A new neural network approach including first-guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observations

F. Aires, C. Prigent

Department of Applied Physics, Columbia University, NASA Goddard Institute for Space Studies, New-York

W.B. Rossow

NASA Goddard Institute for Space Studies, New-York

M. Rothstein

Science Systems and Application, Inc., NASA Goddard Institute for Space Studies, New-York

Abstract. The analysis of microwave observations over land to determine atmospheric and surface parameters is still limited due to the complexity of the inverse problem. Neural network techniques have already proved successful as the basis of efficient retrieval methods for non-linear cases, however, first-guess estimates, which are used in variational methods to avoid problems of solution non-uniqueness or other forms of solution irregularity, have up to now not been used with neural network methods. In this study, a neural network approach is developed that uses a first-guess. Conceptual bridges are established between the neural network and variational methods. The new neural method retrieves the surface skin temperature, the integrated water vapor content, the cloud liquid water path and the microwave surface emissivities between 19 and 85 GHz over land from SSM/I observations. The retrieval, in parallel, of all these quantities improves the results for consistancy reasons. A data base to train the neural network is calculated with a radiative transfer model and a a global collection of coincident surface and atmospheric parameters extracted from the National Center for Environmental Prediction reanalysis, from the International Satellite Cloud Climatology Project data and from microwave emissivity atlases previously calculated. The results of the neural network inversion are very encouraging. The r.m.s. error of the surface temperature retrieval over the globe is 1.3 K in clear sky conditions and 1.6 K in cloudy scenes. Water vapor is retrieved with a r.m.s. error of 3.8 kg/m^2 in clear conditions and 4.9 kg/m^2 in cloudy situations. The r.m.s. error in cloud liquid water path is 0.08 kg/m^2. The surface emissivities are retrieved with an accuracy of better than 0.008 in clear conditions and 0.010 in cloudy conditions. Microwave land surface temperature retrieval presents a very attractive complement to the infra-red estimates in cloudy areas: time record of land surface temperature will be produced.

1. Introduction

Even after 20 years of global microwave satellite observations, the use of microwave data over land for the retrieval of atmospheric and surface parameters is still very limited. While the ocean surface has a low microwave emissivity ~0.5 that produces good contrast of atmospheric phenomena against a low brightness

very sensitive to the presence of thin clouds. The sensitivity of SSM/I to water vapor is very low, except in the most arid areas; so the results do not improve on the first-guess values. With an estimated accuracy of $\sim$0.1 kg/m^2, the SSM/I retrieval does not properly characterize the thinner clouds (the majority) but the cloud structures with higher liquid water content are well delineated.

A further improvement in this variational inversion scheme could be obtained by also retrieving the seven surface emissivities as they undergo small day-to-day changes induced by variations of the soil moisture, the vegetation density, or the snow cover. However, in this case, ten variables would have to be retrieved (Ts, WV, LWP plus the seven emissivities Em_i, where i represents the 7 channels of SSM/I: 19GHz V, 19 GHz H, 22 GHz V, 37 GHz V, 37 GHz H, 85 GHz V and 85 GHz H; V for vertical polarization and H for horizontal polarization) from the seven SSM/I brightness temperatures and additional information would be needed to solve the problem. The monthly-mean emissivity values previously computed could be used as first-guess (or, using more specifically the variational assimilation formalism, the background) estimates of the surface emissivity and the first-guess matrix of error covariances could be calculated. There are several options: The covariance matrix could be calculated globally for a given month, estimated for a given type of surface, or even calculated for each single pixel considering all the monthly mean emissivities for this pixel. The inversion scheme would then rely very heavily on the representativity of such covariance matrices, giving an important weight to the statistical description of the emissivity relationships. Given this difficulty with the retrieval of the surface emissivities with a variational method, another inversion approach is considered.

Neural network techniques have already proved very successful in the development of computationally efficient inversion methods for satellite data and for geophysical applications [*Escobar et al.*, 1993; *Aires et al.*, 1998; *Chevallier et al.*, 2000]. They are well adapted to solve non-linear problems and are especially designed to capitalize on the inherent statistical relationships among the retrieved parameters. Note that variational techniques, as usually implemented, do not account for correlations among the retrieved parameters. However, for many ill-conditioned problems, the use of a first-guess estimate is very important to regularize the inversion process and the first-guess error covariance matrix is also essential in 3D/4D variational assimilation schemes since it controls the impact of the observations

This distance is dependent on the a priori information available on the probability distribution functions of the variables involved. If the observations y are assumed to be Gaussian distributed with zero-mean and without other a priori information, the Mahalanobis distance [*Crone and Crosby*, 1995] is optimal

$$\frac{1}{2}[y(\hat{x}) - y^o]^t < y \cdot y^t >^{-1} [y(\hat{x}) - y^o], \qquad (3)$$

where $< y \cdot y^t >$ is the covariance matrix of the observable quantities without measurement noise, y. This procedure has to be applied to find an optimum solution for each observation separately and can require significant computational resources.

The second approach consists in estimating a general mapping function g_W, with parameters W, that is a global model for y^{-1}. The parameters W are the results of the minimization of a cost function

$$\int D(\hat{x}, x) P(x, \eta), \qquad (4)$$

where $\hat{x} = g_W(y^o) = g_W(y(x) + \eta)$, and P is the joint probability distribution of the physical variables x and the noise η. The distance $D(\hat{x}, x)$ is integrated over the physical states and over the observation noise, so that the model g_W is optimized globally over the range of x and the noise. In practice, to minimize the previous criterion, a data base is created, composed of a statistically representative sample of coincident variables x and observations y^o and the estimation of the parameters W is made once and for all using this dataset. These schemes are called "global" inversions. After this preliminary step for the estimation of W, the inversion of an observation is very fast since it involves only the direct use of the model g_W.

The distances used for localized and global inversion schemes involves different variables. The first one works on the brightness temperature space, the second one on physical variable space. The optimum solution in (4) gives an estimation $\hat{x}$ that is close to the true solution x while the distance in (2) specifies that the brightness temperatures $y(\hat{x})$, associated with the estimated solution $\hat{x}$, are close to the brightness temperatures $y(x)$ associated to the real solution x.

Inverse problems are often ill-posed since the existence and the uniqueness or the stability of the solution is not always known [Vapnik, 1997]. This is especially the case when the "forward" model, $y(x)$, is not linear; in our case the radiative transfer is not linear. To regularize the inversion process, all a priori information available should be used to constrain the solution,

This equation is the only computation required in the operational mode (once the synaptic weights have been determined by the training procedure). A bias term for each neuron has been deliberately omitted to simplify the notation, even if it is used in the neural network.

It has been demonstrated [*Hornik et al.*, 1989; *Cybenko*, 1989] that any continuous function can be represented by a one-hidden-layer MLP with sigmoide functions σ.

2.2.2. Optimization Algorithm: The Back-Propagation of Errors. Given a neural architecture (functions used as transfer functions σ, number of layers, neurons and connections), all the information of the network is contained in the set of all synaptic weights $W = \{w_{ij}\}$. The learning algorithm is an optimization technique that estimates the optimal network parameters W^* by minimizing a cost function $C_1(W)$, approaching as closely as possible the desired function. The criterion usually used to derive W^* is the mean square error in network outputs

$$C_1(W) = \frac{1}{2} \sum_{k \in S_2} \int \int D_E(\hat{x}_k(Y; W), x_k)^2 \, P(Y, x_k) dx_k dY \tag{8}$$

where D_E is the Euclidean distance between x_k, the kth desired output component, and $\hat{x}_k$, the kth neural network output component and S_2 is the output layer of the neural network. Other contrast measures can be used if a priori information is available. $P(Y, x_k)$ is the joint probability distribution function of Y and x_k. This criterion is just the integrated distance between $\hat{x}$ and x introduced in (4).

In practice, the probability distribution function, $P(x_k, Y)$, is sampled in a dataset $\mathcal{B} = \{(Y^e, {x_k}^e), e = 1, \ldots, N\}$ of N input/output couples, and $C_1(W)$ is then approximated by the classical least square criterion:

$$\tilde{C}_1(W) = \frac{1}{2N} \sum_{e=1}^{N} \sum_{k \in S_2} D_E(\hat{x}_k(Y^e; W), {x_k}^e)^2 \tag{9}$$

The Error Back-Propagation algorithm [*Rumelhart et al.*, 1986] is used to minimize $\tilde{C}_1(W)$. It is a gradient descent algorithm that is very well adapted to the MLP hierarchical architecture because the computational cost is linearly related to the number of parameters. Traditional gradient descent algorithms use all the samples of the dataset $\mathcal{B}$ to compute a mean Jacobian of the criterion $\tilde{C}_1(W)$ in equation (9). These algorithms are called deterministic gradient descent. The major inconvenience of this approach is that the descent can be trapped in local minima. In the present application, a stochastic gradient descent algorithm is

radiative transfer equation globally, once and for all and uses the distribution $P(x)$ for this purpose. This model is then valid for all observations (i.e. global inversion).

To minimize this criterion, the neural network method creates a dataset $\mathcal{B} = \{(x^e, y^{oe}, x^{be}); e = 1, \ldots, N\}$ that samples as well as possible all these probability distribution functions. Then, the practical criterion used during the learning stage is given by:

$$\tilde{C}_2(W) = \frac{1}{2N} \sum_{e=1}^{N} D_E(g_W(x^{be}, y^{oe}), x^e))^2. \qquad (13)$$

To create the dataset $\mathcal{B}$, we sample the probability distribution function $P(x)$ by selecting geophysical states (x^e) that cover all natural combinations and their correlations and by calculating $y^e = y(x^e)$ with the physical model (the radiative transfer model in this case). Alternatively we could obtain these relationships from a "sufficiently large" set of colocated and coincident values of y and x. For P_η we need a priori information about the measurement noise characteristics; a physical noise model could be used, but if all we have is an estimation of the noise magnitude, then we have to assume Gaussian distributed noise η that is not correlated among the measurements. For $P(x^b|x)$ there are two situations. If a first-guess dataset $\{x^{be};\ e = 1, \ldots, N\}$ exists, then x^{be} can be used directly. If such a dataset is not available, we have to determine $P(\varepsilon)$ (as it is done in variational assimilation technique), the distribution of errors in the first-guess, $\varepsilon = x^b - x$, and use $x^b = x + \varepsilon$ as input to the network. The balance between reliance on the first-guess and the direct measurements is then made automatically and optimally by the neural network during the training.

Table 1 summarizes the specific features of the neural network scheme with first-guess and the variational assimilation inversion technique.

Table 1

3. Generation of a Data Base to Train the Neural Network

Two neural networks are trained: One for clear scenes (NN1), one for cloudy scenes (NN2). They both retrieve simultaneously the surface temperature Ts, the integrated water vapor content WV, and the seven SSM/I surface emissivities Em_is. In addition to these parameters, NN2 retrieves the cloud liquid water path LWP. Two sources of information are used for this purpose: (1) seven SSM/I brightness temperatures (observations), and (2) a priori information of the state of the surface and atmospheric variables from ancillary datasets. In this study the experimental configuration is

information in the retrieval.

3.1.2. The use of the ISCCP dataset. In the ISCCP data, cloud parameters and related quantities are retrieved from visible (VIS ~0.6 μm wavelength) and infrared (IR ~11 μm wavelength) radiances provided by the set of polar and geostationary meteorological satellites [*Rossow and Schiffer*, 1999]. The ISCCP dataset is used in this study to discriminate between clear and cloudy scenes (selecting NN1 or NN2) and to give estimates of the cloud top temperature and surface skin temperatures. The pixel level dataset (the DX dataset) is selected for its spatial sampling of about 30 km and its sampling interval of 3 hours [*Rossow et al.*, 1996].

3.1.2.1. The surface temperature first-guess: ISCCP provides the surface skin temperature first-guess retrieved from IR radiances under clear conditions. The IR emissivity of the surface is always close to 1 and varies with the land surface type as in the GISS GCM. Instead of selecting the closest-in-time DX image to derive the surface temperature, a linear interpolation between two ISCCP surface temperature estimates to the precise time of the SSM/I overpass is calculated to account for the diurnal cycle. If the ISCCP DX scenes are cloudy, a clear sky compositing procedure is conducted within the ISCCP process to derive an estimation of the surface temperature (see *Rossow and Garder* [1993] for more details). The error associated with the surface temperature is estimated to be 4 K [*Rossow and Garder*, 1993].

3.1.2.2. Cloud a priori information: First the ISCCP data helps discriminate between clear and cloudy scenes. Over the ocean, it has been shown that the VIS and IR observations have a better ability than the microwave measurements to detect clouds [*Lin and Rossow*, 1994]. Given that the sensitivity of the microwave to clouds over land is much lower than over ocean, when a pixel is considered clear by the ISCCP procedure, the LWP is fixed to zero. Two neural networks are used, one for clear scenes another for cloudy scenes. The ISCCP cloud flag directs the retrieval to one network or the other.

For cloudy scenes, the cloud top temperature derived from IR measurements is added to the retrieval process as additional information, to account for the changes in the emission temperature of the cloud and in the cloud liquid water absorption coefficient. In contrast to the ocean case, clouds induce only small variations in the microwave radiation over land and additional cloud information facilitates their detection. *Prigent and Rossow* [1999] showed that the ability to estimate

lengths, has been recently developed [*Pardo et al.*, 2000] but is not used in this study because the differences on SSM/I frequencies are negligible.

Cloud absorption is calculated using the Rayleigh approximation which is valid for most non-precipitating liquid water clouds at SSM/I frequencies. The cloud temperature is assumed to be equal to the air temperature at the same level. The dielectric properties of liquid water are taken from *Manabe et al.* [1987]. Scattering by large particles is not considered meaning that convective clouds and rain are not represented in the data base.

The surface contribution is calculated using the monthly mean emissivities previously calculated [*Prigent et al.*, 1997, 1998] and assuming specular reflection at the surface.

The consistency of the radiative transfer model has been checked. Observed brightness temperatures and simulated Tbs using the ISCCP Ts and LWP, the NCEP WV, and the monthly Em_is have been compared for two months of SSM/I data globally over snow and ice-free pixels: For all channels, the bias is smaller than 0.5 K even for cloudy cases. Thus, the training dataset generated with this radiative transfer model and sources of global data accurately represents the distribution of these parameters that SSM/I observes.

3.3. Statistical Analysis of the Training Data Base

The training data base generated by the RT model applied to the ISCCP, NCEP and monthly Em_is datasets contains the variables to be retrieved (Ts, WV, LWP, and the seven Em_i), the seven simulated brightness temperatures Tb, and a priori information on the cloud top temperature Tc and the temperature of the lowest layer of the atmosphere Ta. An error is associated with most variables that are used as first-guesses. The data base is produced from data collected for January and June 1993 over land between 60°S and 80°N. Snow or ice covered pixels are not considered: The snow and ice information comes from the NOAA operational analysis. 1,391,671 samples are collected, 55% of them correspond to cloudy scenes.

Figure 3 shows the "global" distributions of some of the variables in the training data base. The distributions are non-Gaussian and some of them are truncated. Non-Gaussian distributions are often indicators of non-linear behavior [*Burgers and Stephenson*, 1999; *Palmer*, 1999]. For example, the liquid water path distribution has its maximum at the lowest values and obviously cannot be negative. When retrieving such a variable with

Figure 3

perature with brightness temperature is higher for the vertical polarization than the horizontal one. This can be explained by less variability in the emissivities for the vertical polarization than for the horizontal one. Correlations between surface temperature and brightness temperatures at vertical polarization are similar at all SSM/I frequencies, which was not anticipated. At 22 and 85 GHz, water vapor absorption was expected to impede a direct relation between surface contributions and top of the atmosphere measurements and as a matter of fact, derivatives of the brightness temperatures with surface temperature (sensitivities) are smaller at 22 and 85 GHz than at 19 and 37 GHz [*Prigent et al.*, 1999]. Other authors have also observed large correlations at 85 GHz between surface air temperatures and *Tbs* at 85 GHz. *MacFarland et al.* [1990] investigated the correlation between SSM/I observations and surface air temperature and concluded that 22 and 85 GHz measurements, depending on the surface type, are the most sensitive to the land surface temperature. *Basist et al.* [1998] also proposed a method to retrieve near-surface air temperature from SSM/I that relies heavily on the 85 GHz channels. These results can be explained by two factors. First, for a given polarization, the surface emissivities at 19 GHz are more variable than at other frequencies because of higher sensitivity to surface properties like soil moisture or vegetation water content and structure. These emissivity variations are not correlated with to surface temperatures fluctuations as indicated by the correlation coefficients between Ts and emissivity at 19 GHz (see Table 2). Secondly, the absorption at 22 and 85 GHz actually dumps the effects of emissivity fluctuations, enhancing the relationship between brightness temperatures and surface temperature. The global correlation coefficients in Table 2 may not be representative on a local scale. However, correlation coefficients for Ts have been calculated for three ranges of atmospheric water vapor amount and emissivities and no significant differences were observed in the coefficients.

Global correlations between atmospheric water vapor and brightness temperatures are relatively low especially for vertical polarization because of large surface emissivities reducing the contrast between atmospheric and surface emissions. Even for horizontal polarization, global correlations never exceeds 0.6. However, these global values mask large local differences. Correlation coefficients calculated for different ranges of emissivities and water vapor amounts show that the results are very different, especially for the 85 and 22 GHz channels depending on water vapor amount. As a consequence, the

scribes the variety of the situations to be analyzed.

4. Results from the Neural Network Inversions

Two neural networks have been trained, one for clear pixels (NN1) the other one for cloudy pixels (NN2), both using a priori information. The ISCCP cloud flag discriminates between clear and cloudy pixels. The architecture of the network NN1 is an MLP with 17 inputs coding the 7 SSM/I observations y^0 and the first-guess x_b (Ts, Ta, WV, and 7 Em_i), 30 neurons in the hidden layer, and 9 neurons in the output layer coding the retrieval x (Ts, WV, and 7 Em_i). The number of neurons in the hidden layer is estimated by a heuristic procedure that monitors the generalization errors of the neural network as the configuration is varied. The network NN2 has one additional input, the cloud top temperature Tc, and one additional retrieval, the liquid water path (LWP). The input variables and their associated standard deviation errors are summarized in Table 3. The full matrix of the error covariances is calculated at the end of the training phase (not shown here). This matrix gives the statistical structure of the errors and is of great importance in the assimilation of retrieved products in a Numerical Weather Prediction scheme.

Table 3

For each variable, the distribution of the first-guess error is a Gaussian truncated at 2 standard deviations. In contrast to variational method, where only Gaussian distributions can be used, the neural network method can use any distributions shape. However in the present study, no in situ data are available to calculate the distribution of the first-guess errors, so Gaussian noise is introduced independently on each variable to generate the first-guess. In the operational mode with real first-guesses, the technique will use the structure of the first-guess error correlations and the results should be even better.

Figure 4 presents the learning curves of the neural network for clear and cloudy situations with and without first-guesses. To measure the impact of the introduction of the first-guess information in a neural network inversion scheme, two additional networks have been trained without first-guesses, one for clear conditions and another for cloudy scenes. The architectures of the the networks without first-guess are similar in structures, except that there are only seven inputs, coding the SSM/I observations y^0. For each retrieved variable, the r.m.s error decreases from the first-guess r.m.s. error to a stable value after several iterations. The networks with first-guess input show substantially

Figure 4

exploit the spectral dependence of the first-guess emissivities to provide a more accurate estimate of both the emissivities and the surface temperature.

4.2. Water Vapor

WV is retrieved with a relative error of $\sim$ 30% for both clear and cloudy situations, when using a first-guess. This is a small improvement over the first-guess r.m.s. error of 40%. The errors are not significantly different in presence of clouds. With the variational method [*Prigent and Rossow*, 1999], the retrieval errors were found to increase with decreasing emissivities and to increase in presence of clouds as expected from the sensitivity of the radiative transfer to the various parameters. As observed in Table 2, the correlation between the brightness temperatures and WV is rather low (maximum of $\sim$0.6 globally), and the neural network scheme is likely to exploit water vapor correlation with another variable to extract water vapor information when direct correlation between Tbs and WV is not sufficient. It is worth mentioning that the neural network is trained to minimize the absolute WV error difference not the relative error in WV. Changes could be made to minimize the relative error if this option were preferred.

4.3. Cloud Liquid Water Path

For LWP, the r.m.s error is 0.08 kg/m^2 globally. As expected, the error is larger in areas of high emissivities where the contrast between the land surface and the cloud is smaller. Even in areas of low emissivities ($0.85 \leq$emissivity19H≤ 0.9), the accuracy of the retrieval is not suitable for detection of majority if clouds. The cloud flag from ISCCP is of importance in this case to direct the retrieval toward the appropriate neural network. However, cloud structures with large liquid water path can be detected whatever the surface type; Plate 1 shows several extended and thick clouds that are also present on the ISCCP images (not shown). Plate 1 does not show any evidence of LWP errors (discontinuities) related to strong emissivity gradients.

4.4. Land Surface Emissivities

When using a first-guess, the neural network technique shows a good aptitude for retrieving land surface emissivities with an r.m.s. error lower than 0.008 (0.010) globally for all channels, in clear conditions (cloudy conditions respectively). This is an improvement over the first-guess errors. Unaided by the first-guess estimate, the neural network technique does not

4.6. Analysis of the Neural Network Sensitivities

An interesting capability of the neural network technique is that the adjoint model of the neural network is directly provided [*Aires et al.*, 1999]. The computation of this adjoint model (or neural Jacobians or neural sensitivities) is analytical and very fast. Since the neural network is non-linear, these Jacobians are dependent on the situation x. For example, the neural Jacobians in our example of equation (7) (an MLP network with one hidden layer) are:

$$\frac{\partial x_k}{\partial y_i} = \sum_{j \in S_1} w_{jk} \sigma' \left(\sum_{i \in S_0} w_{ij} y_i \right) w_{ij} \qquad (14)$$

where σ' is the derivative of the transfer function σ. For a more complex MLP network with more hidden layers, there exists a back-propagation algorithm that efficiently computes the neural Jacobians. The neural Jacobian concept is a very powerful tool since it allows for a statistical estimation of the multivariate and non-linear sensitivities between input and output variables in the model under study [*Aires and Rossow*, 2000].

Table 5 gives the mean neural Jacobian values for the variables x_k and y_i for the neural network NN1 with first-guess. The neural Jacobians are normalized by the standard deviation of the respective variables ($\frac{\partial x_k}{\partial y_i} \times \frac{std(y_i)}{std(x_k)}$) to enable comparison of the sensitivities between variables with different variation characteristics. These values indicate the relative contribution of each input in the retrieval of a given output parameter. The numbers correspond to mean global values which may mask rather different behavior in various regions of the globe.

Table 5

Figure 6 presents some of the normalized neural Jacobians for the surface temperature and the water vapor for three ranges of Em at 19 GHz H polarization. Depending on the surface emissivity, the sensitivity of Ts to different inputs changes from larger sensitivity to 19 GHz vertical polarization for high emissivities to larger sensitivity to the 85 GHz observations and the first-guess information at low emissivities (Figure 6a). For WV retrieval, very different regimes are observed for low and high water vapor amounts (Figure 6 b), from larger sensitivity to the 85 GHz channel horizontal polarization for high water vapor amount to smaller sensitivity for low water vapor contents. The same trend is observed at 22 GHz. We have already commented on the differences between local and global correlations in Section 3.3. In contrast to a linear regresssion-type algorithm that fits a mean state mapping between inputs

Figure 6

scheme that include first-guess information. Its potential have been tested in the complex and ill-conditioned problem of inversion of SSM/I microwave observations over land. A data base to train the neural network is derived from a global collection of coincident surface and atmospheric parameters, extracted from the NCEP reanalysis, from the ISCCP data, and from microwave emissivity atlases previously calculated. The introduction of the first-guess information into the neural network has a considerable impact on the results compared to the network without first-guess.

The r.m.s. error of the surface temperature retrieval is 1.3 K in clear sky conditions and 1.6 K in cloudy scenes over the globe. Microwave land surface temperature retrieval presents a very attractive complement to the infra-red estimates in cloudy areas. By combining both measurements as we have done, a complete (clear and cloudy days) time record of land surface temperature can be produced. Water vapor is retrieved with an r.m.s. error of 3.8 kg/m^2 in clear conditions and 4.9 kg/m^2 in cloudy situations. The r.m.s. error in liquid water path is 0.08 kg/m^2. The surface emissivities are retrieved with an accuracy of better than 0.008 in clear conditions and 0.010 in cloudy conditions, both improvements on the original first-guess.

The analysis methodology presented here and compared to the better-known variational assimilation technique provides an illustration of a more general approach to the analysis of high-volume, multi-wavelength satellite observations that may have great potential. The common practice of isolating one variable at a time from such datasets breacks correlations among the measurements and among the retrieved quantities. The variational approach goes a step further by obtaining simultaneous retrievals of many quantities form multiple measurements; however, as usually implemented, the variational analysis still does not account for correlations of variables. The neural network approach is not only able to accomodate strongly non-linear relationships, but also is able to benefit from the correlations to improve the retrievals. The neural network approach also requires much less computation than the variational assimilation approach. That the two methods are conceptually close, as we have shown, puts the neural network approach on the same theoretical ground as the better-studied variational analysis methods. However, the fact that a simple neural network has been shown to provide a statistical fit to any function suggests that what the trained network is doing is simulating (statistically) the equations of the physical model, in this case an inverse radiative transfer model. Thus, despite use

Appendix A: Notation

x	vector of physical variables to retrieve
$\hat{x}$	estimate of x
x^b	first-guess a priori information for x
x_n	nth estimate of x in variational method
ε	$= x^b - x$, first-guess error
$y(x)$	radiative transfer function for the physical variable x (also a vector)
y^o	SSM/I brightness temperature observations
η	SSM/I instrumental noise
P	generic probability distribution function
$P_\eta(\eta)$	probability distribution function of η
$P_\varepsilon(\varepsilon)$	probability distribution function of ε
$H(x)$	derivative of y with respect to x
$A(x)$	covariance matrix of retrieval error estimates in variational method
B	$=< \varepsilon^t \cdot \varepsilon >$, covariance matrix of the first-guess errors
E	$=< \eta^t \cdot \eta >$, covariance matrix of the measurement errors
F	covariance matrix of the radiative transfer model errors
$E[\cdot]$	expectation operator
a_i	activity of neuron i
σ	sigmoide function of the neural network
z_i	output of the neuron i
w_{ij}	synaptic weight between neuron i and neuron j
g_W	neural network model
W	$= \{w_{ij}\}$, the set of the parameters of the neural network
y_i	neural network input value on neuron i
x_k	neural network output value on neuron k
$\mathcal{B}$	dataset sampling the probability distribution functions
D	generic distance
D_E	Euclidean distance
$C_1(W)$	theoretical quality criterion for classical neural network learning phase
$\tilde{C}_1(W)$	practical quality criterion for classical neural network learning phase
$C_2(W)$	theoretical quality criterion for classical neural network learning phase with first-guess information
$\tilde{C}_2(W)$	practical quality criterion for classical neural network learning phase with first-guess information

Appendix B: The 1-D Variational Scheme

This method is described by *Rodgers* [1976] and by *Eyre* [1989]. The unified notation of *Ide et al.* [1997]

Using Bayes theorem, we can rewrite the conditional probability in (B3) as

$$P(x|y^o, x^b) = \frac{P(x, y^o, x^b)}{P(y^o, x^b)} = \frac{P(y^o, x^b|x)P(x)}{P(y^o, x^b)}. \quad \text{(B5)}$$

It is often assumed, even if it is not always the case, that y^o and x^b, the direct and virtual (first-guess) measurements, are independent. In that case, we can expand the corresponding joint probability distribution functions using the Bayes theorem

$$= \frac{P(y^o|x)P(x^b|x)P(x)}{P(y^o)P(x^b)}. \quad \text{(B6)}$$

We want to maximize this probability with respect to x. If the probability distribution $P(x)$ of the physical variables x is available, it is possible to use it in the general context of Bayesian estimation. If this pdf is Gaussian, this would correspond to the addition of a term $\frac{\partial}{\partial x}[x - \bar{x}]^t \bar{B}^{-1}[x - \bar{x}]$ in (B8), where $\bar{x}$ is the mean state of the physical variables and $\bar{B}$ is the covariance matrix of the physical variables. This approach is not used in general in variational assimilation.

If no a priori information on the distribution $P(x)$ is available, this distribution is considered to be uniformly distributed (i.e., no information), so this term can be neglected during the maximization process. The two probabilities $P(y^o)$ and $P(x^b)$ are not dependent on x so they can also be neglected. The maximum likelihood estimator is then obtained at the minimum of minus the log of the two remaining probabilities. Assuming that the minimum is unique, the optimal solution is characterized by

$$-\frac{\partial log[P(y^o|x)P(x^b|x)]}{\partial x} = 0. \quad \text{(B7)}$$

These probabilities need to be rewritten in order to extract the independent random variables involved in the model. Note $P(y^o|x) = P(y^o|y(x))$ since the theoretical radiative transfer function y is not a stochastic function. So $P(y^o|x) = P_\eta(y^o - y)$, where P_η is the probability distribution function of the instrumental noise and the forward model error. Furthermore, $P(y^o|x) = P_\eta[H(x_n)(x - x_n) + (y(x_n) - y^o)]$ using relation (B1). Also $P(x^b|x) = P_\varepsilon(x^b - x)$ where P_ε is the probability distribution function of the first-guess error ε.

Assuming that the errors in the observations, the direct radiative transfer model, and the a priori first-guess information are unbiased, uncorrelated, and have Gaussian distributions, expression (B7) is equivalent to

two pieces of information. If these matrices are not sufficiently precise, or if the variability of the matrices with atmospheric situations is not sufficiently sampled, an "empirical" weight has to be determined.

Acknowledgments. The authors would like to thank Frédéric Chevallier and Jean-Noel Thépaut for their helpful comments and Alain Chédin for fruitfull discussions. This work was partly supported by special funding provided by Dr. Robert J. Curran, NASA Climate and Radiation Program.

References

Aires, F., R. Armante, A. Chédin, and N.A. Scott, Surface and atmospheric temperature retrieval with the high resolution interferometer IASI, *Proc. of the Am. Meteorol. Soc., 98,*, Paris, No P5.46.B, 181–186, 1998.

Aires, F., M. Schmitt, N.A. Scott, and A. Chédin, The Weight Smoothing regularization for Jacobian stabilization. *IEEE Trans. on Neural Networks*, **10**, 6, 1502–1510, 1999.

Aires, F., and W.B. Rossow, Inferring instantaneous, multivariate and nonlinear sensitivities for the analysis of feedback processes in a dynamical system: Lorenz model case study, *J. of Clim.*, submitted, 2000.

Basist, A., N. C. Grody, T. C. Peterson, and C. N. Williams, Using the Special Sensor Microwave /Imager to monitor land surface temperatures, wetness, and snow cover, *J. Atmos. Sci.*, **37**, 888–911, 1998.

Burgers, G., and D.B. Stephenson, The "normality" of El Niño, *Geophy. Res. Lett.*, **26**, 8, 1027–1030, 1999.

F. Chevallier, J.-J Morcrette, F. Chéruy, and N. A. Scott, Use of a neural network-based longwave radiative transfer scheme in the ECMWF atmospheric model. *Q. J. R. Meteor. Soc.*, **126**, 761-776, 2000.

Colton, M. C., and G. A. Poe, Intersensor calibration of DMSP SSM/I's: F-8 to F-14, 1987-1997, *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 418-439, 1999.

Crone, L., and D. Crosby, Statistical applications of a metric on subspaces to satelite meteorology, *Technometrics*, **37**, 324-328, 1995.

Cybenco, G., Approximation by Superpositions of a Sigmoidal Function, *Math. Control Signals Systems*, **2**, 303–314, 1989.

Derber, J., and F. Boutier, A reformulation fo the background error covariance in the ECMWF global data assimilation system, *Tellus*, **51A**, 195–221, 1999.

Duflo, M., Algorithmes stochastiques, Mathématiques et Applications, Springer, 1996.

English, S. J., Estimation of temperature and humidity profile information from microwave radiances over different surface types, J. Appl. Meteorol., **38**, 1526-1541, 1999.

Escobar, J., A. Chédin, F. Chéruy, and N.A. Scott, Réseaux de neurones multicouches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sondeurs verticaux satellitaires, *C. R. Acad. Sci. Paris*, **317**, 2, 911–918, 1993.

Eyre, J. R., Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simula-

Pardo, J.R., J. Cernicharo and E. Serabyn, Atmospheric Transmission at Microwaves (ATM): An Updated Model for mm/submm applications, *J. Quant. Spectrosc. Radiat. Transfer*, submitted, 2000.

Phalippou, L., Variational retrieval of humidity profile, wind speed and cloud liquid water path with SSM/I: Potential for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, **122**, 327-355, 1996.

Prigent C., W. B. Rossow, and E. Matthews, Microwave land surface emissivities estimated from SSM/I observations, *J. Geophys. Res.*, **102**, 21,867-21,890, 1997.

Prigent, C., W. B. Rossow, and E.Matthews, Global maps of microwave land surface emissivities: Potential for land surface characterization, *Radio Science*, **33**, 745-751, 1998.

Prigent, C., and W. B. Rossow, Retrieval of surface and atmospheric parameters over land from SSM/I: Potential and limitations, *Q. J. R. Meteorol. Soc.*, **125**, 2379-2400, 1999.

Prigent, C., A. Aires, W. B. Rossow, and E. Matthews, Joint characterization of the vegetation by satellite observations from visible to microwave wavelengths: A sensitivity analysis, submitted to JGR, 2000.

Rabier, F, McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. and F. Bouttier, The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions, *Q. J. R. Meteorol. Soc.*, **124**, 1809-1829, 1998.

Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Review of Geophysics and Space Physics*, **14**, 609-624, 1976.

Rosenkranz, P. W., Absorption of microwave by atmospheric gases. Atmospheric remote sensing by microwave radiometry, M. A. Janssen Ed., Wiley-Interscience, 1992.

Rossow, W. B. and R. A. Schiffer, ISCCP cloud data products, *Bull. Am. Meteorol. Soc.*, **72**, 2-20, 1991.

Rossow, W. B. and L. C. Garder, Validation of ISCCP cloud detections, *Journal of Climate*, **6**, 2370-2393, 1993.

Rossow, W. B., A. W. Walker, D. E. Beuschel, and M. D. Roiter, International Satellite Cloud Climatology Project (ISCCP): Document on new cloud datasets, Natl. Aeronaut. and Space Admin., Goddard Institute for Space Stud., New York, N. Y., 1996.

Rossow, W.B., and R.A. Schiffer, Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, **80**, 11, 2261-2287, 1999.

Rumelhart, D., G. Hinton, and R. Williams, Learning representations by back-propagating error, volume I, chapter 8, 318-362, 1986.

Thépaut, J.-N., R.N. Hoffman, and P. Courtier, Interactions of Dynamics and Observations in a Four-Dimensional Variational Assimilation, *Mon. Weather Rev.*, **121**, 112, 3393-3414, 1993.

Vapnik, V., The nature of statistical learning theory, Springer-Verlag, 1997.

Table 1. Comparison of the variational and neural network inversion schemes.

	Variational method	Neural network inversion
Observation / measurement	y^o	y^o
First-guess a priori information	x^b	x^b
Retrieved variable	x	x
Direct model used	RTM, used during the inversion	RTM, used during the construction of $\mathcal{B}$, if no collocated dataset exists
Inverse model	linearized locally	non-linear, global
Model	$y(x) = y(x_n) + H(x_n)(x - x_n)$	non-linear: $x = g_W(x^b, y^o)$
Quality criterion	$\int\int x\ P(x\|y^o, x^b)$	$\frac{1}{2}\int\int\int D_E(g_W(x^b, y^o), x)^2\ P(x, y^o, x^b)$
Dataset of observations	Used to estimate the first-guess error covariance matrix B	Used to sample the pdfs
Direct model errors	Assumed to be Gaussian: with error covariance matrix F	Already sampled in the dataset, if $\mathcal{B}$ is simulated by a RT model
First-guess error	Assumed to be Gaussian: with error covariance matrix B	No constraint, simulated using true and first-guess solution datasets
Observation error	Assumed to be Gaussian: with error covariance matrix E	No constraint, depends on instrument, supposed to be Gaussian in this study, E
Inversion type	Local inversion: inversion process for each observation	Global inversion: estimation of the inverse model once and for all

Table 3. R.M.S. error results for first-guess and retrievals.

	observation or first-guess errors	NN1 clear without first-guess	NN1 clear with first-guess	NN2 cloudy without first-guess	NN2 cloudy with first-guess
TbSSMI 19 GHz V (K)	0.60	...	...	...	...
TbSSMI 19 GHz H (K)	0.60	...	...	...	...
TbSSMI 22 GHz V (K)	0.60	...	...	...	...
TbSSMI 37 GHz V (K)	0.60	...	...	...	...
TbSSMI 37 GHz H (K)	0.60	...	...	...	...
TbSSMI 85 GHz V (K)	0.60	...	...	...	...
TbSSMI 85 GHz H (K)	0.60	...	...	...	...
Ta[a](K)	3.00	...	...	...	...
Tc[b](K)	2.00	...	...	...	...
Ts[b](K)	4.00	3.47	1.34	3.31	1.57
LWP[b](kg.m-2)	...	...	...	0.09	0.08
WV[a](kg.m-2)	40 %	5.33	3.83	6.86	4.90
Em 19 GHz V	0.016	0.012	0.004	0.012	0.006
Em 19 GHz H	0.018	0.011	0.004	0.012	0.006
Em 22 GHz V	0.018	0.013	0.005	0.013	0.006
Em 37 GHz V	0.015	0.012	0.004	0.012	0.006
Em 37 GHz H	0.018	0.011	0.005	0.013	0.006
Em 85 GHz V	0.020	0.015	0.006	0.016	0.009
Em 85 GHz H	0.023	0.016	0.008	0.018	0.010

[a]NCEP.
[b]ISCCP.

Table 5. Global mean neural sensitivities.

	Tsurf	Vap-int	Em1	Em2	Em3	Em4	Em5	Em6	Em7
Tsurf	.17	-.13	-.17	-.11	-.16	-.19	-.10	-.12	-.06
Vap-int	-.04	.33	.04	.00	.04	.03	-.02	-.04	-.08
TB1	.21	.18	.58	.02	.47	.13	-.21	-.19	-.17
TB2	.14	.32	-.04	.88	-.17	-.38	.09	-.22	-.30
TB3	.09	-.78	.05	-.09	.16	-.24	-.09	-.57	-.24
TB4	.21	-.04	.17	-.30	.10	.72	.05	.50	-.03
TB5	.28	-.95	-.35	.19	-.26	.04	.79	-.22	.64
TB6	.25	-.20	-.38	-.13	-.30	-.09	-.28	.89	.04
TB7	-.21	2.30	.03	-.22	.08	-.17	-.03	-.21	.36
Em1	-.12	.06	.14	.08	.15	.15	.07	.13	.07
Em2	-.12	-.02	.13	.11	.14	.15	.10	.15	.10
Em3	-.09	.05	.11	.06	.14	.12	.06	.14	.08
Em4	-.10	.02	.11	.07	.12	.14	.08	.14	.07
Em5	-.12	-.05	.12	.10	.14	.16	.11	.16	.12
Em6	-.05	-.05	.06	.05	.08	.08	.05	.17	.11
Em7	-.05	-.15	.06	.06	.09	.09	.08	.20	.19
Tlay	-.03	.07	.00	.00	-.01	-.01	-.01	-.06	-.03

Plate 1. Retrieved fields of a) Ts in K, b) Em 19 GHz horizontal polarization, c) WV in kg/m^2, and d) LWP in kg/m^2 for June 11, 1993, from SSM/I observations with the F10 and F11 satellites.

Plate 1. Retrieved fields of a) Ts in K, b) Em 19 GHz horizontal polarization, c) WV in kg/m^2, and d) LWP in kg/m^2 for June 11, 1993, from SSM/I observations with the F10 and F11 satellites.

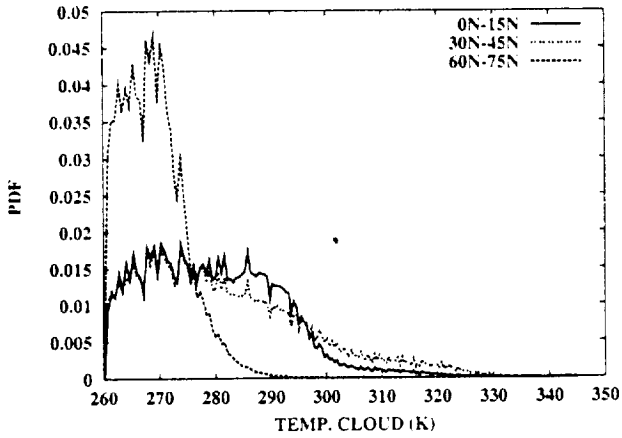


Figure 2.

Figure 2.

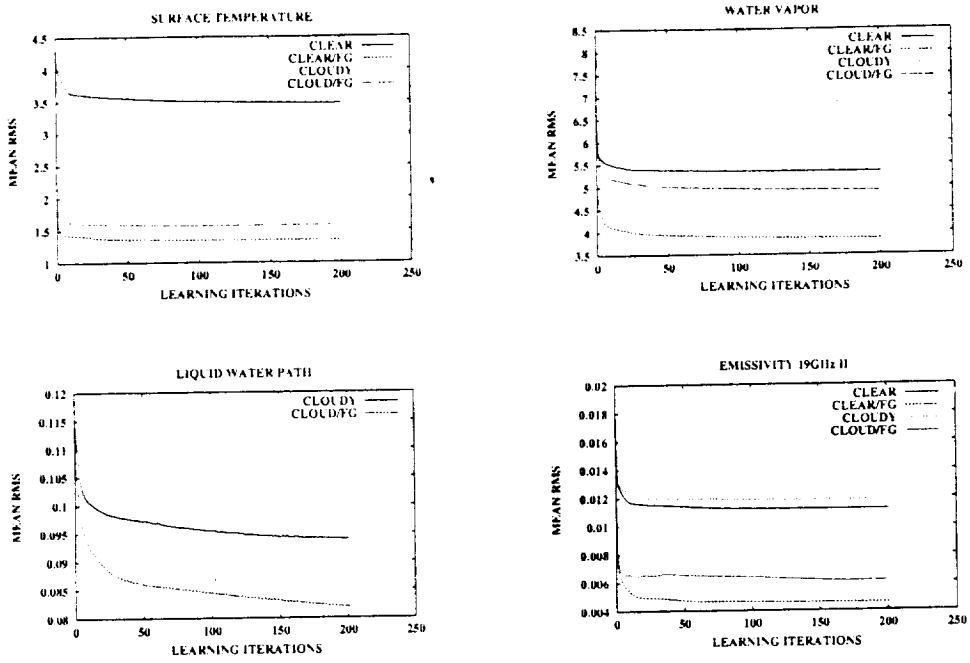
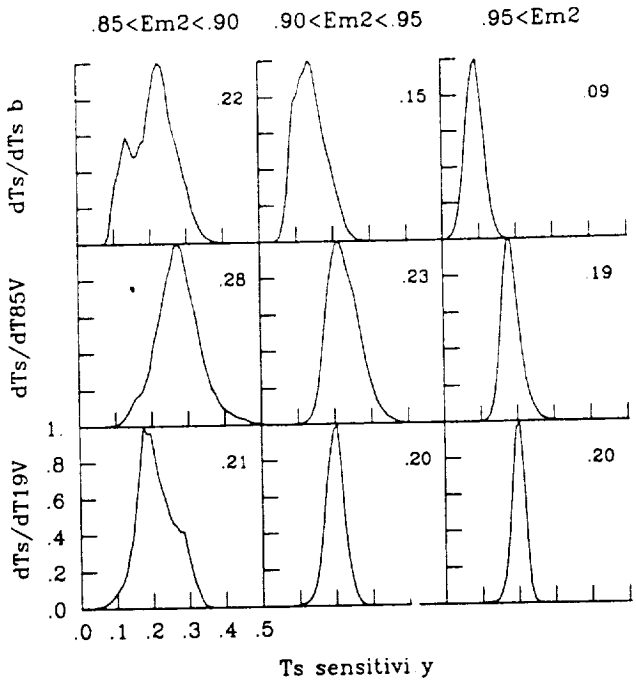
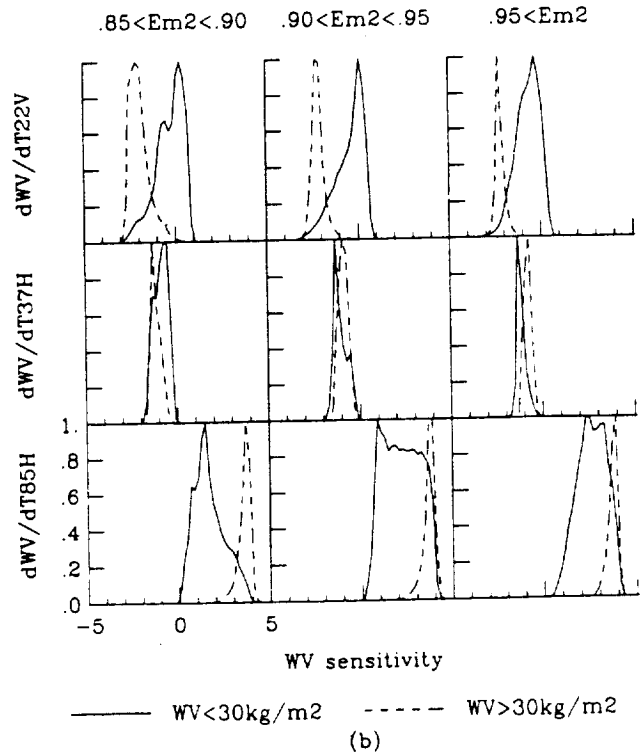


Figure 4.

Figure 4.



(a)

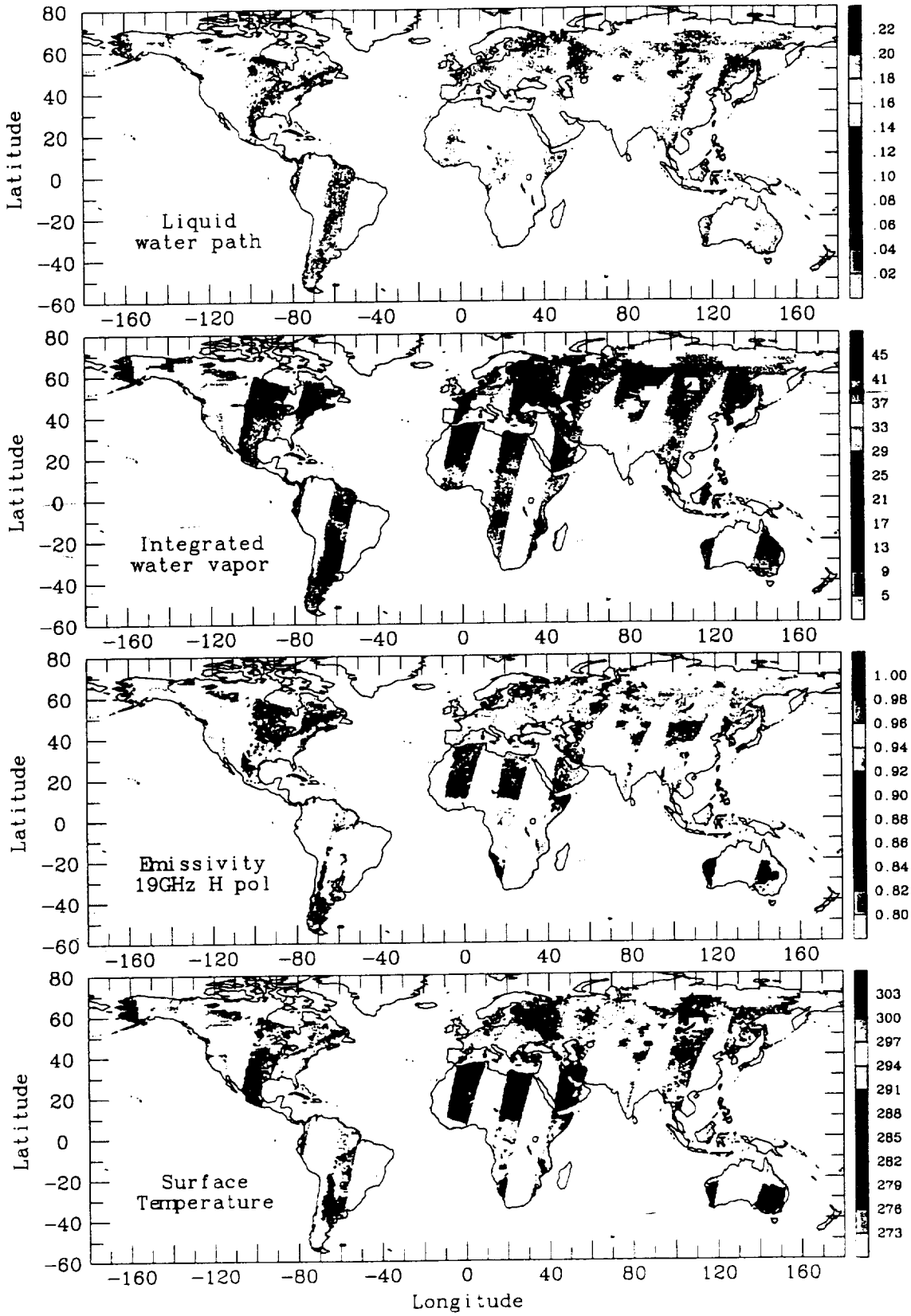


(b)

Figure 6.

Figure 6.

Plate 1



June 11, 1993