern perennially flooded salt flats, such as the Bonnevile Salt Flats of Utah, are white that evaporite facies deposits, in general, should also be white [Clark and Van Hoor, 1981]. Most of the salt plays on Earth that do not receive a seasonal flooding of surface water have reduced albedos (relative to those seasonally flooded). The driest parts of the Atacama Desert, Atacama, and Puna, show bright, medium, and dark salt surfaces (Figure 3), and in space shuttle color photography display reddish and brownish hues. Since an exhumed salt evaporite deposit is likely to have clastic, colluvial, and salt components, its albedo and color would in part be controlled by the nature of its mixed clastic, eolian, and evaporite facies, as...
Figure 3. (Top left) Ground photo of the brown to white salt-encrusted surfaces of the Salar de Atacama, northern Chile (68.8°W, 23.5°S), Oasis of Paine in middle ground, Altiplano in background (RDP, taken July 1982). (Top right) Hand-held space shuttle view of the Salar de Punta Negra, Chile (69.9°W, 24.5°S, NASA 61C-45-018), with bright outer rim of salts and core of dark salt facies. Scale bar is 10 km. (Lower left) Hand-held space shuttle view of evaporite basins in northern Chile with dark central salt facies and bright rims (68.8°W, 25.7°S, NASA STS-23-38-050). (Lower right) Martian crater basin with flat lying dark internal basin fill with a surrounding pediment apron (Viking frames 618A33 and -A35, centered at 322.0°W, 9.0°S). Scale bar is approximately 100 km.

well as the superimposed effects of secondary erosion and weathering of the exhumed deposit. Radiometric data for White Rock indicate higher than Mars-average albedo [McEwen, 1987], with red filter reflectivities higher than blue [Evans, 1979], consistent with that of many terrestrial evaporite facies deposits. Terrestrial (nonmarine) evaporite basin deposits are generally well bedded. As such, aerial and space imagery of partially exhumed deposits (Qidham Basin, western China; Tanerouft Basin, southcentral Sahara; Salt Ranges, Atacama Desert) exhibit a fine concentric to convoluted structure produced by the erosional sculpturing of the internal bedding. A similar fine and convoluted layer structure can also be observed in the Bequeral deposit (Figure 6, lower right). Stereo images of White Rock also reveal that the upper erosional surface has a vertically stratified structure [Williams and Zimbelman, 1994]. The variation of bedding facies in a nonmarine terrestrial evaporite basin can be used to explain an additional attribute of the Martian inselbergs. Both are relict mounds of basinfill deposits isolated
by erosion from the crater-basin margins. Salt-cemented rock (salcrete), while unstable under humid conditions, becomes very stable under arid conditions, and for example, in the hyperarid areas of the Atacama Desert have been used for buildings and roads. Central parts of basins have cores of salt-rich evaporite facies areas originally surrounded by a salt-poor, clastic-rich, marginal evaporite basin facies, which could have been preferentially left behind as winds deflated these basins over time.

**Discussion**

An evaporite basin analog for ancient crater-basins on Mars has several implications for future research and exploration efforts. First, from a climatic perspective, the existence of paleolakes and possible large accumulations of evaporite deposits in small basins (e.g., White Rock) reinforces the interpretations that an equilibrium may have existed between surface, groundwater, and atmospheric water reservoirs.

<table>
<thead>
<tr>
<th>Table 1. Authigenic Minerals of Nevada Playas and Atacama Salars [Paper, 1976; Stowe and Erickson, 1974]</th>
<th>Mineral</th>
<th>Chemical Formula</th>
<th>Mineral</th>
<th>Chemical Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chlorides</strong></td>
<td>Halite</td>
<td>NaCl</td>
<td>Nardon</td>
<td>Na₂CO₃·10(H₂O)</td>
</tr>
<tr>
<td>Sylvite</td>
<td>KCl</td>
<td>Thermonatrite</td>
<td>Na₂CO₃·H₂O</td>
<td></td>
</tr>
<tr>
<td><strong>Sulfates</strong></td>
<td>Gypsum</td>
<td>CaSO₄·2(H₂O)</td>
<td>Pirssonite</td>
<td>Na₂Cu₂CO₃·2(H₂O)</td>
</tr>
<tr>
<td>Basanite</td>
<td>CaSO₄·1/2(H₂O)</td>
<td>Strontianite</td>
<td>SrCO₃</td>
<td></td>
</tr>
<tr>
<td>Anhydrite</td>
<td>CaSO₄</td>
<td>Nalinite</td>
<td>Na₂CO₃·H₂O</td>
<td></td>
</tr>
<tr>
<td>Mirehitite</td>
<td>Na₂SO₄·10(SO₄)</td>
<td>Borax</td>
<td>Na₂S·10H₂O</td>
<td></td>
</tr>
<tr>
<td>Theranlite</td>
<td>Na₂SO₄</td>
<td>Tisocarbonate</td>
<td>Na₂S·5H₂O</td>
<td></td>
</tr>
<tr>
<td>Glauberite</td>
<td>Na₂Ca(SO₄)₂</td>
<td>Ulexite</td>
<td>NaCa₂O₂·2H₂O</td>
<td></td>
</tr>
<tr>
<td>Epsonite</td>
<td>MgSO₄·7(H₂O)</td>
<td>Collophane</td>
<td>Ca₂B₂O₇·11H₂O</td>
<td></td>
</tr>
<tr>
<td>Aphoniladite</td>
<td>Na₂SO₄·4H₂O</td>
<td>Starkeite</td>
<td>Na₂Si₃O₆·(H₂O)</td>
<td></td>
</tr>
<tr>
<td>Bleodite</td>
<td>Na₂MgSO₄·4(H₂O)</td>
<td>Soda Niter</td>
<td>NaNO₃</td>
<td></td>
</tr>
<tr>
<td>Celestine</td>
<td>SrSO₄</td>
<td>Others</td>
<td>Fluorite</td>
<td>CaF₂</td>
</tr>
<tr>
<td><strong>Carbonates</strong></td>
<td>Calcite</td>
<td>CaCO₃</td>
<td>Philippite (other zeolites)</td>
<td></td>
</tr>
<tr>
<td>Aragonite</td>
<td>CaCO₃</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dolomite</td>
<td>CaMg(CO₃)₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaylussite</td>
<td>Na₂Ca(CO₃)₂·5H₂O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnesite</td>
<td>MgCO₃</td>
<td>Potassium Feldspar</td>
<td>KASi₃O₈</td>
<td></td>
</tr>
<tr>
<td>Trooni N3</td>
<td>HCO₃·2(H₂O)</td>
<td>Burkeite</td>
<td>Na₂C₂O₄·SO₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hanksite</td>
<td>Na₆Ca₂KCl·(CO₃)·2SO₄·9H₂O</td>
<td></td>
</tr>
</tbody>
</table>
The potential presence of evaporite deposits on Mars would suggest that, at least locally, groundwater conditions reached saline conditions [Zent and Fanale, 1986]. The possible existence of saline groundwater has major chemical and mechanical consequences on the surface processes that have operated in the past. The densities of salt-saturated brines can be approximately 20% heavier, and their dynamic viscosities two- to threefold higher than that of distilled water. The mechanical impact is an increase in the driving forces for mass wasting of slopes, and an increase in the internal bulk dynamic viscosity of "flows". While the effects of salts on surface and shallow subsurface processes have been discussed [e.g. Clark and Van Hout, 1981; Malin, 1974; Zent and Fanale, 1986], there have generally not acknowledged the possible presence of localized evaporite facies deposits. As such, a salt karst alternative has not been included along with discussions of thermo- and carbonate-karst alternatives for many of the sapping-related features and terrains.

The evaporite basin hypothesis can be tested by several means, both with existing data and with improved data from future missions. XRF results from the Viking landers are interpreted to indicate the presence of various chloride salts on Mars [Ranit et al., 1992], and the present work would suggest that a salt signature might be strongest in the Memnonia region. Even with the limited spatial resolution available to Earth-based instruments, the great spectral resolution now available might be able to detect regional variations in some evaporite concentrations. Viking image data of the cratered highlands outside of Memnonia need to be searched systematically to assess the global distribution of possible fluvial terraces on Mars.

Future missions, particularly the Mars Global Surveyor spacecraft which will fly a large portion of the Mars Observer instruments to Mars, hold great potential for providing further tests to the hypothesis. The 1.4-m resolution of the Mars Observer cameras [Malin et al., 1992] may detect deflation pits associated with the "salt finger" weathering, as is observed on evaporite basins in South America, or layering associated with...
the evaporite deposits themselves. The combined spatial and spectral resolution of the thermal emission spectrometer [Christensen et al., 1992] may be able to detect evaporites at basin margins where eolian erosion has exposed variable albedo materials. Finally, the greatly improved topography which should result from the Mars Observer laser altimeter (Zuber et al., 1992) should be able to determine if terrains within the large crater basins conform to a geopotential surface, as would be expected for a fluvial origin. These new data are crucial to testing a variety of hypotheses related to the history of Mars, such as the evaporation basin hypothesis presented here.

The possible existence of evaporite deposits may at some point in the future have major practical importance for a long-term manned base on Mars. On Earth these deposits contain wide arrays of authigenic sulfates, chlorides, carbonates, borates, and nitrates (Table 1) and are economic sources of sodium chloride, borax, sodium carbonate, sodium sulfate, and lithium, and are potential sources of potash, magnesium, calcium chloride, fluorine, tungsten, uranium, zeolites, nitrates and clays. They may also form basic construction materials for buildings or roads under a hyperarid climatic condition. Finally, in our future efforts to search for past life on Mars, the postulated evaporite basin deposits would be a natural target for paleontologic study.

Acknowledgements. Research was supported by a Smithsonian Senior Fellowship to the first author and aided by resources made available at the Center for Earth and Planetary Studies, National Air and Space Museum. We thank Steve Clifford and an anonymous reviewer for their constructive reviews.

References


Bloom, A. L., Geomorphology, A Museum. We thank Steve Clifford and an anonymous reviewer for their constructive reviews.

Acknowledgements. Research was supported by a Smithsonian Senior Fellowship to the first author and aided by resources made available at the Center for Earth and Planetary Studies, National Air and Space Museum. We thank Steve Clifford and an anonymous reviewer for their constructive reviews.

References


Bloom, A. L., Geomorphology, A Museum. We thank Steve Clifford and an anonymous reviewer for their constructive reviews.

Acknowledgements. Research was supported by a Smithsonian Senior Fellowship to the first author and aided by resources made available at the Center for Earth and Planetary Studies, National Air and Space Museum. We thank Steve Clifford and an anonymous reviewer for their constructive reviews.

References


R. D. Forsythe, Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223 (email: falryth@email.uncc.edu)
J. R. Zimbelman, Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Washington, DC 20560 (email: jrz@exp.mau.edu)
(Received March 1, 1994; revised January 25, 1995; accepted January 25, 1995.)
Page intentionally left blank
Journal of Geophysical Research

Information for Contributors

Original scientific contributions on the physics and chemistry of the Earth, its environments, and the solar system will be considered for publication in the Journal of Geophysical Research. Every part of a paper submitted for publication should convey the author's findings precisely and immediately to the reader. Authors are urged to have their papers reviewed critically by colleagues for scientific accuracy and clarity of presentation.

Typescript. All parts of the paper must be typed double-spaced, preferably triple-spaced on good quality white paper 8½ by 11 inches (21.5 x 28 cm) with at least 1-inch (2.5 cm) margins at top, bottom, and sides. Erasable bond, which smudges easily, and tissue paper are not acceptable. Each page of the typescript should be numbered in the top right-hand corner.

Authors are expected to supply neat, clean copy and to use correct spelling, punctuation, grammar, and syntax. Spelling and hyphenation of compound words follow the unabridged Webster's Third New International Dictionary. The metric system must be used throughout; use of appropriate SI units is encouraged. Following recommended style and usage improves proofreading and reduces the chance of error in the typesetting.

Because text footnotes are expensive to set and distracting to the reader, they should be incorporated into the text or be eliminated completely.

The typescript should be arranged as follows: (1) title page including authors' names and affiliations, (2) abstract, (3) text, (4) reference list, (5) figure legends, and (6) tables.

Abstract. The abstract should be a single paragraph (250 words or fewer) stating the nature of the investigation and summarizing its important conclusions. Lining the contents in terms such as "this paper describes" or "the paper presents" should be avoided. Use of the passive voice often indicates that the author is merely describing the procedure rather than presenting conclusions. References should not be cited in the abstract.

The abstract should be suitable for separate publication in an abstract journal and be adequate for indexing.

When a paper is accepted for publication, the author will be requested to prepare camera-ready copy of the abstract for inclusion in Geophysical Abstracts in Press (GAP). AGU's editing service will accept accepted papers. At the same time, the author will select appropriate index terms for the paper.

Mathematics. All characters available on a standard typewriter must be typewritten in equations as well as in text. The letter I and the numeral 1 and the letter O and the numeral 0, which are identical on most typewriters, should be identified throughout the paper to prevent errors in typesetting. Any symbols that must be drawn by hand should be identified by a note in the margin. Authors will be charged for all corrections in galley proof for material that was handwritten in the typescript. Allow ample space (3½ inch (2 cm) above and below) around equations so that type can be marked for the printer.

Alignment of symbols must be unambiguous. Superscripts and subscripts should clearly be in superior or inferior position. Fraction bars should extend under the entire numerator. Barred and accented characters that are available for machine typesetting may be used. Symbols that are not available and therefore must be avoided are triple dots, any accents (other than bars) that extend over more than one character, and double accents (e.g., a circumflex over a bar). Accents over characters can be eliminated by the use of such symbols as ^, and ] and ] set as superscripts.

If an accent or underscored character is used to designate a special typeface (e.g., boldface for vectors, script for transforms), sans serif for tensors), the type should be specified by a note in the margin.

If the argument of an exponential is complicated or lengthy, "exp" rather than e should be used. Awkward fractional composition can be avoided by the proper introduction of negative powers. In text, scientific fractions (1/5) should be used, and enough conclusions should be included to avoid ambiguity. According to the accepted convention, parentheses, brackets, and braces are in the order: \{ [ ( ) ] \}. Displayed equations should be numbered consecutively throughout the paper; the number (in parentheses) should be to the right of the equation.

Notation. The notation is a list of symbols used in the text as an aid to the reader. It should be set up in this form:

- c rate of soil accumulation, m/yr
- d median grain size of water-deposited material, mm
- D distance of the loci of points, m
- h elevation of the rock stream channel at a particular time, m above base level
- H maximum vertical displacement of the fault associated with an earthquake, m
- H' amount of vertical displacement of the mountain-bounding fault, m
- z vertical coordinate of the model grid system
- p parameter computed by \( p = \frac{1}{\ln(\frac{2}{1})} \)
- \( y \) mean peak flow rate above base flow of stream, m/s

References. A complete and accurate reference list is of major importance. Omissions, discrepancies in the spelling of names, errors in titles, and incorrect dates make citations annoying, if not worthless, to the reader and cast doubt on the reliability of the author as well.

Only works cited in the text should be included in the reference list. References are cited in the text by the last name of the author and the year: (Jones, 1990). If the author's name is part of the sentence, only the year is bracketed. Personal communications and unpublished data or reports are not included in the reference list; they should be shown parenthetically in text: (F. S. Jones, unpublished data, 1990).

References are arranged alphabetically by the last names of authors. Multiple entries for a single author are arranged chronologically. Two or more publications by the same author in the same year are distinguished by a, b, c after the year.
For laboratory, company, or government reports, information should be included on where the report can be obtained. For Ph.D. and M.S. theses the institution granting the degree and its location should be given.

References to papers delivered at meetings should include title of paper, full name of meeting, sponsor, meeting site, and date. Citations of papers presented at meetings have been complicated by the practice of collecting manuscripts from participants, reproducing the manuscripts, and distributing the collections to people attending the meeting. Such collections should not be cited as published works.

References to books should include the page numbers of the material being cited.

Names of periodicals should be either written out in full or abbreviated according to the system employed by the Chemical Abstracts Service. One-word titles should always be given in full: Science. It is permissible to give only the initial page number of a paper but preferable to give the range of pages. Samples:


Refer to recent issues of the Journal of Geophysical Research for more examples.

Tables. Tables should be typed as authors expect them to look in print. Every table must have a title, and all columns must have headings. Column headings must be arranged so that their relation to the data is clear. Footnotes should be indicated by reference marks (*, †, ‡) or by lowercase letters typed as superscripts. Each table must be cited in text. Authors are urged to make supporting data available as supplements, as described below.

Illustrations. One original or glossy set of illustrations suitable for reproduction and three clear copies for the reviewers should accompany the typescript. Illustrations should be original drawings in black ink or, preferably, sharply focused glossy prints. Lettering should be mechanical, not hand drawn. When reduced for printing, lettering should be at least 1.5 mm high; the largest lettering should not exceed 3 mm. The following limiting dimensions apply: page height, 24.1 cm; page width, 17.3 cm, column width, 8.4 cm.

Figures not cluttered with information that could be placed in the legend look neater and are easier to read. All details on the figures should be checked carefully because errors found in galley proof stage (e.g., misspelled words, incorrect values, omitted symbols) must be corrected by the author, and the cost of preparing the new figure for printing will be charged to the author.

Each figure must be cited in numerical order in text and must have a figure legend.

Color figures and foldouts can be accommodated, but the additional costs for these special figures must be borne by the author. The author should retain an identical print for checking color proofs. The AGU Publications office should be consulted for current costs and other special instructions.

If figures that have been published under copyright are to be reproduced in JGR, AGU must have permission from the copyright holders. Copies of letters of permission to republish should accompany the typescript.

Supplementary material: Authors are encouraged to submit papers that are as concise as possible. Supporting material, such as tables of data, computer printouts, appendices, additional graphs, lengthy mathematical derivations, and extended background discussions may be made available through videotape, electronic emanipation, or microfiche. All supplements will be subject to the same peer review as the printed material. Summaries of articles and key figures may be submitted for publication in the journal, the detailed article being published on microfiche. All material published on microfiche is included in the microform editions of JGR (and therefore archived in libraries) and is available to individuals on order. Contact the editors or AGU Publications office for more information on this service.