NASA Technical Memorandum 104606, Vol. 4

Technical Report Series on Global Modeling and Data Assimilation

Max J. Suarez, Editor

Volume 4

Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 1

James Pfaendtner, Stephen Bloom, David Lamich, Michael Seablom,
Meta Sienkiewicz, James Stobie, and Arlindo da Silva

January 1995

LABORATORY FOR ATMOSPHERES
LABORATORY FOR HYDROSPHERIC PROCESSES

DATA ASSIMILATION OFFICE

NASA Technical Memorandum 104606, Vol. 4

Technical Report Series on Global Modeling and Data Assimilation

Max J. Suarez, Editor
Goddard Space Flight Center
Greenbelt, Maryland

Volume 4

Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 1

James Pfaendtner
Arlindo da Silva
Goddard Space Flight Center
Greenbelt, Maryland

Stephen Bloom
David Lamich
Michael Seablom
Meta Sienkiewicz
James Stobie
General Sciences Corporation
Laurel, Maryland

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland
1995

Abstract

This report describes the analysis component of the Goddard Earth Observing System, Data Assimilation System, Version 1 (GEOS–1 DAS). The general features of the data assimilation system are outlined, followed by a thorough description of the statistical interpolation algorithm, including specification of error covariances and quality control of observations. We conclude with a discussion of the current status of development of the GEOS data assimilation system.

The main components of GEOS-1 DAS are an atmospheric general circulation model and an *Optimal Interpolation* algorithm. The system is *cycled* using the Incremental Analysis Update (IAU) technique in which analysis increments are introduced as time independent forcing terms in a forecast model integration. The system is capable of producing dynamically balanced states without the explicit use of initialization, as well as a time-continuous representation of non-observables such as precipitation and radiational fluxes. This version of the data assimilation system was used in the five-year reanalysis project completed in April 1994 by Goddard's Data Assimilation Office (Schubert et al. 1993). Data from this reanalysis are available from the Goddard Distributed Active Archive Center (DAAC) which is part of NASA's Earth Observing System Data and Information System (EOSDIS). For information on how to obtain these data sets contact the Goddard DAAC at (301) 286-3209, EMAIL daac@gsfc.nasa.gov, or consult DAO's Mosaic Home Page URL `http://hera.gsfc.nasa.gov/dao.home_page.html`.

This document is available electronically via anonymous ftp from `dao.gsfc.nasa.gov`, directory `pub/tech_memos`, file `volume_4.ps.Z`.

Contents

List of Figures

List of Tables

1 Introduction

A recent NRC panel report (National Research Council 1991) provides an overview of the issues involved in using data assimilation in climate research. In its report the panel recommends the routine generation of research–quality, model–assimilated and tested geophysical data sets to serve a broad range of national endeavors, including climate and global change research and prediction. In early 1991 the Data Assimilation Office (DAO) was formed within the NASA/Goddard Laboratory for Atmospheres (GLA) and charged with producing research–quality analyses for general Earth Science applications.

In April 1994 the DAO completed a five–year reanalysis for the years 1985 through 1989 (Schubert et al. 1993). While providing an important tool for climate research, this analysis will also serve as a baseline for further assimilation system development. The DAO's longer term goal is to develop an assimilation system capable of making optimal use of the large volume of high–quality observations expected near the turn of the century from NASA's Earth Observing System (EOS) satellites.

The main objective of this Technical Memorandum is to describe the analysis component of the system used for the five–year reanalysis mentioned above. This system has been named the Goddard EOS Version 1, Data Assimilation System, or GEOS–1 DAS. Its immediate predecessor is the optimum interpolation (OI) analysis scheme described in Baker et al. (1987), hereafter denoted B87. The atmospheric general circulation model (AGCM) component of the system is referred to as GEOS–1 GCM, and is documented in Volume 1 of this Technical Memorandum series (Takacs et al. 1994).

This report describes the analysis component of GEOS–1 DAS with particular emphasis on those features which differ significantly from B87. Sections 2–3 present a general outline of the data assimilation system and forecast model, followed by a description of the statistical interpolation algorithm (section 4), quality control of observations (section 5) and localization of the analysis problem (section 6). The specification of forecast and observational error statistics is covered in sections 7 and 8. We conclude with a brief discussion of the improvements made to GEOS DAS since its version 1. The appendix gives detailed derivations of the statistics used in the GEOS–1 DAS and provides a list of acronyms.

This document is available electronically via anonymous ftp from `dao.gsfc.nasa.gov`, directory `pub/tech_memos`, file `volume_4.ps.Z` or from DAO's Mosaic Home Page URL `http://hera.gsfc.nasa.gov/dao.home_page.html`.

2 General Features of GEOS–1 DAS

The GEOS-1 DAS analyzes global sea level pressure and near surface winds over the oceans, as well as geopotential height, vector wind, and water vapor mixing ratio on constant pressure surfaces. The upper air height/wind analyses and the sea level pressure/surface wind analyses are done using multivariate statistical interpolation algorithms in which mass (height/pressure) and wind data affect both the mass and wind analyses. The moisture analysis is done with a univariate statistical algorithm, and only at levels from 1000 hPa to 300 hPa. The basic GEOS-1 DAS configuration consists of a 2° latitude by 2.5° longitude, 14-level analysis (20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 1000 hPa) coupled to a 20-level, 2° by 2.5° model for the troposphere and lower stratosphere. Another version of the system (not used in the 5-year reanalysis) consists of a 2° by 2.5°, 18-level (surface to 0.4 hPa) analysis coupled to a 46-level, 2° by 2.5° model. This later system has been named STRATAN for stratospheric analysis (Coy *et al.* 1994 and references therein). Figure 1 shows the vertical extent and distribution of layers for both the analysis and model components of GEOS-1 DAS.

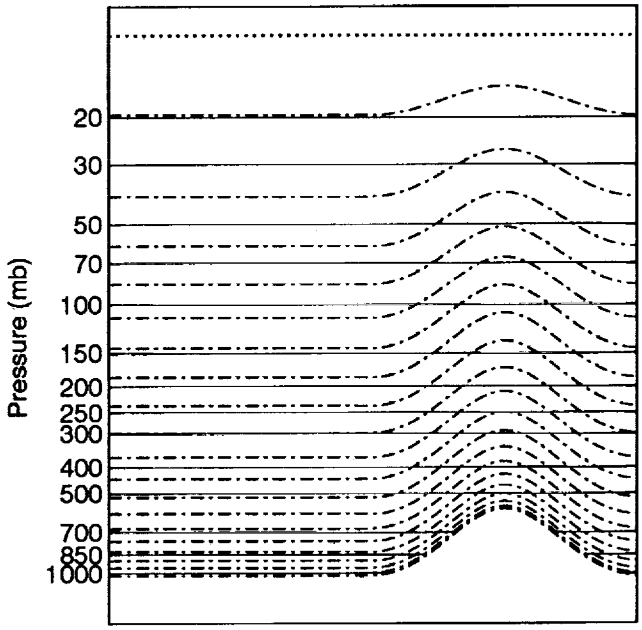

Figure 1: Vertical extent and distribution of layers for both the analysis and model components of GEOS-1 DAS. Solid lines are analysis levels and hatched lines are model sigma levels.

The analysis scheme

The GEOS-1 DAS statistical analysis scheme uses up to 75 observations to analyze all grid-points within a small three-dimensional cluster. The data selection algorithm, which chooses the observations to be used from those that have passed the quality control procedures, is an empirically tuned decision tree which uses *a priori* observation error estimates in making its choices. All observations used in the analysis have passed a two stage quality check (Seablom et al. 1991). The first stage *gross check* makes use of the assimilation's forecast error variance fields to estimate expected innovation vector variances. In a second stage a *buddy check* compares suspect observations with neighboring data by means of a successive correction analysis to the location of the suspect datum. Forecast error correlations are modeled with the damped cosine function as in B87, but the fit parameters have been recalculated using GEOS-1 DAS. Details are given in section 8 and in the Appendix.

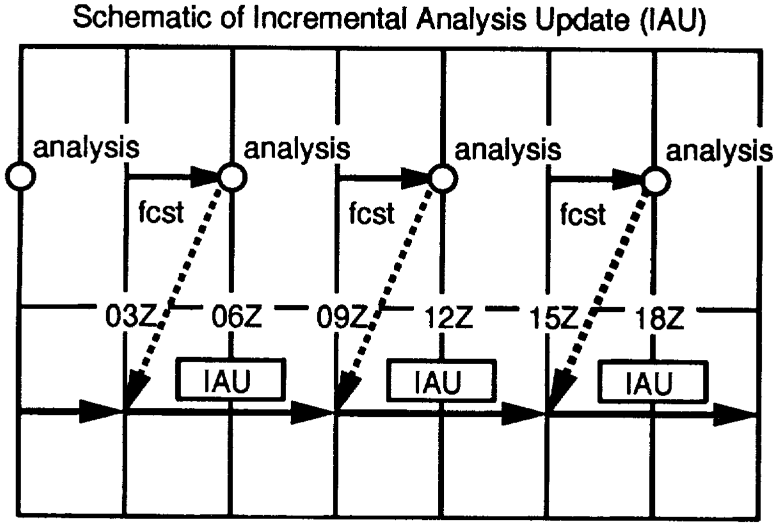

Figure 2: Schematic diagram of Incremental Analysis Update (IAU) procedure.

Incremental Analysis Updates

The assimilation scheme does not include an explicit initialization step but rather relies on the damping properties of a Matsuno time differencing scheme to control initial imbalances generated by the insertion of analysis increments. However, the initial imbalances and spin-up have been greatly reduced over earlier versions by the introduction of an incremental analysis update (IAU) procedure (Bloom et al. 1991). As shown in Fig. 2, in the IAU procedure the analysis increments are computed in a conventional way at the analysis times (00, 06, 12, 18 UTC). These increments are then inserted gradually into the model by restarting the short-term forecast that provided the analysis background and adding a fraction of the analysis increment at each model step. Over the six hour period centered on the analysis time, the full effect of the analysis increment is realized. The final assimilation product thus effectively consists of a model forecast (heavy solid line in Fig. 2) produced using additional heat, momentum, moisture and mass tendency terms which are updated every six hours from observations. This update scheme for the assimilation is similar to the way the forcing tendencies from the model's physical parameterizations and filters are recomputed intermittently (usually every 30 minutes) and gradually introduced into the ongoing integration at every time-step. An important difference between the IAU scheme and the usual Newtonian nudging procedure (Anthes 1974; Stauffer and Seaman 1990) is that the IAU forcing terms are held *constant* over the insertion period, while in Newtonian nudging they are proportional to the changing difference between a target analysis and the instantaneous current model state. Significant improvements in terms of assimilation accuracy, noise control, and the hydrologic cycle spin-up are obtained using the IAU technique.

The actual implementation of IAU in GEOS–1 DAS involves the following steps:

1. The analysis fields must be interpolated from mandatory levels to the sigma-levels required by the GEOS GCM. To minimize the adverse effects of the vertical interpolation, only the analysis increments are vertically interpolated, and then added to the original model first guess in sigma-coordinates. For details of the interpolation algorithm refer to Takacs et al. (1994).
2. The analysis variables must be converted to the GCM prognostic variables. For example, the potential temperature field is first computed from analyzed heights and mixing ratio. Because the model thermodynamic equation is formulated in flux form (Takacs et al. 1994, eq. 9), the *after analysis* potential temperature must be mass-weighted, *i.e.*, the potential temperature field must be multiplied by the factor $\Pi = p_s - p_T$, where p_s is the surface pressure and $p_T = 10$ hPa is the pressure at the top of the model.
3. The IAU forcing terms are formed by subtracting the after analysis prognostic fields from the corresponding first guess fields, and then dividing this difference by the number of seconds in 6 hours. IAU forcing fields are produced for surface pressure, wind components, mass-weighted temperature and mass-weighted moisture.
4. The model integration is restarted 3 hours before the synoptic time and continues for 6 hours with the IAU forcing held constant in time. After these 6 hours, the forcing terms are set to zero and the model integration continues for another 3 hours up to next synoptic time when the model state is used as first guess for the next analysis.

3 The GEOS-1 Model

This section presents a summary of the main features of the GEOS-1 GCM. This model is fully documented by Takacs et al. (1994) which is available electronically via anonymous ftp from `dao.gsfc.nasa.gov`, directory `pub/tech_memos`, file `volume_1.ps.Z` or from DAO's Mosaic Home Page URL `http://hera.gsfc.nasa.gov/dao.home_page.html`.

The GEOS-1 GCM uses the second-order potential enstrophy and energy conserving horizontal differencing scheme on a C–grid developed by Sadourney (1975), and further described by Burridge and Haseler (1977). An 8th-order Shapiro filter with a reduced coefficient is applied to the wind, potential temperature and specific humidity to avoid non–linear computational instability. The reduced–coefficient filter is applied at every step in such a way that the amplitude of the two–grid interval wave is essentially removed in six hours. Applying the filter weakly at each step way eliminates the shock that occurred in earlier assimilations using an intermittent application of the filter. The model also uses a polar Fourier filter to avoid linear instability due to violation of the CFL condition for the Lamb wave and interval gravity waves. This polar filter is applied only to the *tendencies* of the winds, potential temperature, specific humidity and surface pressure. The model's vertical finite differencing scheme is that of Arakawa and Suarez (1983). The dynamics routines are organized into a plug–compatible module called the "ARIES/GEOS dynamical core" which is described in Suarez and Takacs (1995).

The infrared and solar radiation parameterizations follow closely those described by Harshvardhan et al. (1987). In the longwave, water vapor absorption is parameterized as in Chou (1984), the 15 micron band of CO_2 as in Chou and Peng (1983), and ozone absorption as in Rogers (1968) with modifications suggested by Rosenfield et al. (1987). The shortwave follows Davies (1982), as described in Harshvardhan et al. (1987). Shortwave absorption by water vapor uses a k–distribution approach as in Lacis and Hansen (1974). Cloud albedo and transmissivity for the model layers are obtained from specified single–scattering albedo and cloud optical thickness using the delta–Eddington approximation (Joseph et al. 1976; King and Harshvardhan 1986).

The penetrative convection originating in the boundary layer is parameterized using the Relaxed Arakawa–Schubert (RAS) scheme (Moorthi and Suarez 1992) which is a simplified Arakawa–Schubert (1974) type scheme. As an approximation to the full interaction between the different allowable cloud types in the original AS scheme, many clouds are simulated frequently with each modifying the large scale environment some fraction of the total adjustment. The parameterization thereby "relaxes" the large scale environment towards neutrality. In addition to the RAS cumulus convection scheme, the GEOS-1 GCM employs a Kessler-type scheme for the re-evaporation of falling rain (Sud and Molod, 1988). The scheme accounts for the rainfall intensity, the drop size distribution, and the temperature, pressure and relative humidity of the surrounding air.

Super-saturation or large-scale convection is defined in the GEOS-1 GCM whenever the specific humidity in any grid-box exceeds its super-saturation value. The large-scale precipitation scheme rains at super-saturation, and re-evaporates during descent to partially saturate lower layers in a process that accounts for some simple micro-physics.

The GEOS-1 GCM turbulence parameterization consists of elements which handle vertical diffusion (Helfand and Labraga, 1988) and surface fluxes of heat, moisture and momentum (Helfand, et al, 1991, and Helfand and Schubert, 1994). The vertical regime is divided into a free atmosphere, a surface layer, and a viscous sub-layer above the surface roughness elements. The turbulent eddy fluxes are calculated using a variety of methods depending on the vertical location in the atmosphere.

Turbulent eddy fluxes of momentum, heat and moisture in the surface layer are calculated using stability-dependent bulk formulae based on Monin-Obukhov similarity functions. For an unstable surface layer, the chosen stability functions are the KEYPS function (Panofsky, 1973) for momentum, and its generalization for heat and moisture. The function for heat and moisture assures non-vanishing heat and moisture fluxes as the wind speed approaches zero. For a stable surface layer, the stability functions are those of Clarke (1970), slightly modified for the momentum flux. The moisture flux also depends on a specified evapotranspiration coefficient.

Above the surface layer, turbulent fluxes of momentum, heat and moisture are calculated by the Level 2.5 Mellor-Yamada type closure scheme of Helfand and Labraga (1988), which predicts turbulent kinetic energy and determines the eddy transfer coefficients used for vertical diffusion.

4 The Analysis Scheme

As mentioned above, the forcing terms for the IAU assimilation are obtained using a standard statistical analysis methodology. A short term 6–hour forecast to the analysis time, t_k (= 00, 06, 12, or 18 UTC) is used as the background, $\mathbf{w}_k^f$ (refer to Fig. 2). Using all observations, $\mathbf{w}_k^o$, falling within a 6–hour data window centered on the analysis time, the analysis, $\mathbf{w}_k^a$, is defined as the background forecast plus a correction expressed as a quasi-linear combination of the differences between the observations and *a priori* estimates of the observations computed from the background forecast,

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathcal{K}_k(\mathbf{w}_k^o - \mathcal{H}_k\mathbf{w}_k^f). \tag{1}$$

In GEOS–1 DAS the "forward problem" transformation $\mathcal{H}_k$ is factored as $\mathcal{H}_k = H_k\mathcal{S}$ where $\mathcal{S}$ transforms from the model's variables (temperature, wind, specific humidity, surface pressure) and sigma coordinates to the analysis variables (geopotential height, wind, water vapor mixing ratio, sea level pressure) and pressure coordinates. H_k is the multi-linear interpolation of the pressure-level background vector $\tilde{\mathbf{w}}_k^f \equiv \mathcal{S}(\mathbf{w}_k^f)$ to our best *a priori* estimate of $\mathbf{w}_k^o$. Consistent with the factorization of $\mathcal{H}$, the weights $\mathcal{K}_k$ also factor, $\mathcal{K}_k = \mathcal{S}^+ K_k$, where K_k is the usual gain matrix (see eq. 5) used to define the increments for the analysis variables on the constant pressure analysis levels, and $\mathcal{S}^+$ is the transform of these increments back to the model's variables and sigma coordinates. The notation here follows that of Cohn and Parrish (1991).

The forcing vector ξ_k used to modify the model's tendencies during the IAU assimilation from time $t_k - 3$ to time $t_k + 3$ is held constant during that 6–hour interval and is obtained by transforming the pressure-level analysis increment back to the model's sigma coordinates and variables (see Takacs et al. 1994 for details of the vertical interpolation). Thus ξ_k is defined by

$$\xi_k = \frac{1}{N\Delta t}\mathcal{S}^+ K_k(\mathbf{w}_k^o - H_k\tilde{\mathbf{w}}_k^f), \tag{2}$$

where N is the number of model time-steps of length Δt seconds in the 6–hour assimilation update period centered on time t_k. As mentioned above, $\mathcal{S}^+$ represents the pressure to sigma transform for analysis increments. Since there are more model sigma levels than analysis pressure levels, the transform $\mathcal{S}$ is not actually invertible. Hence, $\mathcal{S}^+$ only denotes an approximate inverse for $\mathcal{S}$. It does, however, act in such a way that $\xi_k = 0$ in the interior of any region where $K_k(\mathbf{w}_k^o - H_k\tilde{\mathbf{w}}_k^f) = 0$, *i.e.*, where the analysis increment is zero because there are no nearby data, the relevant components of ξ_k will also be zero.

By minimizing the variance of the analysis error, $\tilde{\epsilon}_k^a \equiv \mathcal{S}(\mathbf{w}_k^t - \mathbf{w}_k^a)$, we are led to a system of linear equations which depends on the expected error characteristics of the background and observations and has the gain matrix K_k as its solution (cf. Jazwinski 1970, Example 7.5),

$$K_k = \left[\tilde{P}_k^f H_k^T + \tilde{J}_k\right]\left[H_k\tilde{P}_k^f H_k^T + R_k + (H_k\tilde{J}_k + \tilde{J}_k^T H_k^T)\right]^{-1}. \tag{3}$$

Using K_k from (3) in (1), we also obtain an equation for the analysis error variance,

$$\operatorname{diag}\{\tilde{P}_k^a\} = \operatorname{diag}\{(I - K_kH_k)\tilde{P}_k^f - K_k\tilde{J}_k^T\}. \tag{4}$$

In (3) and (4), $\tilde{P}_k^a = \left\langle \tilde{\epsilon}_k^a \tilde{\epsilon}_k^{aT} \right\rangle = \mathcal{S} P_k^a \mathcal{S}^T$ is the pressure–coordinate analysis error covariance matrix, $\tilde{P}_k^f = \left\langle \tilde{\epsilon}_k^f \tilde{\epsilon}_k^{fT} \right\rangle = \mathcal{S} P_k^f \mathcal{S}^T$ is the pressure–coordinate background error covariance matrix, $R_k = \left\langle \epsilon_k^o \epsilon_k^{oT} \right\rangle$ is the observation error covariance matrix, and $\tilde{J}_k = \left\langle \tilde{\epsilon}_k^f \epsilon_k^{oT} \right\rangle$ is the pressure–coordinate cross–covariance matrix of background and observation errors. The background error is defined by $\tilde{\epsilon}_k^f \equiv \mathcal{S}(\mathbf{w}_k^t - \mathbf{w}_k^f)$, and the observation error is defined by $\epsilon_k^o \equiv \mathbf{w}_k^o - H_k \tilde{\mathbf{w}}_k^t$, where $\tilde{\mathbf{w}}_k^t \equiv \mathcal{S} \mathbf{w}_k^t$ is the unknown true state projected onto the analysis grid. We have made the assumption that both the background and observation errors are unbiased, *i.e.*, that $\left\langle \tilde{\epsilon}_k^f \right\rangle = \langle \epsilon_k^o \rangle = 0$. The standard assumption will also be made that $\tilde{\epsilon}_k^f$ and ϵ_k^o are uncorrelated, so that $\tilde{J}_k = 0$ (see Daley 1991, for the more general case). Given this last assumption (3) and (4) simplify to

$$K_k = \left[\tilde{P}_k^f H_k^T \right] \left[H_k \tilde{P}_k^f H_k^T + R_k \right]^{-1}, \tag{5}$$

and

$$\mathrm{diag}\{\tilde{P}_k^a\} = \mathrm{diag}\{(I - K_k H_k) \tilde{P}_k^f\}. \tag{6}$$

The computational flow of the assimilation is as follows:

Step 1. The innovation vector, $[\mathbf{w}_k^o - H_k \tilde{\mathbf{w}}_k^f]$, is computed, and the quality control of the observations is performed. See section 5 for details.

Step 2. The components of $\tilde{\mathbf{w}}_k^a$ are partitioned into small geographically localized subsets of related variables. For example, one of these subsets might contain the height and wind components for the analysis grid-points in a small latitude-longitude-pressure box. For each of these subsets of $\tilde{\mathbf{w}}_k^a$ components, a local data selection is made from the quality checked innovation vector. See section 6 for details of this partitioning and the local data selection algorithm.

Step 3. The innovation covariance matrix, $[H_k \tilde{P}_k^f H_k^T + R_k]$, is approximated locally using the types and locations of the selected data. Equation (5) is then solved for those observation weights in K_k that give the contributions of the selected data to the $\tilde{\mathbf{w}}_k^a$ components being computed. Sections 7 and 8 contain details of the approximations for R_k and $H_k \tilde{P}_k^f H_k^T$, respectively.

Step 4. The weights from K_k computed in step 3 are also used in (6) to estimate the analysis error variance components in $\mathrm{diag}\{\tilde{P}_k^a\}$.

Step 5. The forcing vector ξ_k is computed as in (2) and passed to the model for use in the assimilating IAU integration, which starts at time $t_k - 3$ hours and extends to time $t_k + 3$ hours. At $t_k + 3$ hours, ξ_k is discarded and the integration is continued without analysis forcing to t_{k+1} to produce the background, $\mathbf{w}_{k+1}^f$, for use in the next analysis. The forecast error variances, $\mathrm{diag}\{\tilde{P}_{k+1}^f\}$, which are also needed for the next analysis, are estimated using an error growth model together with the analysis error variance estimates for time t_k that were computed in step 4. See section 8 for details of the forecast error growth model.

These steps are further described in the following sections.

5 Quality Control of Observations

5.1 Pre-analysis Data Checks

The GEOS-1 DAS ingests the global conventional observations and the temperature retrievals from the HIRS2/MSU/SSU sounders on the NOAA satellites. Either the NESDIS retrieved temperature profiles or those created by the GLA physical retrieval system are used. Conventional data in *NMC's Office Note 29*, format as well as the NESDIS format retrieval data are unpacked and put into common format data sets of one day each. A *day* corresponds to the four analysis times: 0, 6, 12 and 18Z. Thus, the observations in one file will be from 2100Z of the previous day to 2059Z of the current day.

The unpacking process keeps all data that appear in the original data sets, except those that do not have a realistic time stamp. Reports that are obviously in the wrong synoptic time (late arriving data, for example) are moved to the correct file. These data sets form a complete set of historical observations that are easily manageable on the computer system.

A second preprocessing step standardizes the observations for ingest into the OI. Only observations of quantities to be analyzed are extracted for these data sets: sea level pressure and wind, upper-air height, wind and moisture. The satellite temperature retrievals are converted into thicknesses. The observations are stratified by type (e.g., surface land, surface ship) and in some instances by location. This is necessary to assign the proper observation errors to the data. More stringent checks on the data are performed:

1. Quality marks provided with the observations are used to eliminate bad data outright. Data marked as being of suspicious quality are kept along with the quality flag.
2. Observations with grossly bad values are deleted.
3. A hydrostatic check is performed on rawinsonde data.
4. Satellite profiles are checked for completeness.

Before the observation data sets are ingested by the analysis scheme, a program is run to detect gaps in the observation time series. This program categorizes each observation by type, synoptic time and in the case of retrieved satellite temperature profiles, by the vertical extent of the profile (*stratospheric* and *tropospheric* profiles in the case of NESDIS TOVS data). These counts are then graphically displayed as a series of bar charts as in Fig. 3. In this figure, the number of NESDIS TOVS path A (clear column) temperature retrievals for each synoptic time in the month of August 1985 is shown. The program has identified a significant gap in the TOVS data record from 11 to 17 August. In this case, it is likely that the data were not received properly from the primary archive center. This particular gap was not filled for the reanalysis, but usually such gaps have been filled with additional data received through secondary sources. These charts also serve as a record of the observational data that are available to the analysis at a particular time and are made available to the users.

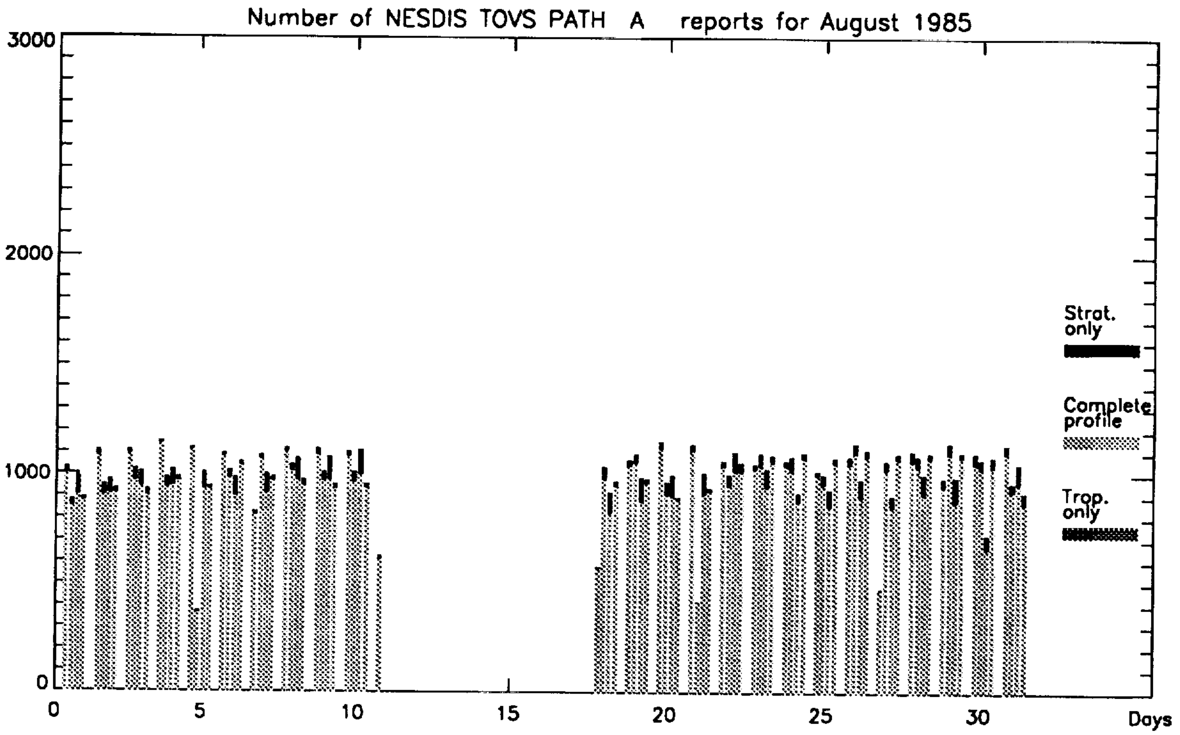

Figure 3: Number of NESDIS TOVS path A (clear column) temperature retrievals for each synoptic time in the month of August 1985. *Stratospheric* retrieval profiles are those that only have levels above 100 hPa; *tropospheric* retrieval profiles are those that only have levels below 100 hPa. *Complete* profiles report at all levels.

5.2 Analysis Data Checks

The data quality control is an important part of any data analysis scheme, as demonstrated in Shaw et al. (1986). More recent studies performed with the GEOS–1 DAS confirm that changes in quality control and subsequent data selection can have a major impact on the assimilation. The quality control technique currently employed in the GEOS–1 DAS consists of two major steps: a *gross* error check and a *buddy* check (Seablom, 1990). The *gross* error check is defined to be:

$$\Delta^2 \leq \left((\sigma^o)^2 + (\sigma^f)^2\right)\tau, \tag{7}$$

where Δ is the difference between an observation and the interpolated background first-guess value, $(\sigma^o)^2$ and $(\sigma^f)^2$ are the observation and forecast error variances, respectively, and τ is a subjectively defined tolerance value which varies with quantity, latitude and height.

The tolerance value is somewhat reduced for all quantities in the tropics and increased slightly for the winds near jet level in the middle latitudes. Those data that fail to satisfy (7) are marked as *suspect.* The *buddy* check involves performing a single pass successive-correction analysis of the data that passed the *gross* check to the locations of the suspect data. The difference between the interpolated value and the suspect value is then compared to the error statistics as in (7) and a decision is then made to reaccept the observation or to reject it. Typical rejection rates are between 5 and 10%.

6 Localization of the Analysis Problem

The GEOS–1 DAS global analyses are performed as a series of localized analyses on smaller regions referred to as *mini-volumes.* These mini-volumes are a set of non-overlapping groups of analysis grid-points. Associated with each mini-volume is an approximately cylindrical *search region*, having 3200 km diameter, from which the data are selected for the generation of the analysis in the mini-volume. The GEOS–1 DAS analysis has three distinct types of mini-volumes, each containing a different number of horizontal grid-points. In the tropics and low latitudes, each mini-volume contains six horizontal grid-points, rectangularly arranged. The mid-latitude region, between 30 degrees and 82 degrees latitude, contains mini-volumes with eight horizontal grid-points, while the polar regions place an entire latitude band of grid-points into each mini-volume. Additionally, each mini-volume contains two vertical layers of grid-points. The total analysis contains nearly 12,000 mini-volumes.

A data search within each *search region* selects the 75 observations closest to the volume midpoint, with at most 60% of these from rawinsondes. The covariance matrix is then formed and the linear system of equations in (5) is solved for the weights of the observations K_k. With the current configuration of mini-volumes, neighboring search regions have an approximate 85% overlap. Because only 75 observations are selected, however, the actual overlap could be substantially less.

The mini-volume approach to performing the global analysis is preferable to a single grid-point approach for two reasons. First, solving a local problem centered on a region rather than on a particular grid-point significantly reduces the number of covariance matrices that need be set up and solved. The data search, also computationally expensive, is minimized by avoiding redundant calculations. Second, the independent nature of the mini-volumes allows the GEOS-1 DAS analyses to operate well in a parallel computing environment. The major drawback is that the limitation of 75 observations per mini-volume reduces the search region overlap, particularly in areas of high data density, leading to the adverse effect of *boxiness* in the analysis increments.

7 Specification of Observation Error Covariances

The elements of $R_k = R = [r_{i,j}]$ are the observation–observation error covariances, *i.e.* $r_{i,j} = \left\langle e_i^o e_j^{oT} \right\rangle = r_{j,i}$. The time index k will now be omitted, and subscripts will be used to denote vector and matrix elements. We write $r_{i,j} = \sigma_i^o \sigma_j^o \mu_{i,j}^o$, where $\sigma_i^{o2} = \left\langle e_i^{o2} \right\rangle$, $\sigma_j^{o2} = \left\langle e_j^{o2} \right\rangle$, and $\mu_{i,j}^o$ is the observation-observation error correlation. In the GEOS-1 DAS implementation σ_i^o and σ_j^o are obtained from tabulated values which vary with instrument type and level. Vertical interpolation using the tabulated values is done as needed. The observation error standard deviations for the instrument types being used in the DAO's baseline five–year assimilation are shown in Table 1.

The error standard deviations shown in Table 1 are in general larger than the standard errors in the measurements. The tabulated values contain an error of representativeness in addition to the instrument error. If $\mathbf{w}_i^o(true)$ is the theoretical result of a perfect measurement corresponding to the real measurement $\mathbf{w}_i^o$, then $[\mathbf{w}_i^o - \mathbf{w}_i^o(true)]$ is the measurement or instrument error. For the analysis, we have defined the corresponding observation error by

$$e^o = \mathbf{w}_i^o - (H\tilde{\mathbf{w}}^t)_i = [\mathbf{w}_i^o - \mathbf{w}_i^o(true)] + \left[\mathbf{w}_i^o(true) - (H\tilde{\mathbf{w}}^t)_i\right]. \qquad (8)$$

The last term in brackets in (8) is what is called the error of representativeness (Lorenc 1986). It depends on such things as grid resolution, subgrid–scale variability, footprint size for satellite instruments, and importantly on errors in the formulation of the "forward problem" operator H. These effects are not explicitly modeled in our system. The values in Table 1 have simply been inflated to account in a crude way for the representativeness error term. It should also be noted that observation error correlations for satellite soundings are probably in large part a reflection of the correlation between the representativeness errors for the observations.

In general, the correlation, μ_{ij}^o, between e_i^o and e_j^o is set to δ_{ij}, the Kronecker delta, so that errors from distinct observations are assumed uncorrelated. However, in two situations μ_{ij}^o is modeled as the product of an isotropic horizontal correlation, $\mu^{o,h}(s_{ij})$, which depends only on the horizontal distance, s_{ij}, separating the observations, and a vertical correlation, $\nu^{o,v}$, which depends only on their vertical separation. The first case occurs when the data are two height observations from the same radiosonde ascent. Here $\mu^{o,h}(s_{ij}) = 1$ and $\nu^{o,v}$ is similar to the function from Lönnberg and Hollingsworth (1986). Our current $\nu^{o,v}$ has been extended in its vertical extent and its eigenstructure has been slightly adjusted, as shown in Table 2. The second case occurs when the data are two height observations from

Table 1: Observation error standard deviations for the instrument types being used in the DAO's baseline five-year reanalysis. Here z is in m, u and v in m s^{-1}, w in g kg^{-1} and p_{sl} in hPa.

		Pressure levels (hPa)														
Observing System		sea level	1000	850	700	500	400	300	250	200	150	100	70	50	30	20
Rawinsondes	(z)	—	8.0	8.2	10.6	14.6	18.0	23.8	26.3	28.1	30.9	41.0	49.8	51.0	60.0	70.0
	(u, v)	—	2.2	2.6	3.2	4.0	6.0	8.0	9.0	10.0	9.5	9.0	8.0	7.0	8.6	9.7
	(w)	—	1.7	1.3	0.9	0.3	0.2	0.1	—	—	—	—	—	—	—	—
Dropwindsondes	(z)	—	8.0	8.2	10.9	14.6	18.0	23.8	26.3	28.1	30.9	41.0	49.8	48.7	—	—
	(u, v)	—	2.2	2.6	3.2	4.0	6.0	8.0	9.0	10.0	9.5	9.0	8.0	7.0	4.1	4.1
	(w)	—	1.7	1.3	0.9	0.3	0.2	0.1	—	—	—	—	—	—	—	—
Rocketsondes	(z)	—	7.5	10.4	13.2	14.6	16.8	18.0	18.5	20.6	26.0	38.7	45.4	53.5	53.5	177.0
	(u, v)	—	5.2	6.3	5.6	7.7	7.8	10.7	10.6	11.8	10.7	8.8	7.5	9.0	9.5	1.1
NAVAIDS	(z)	—	8.0	8.2	10.9	14.6	18.0	23.8	26.3	28.1	30.9	41.0	49.8	48.7	—	—
	(u, v)	—	2.2	2.6	3.2	4.0	6.0	8.0	9.0	10.0	9.5	9.0	8.0	7.0	4.1	4.1
	(w)	—	1.7	1.3	0.9	0.3	0.2	0.1	—	—	—	—	—	—	—	—
Pilot balloons	(u, v)	—	4.4	5.2	7.2	9.0	11.2	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4
ASDAR/AIDS	(u, v)	—	2.2	2.6	3.2	4.0	6.0	8.0	9.0	10.0	9.5	9.0	8.0	7.0	8.6	9.7
Cloud track winds	(u, v)	—	6.0	6.0	6.0	6.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0
TOVS A	(z)	—	20.8	25.9	32.1	42.8	49.9	59.0	64.8	71.9	81.1	93.9	105.3	116.0	132.2	167.1
TOVS B	(z)	—	30.6	35.6	41.7	52.1	59.0	67.9	73.5	80.4	89.3	101.8	112.9	123.3	139.1	173.1
TOVS C	(z)	—	48.4	53.2	58.9	68.9	75.4	84.0	89.3	95.9	104.4	116.4	127.0	136.9	152.0	184.5
Surface stations (land)	(p_{sl})	1.9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Drifting buoys	(p_{sl})	1.9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	(u, v)	2.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Environmental buoys	(p_{sl})	1.9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	(u, v)	2.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Ships	(p_{sl})	1.9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	(u, v)	2.8	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Table 2: The vertical correlation of observation errors for two geopotential height observations from the same radiosonde ascent.

	Pressure level (hPa)													
level	20	30	50	70	100	150	200	250	300	400	500	700	850	1000
20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	0.00	0.00	1.00	0.89	0.88	0.84	0.81	0.73	0.70	0.53	0.43	0.24	0.01	0.00
70	0.00	0.00	0.89	1.00	0.89	0.86	0.83	0.75	0.71	0.56	0.46	0.30	0.04	0.01
100	0.00	0.00	0.88	0.89	1.00	0.88	0.86	0.78	0.73	0.59	0.51	0.36	0.10	0.03
150	0.00	0.00	0.84	0.86	0.88	1.00	0.88	0.81	0.76	0.62	0.54	0.45	0.23	0.07
200	0.00	0.00	0.81	0.83	0.86	0.88	1.00	0.84	0.76	0.61	0.53	0.46	0.24	0.08
250	0.00	0.00	0.73	0.75	0.78	0.81	0.84	1.00	0.84	0.71	0.57	0.40	0.14	0.03
300	0.00	0.00	0.70	0.71	0.73	0.76	0.76	0.84	1.00	0.78	0.61	0.39	0.13	0.05
400	0.00	0.00	0.53	0.56	0.59	0.62	0.61	0.71	0.78	1.00	0.78	0.54	0.14	0.09
500	0.00	0.00	0.43	0.46	0.51	0.54	0.53	0.57	0.61	0.78	1.00	0.70	0.24	0.15
700	0.00	0.00	0.24	0.30	0.36	0.45	0.46	0.40	0.39	0.54	0.70	1.00	0.64	0.36
850	0.00	0.00	0.01	0.04	0.10	0.23	0.24	0.14	0.13	0.14	0.24	0.64	1.00	0.48
1000	0.00	0.00	0.00	0.01	0.03	0.07	0.08	0.03	0.05	0.09	0.15	0.36	0.48	1.00

the same TOVS sounding or from different soundings of the same type. In this case, the vertical correlation (see Table 3) is also a slight modification of that used in B87. The horizontal observation error correlation between NESDIS TOVS soundings of the same type is modeled by $\mu^{o,h}(s_{ij}) = \exp\left[-(s_{ij}/150.0)^2\right]$, where s_{ij} is the separation in km. This function is considerably sharper than the functions which appeared in B87. If the soundings are of different types, then $\mu^o_{ij} = \delta_{ij}$ is used.

8 Specification of Background Error Covariances

The central feature of Kalman filter theory (Jazwinski 1970, Daley 1991, Cohn 1993, and references therein) is its inclusion of a prediction equation for forecast error covariances,

$$P^f_k = A_{k-1} P^a_{k-1} A^T_{k-1} + Q_{k-1}, \tag{9}$$

where A_{k-1} represents a linearization of the atmospheric prediction model and Q_{k-1} is the covariance matrix of model errors, *i.e.*, $Q_{k-1} = \left\langle e^t_{k-1} e^{t\,T}_{k-1} \right\rangle$ with $\mathbf{w}^t_k = A_{k-1}\mathbf{w}^t_{k-1} + e^t_{k-1}$. Optimum interpolation can be viewed as an approximation to the Kalman filter in which (9) is replaced by an empirical model for the evolution of P^f_k (see Todling and Cohn 1994).

Note that (9) gives some guidance in the design of such empirical models in that it shows that P^f_k depends upon: (1) the recent history (contained in P^a_{k-1}) of the distribution and quality of assimilated observations; (2) the model error covariance (contained in Q_{k-1}); and (3) the dynamical evolution of existing errors (governed by A_{k-1}). In the GEOS-1 DAS implementation of OI, the last of these three factors is ignored and the first two are accounted for only in a crude way. Writing the elements of $\tilde{P}^f_k = \left[\tilde{p}^f_{ij}\right]$ as $\tilde{p}^f_{ij} = \sigma^f_i \sigma^f_j \mu^f_{ij}$, we separately model the error standard deviations, σ^f_i and σ^f_j, for the ith and jth components of the background error vector, and the correlation, μ^f_{ij}, between these error components.

Table 3: Vertical correlations of observation errors for two geopotential height observations from NESDIS TOVS retrievals of the same type. The correlation function from the upper triangular part of the table is used for type A (clear) and type B (partially cloudy) retrievals. The function from the lower triangular part of the table is used for the type C (cloudy) microwave retrievals.

	Pressure level (hPa)													
level	20	30	50	70	100	150	200	250	300	400	500	700	850	1000
20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	0.00	0.00	1.00	0.87	0.82	0.67	0.57	0.45	0.31	0.28	0.26	0.22	0.12	0.00
70	0.00	0.00	0.87	1.00	0.83	0.70	0.58	0.45	0.30	0.26	0.23	0.17	0.06	0.00
100	0.00	0.00	0.81	0.85	1.00	0.83	0.70	0.57	0.43	0.38	0.35	0.25	0.16	0.00
150	0.00	0.00	0.68	0.73	0.83	1.00	0.81	0.71	0.59	0.53	0.48	0.33	0.22	0.00
200	0.00	0.00	0.56	0.61	0.73	0.84	1.00	0.85	0.77	0.65	0.55	0.28	0.15	0.00
250	0.00	0.00	0.46	0.50	0.63	0.75	0.85	1.00	0.86	0.77	0.65	0.32	0.16	0.00
300	0.00	0.00	0.35	0.39	0.53	0.65	0.79	0.86	1.00	0.84	0.73	0.37	0.19	0.00
400	0.00	0.00	0.34	0.36	0.47	0.55	0.67	0.77	0.84	1.00	0.85	0.56	0.36	0.00
500	0.00	0.00	0.29	0.31	0.50	0.47	0.55	0.64	0.72	0.85	1.00	0.72	0.52	0.00
700	0.00	0.00	0.24	0.24	0.30	0.31	0.32	0.38	0.47	0.67	0.81	1.00	0.82	0.00
850	0.00	0.00	0.16	0.12	0.19	0.20	0.17	0.20	0.29	0.51	0.67	0.83	1.00	0.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

8.1 Background Error Variances

The background error standard deviations σ_i^f for geopotential height, sea level pressure, and water vapor mixing ratio are estimated in a two-step process. First, the analysis error standard deviations, σ_i^a, for these quantities at time t_{k-1} (square roots of elements along the diagonal of $\tilde{P}_{k-1}^a$) are computed using a localized version of (6). A growth term is then added to simulate the effect of Q_{k-1} on the diagonal of $\tilde{P}_k^f$ (Pfaendtner and Sivakumaran, 1991). The equation used for σ_i^f at time t_k can be written

$$\sigma_i^f = \sigma_i^a + \boxed{\text{growth term}} = \sigma_i^a + \alpha\Big\{1 - \frac{\sigma_i^a}{\beta}\Big\}. \tag{10}$$

The parameters α and β used in the definition of the growth term, which determine the growth rate and saturation values for the errors, respectively, depend on the pressure level. For geopotential height and sea level pressure there is an additional crude dependence on latitude. This latitudinal dependence is defined using tropical, α_{tr} and β_{tr}, and extra-tropical, α_{xt} and β_{xt}, parameter values with a smooth transition between 15° and 45° north and south latitude. For example, for β we have:

$$\beta = \beta(\varphi) = \begin{cases} \beta_{tr}, & |\varphi| \le 15^o \\ \frac{1}{2}(\beta_{tr} + \beta_{xt}) + \frac{1}{2}(\beta_{tr} - \beta_{xt})\cos(\frac{|\varphi|-15}{20}\pi), & 15^o < |\varphi| < 45^o \\ \beta_{xt}, & 45^o \ge |\varphi|\,. \end{cases} \tag{11}$$

The form of this functional dependence on latitude was motivated by an examination of the growth of zonally averaged RMS errors from an ensemble of short-term (to 48 hr.) forecasts. The parameters (see Table 4) have been empirically adjusted by comparing the mean square innovation vector components, $\left\langle \left[\mathbf{w}_i^o - (H\tilde{\mathbf{w}}^f)_i\right]^2 \right\rangle$, to the OI estimate $\langle\sigma_i^o\rangle^2 + \left\langle (H\sigma^f)_i \right\rangle^2$.

An example of the effectiveness of the error growth model is shown in Fig. 4. The figure shows a time series of 500 hPa of forecast error standard deviation estimated from the innovations and observation error standard deviations by the formula

$$\sigma^f \approx \left\langle (\mathbf{w}^o - H\mathbf{w}^f)_i^2 - (\sigma^o)^2 \right\rangle^{1/2}, \tag{12}$$

compared to the forecast error standard deviation produced using the growth model, eq. (10). This figure shows modelled and observed σ^f for a box in the South Pacific (165W–105W,60S–36S), from July 7 to August 4, 1979. Although the magnitude of the modelled σ^f is slightly larger than observed, both curves are highly correlated after July 14. Therefore, the empirical growth model seems to capture the main characteristics of the temporal variations of the forecast error standard deviations.

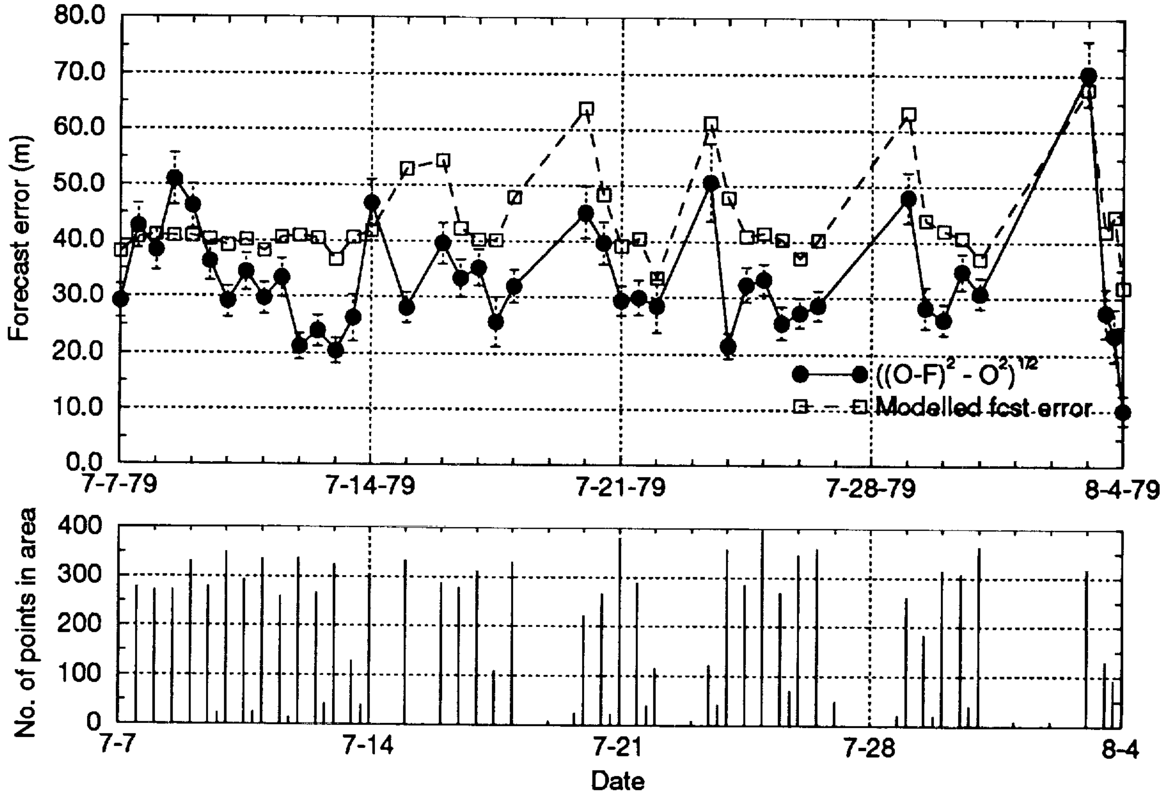

Figure 4: Upper panel: Observed (solid line) and modelled (dashed line) 500 hPa forecast error standard deviations for a box in the South Pacific (165W–105W, 60S–36S) from July 7 to August 4, 1979. Lower panel: the number of observations used to estimate the observed forecast error standard deviations.

Outside of the tropics the standard deviations of the u and v components of the wind errors, σ_i^u and σ_i^v, used in the multivariate upper air height–wind analysis are determined using the geostrophic approximation as in B87 (see also Appendix A),

$$(\sigma_i^u)^2 = (\sigma_i^v)^2 = \frac{g^2(\sigma_i^z)^2}{|f|^2}\left\{-\frac{\partial^2\mu^{z,z}(s)}{\partial s^2}\bigg|_{s=0}\right\}, \tag{13}$$

where g is the gravitational constant, f the Coriolis parameter, σ_i^z the forecast height error standard deviation from (10), and $\mu^{z,z}(s)$ the isotropic separation dependent model function

Table 4: Parameters used in the background error variance growth model for geopotential height, water vapor mixing ratio and sea level pressure.

	Geopotential height (m)				Mixing ratio (g/kg)	
	Tropics	Extra–Trop.	Tropics	Extra–Trop.		
level	α_{tr}	α_{xt}	β_{tr}	β_{xt}	α	β
10	125.00	70.00	147.00	112.50	—	—
30	110.00	59.00	132.00	109.50	—	—
50	95.00	48.00	117.00	106.50	—	—
70	80.00	37.00	105.00	103.00	—	—
100	60.00	35.00	96.00	98.50	—	—
150	25.00	32.00	76.00	95.00	—	—
200	21.00	30.00	56.00	91.50	—	—
250	17.00	28.00	37.00	88.50	—	—
300	14.00	26.50	29.00	85.00	0.05	0.40
400	13.00	23.00	26.90	78.00	0.15	1.20
500	12.00	20.00	24.50	71.00	0.20	1.60
700	10.50	15.00	20.00	58.00	0.65	5.20
850	9.00	12.50	20.00	56.50	0.90	7.20
1000	9.00	10.50	20.00	55.00	1.20	9.60
SLP	1.13	1.31	2.50	6.88	—	—

for the horizontal correlation of forecast height errors, which is discussed in the next section. In the tropical band 15°S $\leq \varphi \leq$ 15°N, σ_ι^u and σ_ι^v are assigned the pressure level dependent value $\sigma_{trop}^{u,v}$ given in Table 5. At points in the subtropical bands, 25°S $\leq \varphi <$ 15°S and 15°N $< \varphi \leq$ 25°N, σ_ι^u and σ_ι^v are defined by

$$(\sigma_\iota^u)^2 = (\sigma_\iota^v)^2 = \left[\frac{25-|\varphi|}{10}\right](\sigma_{\mathrm{trop}}^{u,v})^2 + \left[\frac{|\varphi|-15}{10}\right]\frac{g^2(\sigma_\iota^z)^2}{|f_{25}|^2}\left[-\frac{\partial^2\mu^{z,z}(s)}{\partial s^2}\bigg|_{s=0}\right] \tag{14}$$

where $f_{25} = 2\Omega\sin(25°)$.

Outside of the tropical band, 20°S $\leq \varphi \leq$ 20°N, the standard deviations of the u_s and v_s components of the near surface wind errors over the oceans are defined as in B87 by

$$(\sigma_\iota^{u_s})^2 = (\sigma_\iota^{v_s})^2 = \frac{(\sigma_\iota^p)^2}{a^2}(C_\iota^2 + F_\iota^2)\left[-\frac{\partial^2\mu^{p,p}(s)}{\partial s^2}\bigg|_{s=0}\right], \tag{15}$$

where σ_ι^p is the background sea level pressure error standard deviation from (10), a the Earth's radius, C_ι and F_ι the coefficients for the surface wind model derived in B87, and $\mu^{p,p}(s)$ the isotropic separation dependent model function for the horizontal correlation of background sea level pressure errors discussed in the next section. In the tropical band, 10°S $\leq \varphi \leq$ 10°N, $\sigma_\iota^{u_s}$ and $\sigma_\iota^{v_s}$ are set to 5 ms^{-1}, while in the subtropical bands, 20°N $\leq \varphi <$ 10°S and 10°N $< \varphi \leq$ 20°N, they are defined by

$$(\sigma_\iota^{u_s})^2 = (\sigma_\iota^{v_s})^2 = \left[\frac{20-|\varphi|}{10}\right]\times(5)^2 + \left[\frac{|\varphi|-10}{10}\right]\frac{(\sigma_\iota^p)^2}{a^2}(C_\iota^2 + F_\iota^2)\left\{-\frac{\partial^2\mu^{p,p}(s)}{\partial s^2}\bigg|_{s=0}\right\}. \tag{16}$$

Table 5: Tropical upper air wind forecast error standard deviations.

Tropical Pressure interval			σ^u and σ^v (m s^{-1})
	p <	20	7.1
20	≤ p <	40	7.1
40	≤ p <	60	8.0
60	≤ p <	85	7.3
85	≤ p <	125	7.3
125	≤ p <	175	7.3
175	≤ p <	225	7.4
225	≤ p <	275	7.0
275	≤ p <	350	6.2
350	≤ p <	450	5.2
450	≤ p <	600	4.5
600	≤ p <	775	4.4
775	≤ p <	925	4.4
925	≤ p		4.3

8.2 Background Error Correlations

For all permutation pairs, $[a, b]$, of variables for the upper-air multivariate analysis [geopotential height (z), east-west wind component (u), and north-south wind component (v)] the forecast error correlations, μ^f_{ij}, are assumed *separable i.e.* they are written as the product of a horizontal, $\mu^{a,b}(s_{ij})$, and a vertical, $\nu^{a,b}_{ij}$, error correlation. As in B87, the horizontal height-height error correlations are modeled using a damped cosine function (Thiebaux 1975),

$$\mu^{z,z}(s) = \frac{1}{c_1 + c_3}[c_1 \cos(c_2 s) + c_3][1 + (c_4 s)^2]^{-c_5}. \tag{17}$$

Notice that this function has two length scales, c_2 and c_4. The fitting parameters have been recomputed since B87, and a single function is used globally in contrast to the regionally dependent functions described in B87. The current parameter values ($c_1 = 0.0129928$, $c_2 = 0.00389265$ km^{-1}, $c_3 = 0.694005$, $c_4 = 0.00105123$ km^{-1} and $c_5 = 1.20815$) were obtained by a fit using 500 hPa radiosonde height data from North America for January and February of 1979. The graph of $(c_1 + c_3)\mu^{z,z}(s)$, illustrating its fit to the data, is shown in Fig. 5. The vertical correlation of background height errors, $\nu^{z,z}_{ij}$, is obtained by interpolating from the values appearing in the upper triangular part of Table 6. In GEOS-1, the chord length approximation is used to computed distances (see Appendix 9).

For the correlation of forecast height errors with forecast wind errors as well as the wind-wind forecast error correlations an approximation to the geostrophic assumption is used. Starting with the height-height error covariance model, $\left\langle \tilde{\epsilon}^z_i \tilde{\epsilon}^z_j \right\rangle = \sigma^z_i \sigma^z_j \mu^{z,z}_{ij} \nu^{z,z}_{ij}$, and the geostrophic approximations for the wind component errors,

$$\tilde{\epsilon}^u = -\frac{g}{af}\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \quad \text{and} \quad \tilde{\epsilon}^v = \frac{g}{af\cos\varphi}\frac{\partial \tilde{\epsilon}^z}{\partial \lambda}, \tag{18}$$

Table 6: Vertical correlation of background errors for geopotential height (shown in upper triangular part of table), and for water vapor mixing ratio (shown in lower triangular part of table).

	Pressure level (h Pa)													
level	10	30	50	70	100	150	200	250	300	400	500	700	850	1000
20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
30		1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50			1.00	0.87	0.71	0.29	0.19	0.06	0.00	0.06	0.08	0.10	0.14	0.09
70				1.00	0.75	0.31	0.19	0.05	0.00	0.05	0.06	0.09	0.12	0.09
100					1.00	0.60	0.46	0.30	0.21	0.22	0.21	0.15	0.18	0.12
150						1.00	0.82	0.66	0.56	0.47	0.43	0.25	0.19	0.10
200							1.00	0.81	0.72	0.61	0.52	0.32	0.24	0.14
200								1.00	0.86	0.77	0.71	0.49	0.40	0.25
300									1.00	0.85	0.80	0.59	0.48	0.30
400									0.94	1.00	0.87	0.70	0.59	0.37
500									0.60	0.71	1.00	0.77	0.68	0.43
700									0.07	0.12	0.33	1.00	0.83	0.57
850									0.02	0.04	0.13	0.50	1.00	0.66
1000									0.04	0.05	0.04	-0.01	0.27	1.00

we get:

$$\left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^u \right\rangle = -\frac{g}{af_j} \left\langle \tilde{\epsilon}_i^z \left(\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \right)_j \right\rangle = -\frac{g}{af_j} \frac{\partial \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \varphi_j} \tag{19}$$

$$\left\langle \tilde{\epsilon}_i^u \tilde{\epsilon}_j^z \right\rangle = -\frac{g}{af_i} \left\langle \left(\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \right)_i \tilde{\epsilon}_j^z \right\rangle = -\frac{g}{af_i} \frac{\partial \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \varphi_i} \tag{20}$$

$$\left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^v \right\rangle = \frac{g}{af_j \cos\varphi_j} \left\langle \tilde{\epsilon}_i^z \left(\frac{\partial \tilde{\epsilon}^z}{\partial \lambda} \right)_j \right\rangle = \frac{g}{af_j \cos\varphi_j} \frac{\partial \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \lambda_j} \tag{21}$$

$$\left\langle \tilde{\epsilon}_i^v \tilde{\epsilon}_j^z \right\rangle = \frac{g}{af_i \cos\varphi_i} \left\langle \left(\frac{\partial \tilde{\epsilon}^z}{\partial \lambda} \right)_i \tilde{\epsilon}_j^z \right\rangle = \frac{g}{af_i \cos\varphi_i} \frac{\partial \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \lambda_i} \tag{22}$$

$$\left\langle \tilde{\epsilon}_i^u \tilde{\epsilon}_j^u \right\rangle = \frac{g^2}{a^2 f_i f_j} \left\langle \left(\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \right)_i \left(\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \right)_j \right\rangle = \frac{g^2}{a^2 f_i f_j} \frac{\partial^2 \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \varphi_i \partial \varphi_j} \tag{23}$$

$$\left\langle \tilde{\epsilon}_i^u \tilde{\epsilon}_j^v \right\rangle = -\frac{g^2}{a^2 f_i f_j \cos\varphi_j} \left\langle \left(\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \right)_i \left(\frac{\partial \tilde{\epsilon}^z}{\partial \lambda} \right)_j \right\rangle = -\frac{g^2}{a^2 f_i f_j \cos\varphi_j} \frac{\partial^2 \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \varphi_i \partial \lambda_j} \tag{24}$$

$$\left\langle \tilde{\epsilon}_i^v \tilde{\epsilon}_j^u \right\rangle = -\frac{g^2}{a^2 f_i f_j \cos\varphi_i} \left\langle \left(\frac{\partial \tilde{\epsilon}^z}{\partial \lambda} \right)_i \left(\frac{\partial \tilde{\epsilon}^z}{\partial \varphi} \right)_j \right\rangle = -\frac{g^2}{a^2 f_i f_j \cos\varphi_i} \frac{\partial^2 \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \lambda_i \partial \varphi_j} \tag{25}$$

$$\left\langle \tilde{\epsilon}_i^v \tilde{\epsilon}_j^v \right\rangle = \frac{g^2}{a^2 f_i f_j \cos\varphi_i \cos\varphi_j} \left\langle \left(\frac{\partial \tilde{\epsilon}^z}{\partial \lambda} \right)_i \left(\frac{\partial \tilde{\epsilon}^z}{\partial \lambda} \right)_j \right\rangle = \frac{g^2}{a^2 f_i f_j \cos\varphi_i \cos\varphi_j} \frac{\partial^2 \left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle}{\partial \lambda_i \partial \lambda_j}. \tag{26}$$

In computing the partial derivatives of $\left\langle \tilde{\epsilon}_i^z \tilde{\epsilon}_j^z \right\rangle$ in (19)–(25) we make the additional assump-

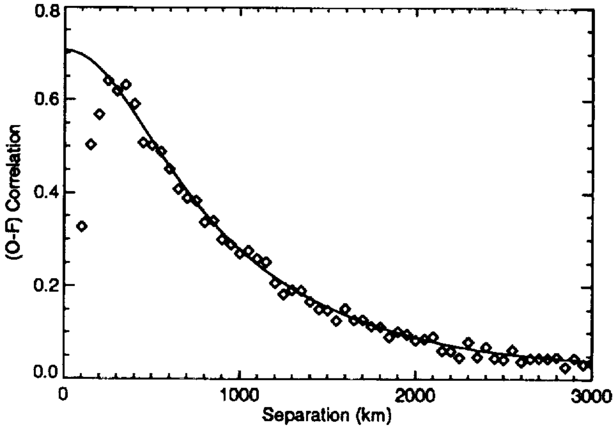

Figure 5: Fit of scaled damped cosine correlation function to radiosonde minus background 500 hPa height difference correlations using North American radiosondes for January and February 1979.

tion of local homogeneity for the height error variances,

$$\frac{\partial \sigma^z}{\partial \varphi} \approx \frac{\partial \sigma^z}{\partial \lambda} \approx 0. \tag{27}$$

Consequences of this *homogeneity* assumption are discussed in Cohn and Morone (1984) and Morone and Cohn (1985). Derivations of effective computational equations for the height–wind and wind–wind error correlations including modifications for the tropics are given in Appendix C. Note that separate methods have been used to decouple tropical wind-wind variances (eq. 14) and tropical wind-mass correlations (eqs. 118,120–123).

9 Current Development Status of GEOS DAS

In the time since GEOS-1 DAS was created to perform the 5 year reanalysis, one new version of the analysis component has been made operational (OI 1.5) and another version, with many more significant changes, is about to become operational (OI 2.0). The most important change in version OI 1.5 is the correction applied to the wind-wind forecast error correlations in the deep tropics. The geostrophically derived correlations are not used across the equator because sign changes introduce spurious divergence in the wind analysis; no changes were introduced in the extratropics (defined as those latitudes poleward of 15°).

For OI 2.0, the code has been generalized to allow up to 300 observations (one *observation* equals one observed variable) to contribute to the analysis of the grid points in each mini-volume. We have also added the ability to assimilate the SSM/I surface wind speeds, SSM/I total precipitable water retrievals, and have experimented with assimilation of the microwave limb sounder temperatures from the Upper Atmosphere Research Satellite (UARS/MLS). The code has been generalized to accept new conventional data types, such as the ARINC Communication Addressing and Reporting System (ACARS) aircraft data, which are winds reported by wide-bodied US commercial aircraft. The large mini-volumes at the poles, which had been a collection of all pole points, have been broken up into mid-latitude sized mini-volumes. This eliminated some discontinuities associated with such large mini-volumes. The observation increment history data set now contains a history of the quality control decisions, and we are carrying the time of each observation through the analysis to begin experimenting with the decoupling of satellite temperature retrievals from different orbits. Work is in progress to retune observation and forecast error statistics consistent with recent improvements in the system.

Since GEOS–1 DAS, there have also been several improvements made to the GEOS GCM. The first improvement entails the development of a fourth-order version of the Sadourney scheme for the momentum equation. The "fourth-order Sadourney" is derived in Suarez and Takacs (1995). This scheme conserves total energy and potential enstrophy for non-divergent flow, and is fourth-order in the sense that the advection of the second-order vorticity reduces to that governed by the fourth-order Arakawa (1966) Jacobian. This scheme was chosen rather than the Arakawa-Lamb (1981) scheme modified to fourth-order by Takano-Wurtele (1982) due to severe polar noise problems generated in three-dimensional GCM calculations from the terms required for potential enstrophy conservation for general flow. Together with the "fourth-order Sadourney" scheme for the momentum equations, the current GEOS GCM also uses fourth-order horizontal advection in the thermodynamic and moisture equations. It uses a scheme developed by Arakawa which has been used for many years in the UCLA GCM.

The dynamical climatology of the GEOS GCM has been thoroughly examined by Takacs and Suarez (1995). They have shown that increasing the resolution and/or the order-of-accuracy in the GEOS GCM has a very significant impact on the zonal mean flow. Running the second-order GEOS-1 GCM at both 4° x 5° and 2° x 2.5° horizontal resolution, an examination of the first and second moments from a five year simulation revealed a strong systematic bias for most fields between the two resolutions, particularly in the southern hemisphere. Most of these biases are related to the simulation of the transient eddies, with increased accuracy enhancing transient transports of heat and momentum and decreasing transports of moisture. These systematic biases are removed, however, when fourth-order accuracy is used. The higher-order and higher-resolution experiments are in closer agreement to the GEOS-1 DAS analysis in the southern hemisphere where the transports are primarily determined by the transient flow.

Another major enhancement to the GEOS GCM is the ability to run both simulations and the GEOS DAS assimilation using a rotated coordinate. In the current system, the geographic placement of the computational pole is arbitrary. We have found that due to variations of the scheme near the poles required from conservation constraints, the computational instability discussed by Hollingsworth et al. (1983) for the nearest grid-point to the pole gives rise to polar noise when confronted with strong cross-polar flow. This instability is proportional to the mean zonal wind speed and the Coriolis parameter. By rotating the computational grid to the geographic equator, however, the instability near the computational pole is removed due to the vanishing Coriolis term. In addition, the geographic pole now using the transformed grid is also free of noise.

Acknowledgments

We would like to thank Larry Takacs for assistance with the model description and update on the current status of development of the GEOS GCM. The development and research documented in this report were supported by the EOS Interdisciplinary Science Program and by the NASA Research and Applications Program. Computer resources and funding were provided by the EOS Project through the Scientific Computational Facility of the Data Assimilation Office.

Appendix A: Chord-length/Great Circle Distance Approximation

The background and observation error correlation models used in the analysis are expressed as isotropic functions of the separation distance between points on the sphere. Given two points $P_i = (\lambda_i, \varphi_i)$ and $P_j = (\lambda_j, \varphi_j)$ on the sphere, we can write their position vectors as:

$$\mathbf{p}_i = a\mathbf{q}_i = a\left[\cos\varphi_i\cos\lambda_i, \cos\varphi_i\sin\lambda_i, \sin\varphi_i\right] \tag{28}$$

and

$$\mathbf{p}_j = a\mathbf{q}_j = a\left[\cos\varphi_j\cos\lambda_j, \cos\varphi_j\sin\lambda_j, \sin\varphi_j\right], \tag{29}$$

where a is the Earth's radius. The true shortest distance arc length along the great circle connecting P_i and P_j is given by

$$s_{ij} = a\arccos(\mathbf{q}_i\cdot\mathbf{q}_j). \tag{30}$$

On the other hand, the square of the length of the chord connecting P_i and P_j is:

$$\begin{aligned}\hat{s}_{ij}^2 &= (\mathbf{p}_i - \mathbf{p}_j)\cdot(\mathbf{p}_i - \mathbf{p}_j)\\ &= 2a^2(1 - \mathbf{q}_i\cdot\mathbf{q}_j).\end{aligned} \tag{31}$$

In GEOS–1 DAS the chord length approximation (31) is used to compute distances. For the sake of completeness, this section will derive equations based on the great circle formula (30) and the chord-length formula (31).

We will need the first and second partial derivative of s_{ij} and $\hat{s}_{ij}$ with respect to φ_i, λ_i, φ_j, and λ_j. Note that if γ and η are any two (possibly the same) of these four independent variables we have

$$\frac{\partial s_{ij}}{\partial\gamma}\sin(\frac{s_{ij}}{a}) = \frac{\partial\hat{s}_{ij}}{\partial\gamma}(\frac{\hat{s}_{ij}}{a}) = -a\frac{\partial(\mathbf{q}_i\cdot\mathbf{q}_j)}{\partial\gamma} \tag{32}$$

and

$$\frac{\partial^2 s_{ij}}{\partial\eta\partial\gamma}a\,\sin(\frac{s_{ij}}{a}) + \cos(\frac{s_{ij}}{a})\frac{\partial s_{ij}}{\partial\eta}\frac{\partial s_{ij}}{\partial\gamma} = \frac{\partial^2\hat{s}_{ij}}{\partial\eta\partial\gamma}\hat{s}_{ij} + \frac{\partial\hat{s}_{ij}}{\partial\eta}\frac{\partial\hat{s}_{ij}}{\partial\gamma} = -a^2\frac{\partial^2(\mathbf{q}_i\cdot\mathbf{q}_j)}{\partial\eta\partial\gamma}. \tag{33}$$

In particular, we need the following partial derivatives for the height–wind and wind–wind error covariance computations.

- For $\langle zu\rangle$,

$$b\frac{\partial s_{ij}}{\partial\varphi_j} = \hat{b}\frac{\partial\hat{s}_{ij}}{\partial\varphi_j} = -ac_{iy}. \tag{34}$$

- For $\langle uz\rangle$,

$$b\frac{\partial s_{ij}}{\partial\varphi_i} = \hat{b}\frac{\partial\hat{s}_{ij}}{\partial\varphi_i} = -ac_{yj}. \tag{35}$$

- For $\langle zv\rangle$,

$$b\frac{\partial s_{ij}}{\partial\lambda_j} = \hat{b}\frac{\partial\hat{s}_{ij}}{\partial\lambda_j} = -ac_{ix}. \tag{36}$$

- For $\langle vz\rangle$,

$$b\frac{\partial s_{ij}}{\partial\lambda_i} = \hat{b}\frac{\partial \hat{s}_{ij}}{\partial\lambda_i} = -ac_{xj}. \tag{37}$$

- For $\langle uu\rangle$,

$$\begin{aligned} b^3\frac{\partial^2 s_{ij}}{\partial\varphi_i\partial\varphi_j} &= -a(b^2 c_{yy} + c_{ij}c_{iy}c_{yj}) \qquad (38)\\ \hat{b}^3\frac{\partial^2 \hat{s}_{ij}}{\partial\varphi_i\partial\varphi_j} &= -a(\hat{b}^2 c_{yy} + c_{iy}c_{yj}). \qquad (39)\end{aligned}$$

- For $\langle uv\rangle$,

$$\begin{aligned} b^3\frac{\partial^2 s_{ij}}{\partial\varphi_i\partial\lambda_j} &= -a(b^2 c_{yx} + c_{ij}c_{yj}c_{ix}) \qquad (40)\\ \hat{b}^3\frac{\partial^2 \hat{s}_{ij}}{\partial\varphi_i\partial\lambda_j} &= -a(\hat{b}^2 c_{yx} + c_{yj}c_{ix}). \qquad (41)\end{aligned}$$

- For $\langle vu\rangle$,

$$\begin{aligned} b^3\frac{\partial^2 s_{ij}}{\partial\lambda_i\partial\varphi_j} &= -a(b^2 c_{xy} + c_{ij}c_{xj}c_{iy}) \qquad (42)\\ \hat{b}^3\frac{\partial^2 \hat{s}_{ij}}{\partial\lambda_i\partial\varphi_j} &= -a(\hat{b}^2 c_{xy} + c_{xj}c_{iy}). \qquad (43)\end{aligned}$$

- For $\langle vv\rangle$,

$$\begin{aligned} b^3\frac{\partial^2 s_{ij}}{\partial\lambda_i\partial\lambda_j} &= -a(b^2 c_{xx} + c_{ij}c_{ix}c_{xj}) \qquad (44)\\ \hat{b}^3\frac{\partial^2 \hat{s}_{ij}}{\partial\lambda_i\partial\lambda_j} &= -a(\hat{b}^2 c_{xx} + c_{ix}c_{xj}). \qquad (45)\end{aligned}$$

Where we have used,

$$\begin{aligned} b &= \sin(\frac{s_{ij}}{a}) = \sqrt{1-c_{ij}^2}, \qquad (46)\\ \hat{b} &= (\frac{\hat{s}_{ij}}{a}) = \sqrt{2(1-c_{ij})}, \qquad (47)\\ c_{ij} &= \mathbf{q}_i\cdot\mathbf{q}_j\\ &= +\cos\varphi_i\cos\lambda_i\cos\varphi_j\cos\lambda_j + \cos\varphi_i\sin\lambda_i\cos\varphi_j\sin\lambda_j + \sin\varphi_i\sin\varphi_j, \qquad (48)\\ c_{ix} &= \mathbf{q}_i\cdot\frac{\partial\mathbf{q}_j}{\partial\lambda_j}\\ &= -\cos\varphi_i\cos\lambda_i\cos\varphi_j\sin\lambda_j + \cos\varphi_i\sin\lambda_i\cos\varphi_j\cos\lambda_j, \qquad (49)\\ c_{iy} &= \mathbf{q}_i\cdot\frac{\partial\mathbf{q}_j}{\partial\varphi_j}\\ &= -\cos\varphi_i\cos\lambda_i\sin\varphi_j\cos\lambda_j - \cos\varphi_i\sin\lambda_i\sin\varphi_j\sin\lambda_j + \sin\varphi_i\cos\varphi_j, \qquad (50)\\ c_{xj} &= \frac{\partial\mathbf{q}_i}{\partial\lambda_i}\cdot\mathbf{q}_j \end{aligned}$$

$$\begin{aligned}
&= -\cos\varphi_i \sin\lambda_i \cos\varphi_j \cos\lambda_j + \cos\varphi_i \cos\lambda_i \cos\varphi_j \sin\lambda_j, && (51)\\
c_{yj} &= \frac{\partial \mathbf{q}_i}{\partial \varphi_i} \cdot \mathbf{q}_j \\
&= -\sin\varphi_i \cos\lambda_i \cos\varphi_j \cos\lambda_j - \sin\varphi_i \sin\lambda_i \cos\varphi_j \sin\lambda_j + \cos\varphi_i \sin\varphi_j, && (52)\\
c_{xx} &= \frac{\partial \mathbf{q}_i}{\partial \lambda_i} \cdot \frac{\partial \mathbf{q}_j}{\partial \lambda_j} \\
&= +\cos\varphi_i \sin\lambda_i \cos\varphi_j \sin\lambda_j + \cos\varphi_i \cos\lambda_i \cos\varphi_j \cos\lambda_j, && (53)\\
c_{xy} &= \frac{\partial \mathbf{q}_i}{\partial \lambda_i} \cdot \frac{\partial \mathbf{q}_j}{\partial \varphi_j} \\
&= +\cos\varphi_i \sin\lambda_i \sin\varphi_j \cos\lambda_j - \cos\varphi_i \cos\lambda_i \sin\varphi_j \sin\lambda_j, && (54)\\
c_{yx} &= \frac{\partial \mathbf{q}_i}{\partial \varphi_i} \cdot \frac{\partial \mathbf{q}_j}{\partial \lambda_j} \\
&= +\sin\varphi_i \cos\lambda_i \cos\varphi_j \sin\lambda_j - \sin\varphi_i \sin\lambda_i \cos\varphi_j \cos\lambda_j, && (55)
\end{aligned}$$

and

$$\begin{aligned}
c_{yy} &= \frac{\partial \mathbf{q}_i}{\partial \varphi_i} \cdot \frac{\partial \mathbf{q}_j}{\partial \varphi_j} \\
&= +\sin\varphi_i \cos\lambda_i \sin\varphi_j \cos\lambda_j + \sin\varphi_i \sin\lambda_i \sin\varphi_j \sin\lambda_j + \cos\varphi_i \cos\varphi_j. && (56)
\end{aligned}$$

Appendix B: Damped Cosine Correlation Model

As in Baker et al. (1987), the horizontal height–height error correlations, as well as the sea level pressure–sea level pressure error correlations, are modeled using a damped cosine function (Thiebaux 1975),

$$\mu = \mu^{zz}(s) = \mu^{pp}(s) = \frac{1}{c_1 + c_3}[c_1 \cos(c_2 s) + c_3][1 + c_4^2 s^2]^{-c_5}. \tag{57}$$

The fitting parameters have been recomputed since Baker et al. (1987), and a single function is used globally in contrast to the regionally dependent functions described in that paper. The current parameter values,

$$\begin{aligned}
c_1 &= 0.0129928, && (58)\\
c_2 &= 0.00389265\,\mathrm{km}^{-1}, && (59)\\
c_3 &= 0.694005, && (60)\\
c_4 &= 0.00105123\,\mathrm{km}^{-1}, && (61)\\
c_5 &= 1.20815, && (62)
\end{aligned}$$

were obtained by N. S. Sivakumaran using a fit to 500 hPa radiosonde height data from North America for January and February of 1979.

If we define

$$\begin{aligned}
\tilde{c}_1 &= \frac{c_1}{c_1 + c_3}, && (63)\\
\tilde{c}_3 &= \frac{c_3}{c_1 + c_3} = 1 - \tilde{c}_1, && (64)
\end{aligned}$$

$$
\begin{aligned}
\beta^2 &= \tilde{c}_1 c_2^2 + 2c_4^2 c_5, && (65)\\
\gamma &= \frac{\tilde{c}_1 c_2^2}{\beta^2}, && (66)\\
\eta &= \frac{2c_4^2 c_5}{\beta^2} = 1 - \gamma, && (67)\\
A &= \left(1 + c_4^2 s^2\right)^{-1}, && (68)\\
A' &= -2c_4^2 A^2 s, && (69)\\
B &= A^{c_5}, && (70)\\
B' &= -\eta\beta^2 ABs, && (71)\\
C &= \frac{\sin(c_2 s)}{(c_2 s)}, && (72)\\
C' &= (1/s)[\cos(c_2 s) - C], && (73)\\
D &= 2c_4^2 As^2 = -\frac{sA'}{A} = -\frac{sB'}{c_5 B} && (74)
\end{aligned}
$$

we can write

$$
\mu = [\tilde{c}_1 \cos(c_2 s) + \tilde{c}_3]B. \tag{75}
$$

Differentiating we get

$$
\begin{aligned}
\mu' &= (\mu/B)B' - \tilde{c}_1 c_2 \sin(c_2 s)B\\
&= -\eta\beta^2 A\mu s - \tilde{c}_1 c_2^2 CBs\\
&= -\beta^2 s[\eta A\mu + \gamma BC], && (76)
\end{aligned}
$$

and differentiating again gives

$$
\begin{aligned}
\mu'' &= -\beta^2\left(\frac{\mu'}{-\beta^2 s}\right) - \beta^2 s[\eta(A\mu' + A'\mu) + \gamma(BC' + B'C)]\\
&= -\beta^2\{\eta[A\mu + sA\mu' + sA'\mu] + \gamma[BC + sBC' + sB'C]\}\\
&= -\beta^2\{A\mu\eta[1 - 2c_4^2 As^2(1 + c_5)] + B\gamma[\cos(c_2 s) - 4c_4^2 c_5 ACs^2]\}\\
&= -\beta^2\{A\mu\eta[1 - (1 + c_5)D] + B\gamma[\cos(c_2 s) - 2c_5 CD]\}. && (77)
\end{aligned}
$$

At the origin we have: $\mu(0) = 1$, $\lim_{s\to 0}(\frac{\mu'(s)}{s}) = -\beta^2$, and $\mu''(0) = -\beta^2$. The graphs of the functions μ, $-\mu'/(s\beta^2)$, and $-\mu''/\beta^2$ are shown in Fig. 6.

Appendix C: Height–Wind and Wind–Wind Background Error Covariances with Damped Cosine Model Function

In this appendix we give the explicit formulas for the correlation functions used in the multivariate wind-height analysis.

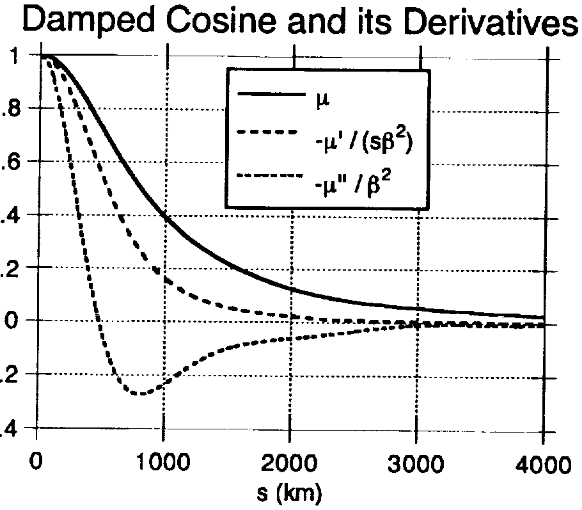

Figure 6: The functions μ, $-\mu'/(s\beta^2)$, and $-\mu''/\beta^2$ calculated using the parameter values given in the text.

C.1 Correlation modeling assumptions for background errors

For all permutation pairs, ab, of the three upper-air analysis variables [geopotential height (z), east–west wind component (u), and north–south wind component (v)] the correlation between background errors in variable a at location $P_i = (\lambda_i, \varphi_i, p_i)$ and background errors in variable b at location $P_j = (\lambda_j, \varphi_j, p_j)$ is defined by

$$\mu(\epsilon_i^a \epsilon_j^b) \equiv \frac{\langle \epsilon_i^a \epsilon_j^b \rangle}{\sigma_i^a \sigma_j^b}. \tag{78}$$

Our first modeling assumption, often referred to as the *separability* assumption, is that each of the $\mu(\epsilon_i^a \epsilon_j^b)$'s can be written as the product of a horizontal, $\mu_{ij}^{ab} = \mu^{ab}(\tilde{s}_{ij})$, and a vertical, $\nu_{ij}^{ab} = \nu^{ab}(\tilde{\delta}_{ij})$, error correlation, where $\tilde{s}_{ij} = \tilde{s}(\lambda_i, \varphi_i, \lambda_j, \varphi_j)$ is a measure of the horizontal distance between points P_i and P_j, and $\tilde{\delta}_{ij} = \tilde{\delta}(p_i, p_j)$ is a measure their vertical separation, *i.e.*,

$$\mu(\epsilon_i^a \epsilon_j^b) \approx \mu^{ab}(\tilde{s}_{ij}) \nu^{ab}(\tilde{\delta}_{ij}) = \mu_{ij}^{ab} \nu_{ij}^{ab}. \tag{79}$$

The second modeling assumption is that the background wind component errors can be related geostrophically to the background height errors, *i.e.* we will make use of the *geostrophic* approximations for the background wind errors,

$$\epsilon_i^u \approx -\frac{g}{a f_i} \frac{\partial \epsilon_i^z}{\partial \varphi_i} \qquad \text{and} \qquad \epsilon_i^v \approx \frac{g}{a f_i \cos\varphi_i} \frac{\partial \epsilon_i^z}{\partial \lambda_i}. \tag{80}$$

The geostrophic assumption implies that we can express all of the $\langle \epsilon_i^a \epsilon_j^b \rangle$'s in terms of $\langle \epsilon_i^z \epsilon_j^z \rangle$ and its partial derivatives with respect to $\lambda_i, \varphi_i, \lambda_j$ and φ_j:

$$\langle \epsilon_i^z \epsilon_j^z \rangle \approx \sigma_i^z \sigma_j^z \mu_{ij}^{zz} \nu_{ij}^{zz} \tag{81}$$

$$\begin{aligned}
\langle\epsilon_i^z\epsilon_j^u\rangle &\approx -\frac{g}{af_j}\langle\epsilon_i^z\frac{\partial\epsilon_j^z}{\partial\varphi_j}\rangle = -\frac{g}{af_j}\frac{\partial\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\varphi_j} && (82)\\
\langle\epsilon_i^u\epsilon_j^z\rangle &\approx -\frac{g}{af_i}\langle\frac{\partial\epsilon_i^z}{\partial\varphi_i}\epsilon_j^z\rangle = -\frac{g}{af_i}\frac{\partial\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\varphi_i} && (83)\\
\langle\epsilon_i^z\epsilon_j^v\rangle &\approx \frac{g}{af_j\cos\varphi_j}\langle\epsilon_i^z\frac{\partial\epsilon_j^z}{\partial\lambda_j}\rangle = \frac{g}{af_j\cos\varphi_j}\frac{\partial\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\lambda_j} && (84)\\
\langle\epsilon_i^v\epsilon_j^z\rangle &\approx \frac{g}{af_i\cos\varphi_i}\langle\frac{\partial\epsilon_i^z}{\partial\lambda_i}\epsilon_j^z\rangle = \frac{g}{af_i\cos\varphi_i}\frac{\partial\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\lambda_i} && (85)\\
\langle\epsilon_i^u\epsilon_j^u\rangle &\approx \frac{g^2}{a^2f_if_j}\langle\frac{\partial\epsilon_i^z}{\partial\varphi_i}\frac{\partial\epsilon_j^z}{\partial\varphi_j}\rangle = \frac{g^2}{a^2f_if_j}\frac{\partial^2\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\varphi_i\partial\varphi_j} && (86)\\
\langle\epsilon_i^u\epsilon_j^v\rangle &\approx -\frac{g^2}{a^2f_if_j\cos\varphi_j}\langle\frac{\partial\epsilon_i^z}{\partial\varphi_i}\frac{\partial\epsilon_j^z}{\partial\lambda_j}\rangle = -\frac{g^2}{a^2f_if_j\cos\varphi_j}\frac{\partial^2\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\varphi_i\partial\lambda_j} && (87)\\
\langle\epsilon_i^v\epsilon_j^u\rangle &\approx -\frac{g^2}{a^2f_if_j\cos\varphi_i}\langle\frac{\partial\epsilon_i^z}{\partial\lambda_i}\frac{\partial\epsilon_j^z}{\partial\varphi_j}\rangle = -\frac{g^2}{a^2f_if_j\cos\varphi_i}\frac{\partial^2\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\lambda_i\partial\varphi_j} && (88)\\
\langle\epsilon_i^v\epsilon_j^v\rangle &\approx \frac{g^2}{a^2f_if_j\cos\varphi_i\cos\varphi_j}\langle\frac{\partial\epsilon_i^z}{\partial\lambda_i}\frac{\partial\epsilon_j^z}{\partial\lambda_j}\rangle = \frac{g^2}{a^2f_if_j\cos\varphi_i\cos\varphi_j}\frac{\partial^2\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\lambda_i\partial\lambda_j}. && (89)
\end{aligned}$$

The third modeling assumption is that the height error standard deviation is horizontally homogeneous. This *homogeneity* assumption (see Morone and Cohn 1985) is expressed through the approximations,

$$\frac{\partial\sigma^z}{\partial\varphi} \approx \frac{\partial\sigma^z}{\partial\lambda} \approx 0, \tag{90}$$

and will be used to simplify the partial derivatives of $\langle\epsilon_i^z\epsilon_j^z\rangle$ which appeared in (82) through (89).

C.2 Derivation of wind component error variances

In order to derive expressions for σ^u and σ^v which are consistent with our correlation modeling assumptions, we will first examine $\langle\epsilon_i^u\epsilon_j^u\rangle$ and $\langle\epsilon_i^v\epsilon_j^v\rangle$ in detail. From the ***separability*** assumption we have:

$$\sigma_i^u\sigma_j^u\mu_{ij}^{uu}\nu_{ij}^{uu} \approx \langle\epsilon_i^u\epsilon_j^u\rangle. \tag{91}$$

Using our *geostrophic* approximation and separability a second time, this becomes

$$\begin{aligned}
\sigma_i^u\sigma_j^u\mu_{ij}^{uu}\nu_{ij}^{uu} &\approx \frac{g^2}{a^2f_if_j}\frac{\partial^2\langle\epsilon_i^z\epsilon_j^z\rangle}{\partial\varphi_i\partial\varphi_j}\\
&\approx \frac{g^2}{a^2f_if_j}\frac{\partial^2(\sigma_i^z\sigma_j^z\mu_{ij}^{zz}\nu_{ij}^{zz})}{\partial\varphi_i\partial\varphi_j}.
\end{aligned} \tag{92}$$

With two applications of the *homogeneity* assumption, we arrive at

$$\begin{aligned}
\sigma_i^u\sigma_j^u\mu_{ij}^{uu}\nu_{ij}^{uu} &\approx \frac{g^2}{a^2f_if_j}\frac{\partial}{\partial\varphi_i}\Big\{\sigma_i^z\sigma_j^z\frac{\partial\mu_{ij}^{zz}}{\partial\varphi_j}\nu_{ij}^{zz}\Big\}\\
&\approx \frac{g^2}{a^2f_if_j}\sigma_i^z\sigma_j^z\Big\{\frac{\partial^2\mu_{ij}^{zz}}{\partial\varphi_i\partial\varphi_j}\Big\}\nu_{ij}^{zz}
\end{aligned} \tag{93}$$

We now use the chain rule to rewrite the expression in brackets to get

$$\sigma_i^u \sigma_j^u \mu_{ij}^{uu} \nu_{ij}^{uu} \approx \frac{g^2}{a^2 f_i f_j} \sigma_i^z \sigma_j^z \Big\{ \frac{\partial^2 \mu_{ij}^{zz}}{\partial \tilde{s}^2} \frac{\partial \tilde{s}}{\partial \varphi_i} \frac{\partial \tilde{s}}{\partial \varphi_j} + \frac{\partial \mu_{ij}^{zz}}{\partial \tilde{s}} \frac{\partial^2 \tilde{s}}{\partial \varphi_i \partial \varphi_j} \Big\} \nu_{ij}^{zz}. \tag{94}$$

Taking the limit as $j \to i$, or equivalently as $\tilde{s} \to 0$, we have

$$(\sigma_i^u)^2 = \frac{g^2 (\sigma_i^z)^2}{a^2 f_i^2} \lim_{j \to i} \Big\{ \frac{\partial^2 \mu_{ij}^{zz}}{\partial \tilde{s}^2} \frac{\partial \tilde{s}}{\partial \varphi_i} \frac{\partial \tilde{s}}{\partial \varphi_j} + \frac{\partial \mu_{ij}^{zz}}{\partial \tilde{s}} \frac{\partial^2 \tilde{s}}{\partial \varphi_i \partial \varphi_j} \Big\}. \tag{95}$$

If μ_{ij}^{zz} is modeled with the damped cosine function described in the previous section, we can use equations (75), (76), (34) and (35) to simplify and get

$$(\sigma_i^u)^2 = -\frac{g^2 \beta^2 (\sigma_i^z)^2}{a^2 f_i^2} \lim_{j \to i} \Big\{ \frac{a^2 c_{iy} c_{yj}}{\tilde{b}^2} \Big\}, \tag{96}$$

where β^2 depends on the damped cosine fit parameters as given in (58)–(62); c_{iy} and c_{yj} are defined in (50) and (52); and $\tilde{b} = b$ from (46) if $\tilde{s}$ is the true arc length, or $\tilde{b} = \hat{b}$ from (47) if $\tilde{s}$ is the chord length distance approximation. In any case, the limit in (96) is equal to $-a^2$ and we have:

$$(\sigma_i^u)^2 = \frac{g^2 \beta^2 (\sigma_i^z)^2}{f_i^2}, \tag{97}$$

as appeared in Baker et al. (1987). This clearly needs to be modified near the equator. These modifications, the so–called *decoupling* of the geostrophic approximation in the tropics, will be discussed in subsection C.4 below.

The analog of (94) for $\langle \epsilon_i^v \epsilon_j^v \rangle$ is

$$\sigma_i^v \sigma_j^v \mu_{ij}^{vv} \nu_{ij}^{vv} \approx \frac{g^2}{a^2 f_i f_j \cos\varphi_i \cos\varphi_j} \sigma_i^z \sigma_j^z \Big\{ \frac{\partial^2 \mu_{ij}^{zz}}{\partial \tilde{s}^2} \frac{\partial \tilde{s}}{\partial \lambda_i} \frac{\partial \tilde{s}}{\partial \lambda_j} + \frac{\partial \mu_{ij}^{zz}}{\partial \tilde{s}} \frac{\partial^2 \tilde{s}}{\partial \lambda_i \partial \lambda_j} \Big\} \nu_{ij}^{zz}. \tag{98}$$

Taking the limit as $j \to i$ and using equations (75), (76), (36) and (37) we get:

$$(\sigma_i^v)^2 = -\frac{g^2 \beta^2 (\sigma_i^z)^2}{a^2 f_i^2} \lim_{j \to i} \Big\{ \frac{a^2 c_{ix} c_{xj}}{\tilde{b}^2 \cos\varphi_i \cos\varphi_j} \Big\}, \tag{99}$$

where c_{ix} and c_{xj} are defined in (49) and (50). The limit in (99) is also equal to $-a^2$. Hence, $(\sigma_i^v)^2$ is the same as $(\sigma_i^u)^2$ as given in (97). Notice that σ^u equals σ^v only under the assumption of homogeneous variances expressed by (90).

C.3 Formulas for computing background error correlations for the upper–air analysis

Now that we have formulas for the background wind error variance, we can substitute them into our separable background error covariance models to get effective formulas for computing the various error cross–correlation functions. We start with μ_{ij}^{uu}.

Substituting σ_i^u and σ_j^u as given by (97) into (94), and noting that ν_{ij}^{uu} must equal ν_{ij}^{zz}, we get:

$$\mu_{ij}^{uu} = \frac{\operatorname{sign}(\varphi_i)\operatorname{sign}(\varphi_j)}{a^2\beta^2}\Big\{\frac{\partial^2\mu_{ij}^{zz}}{\partial\tilde{s}^2}\frac{\partial\tilde{s}}{\partial\varphi_i}\frac{\partial\tilde{s}}{\partial\varphi_j} + \frac{\partial\mu_{ij}^{zz}}{\partial\tilde{s}}\frac{\partial^2\tilde{s}}{\partial\varphi_i\partial\varphi_j}\Big\}. \tag{100}$$

Substitution using (73) and (74) allows us to rewrite this as

$$\mu_{ij}^{uu} = \operatorname{sign}(\varphi_i)\operatorname{sign}(\varphi_j)\Big\{-\frac{\hat{s}}{a^2}\Big[\frac{\mu_{ij}^{zz\prime}(\tilde{s})}{-\beta^2\tilde{s}}\Big]\frac{\partial^2\tilde{s}}{\partial\varphi_i\partial\varphi_j} - \frac{1}{\tilde{b}^2}\Big[\frac{\mu_{ij}^{zz\prime\prime}(\tilde{s})}{-\beta^2}\Big]c_{iy}c_{yj}\Big\}. \tag{101}$$

If we are using the chord length approximation for horizontal separation, *i.e.*, $\tilde{s} = \hat{s}$, we can substitute using (39) to obtain,

$$\begin{aligned}
\mu_{ij}^{uu} &= \operatorname{sign}(\varphi_i)\operatorname{sign}(\varphi_j)\Bigg\{\Big[\frac{\mu_{ij}^{zz\prime}(\tilde{s})}{-\beta^2\tilde{s}}\Big]c_{yy} + \frac{1}{(\hat{s}/a)^2}\Bigg[\Big[\frac{\mu_{ij}^{zz\prime}(\tilde{s})}{-\beta^2\tilde{s}}\Big] - \Big[\frac{\mu_{ij}^{zz\prime\prime}(\tilde{s})}{-\beta^2}\Big]\Bigg]c_{iy}c_{yj}\Bigg\} \\
&= \operatorname{sign}(\varphi_i)\operatorname{sign}(\varphi_j)\Big\{\widehat{R}(\hat{s})c_{yy} + \widehat{T}(\hat{s})c_{iy}c_{yj}\Big\},
\end{aligned} \tag{102}$$

where

$$\widehat{R}(\hat{s}) = \Big[\frac{\mu_{ij}^{zz\prime}(\tilde{s})}{-\beta^2\tilde{s}}\Big], \tag{103}$$

and

$$\widehat{T}(\hat{s}) = \frac{1}{(\hat{s}/a)^2}\Big\{\widehat{R}(\hat{s}) - \Big[\frac{\mu_{ij}^{zz\prime\prime}(\tilde{s})}{-\beta^2}\Big]\Big\}. \tag{104}$$

On the other hand, if we are using the true arc length for horizontal separation, *i.e.*, $\tilde{s} = s$, we can substitute using (38) to obtain,

$$\begin{aligned}
\mu_{ij}^{uu} &= \operatorname{sign}(\varphi_i)\operatorname{sign}(\varphi_j)\Bigg\{\frac{s/a}{\sin(s/a)}\Big[\frac{\mu_{ij}^{zz\prime}(s)}{-\beta^2 s}\Big]c_{yy} \\
&\quad + \frac{1}{\sin^2(s/a)}\Bigg[\frac{s/a}{\sin(s/a)}\Big[\frac{\mu_{ij}^{zz\prime}(s)}{-\beta^2 s}\Big]\cos(s/a) - \Big[\frac{\mu_{ij}^{zz\prime\prime}(s)}{-\beta^2}\Big]\Bigg]c_{iy}c_{yj}\Bigg\} \\
&= \operatorname{sign}(\varphi_i)\operatorname{sign}(\varphi_j)\Big\{R(s)c_{yy} + T(s)c_{iy}c_{yj}\Big\},
\end{aligned} \tag{105}$$

where

$$R(s) = \frac{s/a}{\sin(s/a)}\Big[\frac{\mu_{ij}^{zz\prime}(s)}{-\beta^2 s}\Big], \tag{106}$$

and

$$T(s) = \frac{1}{\sin^2(s/a)}\Big\{R(s)\cos(s/a) - \Big[\frac{\mu_{ij}^{zz\prime\prime}(s)}{-\beta^2}\Big]\Big\}. \tag{107}$$

The computational algorithms for all of the various correlations will require the exact computation of $\tilde{s}_{ij} = (s_{ij} \text{ or } \hat{s}_{ij})$ and then table look-ups for $[\mu^{zz}(\hat{s}_{ij}),\ \widehat{R}(\hat{s}_{ij}),\ \widehat{T}(\hat{s}_{ij})]$ or $[\mu^{zz}(s_{ij}),\ R(s_{ij}),\ T(s_{ij})]$ as appropriate. Note that in Appendix B we showed that $\lim_{s\to 0}\mu^{zz}(s) = \lim_{\hat{s}\to 0}\mu^{zz}(\hat{s}) = 1$, and also that $\lim_{s\to 0}R(s) = \lim_{\hat{s}\to 0}\widehat{R}(\hat{s}) = 1$. We can also show that $\lim_{s\to 0}T(s) = \lim_{\hat{s}\to 0}\widehat{T}(\hat{s}) = 2(ac_4)^2\eta + (ac_2)^2\gamma + (a\beta)^2\eta(1 + \frac{\gamma}{3})$ where β, γ

and η are defined in (65) – (67). This limit at $\tilde{s} = 0$ can be used when tabulating $T(s)$ or $\widehat{T}(\hat{s})$.

$$\lim_{\hat{s}\to 0} \widehat{T}(\hat{s}) = \lim_{\hat{s}\to 0} \frac{a^2}{\hat{s}^2}\Big\{\widehat{R}(\hat{s}) - \Big[\frac{\mu_{ij}^{zz\,\prime\prime}(\tilde{s})}{-\beta^2}\Big]\Big\},$$

which according to (75) and (76) can be written

$$= \lim_{\hat{s}\to 0} \frac{a^2}{\hat{s}^2}\Big\{\gamma BC + A\mu\eta(1+c_5)D - B\gamma\cos c_2\hat{s} + 2\gamma c_5 BCD\Big\}.$$

Using (73) and (74) this can be written

$$= \lim_{\hat{s}\to 0} 2a^2c_4^2A\Big\{A\mu\eta(1+c_5) + 2\gamma c_5 BC\Big\} - \lim_{\hat{s}\to 0}\frac{a^2}{\hat{s}}\gamma BC',$$

and, since $\lim_{\hat{s}\to 0}[C'/\hat{s}] = -\frac{1}{3}c_2^2$, we get

$$\begin{aligned}\lim_{\hat{s}\to 0} \widehat{T}(\hat{s}) &= 2a^2c_4^2[\eta(1+c_5) + 2\gamma c_5] + \frac{\gamma}{3}a^2c_2^2 \\ &= 2(ac_4)^2\eta + (ac_2)^2\gamma + (a\beta)^2\eta(1+\frac{\gamma}{3}). \qquad (108)\end{aligned}$$

For the damped cosine parameters given in (58) – (62), we get $\lim_{\hat{s}\to 0}\widehat{T}(\hat{s}) \approx 219.2$. Note, however, that in (102) and (105) $\lim_{j\to i} c_{iy}c_{yj} = 0$ and $\lim_{j\to i} c_{yy} = 1$, so that $\lim_{j\to i}\mu_{ij}^{uu} = 1$.

In a manner completely analogous to our treatment of μ_{ij}^{uu}, we obtain the following formulas from (81) – (89) for the background error correlations:

$$\begin{aligned}
\mu_{ij}^{zz} &= \mu^{zz}(\tilde{s}_{ij}) && (109)\\
\mu_{ij}^{zu} &= -\operatorname{sign}(\varphi_j)a\beta\widetilde{R}(\tilde{s}_{ij})c_{iy} && (110)\\
\mu_{ij}^{uz} &= -\operatorname{sign}(\varphi_i)a\beta\widetilde{R}(\tilde{s}_{ij})c_{yj} && (111)\\
\mu_{ij}^{zv} &= +\operatorname{sign}(\varphi_j)a\beta\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{ix}}{\cos\varphi_j}\Big] && (112)\\
\mu_{ij}^{vz} &= +\operatorname{sign}(\varphi_i)a\beta\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{xj}}{\cos\varphi_i}\Big] && (113)\\
\mu_{ij}^{uu} &= +\operatorname{sign}(\varphi_i\varphi_j)\Big\{\widetilde{R}(\tilde{s}_{ij})c_{yy} + \widetilde{T}(\tilde{s}_{ij})c_{iy}c_{yj}\Big\} && (114)\\
\mu_{ij}^{uv} &= -\operatorname{sign}(\varphi_i\varphi_j)\Big\{\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{yx}}{\cos\varphi_j}\Big] + \widetilde{T}(\tilde{s}_{ij})c_{yj}\Big[\frac{c_{ix}}{\cos\varphi_j}\Big]\Big\} && (115)\\
\mu_{ij}^{vu} &= -\operatorname{sign}(\varphi_i\varphi_j)\Big\{\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{xy}}{\cos\varphi_i}\Big] + \widetilde{T}(\tilde{s}_{ij})\Big[\frac{c_{xj}}{\cos\varphi_i}\Big]c_{iy}\Big\} && (116)\\
\mu_{ij}^{vv} &= +\operatorname{sign}(\varphi_i\varphi_j)\Big\{\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{xx}}{\cos\varphi_i\cos\varphi_j}\Big] + \widetilde{T}(\tilde{s}_{ij})\Big[\frac{c_{xj}}{\cos\varphi_i}\Big]\Big[\frac{c_{ix}}{\cos\varphi_j}\Big]\Big\}. && (117)
\end{aligned}$$

C.4 Decoupling of the geostrophic assumption in the tropics

So far we have made uniform use of the geostrophic assumption as expressed in (80). This assumption is clearly not valid near the equator, since it requires a division by $f = 2\Omega\sin\varphi$. This leads to the ratio of wind to height errors, $\sigma^u/\sigma^z = g\beta/|f|$, in (97) becoming unbounded near the equator, and to discontinuities at the equator in the definitions of the correlation functions given in (110) - (117).

In equations (110) - (117) for the height-wind error correlations, the discontinuity of the factor sign(φ) at the equator is removed, as in Baker et al. (1987), by making the approximation

$$(a\beta)\,\mathrm{sign}(\varphi) \approx H(\varphi) = (a\beta)\,\mathrm{sign}(\varphi)\Big\{1 - \exp\Big(-\Big|\frac{\varphi}{\pi/9}\Big|\Big)\Big\}. \tag{118}$$

For the wind-wind correlations in (114) - (117) the factor sign($\varphi_i\varphi_j$) is replaced with 1, *i.e.*, the value it has when the two points are in the same hemisphere. Our final equations for the height-height, height-wind, and wind-wind error correlations for the upper-air analysis in the tropics ($15\mathrm{S} \le \varphi \le 15\mathrm{N}$) are[1]

$$\mu_{ij}^{zz} = \mu^{zz}(\tilde{s}_{ij}) \tag{119}$$

$$\mu_{ij}^{zu} = -H(\varphi_j)\widetilde{R}(\tilde{s}_{ij})c_{iy} \tag{120}$$

$$\mu_{ij}^{uz} = -H(\varphi_i)\widetilde{R}(\tilde{s}_{ij})c_{yj} \tag{121}$$

$$\mu_{ij}^{zv} = +H(\varphi_j)\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{ix}}{\cos\varphi_j}\Big] \tag{122}$$

$$\mu_{ij}^{vz} = +H(\varphi_i)\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{xj}}{\cos\varphi_i}\Big] \tag{123}$$

$$\mu_{ij}^{uu} = +\Big\{\widetilde{R}(\tilde{s}_{ij})c_{yy} + \widetilde{T}(\tilde{s}_{ij})c_{iy}c_{yj}\Big\} \tag{124}$$

$$\mu_{ij}^{uv} = -\Big\{\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{yx}}{\cos\varphi_j}\Big] + \widetilde{T}(\tilde{s}_{ij})c_{yj}\Big[\frac{c_{ix}}{\cos\varphi_j}\Big]\Big\} \tag{125}$$

$$\mu_{ij}^{vu} = -\Big\{\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{xy}}{\cos\varphi_i}\Big] + \widetilde{T}(\tilde{s}_{ij})\Big[\frac{c_{xj}}{\cos\varphi_i}\Big]c_{iy}\Big\} \tag{126}$$

$$\mu_{ij}^{vv} = +\Big\{\widetilde{R}(\tilde{s}_{ij})\Big[\frac{c_{xx}}{\cos\varphi_i\cos\varphi_j}\Big] + \widetilde{T}(\tilde{s}_{ij})\Big[\frac{c_{xj}}{\cos\varphi_i}\Big]\Big[\frac{c_{ix}}{\cos\varphi_j}\Big]\Big\} \tag{127}$$

Note that if we set $\lambda_i = \lambda_i^*$ and $\varphi_i = \varphi^* = -90°$ or $90°$, and let $(\varphi_j, \lambda_j) \to (\varphi^*, \lambda_j^*)$ we obtain the following limiting values $\mu_{ij}^{zz} = 1$, $\mu_{ij}^{zu} = \mu_{ij}^{uz} = \mu_{ij}^{zv} = \mu_{ij}^{vz} = 0$, $\mu_{ij}^{uu} = \mu_{ij}^{vv} = \cos(\lambda_i^* - \lambda_j^*)$, $\mu_{ij}^{uv} = \mathrm{sign}(\varphi^*)\sin(\lambda_i^* - \lambda_j^*)$ and $\mu_{ij}^{vu} = -\,\mathrm{sign}(\varphi^*)\sin(\lambda_i^* - \lambda_j^*)$. Contour plots of the correlation functions defined in (120) - (127) for $\lambda_i = 0°$ and $\varphi_i \in \{10°, 40°, 90°\}$ are shown as functions of φ_j and λ_j in figures 7—12.

The wind component error standard deviations given by equation (97) are modified in the tropics as follows. In the tropical band, $15°\mathrm{S} \le \varphi \le 15°\mathrm{N}$, σ_i^u and σ_j^v are assigned the value $\sigma_{trop}^{u,v}$, given in Table 5. At points in the subtropical bands, $25°\mathrm{S} \le \varphi < 15°\mathrm{S}$ and $15°N < \varphi \le 25°\mathrm{N}$, σ_i^u and σ_j^v are defined by

$$(\sigma_i^u)^2 = (\sigma_i^v)^2 = \Big[\frac{25 - |\varphi|}{10}\Big](\sigma_{trop}^{u,v})^2 + \Big[\frac{|\varphi| - 15}{10}\Big]\Big\{\frac{g^2\beta^2(\sigma_i^z)^2}{|f_{25}|^2}\Big\}, \tag{128}$$

where $f_{25} = 2\Omega\sin(25°)$.

---

[1] Inadvertently, the frozen system that produced the five-year reanalysis did not remove the discontinuity across the equator associated with the factor sign(φ) in the wind-wind correlations. This problem has been corrected in subsequent versions of the system (see section 9).

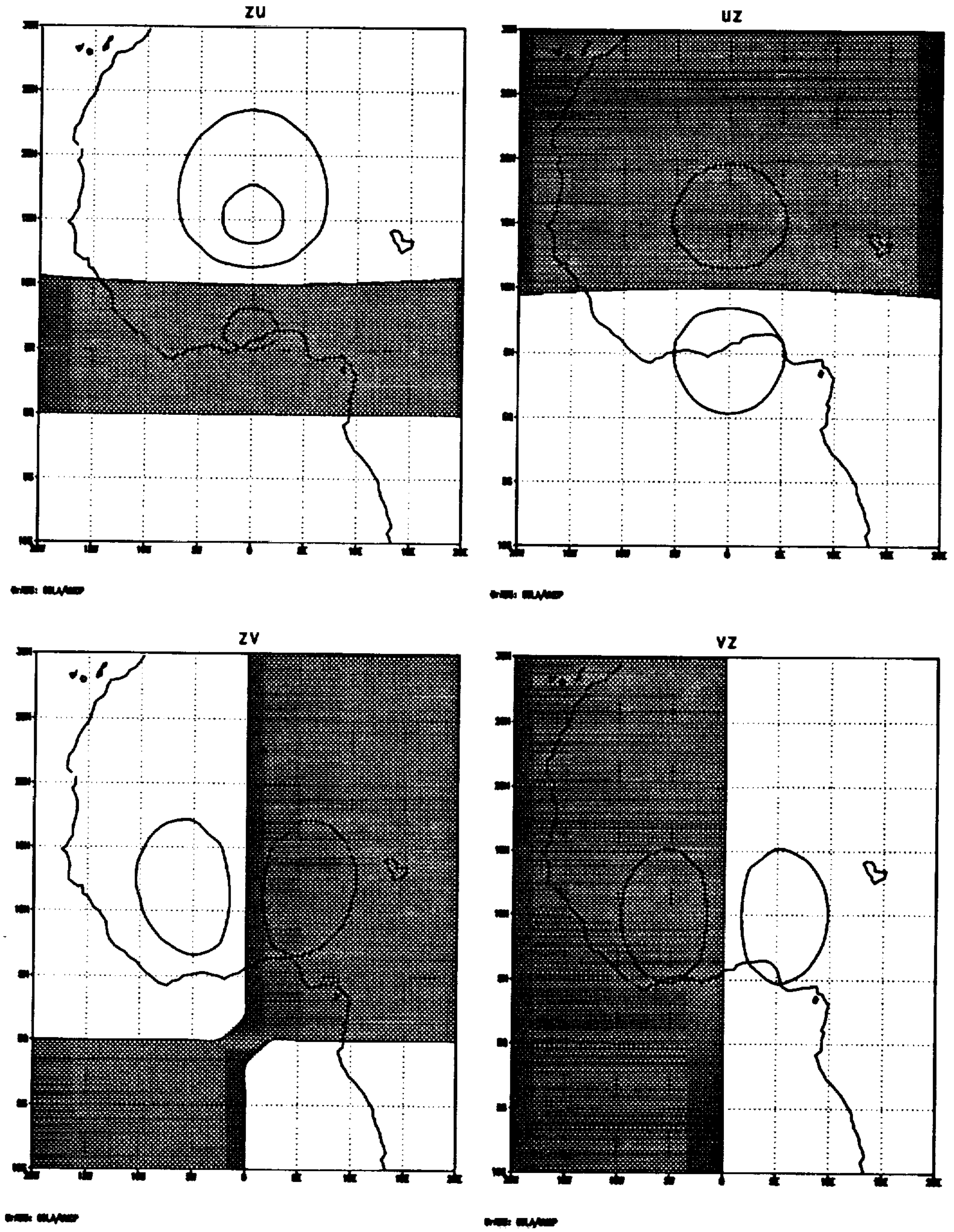

Figure 7: Graphs of the functions μ_{ij}^{zu}, μ_{ij}^{uz}, μ_{ij}^{zv}, and μ_{ij}^{vz} for $\varphi_i = 10^\circ$ and $\lambda_i = 0^\circ$.

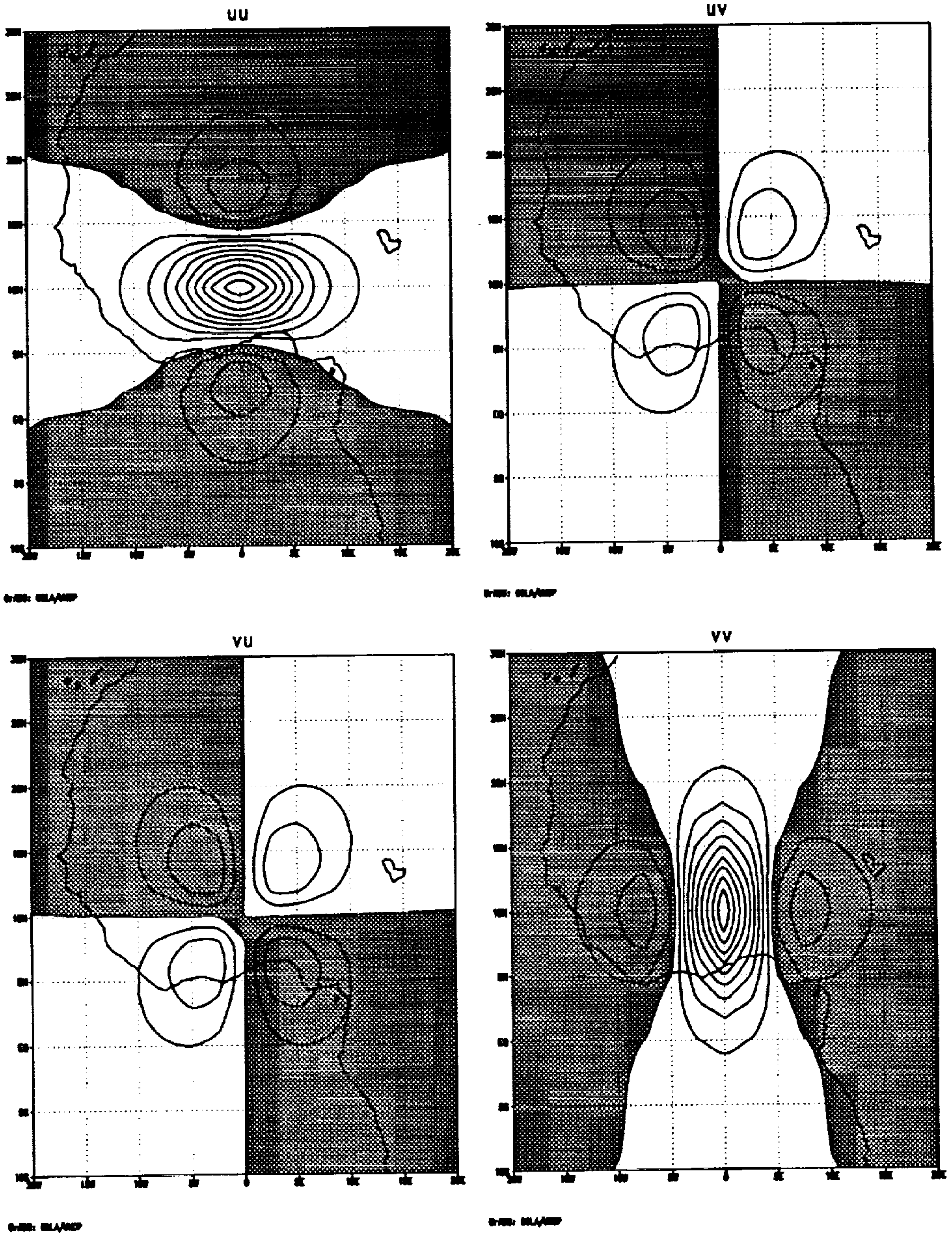

Figure 8: Graphs of the functions μ_{ij}^{uu}, μ_{ij}^{uv}, μ_{ij}^{vu}, and μ_{ij}^{vv} for $\varphi_i = 10^\circ$ and $\lambda_i = 0^\circ$.

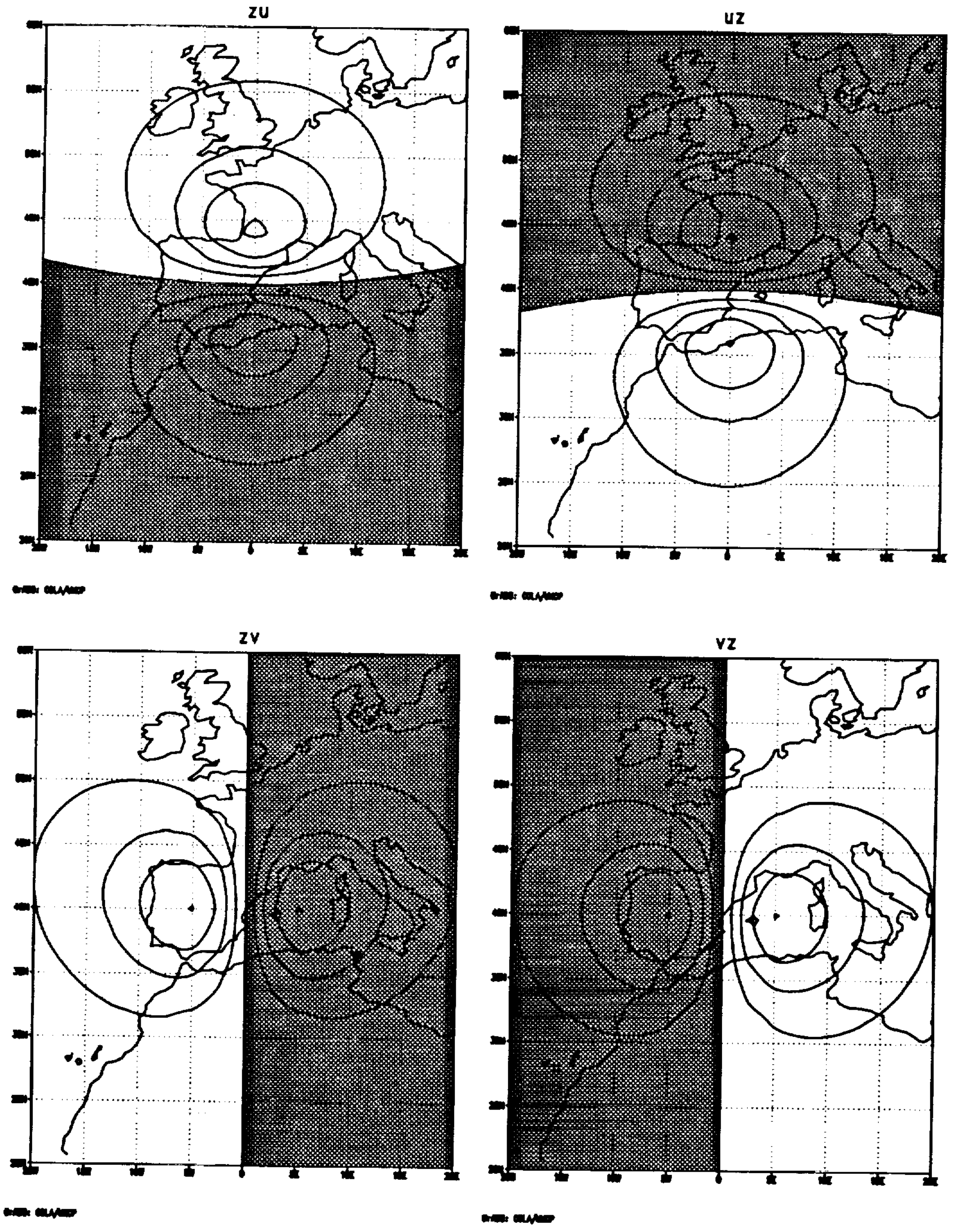

Figure 9: Graphs of the functions μ_{ij}^{zu}, μ_{ij}^{uz}, μ_{ij}^{zv}, and μ_{ij}^{vz} for $\varphi_i = 40^\circ$ and $\lambda_i = 0^\circ$.

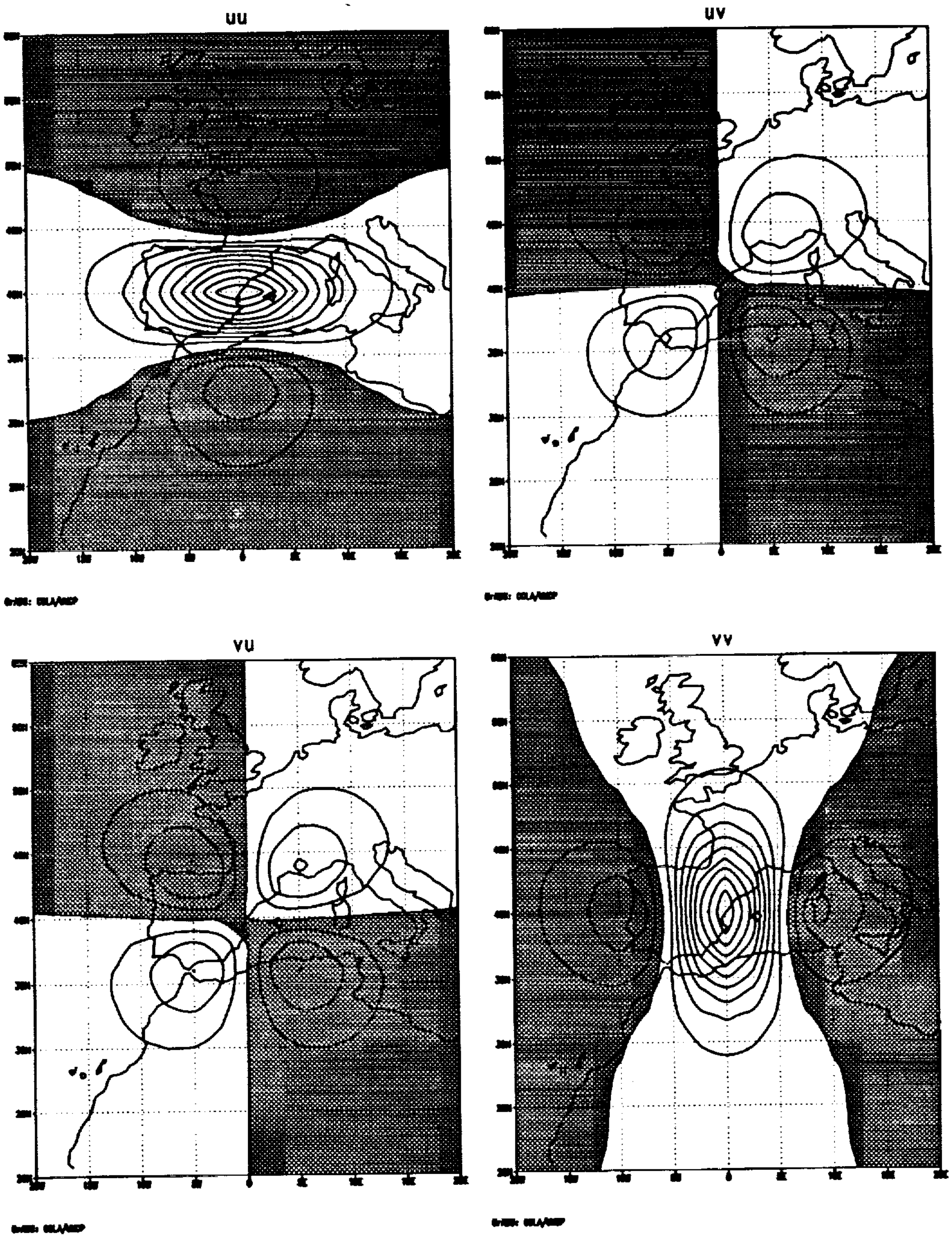

Figure 10: Graphs of the functions μ_{ij}^{uu}, μ_{ij}^{uv}, μ_{ij}^{vu}, and μ_{ij}^{vv} for $\varphi_i = 40^\circ$ and $\lambda_i = 0^\circ$.

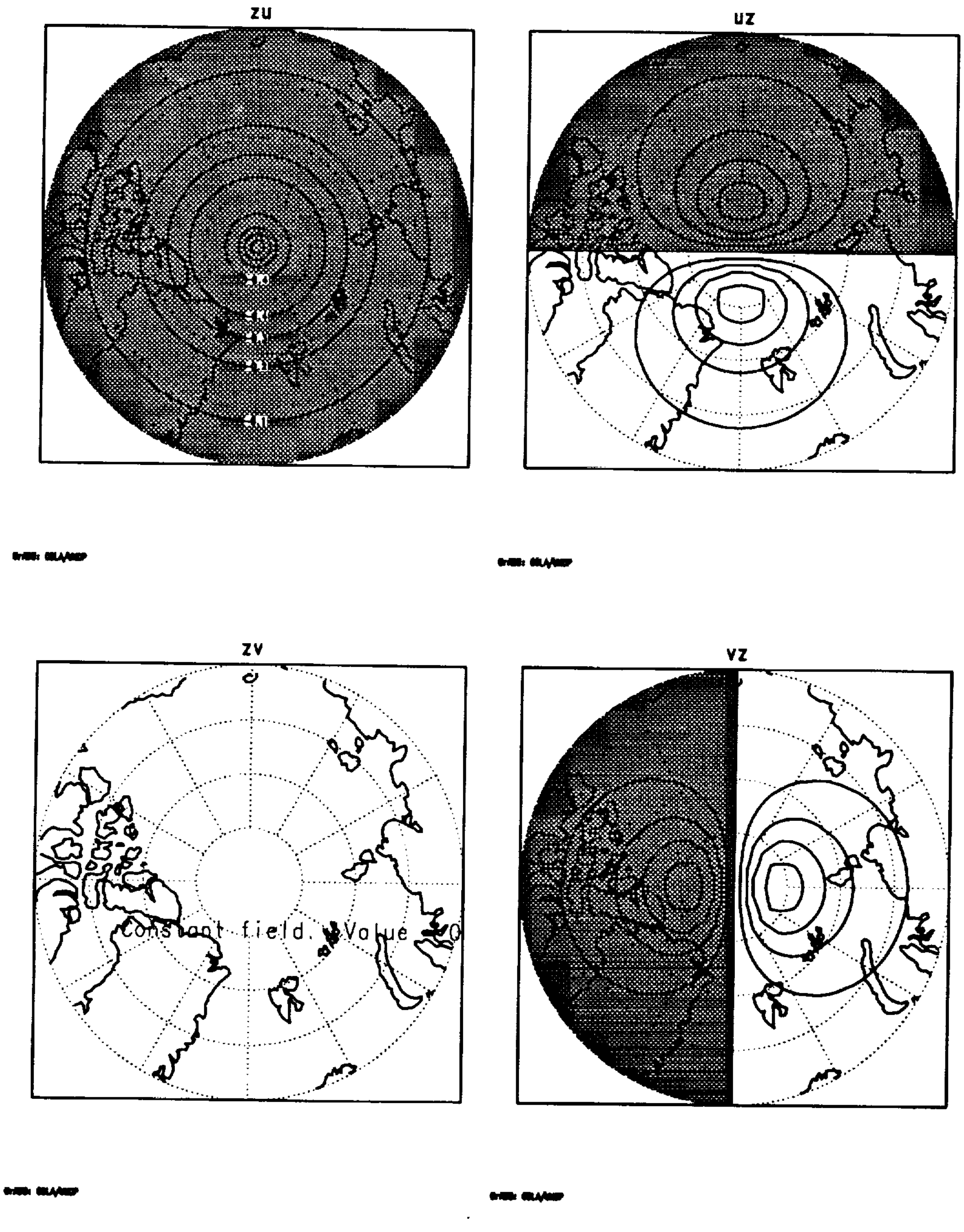

Figure 11: Graphs of the functions μ_{ij}^{zu}, μ_{ij}^{uz}, μ_{ij}^{zv}, and μ_{ij}^{vz} for $\varphi_i = 90^\circ$ and $\lambda_i = 0^\circ$.

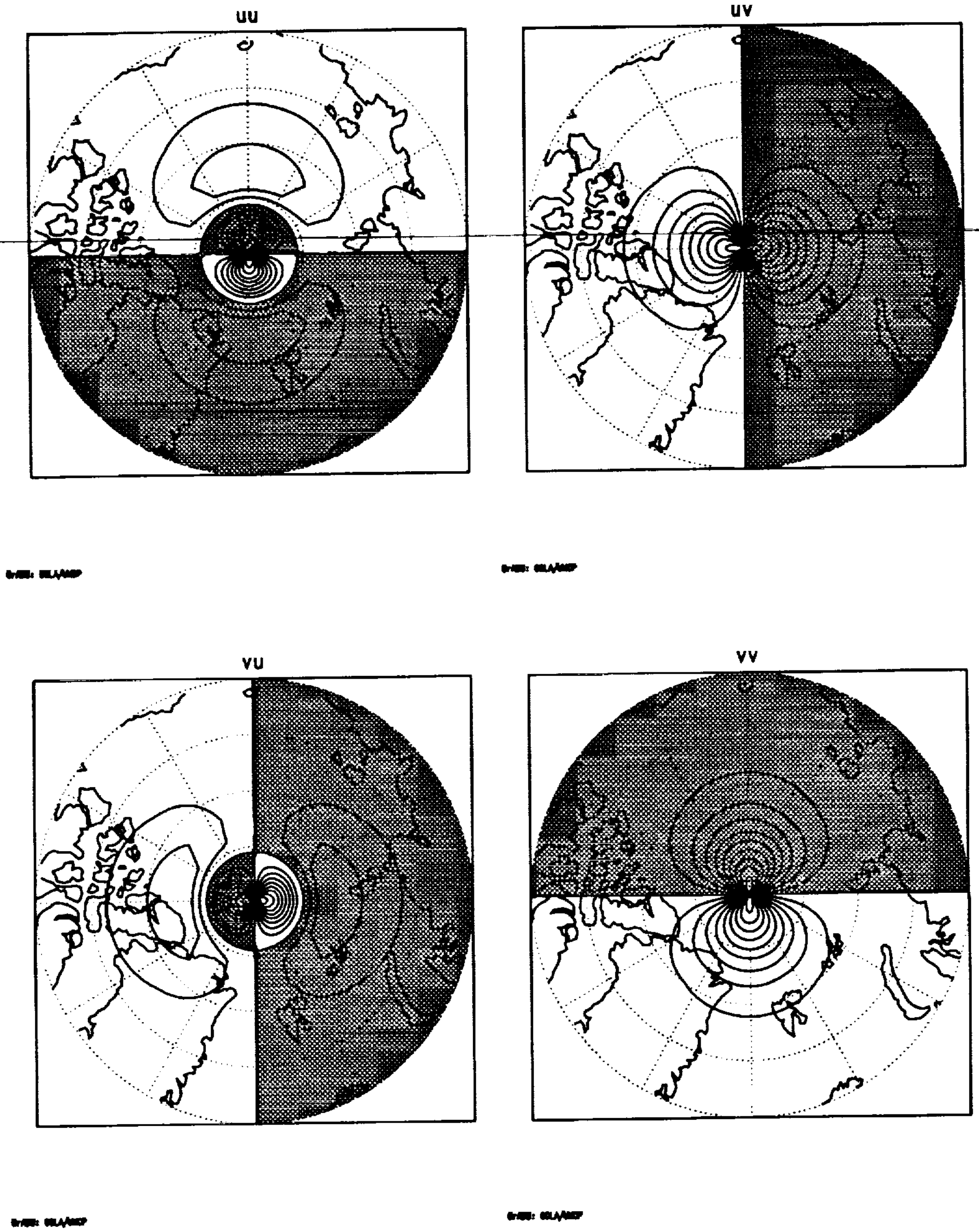

Figure 12: Graphs of the functions μ_{ij}^{uu}, μ_{ij}^{uv}, μ_{ij}^{vu}, and μ_{ij}^{vv} for $\varphi_i = 90°$ and $\lambda_i = 0°$.

Appendix D: List of Acronyms

ACARS	ARINC Communications Addressing and Reporting System
AGCM	Atmospheric General Circulation Model
AIDS	Aircraft Integrated Data System
ARINC	Aeronautical Radio, Incorporated
ASDAR	Aircraft to Satellite Data Relay
CFL	Courant-Friedrichs-Levy (commonly used computational stability condition)
DAAC	Distributed Active Archive Center
DAO	Data Assimilation Office
DAS	Data Assimilation System
EOS	Earth Observing System
GCM	General Circulation Model
GEOS-1 DAS	Goddard EOS-Version 1 Data Assimilation System
GEOS-1 GCM	Goddard EOS-Version 1 General Circulation Model
GLA	Goddard Laboratory for Atmospheres
HIRS	High Resolution Infrared Sounder
IAU	Incremental Analysis Update
LAPACK	Linear Algebra Package
MLS	Microwave Limb Sounder
MSU	Microwave Sounding Unit
NAVAIDS	Navigation Aids (in Table 1 this refers to a special class of pilot balloons)
NESDIS	National Environmental Satellite, Data, and Information Service
NMC	National Meteorological Center
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
OI	Optimal Interpolation
PBL	Planetary Boundary Layer
RAS	Relaxed Arakawa-Schubert
RMS	Root Mean Square
SSM/I	Special Sensor Microwave Imager
SSU	Stratospheric Sounding Unit
STRATAN	Stratospheric Analysis system
TIROS	Television Infrared Observing Satellite
TOVS	TIROS Operational Vertical Sounder
TOVS A	TOVS clear sky retrieval
TOVS B	TOVS partly cloudy sky retrieval
TOVS C	TOVS cloudy sky retrieval
UARS	Upper Atmospheric Research Satellite
UTC	Universal Time Coordinated

References

Anthes, R. A., 1974: Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.*, **31**, 702–719.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Comput. Phys.*, **1**, 119-143.

———, and V.R. Lamb, 1981: A potential enstrophy and energy conserving scheme for the shallow water equations. *Mon. Wea. Rev.*, **109**, 18-36.

———, and W. H. Schubert, 1974: Interaction of a cumulus ensemble with the large–scale environment, Part I. *J. Atmos. Sci.*, **31**, 674–701.

———, and M. J. Suarez, 1983: Vertical differencing of the primitive equations in sigma coordinates. *Mon. Wea. Rev.*, **111**, 34–45.

Baker, W. E., S. C. Bloom, J. S. Woollen, M. S. Nestler, E. Brin, T. W. Schlatter, and G. W. Branstator, 1987: Experiments with a three–dimensional statistical objective analysis scheme using FGGE data. *Mon. Wea. Rev.*, **115**, 273–296.

Bloom S. C., L. Takacs and E. Brin, 1991: A scheme to incorporate analysis increments gradually in the GLA assimilation system, *Preprint Volume, Ninth Conference on Numerical Weather Prediction*, October 14–18, 1991, Denver, CO, AMS, 110–112.

Burridge, D. M. and J. Haseler, 1977: A model for medium range weather forecasting. Tech. Rep. No. 4, European Centre for Medium Range Weather Forecasts, Bracknell, Berks., U. K., 46pp.

Chou, M. D., 1984: Broadband water vapor transmission functions for atmospheric IR flux computations. *J. Atmos. Sci.*, **41**, 1775–1778.

———, and L. Peng, 1983: A parameterization of the absorption in the 15μ CO_2 spectral region with application to climate sensitivity studies. *J. Atmos. Sci.*, **40**, 2183–2192.

Clarke, R. H., 1970: Observational studies in the atmospheric boundary layer. *Quart. J. Roy. Meteor. Soc.*, **96**, 91-114.

Cohn, S. E., 1993: Dynamics of short–term univariate forecast errors. *Mon. Wea. Rev.*, **121**, 3123–3149.

———, and D.Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757- 1785.

———, and L. L. Morone, 1984: The effect of horizontal gradients of height–field forecast error variances upon OI forecast error statistics. *Office Note 296*, National Meteorological Center, Washington, DC, October 1984, 37 pp.

Coy, L., R. Rood, P. Newman, 1994: A Comparison of Winds from the STRATAN Data Assimilation System to Balanced Wind Estimates. *J. Atmos. Sci.*, **51**, 2309–2315.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Davies, R., 1982: Documentation of the solar radiation parameterization in the GLAS climate model. NASA Tech. Memo. 83961, 57pp., Goddard Space Flight Center, Greenbelt, MD 20771.

Helfand, H. M. and J. C. Labraga, 1988: Design of a non-singular level 2.5 second-order closure model for the prediction of atmospheric turbulence. *J. Atmos. Sci.*, **45,** 113–132.

———, M. Fox-Rabinovitz, L. Takacs, and A. Molod, 1991: Simulation of the planetary boundary layer and turbulence in the GLA GCM. Proceedings of the AMS Ninth Conference on Numerical Weather Predicition, 21-25 October 1991, Denver, CO, 514-517.

———, and S. D. Schubert, 1994: Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. *J. Climate*, **submitted.**

Hollingsworth, A., P. Kållberg,, V. Renner, and D.M. Burridge, 1983: An internal symmetric computational instability, *Quart. J.R. Met. Soc.*, **109**, 417-428.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* New York, Academic Press, 376 pp.

Joseph, J. H., W. J. Wiscombe and J. E. Weinman, 1976: The delta–Eddington approximation for radiative flux transfer. *J. Atmos. Sci.*, **33,** 2452–2459.

King, M. D. and Harshvardhan, 1986: Comparitive accuracy of selected multiple scattering approximations. *J. Atmos. Sci.*, **43,** 784–801.

Lacis, A. A. and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the Earth's atmosphere. *J. Atmos. Sci.*, **31,** 118–133.

Lönnberg, P. and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from rawinsonde data. Part II: the covariance of height and wind errors. *Tellus*, **38A**, 137–161.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Met. Soc.*, **112,** 1177–1194.

Mitchell, Herschel L., Cecilien Charette, Steven J. Lambert, Jacques Halle and Clement Chouinard, 1993: The Canadian global data assimilation system: Description and evaluation. *Mon. Wea. Rev.*, **121**, 1467-1492.

Moorthi, S. and M. J. Suarez, 1992: Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120,** 978–1002.

Morone, L. L., and S. E. Cohn, 1985: The spatial variability of height–field forecast errors and its effect upon OI forecast error statistics. *Proc. of Seventh Conf. on Numerical Weather Prediction*, June 17–20, 1985, Montreal, AMS, 27–33.

National Research Council, 1991: Four–dimensional model assimilation data: A strategy for the earth system sciences. *Report from the Panel on Model–Assimilated Data Sets for Atmospheric and Oceanic Research*, National Academy Press, Washington, D.C., 1991, 78 pp.

Panofsky, H. A., 1973: Tower Micrometeorology. In Workshop on Micrometeorology, D. A. Haugen (ed.), American Meteorological Society, Boston, 392 pp.

Pfaendtner, J. and N. Sivakumaran, 1991: Optimal Interpolation with Dynamically Dependent Error Variance Estimates. *Preprint Volume, Ninth Conference on Numerical Weather Prediction*, October 14–18, 1991, Denver, CO, AMS, 103–106.

Rodgers, C. D., 1968: Some extensions and applications of the new random model for molecular band transmission. *Q. J. R. Met. Soc.*, **94**, 99–102.

Rosenfield, J. E., M. R. Schoeberl and M. A. Geller, 1987: A computation of the stratospheric diabatic circulation using an accurate radiative transfer model. *J. Atmos. Sci.*, **44**, 859–876.

Sadourney, R., 1975: The dynamics of finite difference models of the shallow water equations. *J. Atmos. Sci.*, **32**, 680–689.

Schubert, S. D., R. B. Rood and J. Pfaendtner, 1993: An assimilated data set for earth science applications. *Bul. Amer. Meteor. Soc.*,**74**, 2331–2342.

Seablom, M. S., 1990: Experiments with new quality control techniques in the NASA Optimum interpolation analysis system. it Preprints, International Symposium on Assimilation of Observations in Meteorology and Oceanography, Clermont–Ferrand, France WMO, 628- 630.

———, J. W. Pfaendtner and P. E. Piraino, 1991: Quality control techniques for the interactive GLA retrieval/assimilation system. *Preprint Volume, Ninth Conference on Numerical Weather Prediction*, October 14–18, 1991, Denver, CO, AMS, 28–29.

Shaw, D. B., P. Lönnberg, A. Hollingsworth, and P. Unden, 1986: Data assimilation 1984/1985 revisions of the ECMWF mass and wind analysis. *Q. J. R. Met. Soc.*, **113**, 533-1566.

Stauffer, D. R. and N. L. Seaman, 1990: Use of four–dimensional data assimilation in a limited–area mesoscale model. Part I: Experiments with synoptic–scale data. *Mon. Wea. Rev.*, **118**, 1250–1277.

Suarez, M. and L. Takacs, 1995: Documentation of the Aries/GEOS Dynamical Core Version 2, NASA Tech Memorandum 104606–Vol (in press), Goddard Space Flight Center, Greenbelt, MD.

Sud, Y. and A. Molod, 1988: The roles of dry convection, cloud–radiation feedback processes and the influence of recent improvements in the parameterization of convection in the GLA AGCM. *Mon. Wea. Rev.*, **116**, 2366–2387.

Takacs, L., A. Molod, and T. Wang, 1994: Documentation of the Goddard Earth Observing System (GEOS) General Circulation Model - Version 1, NASA Tech. Memorandum 104606, Vol. 1, NASA, Goddard Space Flight Center, Greenbelt, MD.

Takacs, L.L., and M.J. Suarez, 1995: Dynamical Climatology of the GEOS GCM and its Dependence on Horizontal Differencing. *In preparation.*

Takano, K. and M.G. Wurtele, 1982: A fourth-order energy and potential enstrophy conserving difference scheme, Air Force Geophysics Laboratory Report, AFGL-TR-82-0205, (NTIS AD-A126626), AFGL Hanscom AFB, Massachusetts, 01731, 85 pp.

Thiebaux, J., 1975: Experiments with correlation representations of objective analysis. *Mon. Wea. Rev.*, **103,** 617–627.

Todling, R. and S. E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.*, , **122**, 2530–2557.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** January 1995	**3. REPORT TYPE AND DATES COVERED** Technical Memorandum

4. TITLE AND SUBTITLE
Technical Report Series on Global Modeling and Data Assimilation
Volume 4 - Documentation of Goddard Earth Observing System
(GEOS) Data Assimilation System - Version 1

5. FUNDING NUMBERS
C - NAS5-32332
Code 910.3

6. AUTHOR(S)
James Pfaendtner, Stephen Bloom, David Lamich, Michael Seablom,
Meta Sienkiewicz, James Stobie, and Arlindo da Silva

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS (ES)

Laboratory for Atmospheres
Data Assimilation Office
Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PEFORMING ORGANIZATION REPORT NUMBER
95B00040

9. SPONSORING / MONITORING ADGENCY NAME(S) AND ADDRESS (ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING ADGENCY REPORT NUMBER
TM-104606, Vol. 4

11. SUPPLEMENTARY NOTES
Pfaendtner and da Silva: Goddard Space Flight Center, Greenbelt, Maryland;
Bloom, Lamich, Seablom, Sienkiewicz, and Stobie: General Sciences Corporation, Laurel, Maryland

12a. DISTRIBUTION / AVAILABILITY STATMENT
Unclassified - Unlimited
Subject Category 46
This publication is available from the NASA Center for AeroSpace Information, 800 Elkridge Landing Road, Linthicum Heights, MD 21090-2934, (301)621-0390.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report describes the analysis component of the Goddard Earth Observing System, Data Assimilation System, Version 1 (GEOS-1 DAS). The general features of the data assimilation system are outlined, followed by a thorough description of the statistical interpolation algorithm, including specification of error covariances and quality control of observations. We conclude with a discussion of the current status of development of the GEOS data assimilation system. The main components of GEOS-1 DAS are an atmospheric general circulation model and an *Optimal Interpolation* algorithm. The system is *cycled* using the Incremental Analysis Update (IAU) technique in which analysis increments are introduced as time independent forcing terms in a forecast model integration. The system is capable of producing dynamically balanced states without the explicit use of initialization, as well as a time-continuous representation of non-observables such as precipitation and radiational fluxes. This version of the data assimilation system was used in the five-year reanalysis project completed in April 1994 by Goddard's Data Assimilation Office (DAO) (Schubert et al., 1993) Data from this reanalysis are available from the Goddard Distributed Active Center (DAAC), which is part of NASA's Earth Observing System Data and Information System (EOSDIS). For information on how to obtain these data sets, contact the Goddard DAAC at (301) 286-3209, EMAIL daac@gsfc.nasa.gov, or consult DAO's Mosaic Home Page **URL http://hera.gsfc.nasa.gov/dao.home_page.html.**
This document is available electronically via anonymous ftp from **dao.gsfc.nasa.gov**, directory **pub/tech_memos**, file **volume_4.ps.Z.**

14. SUBJECT TERMS
Optimum interpolation, Data assimilation, Analysis, EOS, Statistical interpolation, Observations, GEOS

15. NUMBER OF PAGES
58

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)

National Aeronautics and
Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

Official Business
Penalty for Private Use, $300