U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 379

"THE NATIONAL METEOROLOGICAL CENTER'S SPECTRAL STATISTICAL INTERPOLATION ANALYSIS SYSTEM"

DAVID F. PARRISH AND JOHN C. DERBER
DEVELOPMENT DIVISION

APRIL 1991

THE NATIONAL METEOROLOGICAL CENTER'S SPECTRAL STATISTICAL INTERPOLATION ANALYSIS SYSTEM

David F. Parrish and John C. Derber
Development Division
National Meteorological Center
Washington, DC 20233

THE NATIONAL METEOROLOGICAL CENTER'S SPECTRAL STATISTICAL INTERPOLATION ANALYSIS SYSTEM

ABSTRACT

At the National Meteorological Center (NMC), a new analysis system is being extensively tested for possible use in the operational Global Data Assimilation System. This analysis system is called the Spectral Statistical Interpolation (SSI) Analysis system because the spectral coefficients used in the NMC spectral model are analysed directly using the same basic equations as statistical (optimal) interpolation. Results from several months of parallel testing with the NMC spectral model have been very encouraging. Favorable features include smoother analysis increments, greatly reduced changes from initialization, and significant improvement of 1-5 day forecasts. Although the analysis is formulated as a variational problem, the objective function being minimized is formally the same one that forms the basis of all existing optimal interpolation schemes. This objective function is a combination of forecast and observation deviations from the desired analysis, weighted by the inverses of the corresponding forecast and observation error covariance matrices. There are two principal differences in how the SSI implements the minimization of this functional as compared to the current OI used at NMC. First, the analysis variables are spectral coefficients instead of gridpoint values. Second, all observations are used at once to solve a single global problem. No local approximations are made, and there is no special data selection. Because of these differences, it is straightforward to include unconventional data, such as radiances, in the analysis. Currently temperature, wind, surface pressure, mixing ratio and SSM/I total precipitable water are used as the observation variables. Soon to be added is the scatterometer surface winds. In this paper, we provide a detailed description of the SSI and present a few results.

1. INTRODUCTION

Most of the major operational NWP centers in the world now assimilate observations into forecast models using some form of statistical or optimum interpolation (OI). These systems are based on the ideas of Gandin (1963) and Eliassen (1954, reproduced in Ghil et al. eds), who each introduced statistical considerations to the meteorological data assimilation community. While these OI analysis systems are in widespread use, much interest is now centered on variational methods, in particular the 4-dimensional "adjoint" procedures (Le Dimet and Talagrand, 1986, Lewis and Derber, 1985, Talagrand and Courtier, 1987, Courtier and Talagrand, 1987, Derber, 1989). OI is derived in terms of probability and statistics, while variational methods are based on combining model dynamics with data with the relative weighting defined in an *ad hoc* manner. However, as is illustrated in a review of analysis methods by Lorenc (1986), variational and statistical analysis methods do have a common basis and can be made equivalent by proper definition of weights. Using a Bayesian approach, Lorenc (1986) derived an objective function which can be used as the starting point for both existing OI procedures and any variational schemes. The derived objective function is a combination of deviations of the desired analysis from a forecast and from observations, weighted by the inverses of the corresponding forecast and observation error covariance matrices. The differences between schemes reduce to the specific practical approximations made in the solution of the analysis problem. Earlier references to the work leading to this system refer to the analysis technique as spectral optimum interpolation (Parrish, 1987, 1988, Parrish and Derber, 1988). However, to reduce confusion and to emphasize the differences from conventional implementations based on local approximations to the full problem, we have chosen to refer to the new analysis technique as Spectral Statistical Interpolation (SSI).

The idea of directly analysing data in terms of spectral coefficients is not new. The Hough spectral analysis (Flattery, 1970), used operationally at NMC from 1974–79, also minimized a globally defined objective function. The SSI system described here has some similarities to the Hough analysis, but differs in the use of a background (first guess) and the statistical considerations of OI. In fact, the first two-dimensional version of the SSI (Parrish, 1988) also used a Hough function representation of the analysis variables, but the current version of the SSI system uses a more satisfactory basis for the analysis variables.

Compared to the existing NMC OI analysis systems (Dey and Morone, 1985, DiMego, 1988, Kanamitsu, 1989), there are two principal differences in how SSI approximates the minimum of the desired objective function. First, the analysis variables are closely related to the sigma coordinate coefficients of the spherical harmonic expansions of vorticity, divergence, temperature, logarithm of surface pressure, and mixing ratio used by the NMC spectral model. The operational OI, on the other hand, uses grid point values of heights, winds, and mixing ratio on isobaric surfaces as the analysis variables. Because the analysis variables are spectral, the forecast error covariance must be defined in terms of these spectral variables. Some thought was given to this problem by Halem and Kalnay (1983), but Phillips (1986) was the first to investigate the behavior of forecast errors in terms of normal modes of a simple model. His work provided the initial inspiration for the SSI by demonstrating a simple plausible model for forecast error covariance in terms of mode (or spectral) variables. By assuming errors to be equally distributed among model slow modes only, and uncorrelated between modes, he was able to derive physical space correlations which gave surprisingly good agreement with empirically derived results from the ECMWF model (Hollingsworth and Lönnberg, 1986, Lönnberg and Hollingsworth, 1986). In this paper, the error covariance model is slightly less restrictive than the one used by Phillips (1986), with the analysis not restricted to slow modes and not equally distributed among the modes. However, the error statistics in this paper are currently still assumed to be uncorrelated between modes. (This assumption results in unrealistic background error variances, a weakness for SSI that will be addressed in future investigations. See discussion.)

The second principal difference between the SSI and the operational OI is that in the new system, all observations are used at once to perform the analysis globally. Because the SSI analysis variables are spectrally defined, the analysis must be solved as a single problem and not approximated locally as is done in all current operational systems. Also, the analysis increments can then be found directly with no intermediate solution for weights. Performing the analysis globally has the advantages of not producing discontinuities in the solution resulting only from data selection and eliminating the need for the expensive procedure of data sorting and selection. Traditional OI techniques could theoretically use all observations at once. However, this is generally considered to be too computationally unstable and expensive because it is necessary to invert a matrix of interobservation correlations, which has a dimension equal to the number of observations.

These differences confer several additional advantages to the SSI system. First, since the analysis is performed globally, no difficulties are encountered when using temperature rather than height observations. There is usually a problem in relating the changes in the temperature field to a balanced change in the velocity field in operational analysis systems because the analyses are quasi-horizontal. For this reason, height is frequently the analysis variable. Since the analysis is done 3-dimensionally in the SSI, the corresponding height changes can be calculated from the temperature (and surface pressure) changes to create a corresponding change to the velocity field. In fact, it is much more straightforward to include most types of observations in the SSI system. This will also be demonstrated by the straightforward inclusion of the Special Sensor Microwave/Imager (SSM/I) total precipitable water in the moisture analysis. Second, it is possible to obtain an analysis increment which looks deceptively smooth, but actually yields as good or better rms fit to the observations when compared to the local method of the operational OI. This, combined with a better overall balance, results also in significantly smaller changes caused by initialization.

To be fair, we must point out that the current implementation of the SSI system does not actually use the observations directly. Primarily for reasons of computational convenience, the observation residuals are combined to "superobs" at the closest spectral model grid points. But the observation residuals are first formed by interpolating the background fields directly to the location of each observation. In practice, some form of superobing will always be required to deal with representativeness error. Superobing to the model grid seems like a natural thing to do.

In the next section, we present an outline of SSI, starting with the objective function. Section 3 contains a detailed discussion of the representation of forecast and observation error covariances. Next follow some results of individual analyses and long term data–assimilation runs and resulting forecasts. Finally, the results and present plans for future work will be summarized.

2. THE ANALYSIS PROCEDURE

Both SSI as well as conventional OI minimize the same objective function (Kimeldorf and Wahba, 1970, Lorenc, 1986). This objective function is given by:

$$J = \frac{1}{2}[x^T B^{-1} x + (Lx - y)^T (F + O)^{-1} (Lx - y)] \tag{2.1}$$

where

x is an N component vector of analysis increments,
B is the NxN forecast error covariance matrix,
O is the MxM observational error covariance matrix,
F is the MxM representativeness error covariance matrix,
L is a linear tranformation operator which converts the analysis variables to the observation type and location. Note that the linearity in L is not required but is currently assumed for simplicity.
y is an M component vector of observational increments,
i.e., $y = y_{obs} - Lx_{guess}$,
N is the number of degrees of freedom in the analysis,
M is the number of observations.

An expression for the minimizing solution can be found by differentiating J with respect to x and setting the result equal to zero, which gives

$$B^{-1}x + L^T(F + O)^{-1}(Lx - y) = 0 \ . \tag{2.2}$$

Multiplying through by B and rearranging the terms results in

$$(I + BL^T(F + O)^{-1}L)x = BL^T(F + O)^{-1}y \ . \tag{2.3}$$

To ensure symmetry in the matrix which multiplies x (required by the solution algorithm), a matrix C is defined such that $C^T C = B$. Since B will be a diagonal matrix with positive elements along the diagonal (see next section), C is also a diagonal matrix with diagonal elements equal to the square root of the diagonal elements of B. Defining a new variable $w = C^{-1}x$ and multiplying (2.3) by C^{-1} results in

$$(I + CL^T(F + O)^{-1}LC)w = CL^T(F + O)^{-1}y \tag{2.4}$$

or

$$Aw = f \tag{2.5}$$

where

$$A = I + CL^T(F+O)^{-1}LC \tag{2.6}$$

and

$$f = CL^T(F+O)^{-1}y \tag{2.7}$$

This is the primary analysis equation that must be solved to produce the analysis. Note that once w is found, the actual analysis increments are found by multiplying by C. Note also that the scaling by C to obtain (2.5) has been done to improve the condition of the matrix A, making (2.5) easier to solve.

The above derivation is general and is applicable to any analysis variable and any observations (as long as pseudo–observations can be derived from the analysis variables through the L operator). The ideal analysis variables would be the amplitudes of the eigenvectors of the background error covariance matrix specified in terms of the model variables. In this case the B matrix would then be diagonal with the eigenvalues of the background error covariance matrix in terms of the model variables along the diagonal. But since the computational expense of creating these eigenvectors is prohibitive, another approximate representation has been chosen.

In defining the analysis variables, the balanced components of the mass and momentum fields have been combined into a single variable. This allows the balance between the mass and momentum fields to be implicitly included. Following the convention used for normal mode initialization of sigma coordinate models, a mass variable is defined as

$$H = GT + \frac{R\overline{T}}{g} ln(p_{sfc}) \tag{2.8}$$

where G is a finite difference representation of the hydrostatic integral (see appendix A), and $\overline{T}$ is a mean temperature profile which depends only on sigma. H is partitioned into balanced (slow) and unbalanced (fast) parts, H_s and H_f respectively, using the linear balance equation to define H_s in terms of the relative vorticity. The analysis variables are then given by the relative vorticity ζ, the divergence D, the unbalanced height variable H_f and the specific humidity q. The transformation from these analysis variables to the model variables, ζ, D, T, $\ln(p_{sfc})$, and q is given by

$$\zeta = \zeta \ , \tag{2.9}$$

$$D = D \ , \tag{2.10}$$

$$H = \frac{a^2}{g}(\nabla \cdot f\nabla)(\nabla^{-2}\zeta) + H_f \ , \tag{2.11}$$

$$T = QH \ , \tag{2.12}$$

and

$$\ln(p_{sfc}) = W^T H \ . \tag{2.13}$$

In (2.9–13) f is the coriolis parameter, a the radius of the earth, and ∇^{-2} is unambigously defined because the domain is the whole earth. Q is a square matrix of dimension equal to the number of sigma layers in the model, and W is a vector of similar length. Q and W arise from attempting to invert (2.8). Because H is defined at the same levels as T, one degree of freedom is lost going from T and $\ln(p_{sfc})$ to H using (2.8). To recover the extra variable, (2.8) is inverted subject to the constraint that the second derivative of T is minimized at each layer in the vertical(see appendix A). This approximation appears to work well. Finally, the variables ζ, D, H, and q are represented in the horizontal with the same spherical harmonic expansions as used by the spectral model.

Initially, the discretization in the vertical was done in terms of the model's vertical modes. The intent was to simulate as closely as possible the decomposition into slow and fast components by adapting the implicit normal mode formulation of Temperton (1989), thus allowing direct control over the projection of the analysis increment onto slow and fast modes. However the use of the model's normal modes in the vertical resulted in an overemphasis of the upper levels of the model and created difficulties in the partitioning of the analysis increment between temperature and surface pressure. Also, the model vertical modes are the same for vorticity, divergence, and height variable, while it is clear that at least for divergence, these are inappropriate functions (the external mode, for example, has the same sign over the entire depth of the atmosphere, which is undesirable for the divergence). Thus, the model vertical modes were replaced with empirical orthogonal functions (EOFs) defined from a vertical error covariance matrix (this has no effect on the slow modes, as defined by Temperton's scheme, because they satisfy the linear balance equation). Ideally the vertical covariance matrix would be calculated by comparing the true state to the background field. But the truth is not available, so approximate vertical covariance matrices have been defined using the difference between a 24 hour forecast and a verifying initialized analysis defined on the model's Gaussian grid and averaged globally.

It is not obvious that the higher vertical EOFs for the streamfunction will contribute positively to reducing the variance in the perturbation height modes. Through experimentation it was found that 6 vertical EOFs of vorticity were optimal in defining the balanced part of the height variable. Thus, when calculating the EOFs, streamfunction EOFs are obtained, then the first 6 are used with the linear balance relationship to get balanced height errors which are removed from the total height error. Finally EOFs are obtained for the unbalanced part of the height error, the divergence error, and the mixing ratio error. Note that for each variable the EOF's are defined independently. The first four of these modes for each of the four variables are shown in Fig. 1.

A complete definition of the transformation operator L as used in the current version of the SSI is given in appendix A. Ideally L would be defined such that it first transforms from the analysis variables to T, p_{sfc}, u, v, q and P_w (total precipitable water) on a sigma coordinate Gaussian grid and then interpolate these values to the observation locations. However, for computational reasons an approximation has been introduced into the L operator. Instead of performing the final interpolation to the observation locations, the observational increments and errors have been approximated as "superobservations" (superobs) defined at the Gaussian grid points. Thus the L operator only transforms the analysis variable to the Gaussian grid. Some of the effects of the horizontal interpolation to the observation locations were included in the superobs by using the transpose of the linear interpolation operator to define the distribution of the observation information to the surrounding eight vertices (or four vertices for surface pressure) of the grid volume. For profile data such as radiosonde data, the data are first interpolated to the guess sigma coordinate levels in the vertical to prevent the unrepresentative significant level information from dominating the other data. Note that in calculating the initial increments between the observations and the background field (the y vector), the L operator still includes a horizontal and vertical interpolation to the observation location before the formation of the superobs on the gaussian grid. Thus, the observation increments are calculated at the observation locations before superobing to the Gaussian grid points.

Satellite temperature superobs were created in a slightly different manner to partially account for the strong spatially correlated errors in the observations. In the creation of these superobs, three classes of temperature soundings were available; clear, partly cloudy and cloudy retrievals. It was decided that for each superob, only the satellite retrievals containing the minimum amount of clouds will be used. Thus, if only one clear sounding was available within the grid volume along with many cloudy soundings, only the clear sounding is used. All the temperature increments and observational error variances for the soundings containing the minimum cloudiness are averaged at the nearest grid point. The averaging of the observational error variances reduces the weight given the observations in the analysis procedure commensurate with an assumption of perfect correlation of the observational error within the box surrounding the grid point.

To include the precipitable water (P_w) observations in the analysis required one additional step. Since precipitable water is not a basic variable defined on the grid, it is necessary to include in the L operator an integration of the specific humidity.

$$P_w = p_s/g \sum_{\sigma=1}^{18} q_\sigma \Delta\sigma \tag{2.14}$$

The surface pressure is approximated in this equation by the background surface pressure from the six hour forecast guess inorder to keep L a linear operator. Thus, the direct inclusion of the SSM/I total precipitable water only requires the inclusion of the integration, given by (2.14), in the L operator and the inclusion of the transpose of this integration in the L^T operator.

In solving the analysis equations in conventional OI, a set of weights are found first and then used to interpolate the observation increments to the grid points. The intermediate process of finding the weights can become ill–conditioned when two observations are located close to each other. For the SSI analysis system, no difficulties are encountered with collocated observations since the analysis is done directly in spectral space.

The solution to (2.5) is found using a standard linear conjugate gradient algorithm (e.g., Gill et al, 1981). The current version of the SSI uses 50 iterations for both the moisture analysis and the dynamical variables. Most of the expense in each iteration comes from the application of the L and L^T operators. This is not suprising since these operators contain the transforms to and from the Gaussian grid. Note that currently the moisture analysis is completely independent from the analysis of the other variables. The gradient of (2.1) is reduced by 3–5 orders of magnitude. Only small changes in the analysis are noted after the gradient has been reduced by 2 orders of magnitude. Despite the large amount of computation, the SSI system runs faster than the current operational analysis system. This is primarily because of the removal of the sorting and selection algorithms, but there is also some advantage which results from forming superobs on the model grid.

3. THE SPECIFICATION OF THE FORECAST AND OBSERVATION ERROR COVARIANCES:

The forecast and observation errors define the relative weight each observation is given along with the relative amounts of information projected onto the analysis variables. For this reason they are obviously of vital importance to the analysis procedure. A set of statistics which appears to work reasonably well has been developed, but this is an area of continuing research. In the following three subsections, the current (February 1991) state of the statistics will be described.

a. The forecast error covariances:

The forecast error covariances (the B matrix) were estimated by first forming differences between the operational NMC T80 spectral model sigma coefficients for 24 hour forecast and initialized analyses verifying at the same time. Thirty consecutive days of 0000UTC differences were used. More would be desirable, but this number is determined by computational considerations. These errors in coefficients of vorticity, divergence, temperature, logarithm of surface pressure, and specific humidity were then transformed to the analysis variables as outlined by (2.9–13). Finally, the variances of these transformed variables were computed from the set of 30 cases. All off–diagonal elements of the full forecast error covariance matrix were assumed to be zero. Finally, the variances are rescaled to convert from a 24 hour to a 6 hour estimate of the forecast error variance (the NMC global data assimilation system ingests data at 6 hour intervals). This rescaling parameter was found empirically. This is a very crude first step in specifying the forecast error covariance. A brief discussion of possible future improvements appears at the end of the paper.

To see what are the effective forecast error correlations in physical space that result from the above spectral forecast error covariance model, the output of the SSI system for a single observation can be examined. These result from a set of forecast error statistics which were determined from 30 days of cases for January, 1991. The statistics as currently computed may have some seasonal dependence, but this has not been investigated yet. Reruns of summer cases using these statistics indicate that the SSI is robust with regard to the choice of forecast error statistics (see next section). Figure 2 shows the resultant analyses of temperature and wind for a single temperature with 1 degree residual at sigma level 5 (about 850mb) and 45N, 100W. Note that the fields are qualitatively similar to that which would be produced by current operational systems. The horizontal scale appears to be broader than that from the current operational system. However, tests performed in parallel show that the use of these statistics consistently produce fits to radiosonde observations as good or better than the operational system(see next section). Fig. 3 shows a similar result for a temperature observation at 0N, 100W. Because of the spectral representation and the linear balance constraint between mass and momentum fields, the SSI result is very different from that which would be produced from the operational NMC OI analysis, for which a mass observation at the equator produces no wind correction.

By allowing only the diagonal elements of the forecast error covariance to be nonzero, it is not possible to include the spatial variability in the forecast error variance which results from the inhomogeneous distribution of observations. This variability can be partially included by modifying the observational error variances. At this time we are not including this spatially varying component of the error. Proper inclusion of this spatially varying component of the error will undoubtedly result in further improvement of the results. Note however that some latitudinal dependence of the error is implicit in the system because of the linear balance constraint between mass and momentum fields. The wind error variance is homogeneous over the globe, but the balanced part of the mass errors become very small as the equator is approached.

b. The observation error variances:

The definition of the observational error covariances controls the relative weighting of the various observations. In this section, only the variances will be considered. The inclusion of correlated error will be discussed in the following sub-section. The basic observational error variances are given in Table 1. To the extent possible the current operational values (DiMego, 1988; Dey and Morone, 1985) were used. These basic variances can be modified based on the initial observation – guess increment, the extent of extrapolation (if any), and possibly to account indirectly for spatially varying forecast error variance.

As a simple quality control procedure, the observational error variance is increased if the observation residual is greater than 3 times the assumed error standard deviation (the square root of the variance). In these cases the observational error standard deviation is increased by the absolute value of the observational increment. When the initial increment is greater than 5 times the assumed error standard deviation, the observational error is assumed to be infinite and thus the effect of the observation is removed.

The observational error variance is also modified depending on the amount of vertical extrapolation if the data lies above or below the top or bottom sigma layer of the model. This is done to allow data which are slightly outside the model grid to be used, but the data far outside the grid are given little weight. This is especially important for surface pressure data. The model topography is often higher than the station elevation so that the surface observation lies below the model sigma domain. A correction is applied to the surface pressure observation to account for the terrain difference. But, the larger the terrain difference, the greater the potential error introduced by the correction.

c. The observation error correlations:

Observational errors are correlated to varying degrees. These error correlations are best defined for satellite data, which are both horizontally and vertically correlated. However other types of data such as radiosonde observations also contain vertical correlations in their errors. The technique used by the SSI system to solve the analysis equation (2.5) requires the inverse of the observation error covariance matrix. The inverse could be defined directly, but it is difficult to know how to do this properly, and once again we have potentially very large matrices to work with. If the correlations are all local, then the inversion problem can be approximated as a series of small matrix inversions, as in current OI schemes. However if the correlations are broad, as is the case for horizontal correlation of satellite data, the direct inversion of the matrix can be difficult. The current version of the SSI system only includes horizontal correlation of satellite temperature retrievals. This is done by creating another spectral representation, this time for the observation error. Multiplication by the inverse of the observation error covariance matrix for satellite data reduces to the application of a series of inexpensive operators, similar to the spectral representation of the forecast error covariance.

The observation error covariance matrix is defined by

$$O = ECE \ , \tag{3.1}$$

$$C = RS\hat{C}S^TR \tag{3.2}$$

where

C = observational correlation matrix,

$\hat{C}$ = spectral transform of the correlation matrix (assumed diagonal),

R = diagonal normalization matrix,

S = spherical harmonic transform matrix, and

E = diagonal matrix of observational error standard deviations.

A detailed definition of the diagonal spectral correlation matrix $\hat{C}$ appears in appendix B.

The analysis equation (2.5) requires the inverse of the observation error covariance matrix. In terms of (3.1–2), this is given by

$$O^{-1} = E^{-1}C^{-1}E^{-1} \ , \tag{3.3}$$

$$C^{-1} = R^{-1}S^{-T}\hat{C}^{-1}S^{-1}R^{-1} \tag{3.4}$$

Each of the terms defining O^{-1} can be easily applied as an operator. Thus to include the correlated error involves the inclusion of an additional string of operators when applying the complete operator A in (2.5).

4. RESULTS

The system described in the previous section has been subjected to a long period of testing and evaluation. In this section some of the characteristics of the system will be presented, with emphasis on the differences from the current operational system. The discussion will be divided into three main subsections. First, the analyses from both the operational and SSI systems will be compared. These comparisons will include analyses produced using the same background (first guess) fields along with results from a long term independent assimilation. To evaluate the extent of imbalances in the resultant analyses, the changes made by the initialization procedure will also be presented. Finally, evaluations of the forecast skill from the operational and SSI assimilation systems are shown.

a. Analysis comparisons

The current NMC global operational objective analysis technique (Dey and Morone,1985; DiMego, 1988; Kanamitsu, 1989) has produced good quality results over many years of usage. The results from the operational analysis system will be compared to those from the SSI system in two ways. First, the same background (first guess) field will be inserted into both the analysis systems, along with comparable observational data sets. Note that some differences exist between the datasets primarily because the SSI system uses temperature observations while the operational system uses height observations. Also, the quality control systems are not the same in both systems. Thus, small differences will exist because different observations will be rejected by the quality control. The differences in the analyses resulting from the different datasets are small when compared to those from the differences in the analysis procedures. The second set of comparisons will be between analyses produced after long separate assimilation periods. The previously mentioned differences between the input data would also apply for this comparison over the long assimilation period. In the assimilation mode, the possibility of good or bad characteristics of the analysis systems feeding back into the forecast and thus the next analysis is present.

In Figures 4 and 5, the 250mb height and wind increments (analysis –

background) are shown for 0000UTC Mar. 6,1991. These results are typical of those found on any day or at any vertical level. Both were created using the operational 6 hour forecast from 1800UTC Mar. 5, 1991 as a background field. The SSI analysis is done in the model's sigma coordinates. Thus, it was necessary to integrate the temperature field to produce heights and to perform a vertical interpolation to create comparable analyses. The height and wind increments from the SSI analysis are smoother and smaller than the operational increments. While some of the difference results from the background error covariances in the SSI emphasizing the larger scales, some of the smoothness is also due to the global use of all data. The changes in the data used from one point to the next in the operational system introduces noise into the analysis. Despite the fact that the analysis increments are smoother and smaller in the SSI system, the resulting analysis produces a comparable fit to the data. The fit of the analyses to the radiosonde data is presented in Table 2 for the same day. The differences between the operational and SSI analyses are of well within the variability from one day to the next with some days showing a closer fit from the operational and some days the SSI fits best. Finally, note the much smaller changes in the tropics apparent in the height increments of the SSI analysis. One would expect the tropical height field to be reasonably smooth without large changes from one time period to the next. As will be seen in the next subsection, many of the changes introduced in the tropics by the operational analysis are removed by the initialization procedure.

Figures 6 and 7 show the 250mb height and wind differences from the operational analyses and the SSI analyses for different backgrounds. The upper panel in each figure is the result from the operational background fields (as in the previous paragraph), while the lower panel is the result from a background obtained from an assimilation cycle which applied the SSI analysis every 6 hours. This assimilation has been running independently from the operational system for 80 days. As can be seen from the results, the differences are larger in the tropics and southern hemisphere as might be expected. However, the differences between the analyses are largest over the poorest observed regions and appear to be reasonable when compared to the expected error in the analyses. The fit of the analysis to the data for the cycled version of the SSI system is also comparable to those from the operational system, as can be seen from inspection of Table 2.

The time mean analysis differences between the cycled SSI and operational analysis was also examined. The two most significant differences, weaker tropical precipitation and a weaker mean Hadley circulation, are related to each other. The zonal mean v–component for Mar. 6, 1991 is shown in Fig. 8. The operational assimilation system tends to produce a stronger Hadley circulation which decays with time. The SSI Hadley circulation is similar to the model Hadley circulation after a 5 day forecast and does not decay with time. Thus, the strength of the SSI Hadley circulation appears not to be controlled by the analysis system but rather the model dynamics and physics. This difference is significant and is currently under investigation.

b. Initialization

In most operational assimilation systems, the analyses are initialized before forecasts are made. This is necessary because imbalances between the mass, momentum and diabatic fields can produce large amplitude gravity waves that can amplify in an assimilation system. Unfortunately, the initialization is done as an independent step after the analysis and thus usually adjusts the fields away from the data. Ideally, the necessary balance would be imposed by the analysis procedure making the initialization unnecessary (see Williamson and Daley, 1983). Thus, the magnitude of the adjustment by the initialization is a measure of the quality of the balance imposed in the analysis procedure.

In Figures 9 and 10, the initialization increments (initialized fields minus analysis) for the 250mb heights and winds are shown for the same case. The SSI increments are created using the same background field as the operational system, but similar results are found using the long term assimilation. As can be seen from the figures, the initialization makes very small changes to the SSI analyses, while the operational analyses are substantially altered. Often in the tropics, the initialization removes features introduced by the analysis. The SSI analysis does not produce these features, so it is not necessary to remove them. In Table 2, the fits to the data before and after initialization are closer than those from the operational system. The very small changes by the initialization for the SSI system suggests the possibility of removing the initialization step. We have completed a one week assimilation with no initialization. No harmful effects of removing the initialization have been found but it is necessary to perform further experiments for longer periods.

c. Forecast results

The most important operational test of analysis quality is the resultant accuracy of the forecasts. Five day forecasts have been produced from the SSI assimilation results in parallel with the operational system for an 80 day assimilation period. These forecasts were produced in real time using the operational database. In addition a 30 day case from Aug. 1990 was examined in a retrospective mode. The quality of a forecast can vary greatly from one day to another, so we will present average results for a 29 day period from February and March 1991. These results are similar to those found in the previous two months and those from Aug. 1990. The forecasts are all verified against the operational analyses.

Figures 11–14 show average anomaly correlation scores over the 29 day period for zonal waves 1–20, verified against the operational analysis. Figure 11 shows the result for 1000mb in the northern hemisphere. The improvement by day 5 is about 3%. Similar, but slightly smaller improvement is evident at 500mb for the northern hemisphere (Figure 12). The southern hemisphere results are also encouraging. The low correlation at day 0 for 1000mb and 500mb (Figures 13 and 14) indicates that the two systems have departed from each other significantly. Still, by day 3, even with the handicap of verifying against the operational system, the SSI surpasses the operational. Similar results are found by looking at other measures of the forecast skill.

In addition to the subjective scores, the analyses and forecasts are currently undergoing a subjective evaluation from the operational forecasters in NMC's Meteorlogical Operations Division (MOD). The preliminary results from this evaluation are encouraging but the evaluation is not yet completed.

5. DISCUSSION

The SSI system as currently formulated has potential for substantial improvement. However, even without further improvements the system compares very favorably with the current operational system and can be solved faster than the current operational analysis. The increments created by this analysis are smaller, smoother and better balanced than those from the operational system but still fit the data equally well. The forecast results from this system have also been encouraging, with consistently better results coming from the SSI system.

One of the potential future improvements to the SSI system is in the representation of the forecast error covariance B. B is difficult to obtain in practice, especially for a spectral model. It is dependent on the error due to model misrepresentation of the "true" atmospheric state, and on the errors of all past observations assimilated by the evolving model forecast. In general it is a full matrix of dimension equal to the number of model variables, currently $O(10^6)$, and will always be much too large to keep in its entirety. For small systems ($O(10^3)$), the engineering community uses the Kalman filter (Kalman, 1960, Kalman and Bucy, 1961), in which B is explicitly computed. The Kalman filter cannot be applied directly to NWP systems because of the large number of variables. However, important progress is being made using Kalman filters for one– and two–dimensional NWP models to determine expected properties of B for realistic situations, and what approximations to B are useful (see Cohn and Parrish, 1991; Daley, 1991 for a detailed exposition of this subject). One area for future work is to utilize experience with idealized models and the Kalman filter in improvements to the forecast error representation.

The results can be further enhanced by improving the current crude quality control. Recently an improved quality control system has been developed by Woollen (1991). The results presented in this paper did not utilize the new quality control because it was designed for heights and winds, while SSI works from temperature and wind observations. Changes to the quality control system are currently being made and no serious difficulties are anticipated.

The spectral model's resolution was increased from T80 to T126 on Mar. 6,

1991. New forecast error statistics were gathered from a parallel run of the T126 model. Parallel tests with this higher resolution model have begun and no serious difficulties have been encountered. At this point in time not enough cases have been accumulated to draw any conclusions . However it is anticipated that the SSI system will work even better in the T126 model since the analysis is performed in sigma coordinates and the T126 model's orography is much closer to reality.

One of the advantages of the SSI system is its ability to incorporate nonconventional observations. This will be a major area of emphasis for future improvements. The first and easiest nonstandard observations to include are satellite observations of total precipitable water and surface wind speeds. The precipitable water has been successfully incorporated. A longer–term goal is to use satellite radiances directly in the analysis. In principal this is similar to adding other forms of non–conventional data. However, determining the operator which goes from model variables to observed variables is very difficult. Significant progress is being made at other centers on this problem (Eyre, 1989, for example).

The constraint between the mass and momentum fields is currently a linear balance relationship. While this is somewhat better than a geostrophic constraint, it still can be improved. As discussed by Young (1990), the inclusion of friction in the constraint is vital for properly analysing the boundary layer winds. The addition of this term to the linear balance equation is relatively straightforward, and an attempt to examine the effects of this term will be undertaken in the near future.

Finally, the SSI is a 3–dimensional analysis. However, its formulation makes it possible to mesh it with a 4–dimensional "adjoint" variational system (LeDimet and Talagrand, 1986, Derber, 1989, Talagrand and Courtier, 1987). This would allow a large increase in the effective database since the data over an entire time interval could be included. Thus the poorly known statistics would be of relatively less importance. The introduction of a full 4–D operational variational system is obviously not possible in the near future. However, the successful development of a 3–D variational system, with many potential enhancements, is the first step towards large improvements in future analysis and assimilation systems.

Acknowledgements

The authors would like to thank the many people from NMC that have made this project possible. The assistance of Jack Woolen, Masao Kanamitsu, Robert Kistler, Peter Caplan, Glenn White, Dennis Deaven, and Dennis Keyser in setting up and evaluating the system was particularly appreciated. Many useful comments on this paper and project were provided by Eugenia Kalney, Lev Gandin and Paul Long.

APPENDIX A. DETAILED DESCRIPTION OF THE FORWARD OPERATOR L

This appendix describes the $y = Lx$ operation, the conversion from analysis to pseudo–observations. For the current version of SSI, L is the product of five operators, $L = L_5L_4L_3L_2L_1$. Let $x_0 \equiv x$, $x_1 = L_1x_0, \ldots, y = L_5x_4$. Then each operator is presented in the following five subsections:

a. $x_1 = L_1x_0$:

The first operator L_1 converts from scaled vorticity, divergence, and unbalanced height to physical units. This scaling is used in spectral normal mode initialization schemes to symmetrize the linear model operator from which the normal modes are computed. Although we do not use normal modes in SSI, this scaling still proves useful because it improves the conditioning of the analysis equation (2.5). For L_1,

$$\zeta^l_{1nm} = a^{-1}[n(n+1)]^{\frac{1}{2}}\zeta^l_{0nm} \tag{A.1a}$$

$$D^l_{1nm} = a^{-1}[n(n+1)]^{\frac{1}{2}}D^l_{0nm} \tag{A.1b}$$

$$H^l_{1nm} = \left(\frac{\bar{h}}{g}\right)^{\frac{1}{2}} H^l_{0nm} \tag{A.1c}$$

$$q^l_{1nm} = q^l_{0nm} \; . \tag{A.1d}$$

In the above, l and n are spherical harmonic indices, where l is longitudinal wave number, $-J \leq l \leq J$, and n is the two–dimensional wave number, $|l| \leq n \leq J$. The third index m is the vertical mode number, $1 \leq m \leq N_\sigma$. J is the triangular wavenumber cutoff for horizontal resolution, and N_σ is the number of model sigma layers. Finally, $a = 6.37x10^6$m , $g = 9.8$ms^{-1} , and $\bar{h} = 3000$m . The scale depth is arbitrary here. The value 3000 gives approximate equivalence between scaled mass and wind variables.

b. $x_2 = L_2 x_1$:

L_2 represents the vertical transform in terms of EOFs. Each variable has its own EOF representation in the vertical. Thus,

$$\zeta^l_{2nk} = \sum_{m=1}^{N_\sigma} V^{\zeta}_{km} \zeta^l_{1nm} \tag{A.2a}$$

$$(\zeta H)^l_{2nk} = \sum_{m=1}^{6} V^{\zeta}_{km} \zeta^l_{1nm} \tag{A.2b}$$

$$D^l_{2nk} = \sum_{m=1}^{N_\sigma} V^{D}_{km} D^l_{1nm} \tag{A.2c}$$

$$H^l_{2nk} = \sum_{m=1}^{N_\sigma} V^{H}_{km} H^l_{1nm} \tag{A.2d}$$

$$q^l_{2nk} = \sum_{m=1}^{N_\sigma} V^{q}_{km} q^l_{1nm} \ . \tag{A.2e}$$

Notice that $(\zeta H)^l_{2nk}$ in (A.2b) is a partial sum of just the first six vorticity vertical modes. This is used in the next transform, L_3, to get the contribution of vorticity to total mass variable via the linear balance equation. The indices n and l are as before, but k is the model sigma layer index.

c. $x_3 = L_3 x_2$:

The L_3 transform involves latitude sums of various combinations of spherical harmonics. At this point, vorticity and divergence are converted to u and v, and the linear balance equation (cf. (2.11)) is incorporated to get the contribution of vorticity to mass variable. This results in,

$$u^l_{3jk} = \sum_{n=|l|}^{J} [\zeta^l_{2nk} R^l_{jn} - i D^l_{2nk} Q^l_{jn}] \tag{A.3a}$$

$$v^l_{3jk} = \sum_{n=|l|}^{J} [-i\zeta^l_{2nk} Q^l_{jn} - D^l_{2nk} R^l_{jn}] \tag{A.3b}$$

$$H^l_{3jk} = \sum_{n=|l|}^{J} [(\zeta H)^l_{2nk} B^l_{jn} + H^l_{2nk} P^l_{jn}] \tag{A.3c}$$

$$q_{3jk} = \sum_{n=|l|}^{J} q^l_{2nk} P^l_{jn} \ . \tag{A.3d}$$

In the above, $i\overline{\overline{\sqrt{-1}}}$ and $P^l_{jn} \equiv P^l_j(\phi_j)$ are associated Legendre polynomials defined with the same normalization used by the NMC spectral model, viz.,

$$\int_{-\pi/2}^{\pi/2} (P^l_n(\phi))^2 \cos\phi d\phi = 1 \ . \tag{A.3e}$$

The functions which convert vorticity and divergence to winds are

$$Q^l_{jn} \equiv Q^l_n(\phi_j) = a[n(n+1)]^{-1} l (cos\phi_j)^{-1} P^l_n(\phi_j) \tag{A.3f}$$

and

$$R^l_{jn} \equiv R^l_n(\phi_j) = a[n(n+1)]^{-1} \frac{dP^l_n}{d\phi}(\phi_j) \ . \tag{A.3g}$$

Note that the wind components are not scaled by $\cos(\phi_j)$ as is standard practice in spectral models. Also, the Gaussian grid used in SSI is augmented by adding north and south pole points as an aid for interpolation to observation locations when obtaining observation residuals. The apparent division by zero at the pole points which appears in (A.3f) is taken care of by noting that $P^l_n(\phi)$ has a factor $\cos^l \phi$.

The function which gives a balanced mass variable from vorticity is

$$B^l_{jn} \equiv B^l_n(\phi_j) = -\frac{2\Omega a^2}{g} [(n+1)^{-2} \epsilon^l_{n+1} P^l_{n+1}(\phi_j) + (1-\delta_{1n}) n^{-2} \epsilon^l_n P^l_{n-1}(\phi_j)] \tag{A.3h}$$

where

$$\epsilon_n^l = \left[\frac{(n+l)(n-l)}{(2n+1)(2n-1)} \right]^{\frac{1}{2}} . \tag{A.3i}$$

and δ_{jk} is the Kroneker delta, $\delta_{jk} = 1$ for j=k, but is otherwise zero. Finally $\Omega = 7.292x10^{-4}s^{-1}$.

d. $x_4 = L_4 x_3$:

The L_4 operator is the Fourier sum in longitude which is the same for each of the four variables and is illustrated just for the specific humidity:

$$q_{4sjm} = \sum_{-J}^{J} q_{3jm}^l e^{il\lambda_s} \tag{A.4}$$

The longitudes λ_s are equally spaced and the sum is accomplished with a fast Fourier transform (FFT).

e. $y = L_5 x_4$:

The final operator L_5 converts the total mass variable H_4 to grid values of T and $\ln(p_{sfc})$. Because $N_\sigma + 1$ output quantities are required with only N_σ input quantities given, some additional constraint is required. After some experimentation, it appeared that the best condition to apply was to minimize the 2–grid component in the vertical of the derived temperatures. The problem is to invert equation (2.8),

$$H = GT + \frac{R\overline{T}}{g} \ln(p_{sfc}) \tag{A.5a}$$

This is accomplished by minimizing the following objective function for temperature:

$$J = T^T S^T S T \tag{A.5b}$$

where the matrix S is an $(N_\sigma - 2)\ x\ N_\sigma$ matrix with all zeros except for the three diagonals

$$S_{jj} = 1 \ , \ S_{j,j+1} = -2 \ , \ S_{j,j+2} = 1 \ , \ 1 \le j \le N_\sigma - 2 \ . \tag{A.5c}$$

The matrix S applies 2nd derivatives to the temperatures in the vertical. First (A.5a) is used to eliminate T from (A.5b) and then the resulting equation is minimized with respect to $\ln(p_{sfc})$. Finally we use this value of $\ln(p_{sfc})$ and solve (A.5a) for T. The result is

$$T = QH_4 \tag{A.5d}$$

and

$$\ln(p_{sfc}) = W^T H \tag{A.5e}$$

where the matrix Q is given by

$$Q = G^{-1}[I - (D^T D)^{-1}\overline{T} D^T C] \tag{A.5f}$$

and the vector W is given by

$$W = (\frac{g}{R})(D^T D)^{-1} C^T D \ . \tag{A.5g}$$

The matrix C and the vector D are given by

$$C = SG^{-1} \tag{A.5h}$$

and

$$D = SG^{-1}\overline{T} \ . \tag{A.5i}$$

Finally, the hydrostatic matrix G is here defined as follows:

$$G_{jj} = \alpha_j \ , \ 1 \le j \le N_\sigma \ , \tag{A.5j}$$

$$G_{kj} = \alpha_j + \alpha_{j+1} \ , \ 1 \le j \le N_\sigma - 1 \ , \ j+1 \le k \le N_\sigma \ , \tag{A.5k}$$

$$G_{kj} = 0 \ , \ 2 \le j \le N_\sigma \ , \ 1 \le k \le j-1 \ , \tag{A.5l}$$

$$\alpha_1 = -(\frac{R}{g}) ln(\sigma_1) \ , \tag{A.5m}$$

and

$$\alpha_j = -(\frac{R}{2g}) \ln(\frac{\sigma_j}{\sigma_{j-1}}) \ , \ 2 \le j \le N_\sigma \ . \tag{A.5n}$$

APPENDIX B. DESCRIPTION OF SATELLITE ERROR CORRELATION MODEL

Jerry Sullivan (NESDIS, pers. comm.) has provided the following correlation model for satellite soundings:

$$c(r) = (1 + \frac{r}{L})e^{-\frac{r}{L}} \tag{B.1}$$

where r is separation between observations and L is currently 400km.

To obtain the elements of the diagonal matrix $\hat{C}$ in (3.2), we apply the following derivation for the representation of a correlation function on the sphere. A general correlation function between two points on a sphere has the following spherical harmonic representation:

$$c(\phi_1, \lambda_1; \phi_2, \lambda_2) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{l_1=-n_1}^{n_1} \sum_{l_2=-n_2}^{n_2} \hat{c}_{n_1 n_2}^{l_1 l_2} P_{n_1}^{|l_1|}(\phi_1) P_{n_2}^{|l_2|}(\phi_2) e^{i(l_1\lambda_1 - l_2\lambda_2)} \tag{B.2}$$

A special form which yields a homogeneous and isotropic correlation is

$$c(\phi_1, \lambda_1; \phi_2, \lambda_2) = \sum_{n=0}^{\infty} \hat{c}_n \sum_{l=-n}^{n} P_n^{|l|}(\phi_1) P_n^{|l|}(\phi_2) e^{il(\lambda_1 - \lambda_2)} \tag{B.3}$$

Using the addition theorem for spherical harmonics (see e.g. Korn and Korn, p874), (B.3) reduces to a form which directly illustrates the homogeniety and isotropy:

$$c(\gamma) = \sum_{n=0}^{\infty} \left(\frac{2n+1}{2} \right)^{\frac{1}{2}} \hat{c}_n P_n^0(\frac{\pi}{2} - \gamma) \tag{B.4}$$

where γ is the spherical distance in radians between the points (ϕ_1, λ_1) and (ϕ_2, λ_2) being correlated:

$$\cos\gamma = \sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2) \tag{B.5}$$

To obtain $\hat{c}_n$ for a specified function $c(\gamma)$ we must evaluate the following integrals:

$$\hat{c}_n = \left(\frac{2}{2n+1} \right)^{\frac{1}{2}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} c(\frac{\pi}{2} - \phi) P_n^0(\phi) \cos \phi d\phi \tag{B.6}$$

(B.6) is evaluated numerically using Gaussian quadrature.

REFERENCES

Cohn, S., and D. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. Accepted for publication in *Mon. Wea. Rev.*

Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Q.J.R. Meteorol. Soc.*, **113**, 1329–1347.

Daley, R., 1991: The lagged innovation covariance: a performance diagnostic for atmospheric data assimilation. Submitted to *Mon. Wea. Rev.*

Dee, D.P., 1990b: Simplification of the Kalman filter for meteorological data assimilation. *Q.J.R. Meteorol. Soc.,* in press.

Derber, J.C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.

Derber, J., D. Parrish, J. Sela, I.M. Navon, and X. Zou, 1990: Variational data assimilation experiments at NMC. Extended Abstracts, Second Annual BMRC Numerical Modelling Workshop. Sept. 19–21, 1990. Bureau of Meteorology Research Center, Melbourne, Australia.

Dey, C.H., and L.L. Morone, 1985: Evolution of the National Meteorological Center global data assimilation system: January 1982–December 1983. *Mon. Wea. Rev.*, **113**, 304–318.

Dimego, G.J., 1988: The National Meteorological Center Regional Analysis System. *Mon. Wea. Rev.*, **116**, 977–1000.

Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Reprinted in *Dynamic Meteorology Data Assimilation Methods*, L. Bengtsson, M. Ghil, E. Källèn, eds., *Applied Mathematical Sciences*, 36, Springer–Verlag New York, 1981, pp319–330.

Eyre, J.R., 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. *Q.J.R. Meteorol. Soc.*, **113**, 1001–1037.

Flattery, T.W., 1970: Spectral models for global analysis and forecasting. Proceedings of the Sixth AWS Technical Exchange Conference, U.S., Annapolis, Maryland. Air Weather Service Technical Report 242, pp 42–54.

Gandin, L.S., 1963: Objective analysis of meteorological fields. *Gidrometeorologicheskoe Izdatelstvo*, Leningrad. English translation by: Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp. [NTIS N6618047, Library of Congress QC996.G3313].

Gill, P.E., W. Murray, and M.H. Wright, 1981: *Practical Optimization.* Academic Press, 401 pp.

Halem, M. and E. Kalnay, 1983: Variational global analysis of satellite temperature soundings. Extended Abstracts, Ninth Conference on Aerospace and Aeronautical Meteorology. June 6–9, 1983, Omaha, Nebr. Published by the American Meteorological Society, Boston, MA.

Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136.

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. *Weather and Forecasting*, **4**, 335–342.

Kalman, R.E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.*, **82**, 35–45.

Kalman, R.E., and R.S. Bucy, 1961: New results in linear filtering and prediction theory. *Trans. ASME, Ser. D, J. Basic Eng.*, **83**, 95–108.

Kimeldorf, G. and G. Wahba, 1970: A correspondence between Bayesian estimation of stochastic processes and smoothing by splines. *Ann. Math. Statist.*, **41**, 495–502.

Korn, G.A., and T.M. Korn, 1968: *Mathematical Handbook for Scientists and Engineers.* Second edition. McGraw–Hill, 1130 pp.

Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, **38A**, 97–110.

Lewis, J.M., and J.C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309–322.

Lönnberg , P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde 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. Meteorol. Soc.*, **112**, 1177–1194.

Parrish, D.F., and S.E. Cohn, 1985: A Kalman filter for a two-dimensional shallow-water model: formulation and preliminary experiments. *Office Note 304*, National Meteorological Center, Washington, DC 20233, 64 pp.

Parrish, D.F., 1987: Successive correction statistical interpolation vs. optimum interpolation––a comparison. Extended Abstracts, Tenth Conference on Probability and Stataistics in Atmospheric Sciences. October 6–8, 1987, Edmonton, Alberta. Published by the American Meteorological Society, 45 Beacon St., Boston, MA 02108.

Parrish, D.F., 1988: The introduction of Hough functions into optimum interpolation. Extended abstracts, Eighth Conference on Numerical Weather Prediction. February 22–26, 1988, Baltimore, Maryland. Published by the American Meteorological Society, 45 Beacon St., Boston, MA 02108.

Parrish, D.F., and J. Derber, 1988: Spectral representation of error statistics for the NMC global spectral model with application to optimum interpolation. Research Activities in Atmospheric and Oceanic Modelling, report No. 11, November, 1988. G.J. Boer, ed. Published by the World Meteorological Organization, WMO/TD No. 263.

Parrish, D. and J. Derber, 1990: Direct analysis of model normal modes using optimum interpolation. Extended Abstracts, International Symposium on Assimilation of Observations in Meteorology and Oceanography, July 9–13, 1990, Clermont–Ferrand, France. Published by the World Meteorological Organization.

Phillips, N.A., 1986: The spatial statistics of random geostrophic modes and first–guess errors. *Tellus*, **38A**, 314–332.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q.J.R. Meteorol. Soc.*, **113**, 1311–1328.

Temperton, C., 1989: Implicit normal mode initialization for spectral models. *Mon. Wea. Rev.*, **117**, 426–451.

Williamson, D. L., and R. Daley, 1983: A unified analysis–initialization technique. *Mon. Wea. Rev.*, **111**, 1517–1536.

Woollen, J., 1991: New NMC operational quality control. To appear in extended abstracts, Ninth Conference on Numerical Weather Prediction. October, 1991, Denver, Colorado. Published by the American Meteorological Society, 45 Beacon St., Boston, MA 02108.

Young, J. A. ,1990: Problems of surface wind assimilation for the tropical Pacific. *Air–Sea interaction in tropical western Pacific: proceedings of US–PRC international TOGA symposium, 1988 Beijing,* China Ocean Press. Beijing, 3–13.

Table 1. Observation error variances.

wind errors (ms^{-1})						
	1000mb	700mb	500mb	300mb	100mb	50mb
radiosonde	1.4	2.4	2.8	3.4	2.5	2.7
AIREP	4.5	4.5	4.5	4.5	4.5	4.5
dropsonde	1.4	2.4	2.8	3.4	2.5	2.7
ACARS	1.4	2.4	2.8	3.4	2.5	2.7
low cloud drift	3.9	3.9	3.9	3.9	3.9	3.9
high cloud drift	6.1	6.1	6.1	6.1	6.1	6.1
surface	2.5	2.5	2.5			

temperature errors (K)						
	1000mb	700mb	500mb	300mb	100mb	50mb
radiosonde	1.8	1.3	1.3	2.0	3.1	4.0
AIREP	2.7	2.7	2.9	3.4	4.6	4.6
dropsonde	1.8	1.3	1.3	2.0	3.1	4.0
ACARS	1.8	1.3	1.3	2.0	3.1	4.0
clear satellite retrievals	4.7	3.9	4.0	4.5	4.0	4.0
cloudy satellite retrievals	5.6	4.6	4.6	5.0	4.5	4.5
ocean surface	2.2					
land surface	3.2	2.9	3.0			

other observations	
radiosonde moisture	2% relative humidity (all levels)
radiosonde surface pressure	1mb
dropsonde surface pressure	2mb
ocean surface pressure	1.6mb
land surface pressure	1.0mb
bogus surface pressure	3.0mb
SSM/I precipitable water	4mm

Table 2. Fits of operational and SSI fields to all radiosonde observations within 25mb of 250mb (975 temperature, 970 wind observations).

	mean T error	rms T error	rms vector wind error
operational			
background	–.91	2.26	8.79
analysis	–.64	1.77	5.60
initialized	–.65	1.79	5.85
SSI (operational background)			
background	–.91	2.26	8.79
analysis	–.46	1.68	5.86
initialized	–.49	1.69	5.98
SSI (SSI background)			
background	–.94	2.27	8.65
analysis	–.52	1.73	5.93
initialized	–.54	1.74	6.07

FIGURE CAPTIONS

Fig. 1. First 4 vertical EOFs for (a) scaled vorticity, (b) scaled divergence, (c) unbalanced height variable, and (d) specific humidity.

Fig. 2. Analysis for a single temperature observation at layer 5 and 45N, 100W. (a) temperature increment at layer 5, (b) u increment at layer 7, and (c) v increment at layer 7.

Fig. 3. As in Fig. 2, but for a single observation at 0N, 100W.

Fig. 4. 250mb height analysis increments (analysis – background) for (a) SSI and (b) operational. Both were created using operational background. Contour interval is 10m.

Fig. 5. Same as Fig. 4, but for 250mb wind analysis increments.

Fig. 6. 250mb height difference (operational – SSI) for (a) both with same background, and (b) SSI background from SSI assimilation for 60 days. Contour interval is 20m.

Fig. 7. Same as Fig. 6, except for wind differences.

Fig. 8. Zonal mean cross sections from (a) SSI from assimilation and (b) operational analyses for 0000UTC Mar. 6, 1991.

Fig. 9. Same as Fig. 4, except for initialization increments (initialized – analysis).

Fig. 10. Same as Fig. 5, except for initialization increments.

Fig. 11. Anomaly correlation scores for 1–5 day northern hemisphere 1000mb height forecasts, verified against operational analyses, averaged over the period 15Jan–14Feb 1991. Solid line is from operational assimilation, dashed SSI.

Fig. 12. Same as Fig. 11, but for northern hemisphere 500mb height forecasts.

Fig. 13. Same as Fig. 11, but for southern hemisphere 1000mb height forecasts.

Fig. 14. Same as Fig. 11, but for southern hemisphere 500mb height forecasts.

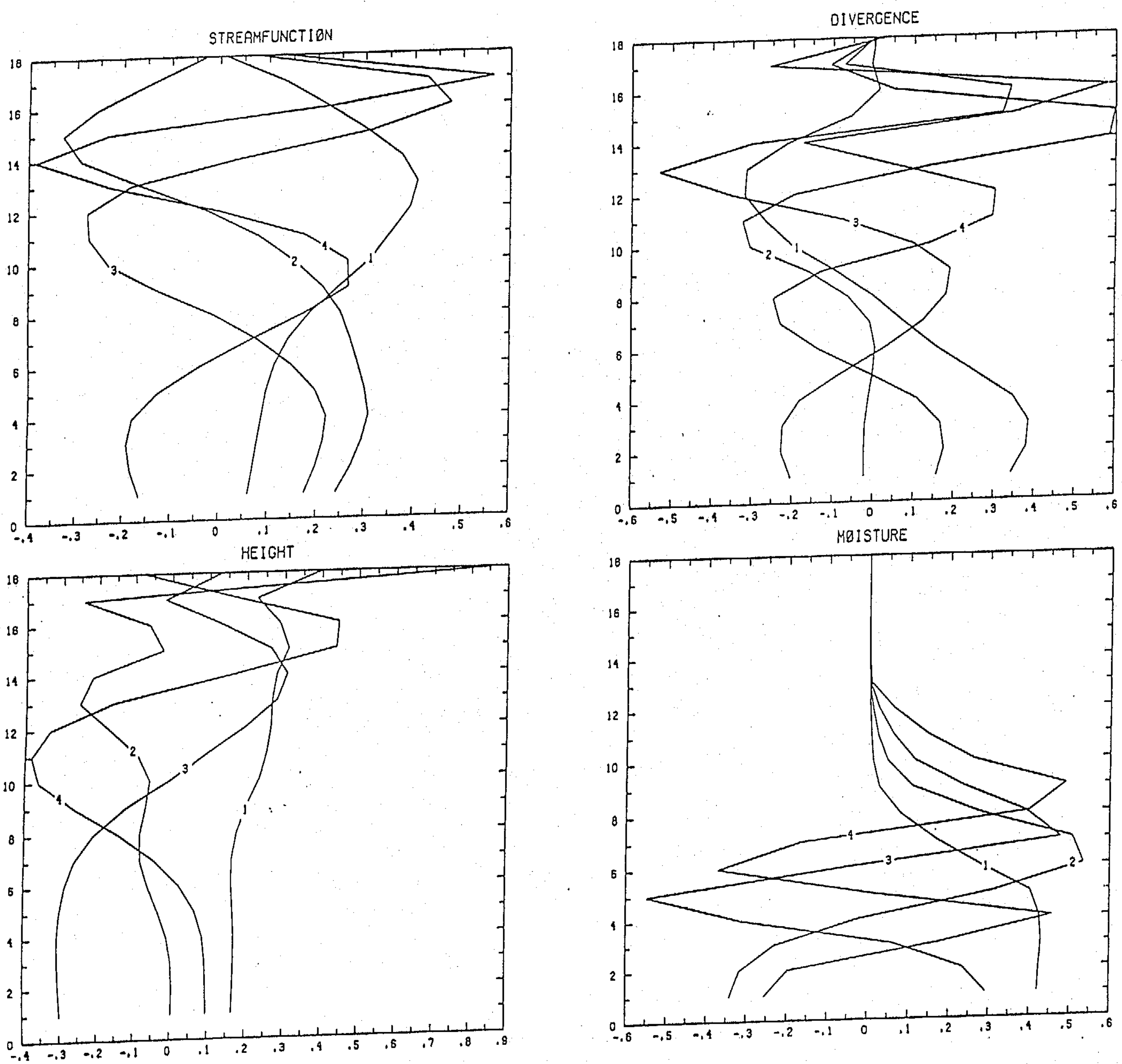
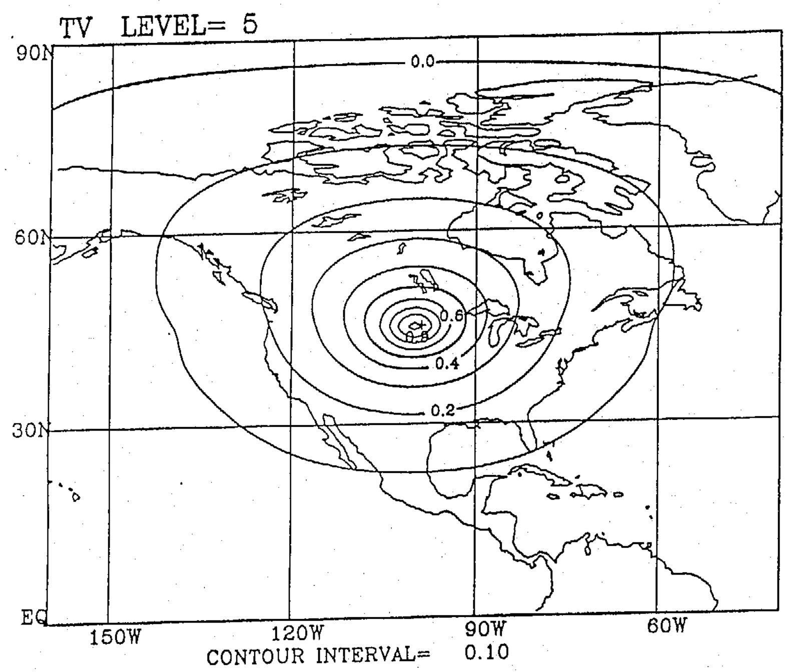
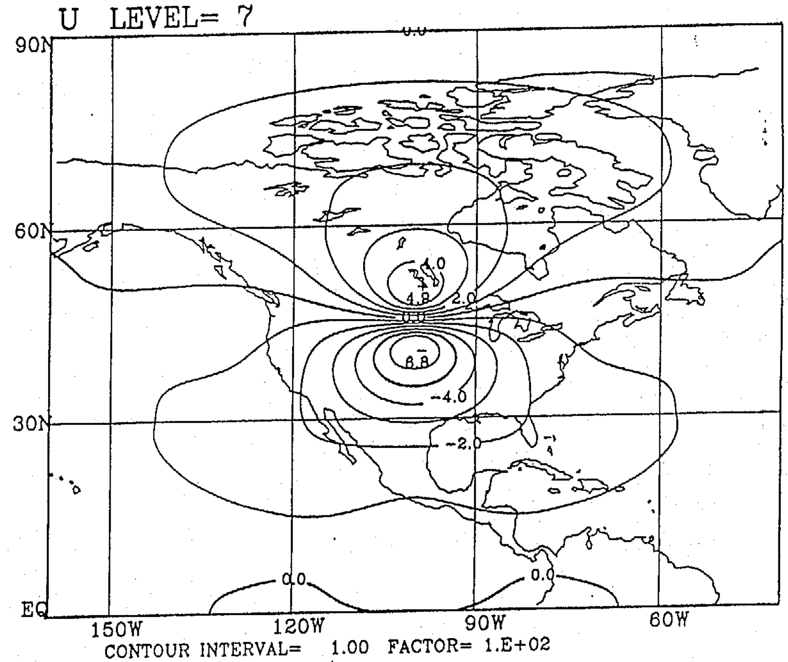


Fig. 1. First 4 vertical EOFs for (a) scaled vorticity, (b) scaled divergence, (c) unbalanced height variable, and (d) specific humidity.





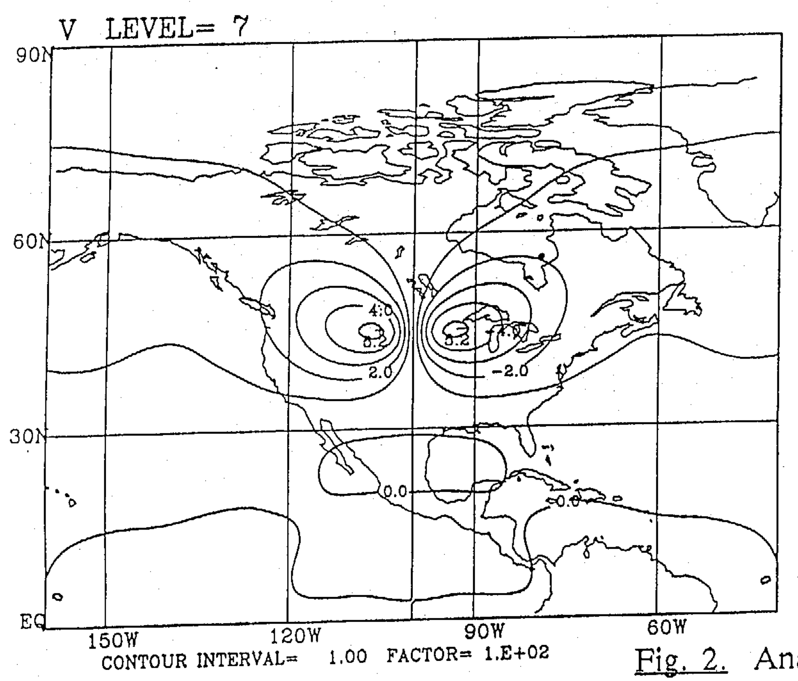


Fig. 2. Analysis for a single temperature observation at layer 5 and 45N, 100W. (a) temperature increment at layer 5, (b) u increment at layer 7, and (c) v increment at layer 7

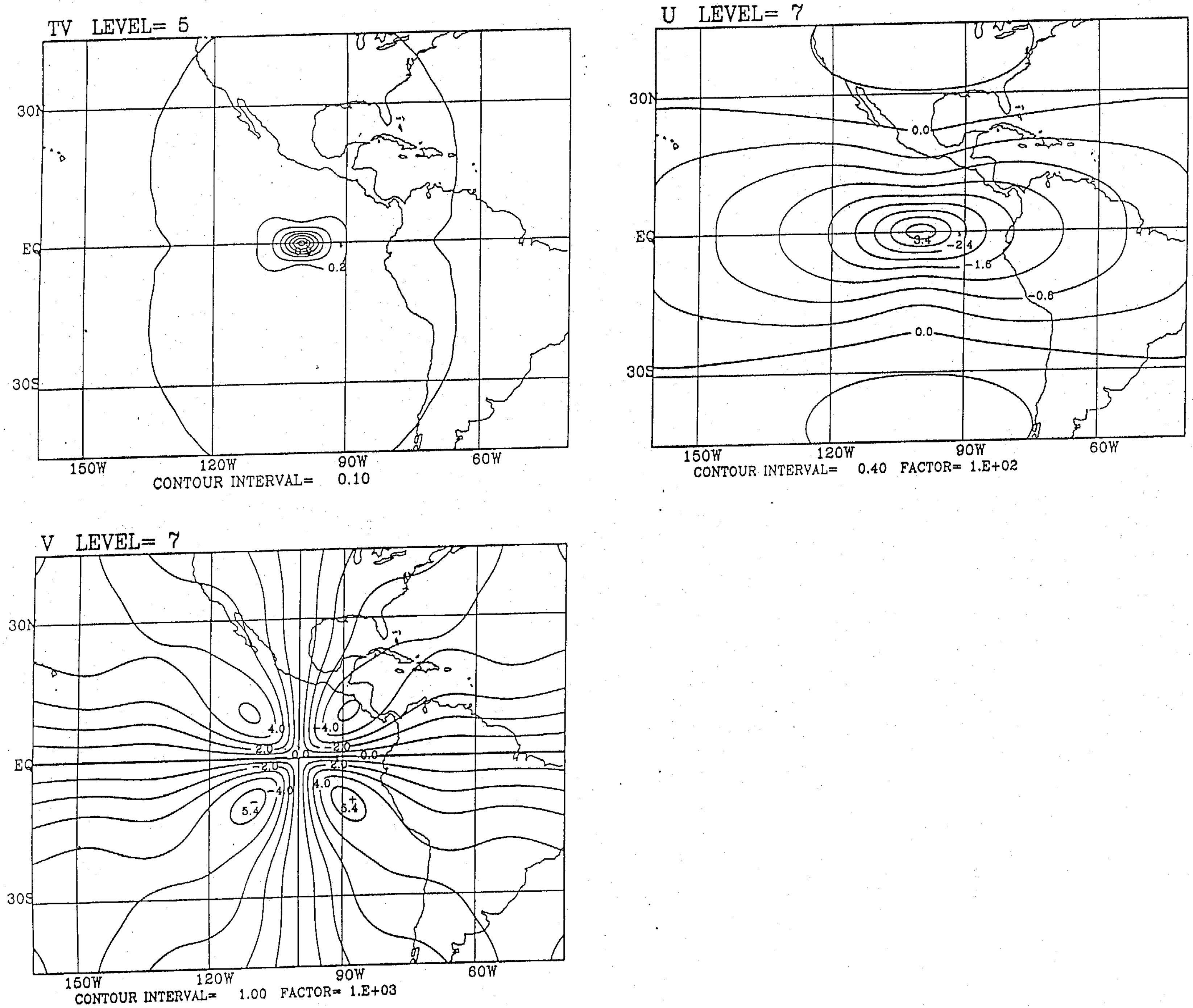


Fig. 3. As in Fig. 2, but for a single observation at 0N, 100W.

(a)

(b)

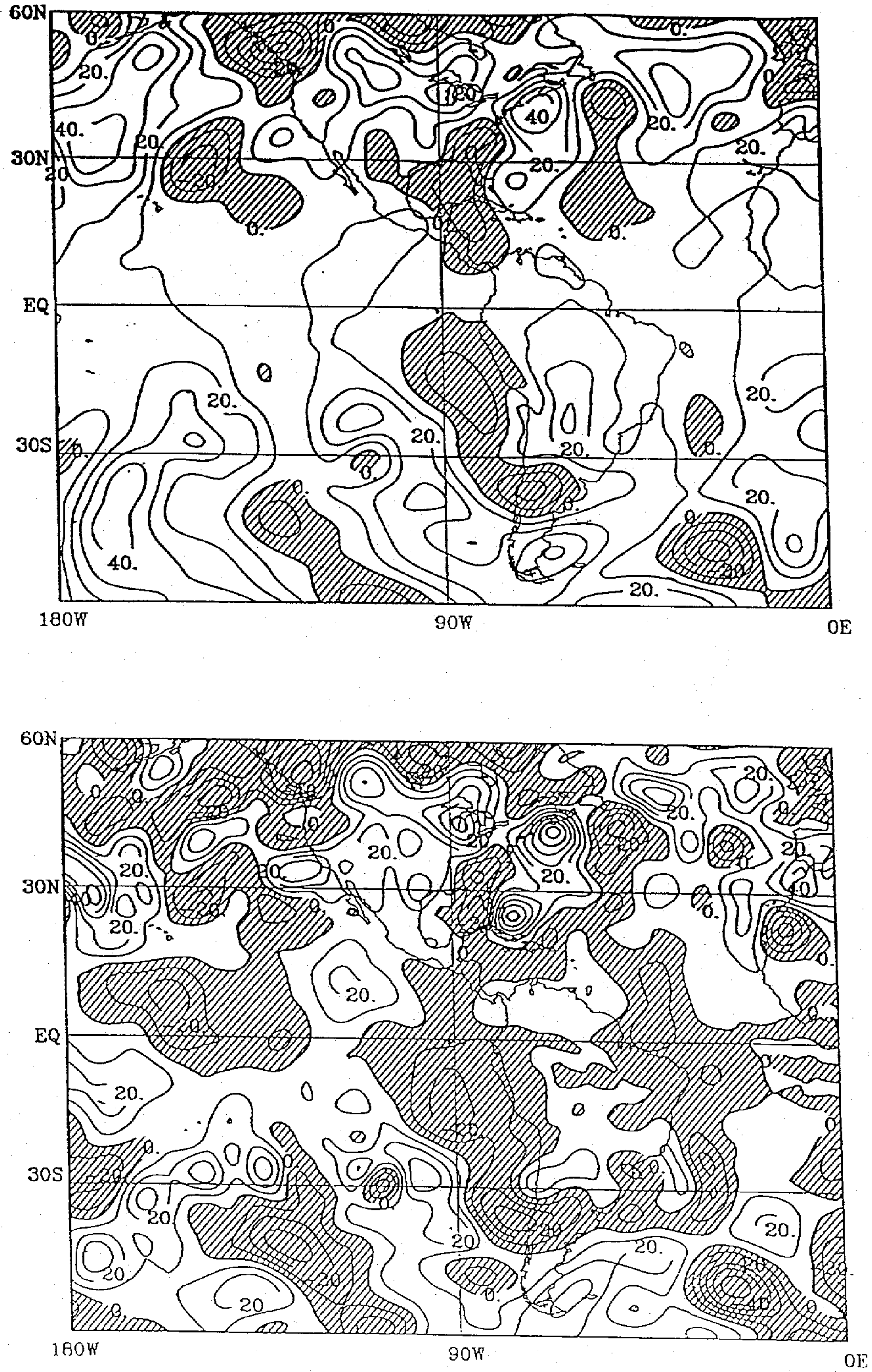


Fig. 4. 250mb height analysis increments (analysis – background) for (a) SSI and (b) operational. Both were created using operational background. Contour interval is 10m.

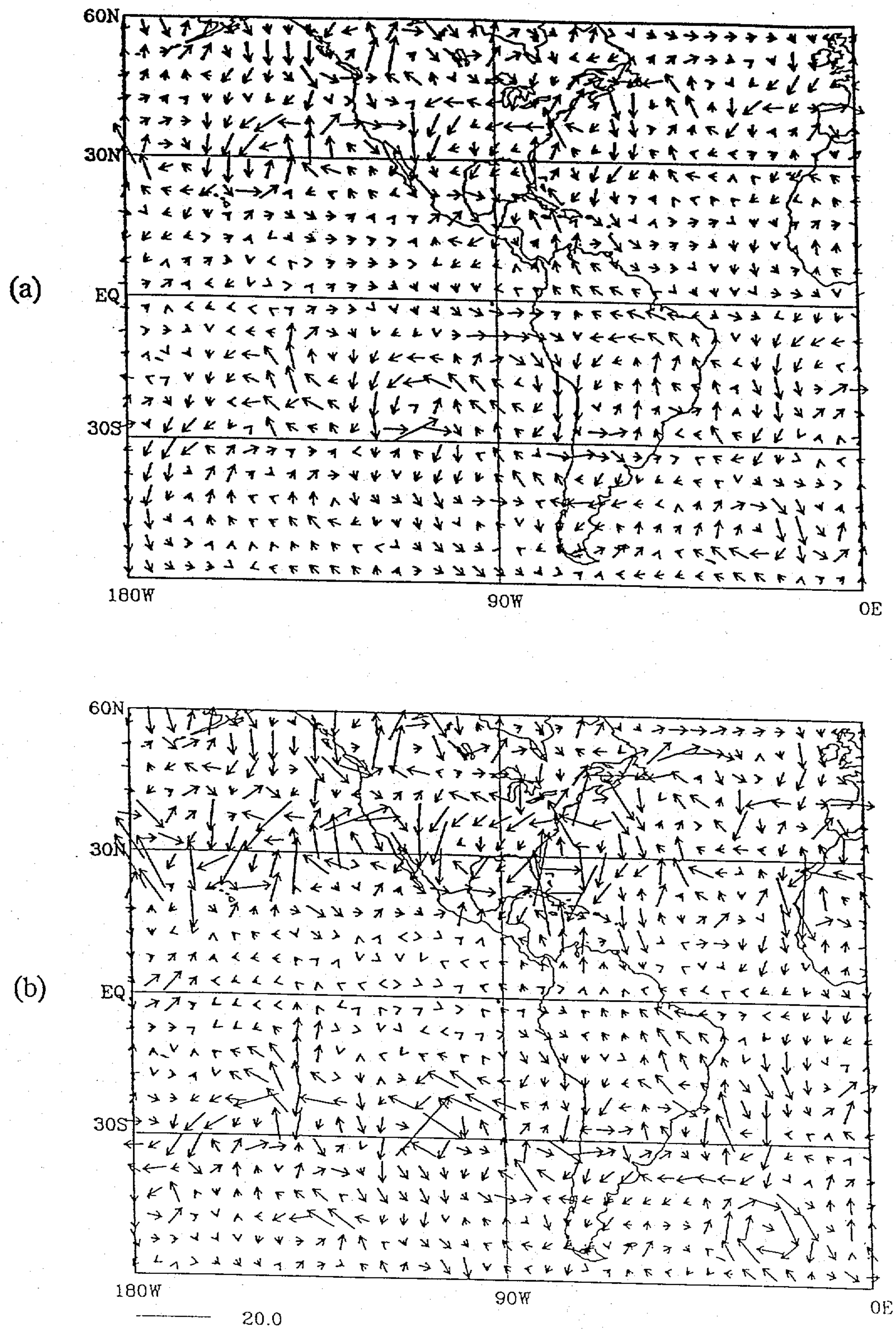


Fig. 5. Same as Fig. 4, but for 250mb wind analysis increments.

(a)

(b)

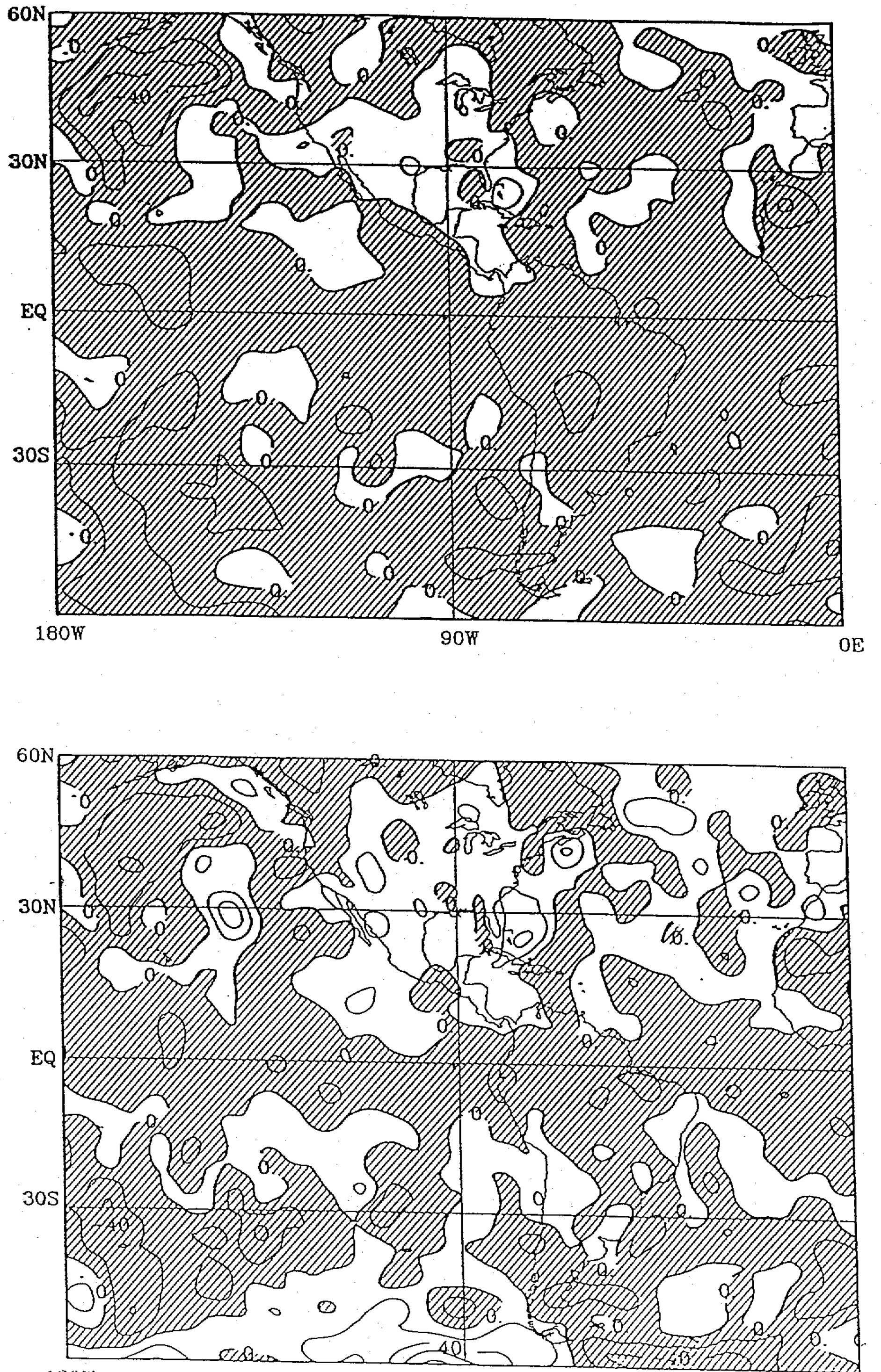


Fig. 6. 250mb height difference (operational – SSI) for (a) both with same background, and (b) SSI background from SSI assimilation for 60 days. Contour interval is 20m.

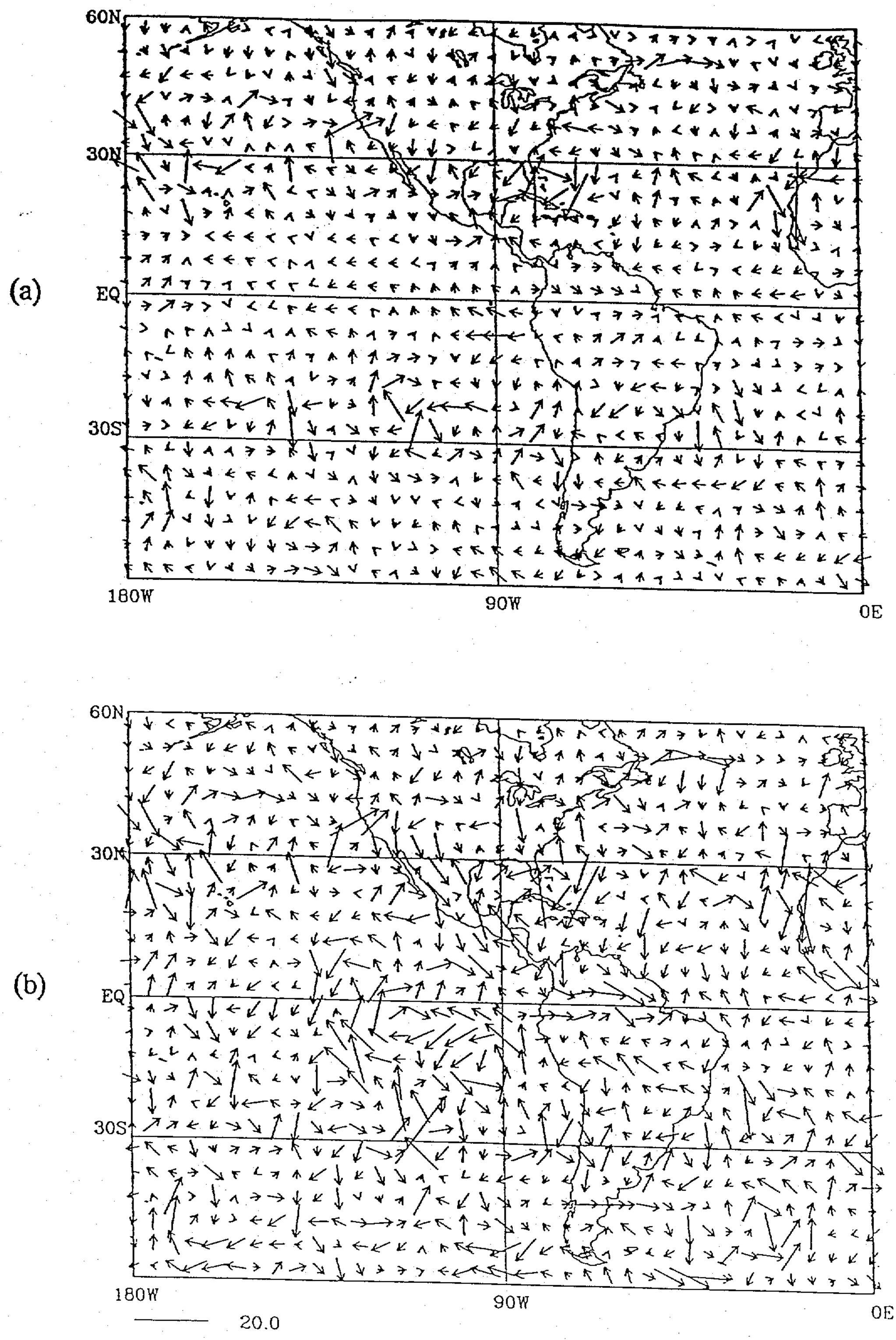


Fig. 7. Same as Fig. 6, except for wind differences.

(a)

(b)

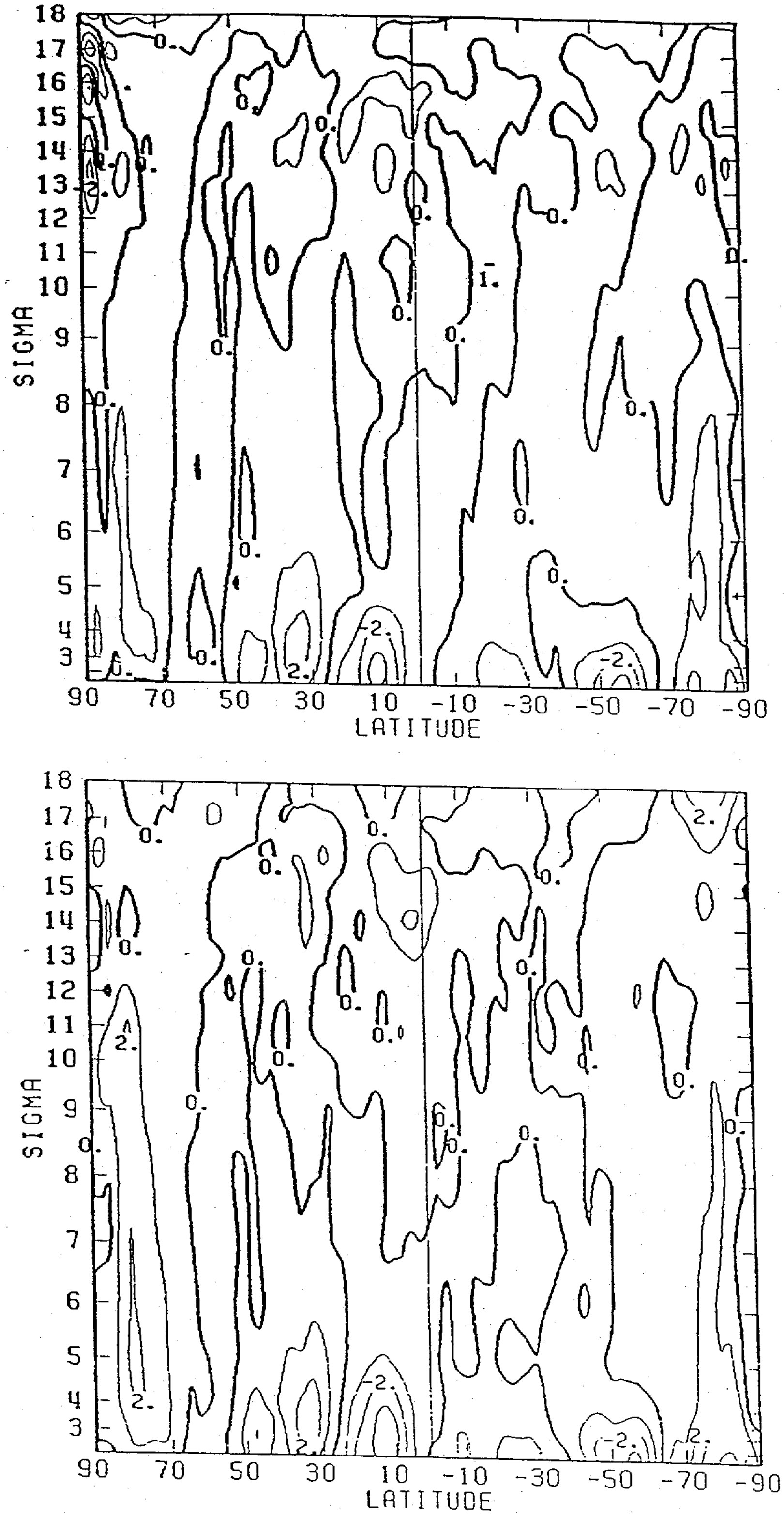


Fig. 8. Zonal mean cross sections from (a) SSI from assimilation and (b) operational analyses for 0000UTC Mar. 6, 1991.

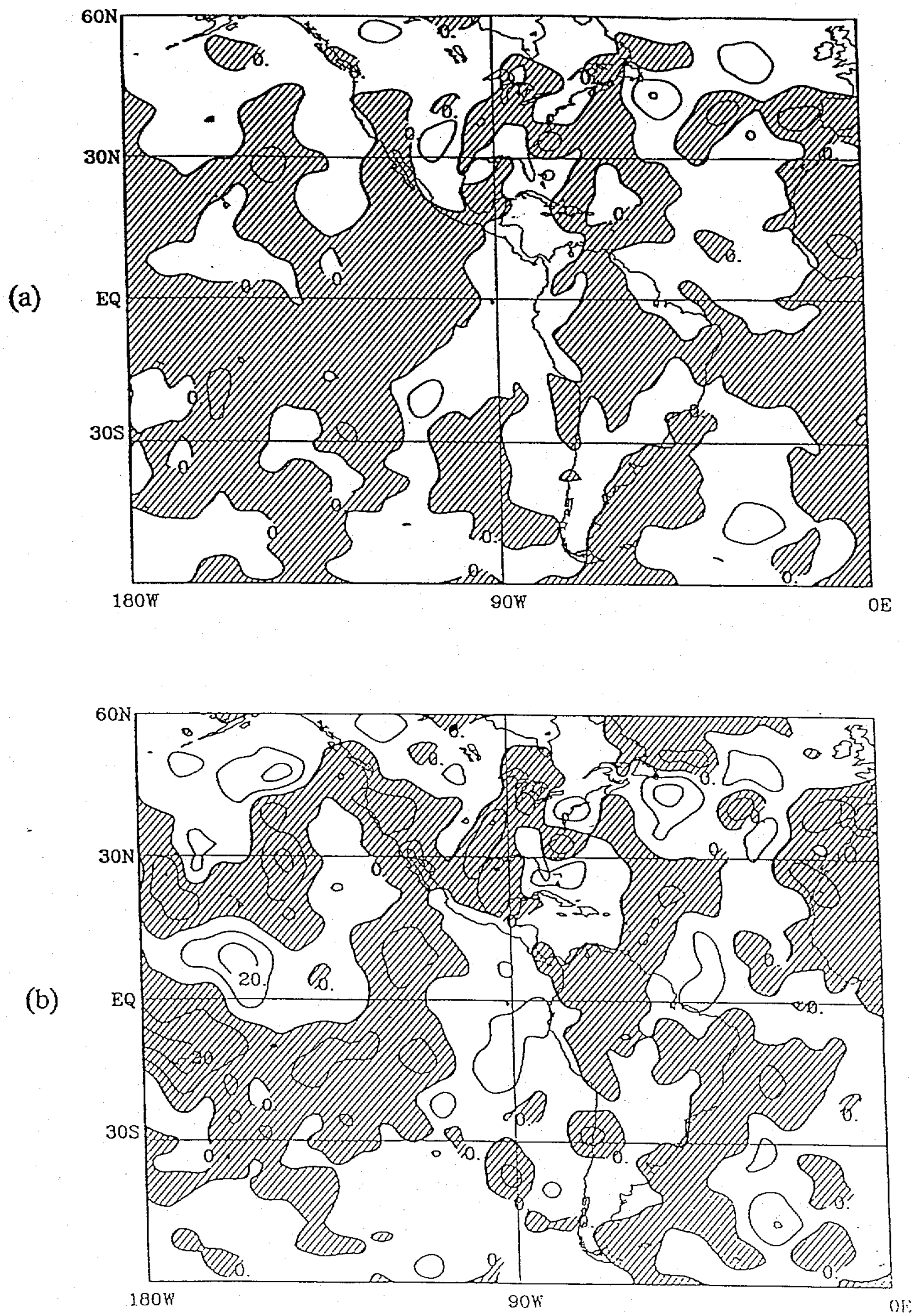


Fig. 9. Same as Fig. 4, except for initialization increments (initialized - analysis).

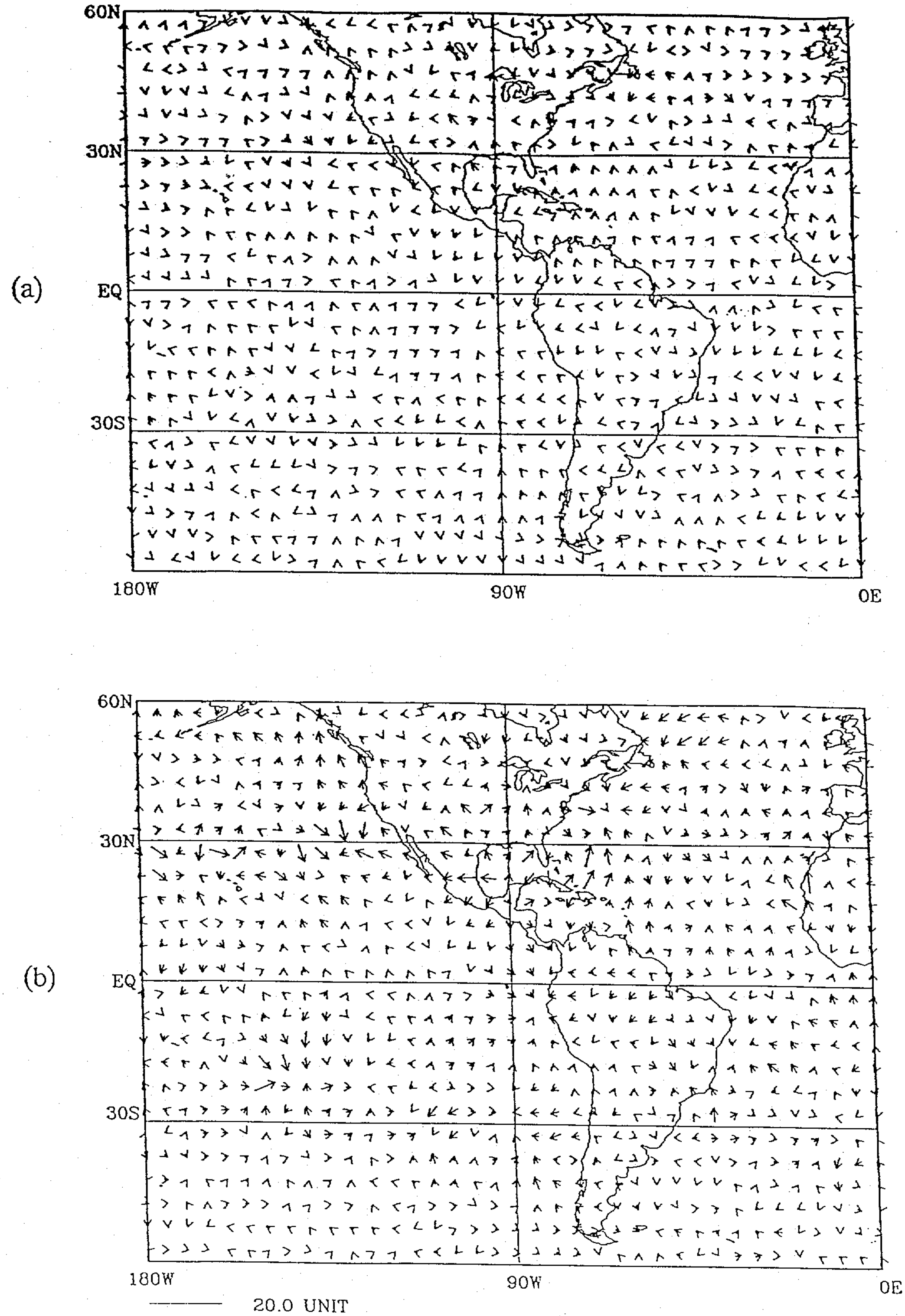
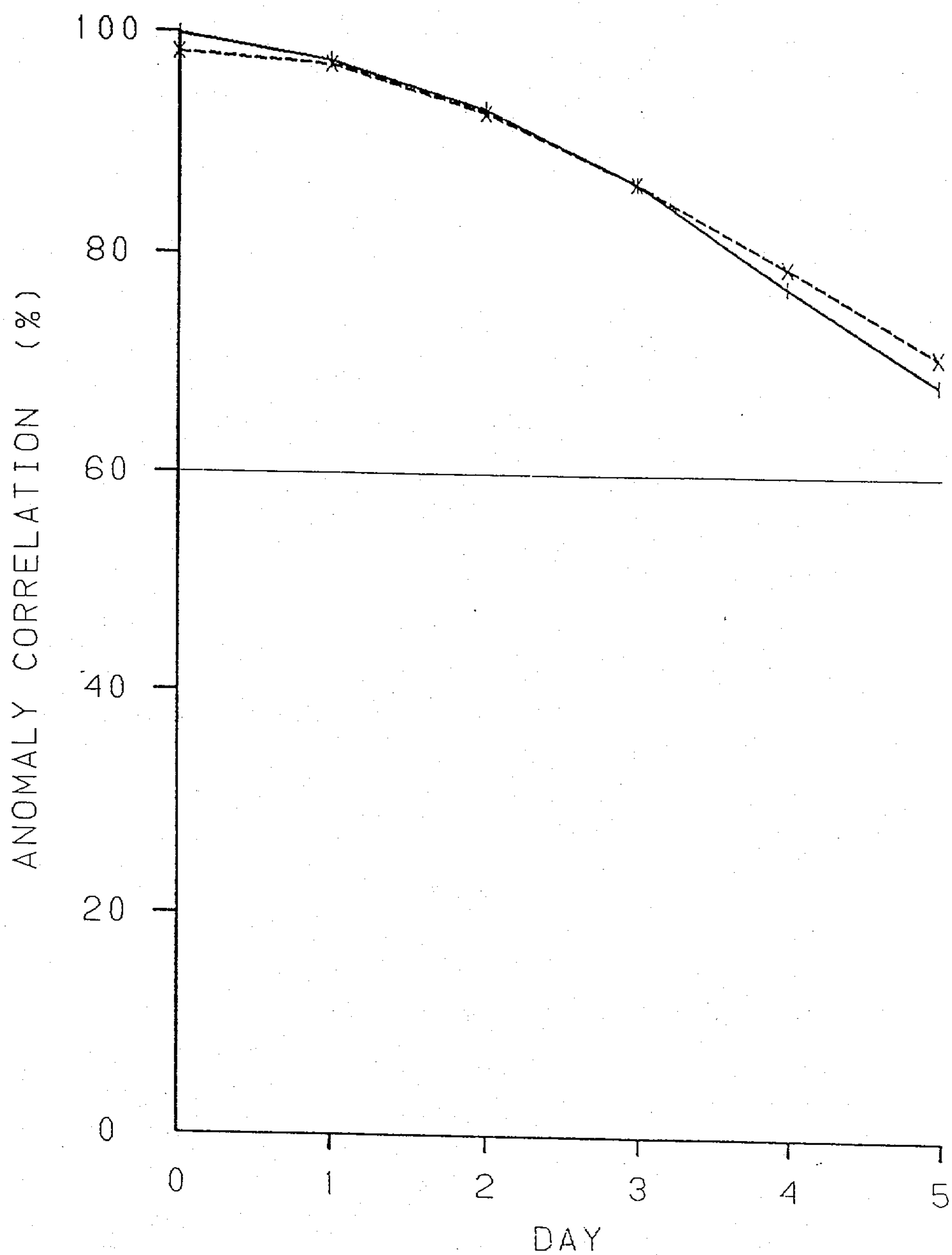


Fig. 10. Same as Fig. 5, except for initialization increments.



<u>Fig. 11.</u> Anomaly correlation scores for 1–5 day northern hemisphere 1000mb height forecasts, verified against operational analyses, averaged over the period 15Jan–14Feb 1991. Solid line is from operational assimilation, dashed SSI.

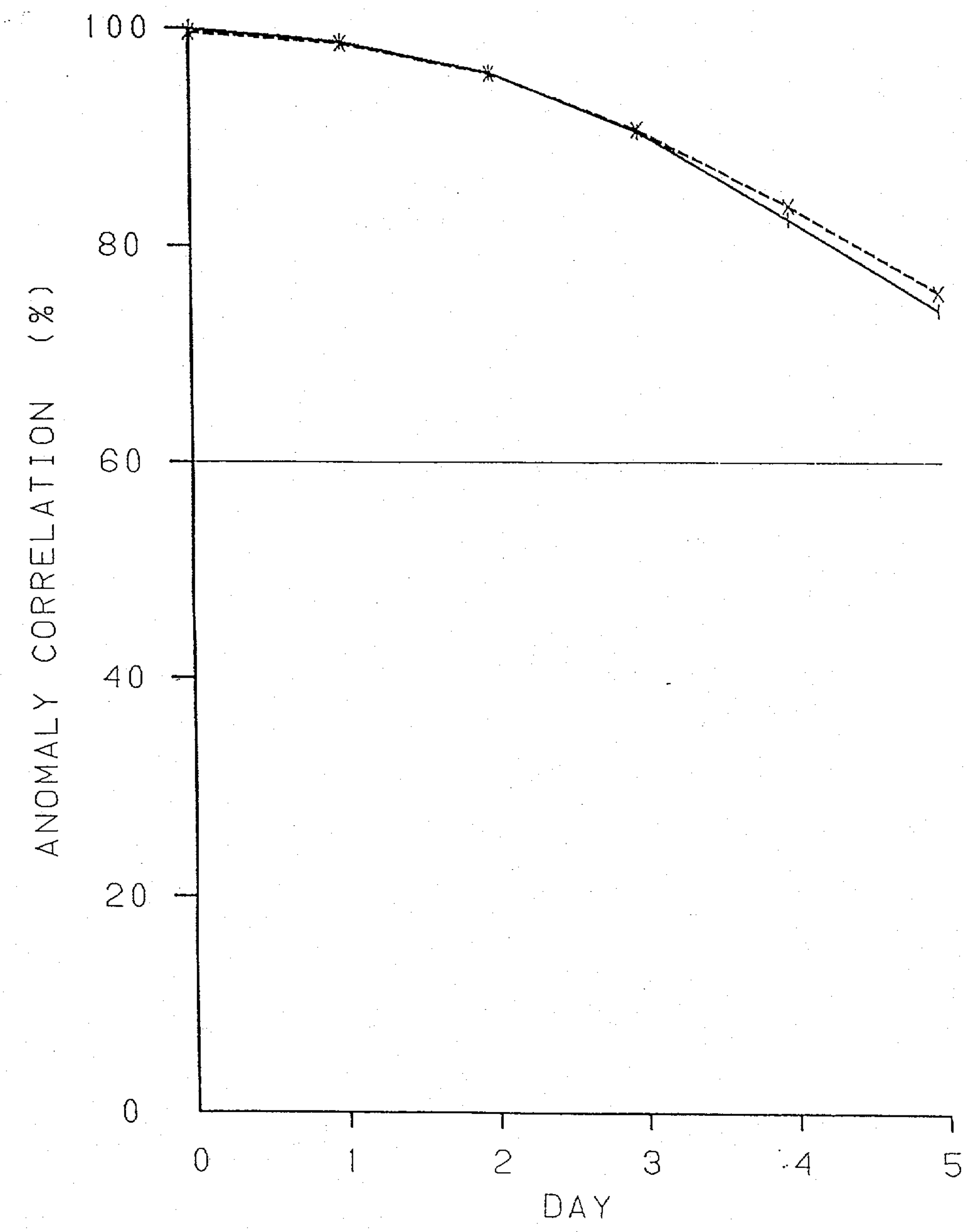


Fig. 12. Same as Fig. 11, but for northern hemisphere 500mb height forecasts.

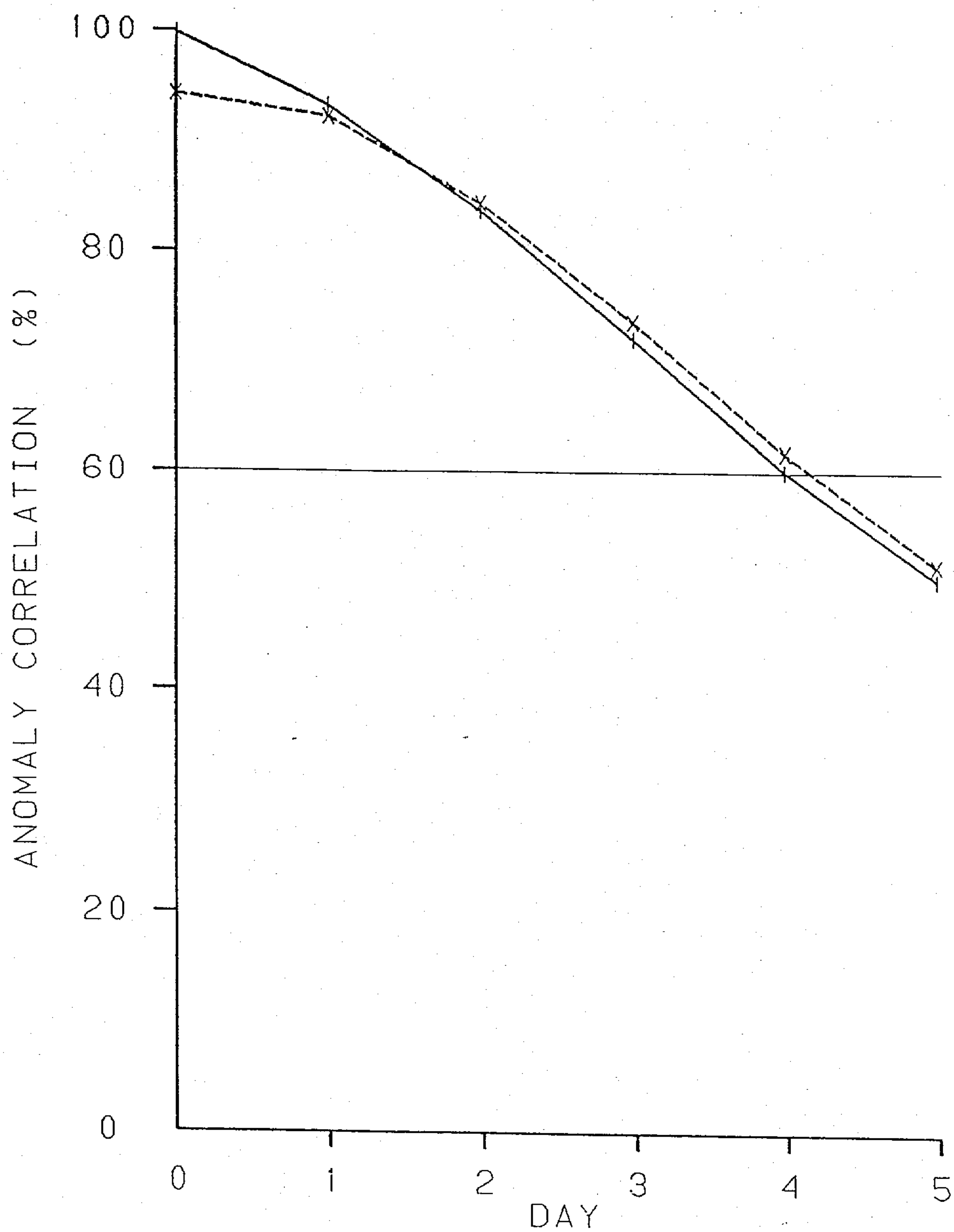


Fig. 13. Same as Fig. 11, but for southern hemisphere 1000mb height forecasts.

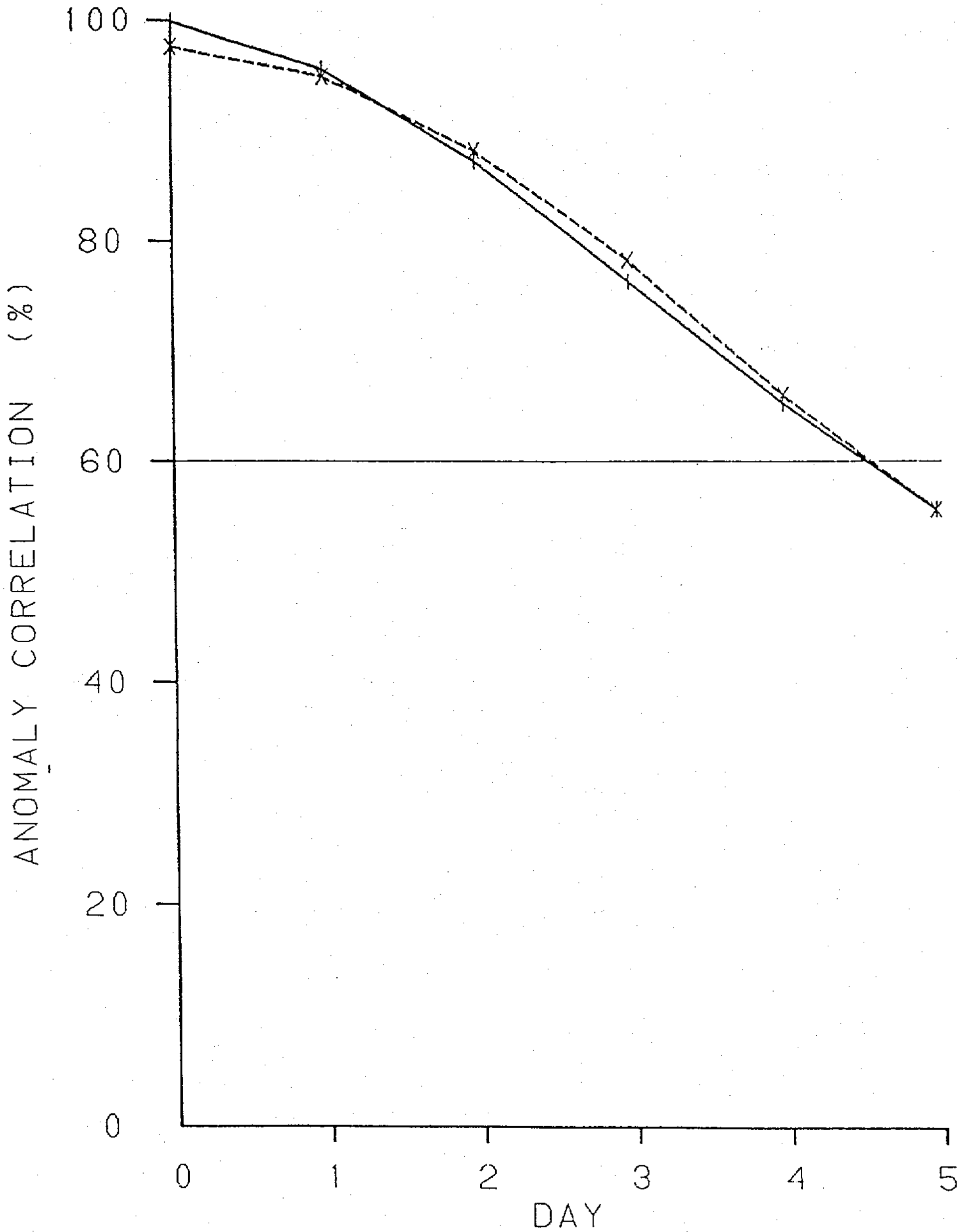


Fig. 14. Same as Fig. 11, but for southern hemisphere 500mb height forecasts.