Revised Interpolation Statistics for the Canadian Data Assimilation Procedure: Their Derivation and Application

HERSCHEL L. MITCHELL, CÉCILIEN CHARETTE, CLÉMENT CHOUINARD AND BRUCE BRASNETT*

Recherche en Prévision Numérique, Atmospheric Environment Service, Dorval, Québec, Canada

(Manuscript received 7 August 1989, in final form 29 January 1990)

ABSTRACT

The first part of this paper presents the results of a study of the structure of the observed residuals, or differences, between radiosonde data and the short-range forecasts that are used as trial fields in an operational hemispheric data assimilation scheme. The study is based on fitting appropriate functional representations to horizontal correlations of observed height and wind residuals. Rather than represent the height residuals by the sum of a degenerate second-order autoregressive function and an additive constant to account for long-wave error, as in a previous study, we use a representation consisting of a sum of two degenerate third-order autoregressive functions of the form $(1 + cr + c^2r^2/3)\exp(-cr)$, where r represents radial distance. For the wind residuals, we use the functional form that follows by geostrophy. In addition to examining the structure of the horizontal and vertical correlations, we also present other statistics relating to the performance of the data assimilation procedure, such as vertical profiles of the magnitude of the observed wind and height residuals for various regions.

In the second part of the paper, the results of the study are used as a basis for specifying interpolation statistics for the objective analysis. To evaluate the impact of the new interpolation statistics, various objective measures of analysis performance are examined and parallel 48-h forecasts are performed. It is found that significant improvements result when the new interpolation statistics are used in the data assimilation procedure.

1. Introduction

Since the pioneering work of Rutherford (1972, 1976), operational data assimilation at the Canadian Meteorological Centre (CMC) has been performed using hemispheric, multivariate, three-dimensional, statistical (or optimum) interpolation. Rather than interpolating the observations directly, the quantity actually interpolated is the total perceived forecast error; i.e., the difference between a short-range forecast, used as trial field, and the observations. Because the observations themselves are in error, we will generally refer to these differences as observed residuals.

The data assimilation procedure at CMC operates in a 6-h cycle. A hemispheric, primitive equations spectral model with nonlinear normal mode initialization is used to generate the required 6-h forecasts. The forecast model has evolved since it was described by Daley et al. (1976) with greater vertical and horizontal resolution and the implementation of a fairly complete physics package. The analysis step of the procedure produces analyses of horizontal wind components (u, v), height (z), and temperature (T) at 11 isobaric levels and dewpoint depression ($T - T_d$) at 6 isobaric levels. As described by Rutherford (1976), horizontal/vertical separability is used, not only in modelling the prediction error covariances, but also in the analysis procedure, which consists of (i) a vertical statistical interpolation at each data location to fill in any missing levels, followed by (ii) a set of horizontal analyses in which vertical interpolation errors associated with vertically interpolated data enter as observation errors.

Increases in computer power over the past few years have served as a catalyst for optimizing and restructuring the objective analysis (o.a.). The more flexible and efficient program that resulted permitted several improvements to the scheme. These have been described by Mitchell et al. (1988) and include a better data selection procedure, use of a larger number of observed residuals in calculating each analyzed value, and use of a finer analysis grid in the horizontal (resolution $\approx 2°$). These improvements resulted in stronger and sharper jets and a better defined vorticity field having more intense centers. Verification of the ensuing forecasts revealed the superiority of forecasts produced from the upgraded analyses, with the magnitude of the improvement increasing with the length of the forecast period.

* Present affiliation: Canadian Meteorological Centre, Atmospheric Environment Service, Dorval, Québec, Canada.

Corresponding author address: Dr. Herschel L. Mitchell, Recherche en prévision numérique, Atmospheric Environment Service, Dorval, Québec, Canada H9P 1J3.

The resolving capability of the analysis depends not only on the mesh length of the analysis grid and the data density but also on the structure functions and other statistics that are basic to a statistical interpolation procedure. Since statistical interpolation for operational numerical weather prediction (NWP) works in terms of observed residuals rather than the observations themselves, the structure functions and other statistics should be recalculated periodically to keep them appropriate to the current forecast system, various components of which are changed from time to time. This has been discussed by Thiébaux et al. (1986; hereafter TMS). Thus a scheme for deriving the analysis statistics is also a basic component of a statistical interpolation procedure.

The objective of this paper is to report upon important changes recently made to the o.a. in operation at CMC: a new algorithm for deriving the statistics was implemented and this has resulted in a major change to the statistics used in the o.a. Included in these changes is a new functional representation for horizontal correlations and a revised method of calculating vertical covariance structure. Prior to this change, the method used operationally to separate total observed residuals into observation and prediction errors and to determine vertical covariance structure was based on the work of Hollett (1975). These statistics varied in time through a system whereby, every 12 hours, for each variable, level and latitude band, running time means of the total observed residuals were updated on the basis of current radiosonde observations and partitioned on the basis of Hollett's results. Horizontal prediction error correlations were modeled by negative squared exponential (or "Gaussian") functions, as discussed by Rutherford (1976). The function widths had not been changed in many years.

Following Rutherford (1972), Hollett (1975), Hollingsworth and Lönnberg (1986; hereafter HL), Lönnberg and Hollingsworth (1986; hereafter LH) and TMS, the revised o.a. statistics are derived from an evaluation of the structure of observed residuals as determined from radiosonde data. This evaluation has been carried out, and the resulting o.a. statistics have been produced, at more vertical levels than are currently being used operationally, in anticipation of an increase in the vertical resolution of the o.a.

The first stage of this project, described in TMS, was a study of the horizontal correlation structure of observed residuals of height and temperature. These residuals had been obtained from the hemispheric CMC data assimilation cycle at 11 isobaric levels ranging from 1000 to 50 mb. A regional comparison of correlation structure at 500 mb revealed significant differences from region to region with generally larger height correlations for a given station-separation over North America than over Eurasian regions at the same latitude. In general, the width of correlation structure functions for both height and temperature residuals was found to increase equatorward on a given isobaric level [in agreement with the theoretical model of Balgovind et al. (1983)] and with decreasing pressure in a given latitude band. Correlations of temperature residuals were found to have narrower structure than those of the corresponding height residuals [in agreement with the theoretical model of Phillips (1986)]. The importance of anisotropic and seasonal variations was found to be small in comparison with other sources of statistical variation.

Other aspects of correlation structure were examined in HL and LH. The authors focused their attention on the structure of wind and height observed residuals in winter mainly over the North American region between 30° and 60°N. Using the formulation introduced by Daley (1985), they derived the kinematical partial differential equations for two-dimensional homogeneous turbulence and then used this framework to examine the wind residuals. Using observed residuals obtained from the European Centre for Medium Range Weather Forecasts (ECMWF) global data assimilation cycle, they examined, for both wind and height, the magnitude of the residuals as a function of position as well as the structure of their horizontal and vertical correlations. This work formed the basis for a revised set of interpolation statistics, including a Bessel series expansion for prediction error correlation structure, for the ECMWF data assimilation procedure (Shaw et al. 1987). The theoretical model of Phillips (1986) successfully reproduced many of the features observed by HL and LH.

The method used to study the structure of CMC observed residuals is described in sections 2 and 3. The study yields various statistics relating to the performance of the CMC operational data assimilation scheme. These are presented in section 3. In section 4 we discuss how this information about the structure of the observed residuals is used in the specification of the analysis statistics. Section 5 deals with the impact of these changes to the o.a. statistics on the analyses and the ensuing forecasts. In section 6 we present a summary and concluding discussion.

2. The procedure used to derive the o.a. statistics

a. The observed residuals

The observed residuals studied here were obtained from the CMC hemispheric data assimilation cycle using North American radiosonde data at 0000 at 1200 UTC during a winter period: 15 December 1987–15 March 1988. We will also briefly consider the residuals for a summer period: 1 June 1988–31 August 1988. Throughout these periods a T59 spectral forecast model with 20 levels in the vertical was used to drive the o.a. The analyses were being produced on a 180 × 45 Gaussian grid, the transform grid of the model, at 11 isobaric levels ranging from 1000 to 50 mb. The 50 mb analysis, which had been added to Rutherford's

10-level scheme in 1980, was a two-dimensional (horizontal) analysis, not directly coupled to the other analysis levels by the vertical analysis step.

In the fall of 1987 the CMC o.a. began using an expanded data file. For radiosonde reports this file contains not only data at mandatory levels (i.e., 1000 mb, 850 mb, 700 mb, etc.), but also, below 100 mb in layers where significant level data is available, values at 25 mb intervals interpolated linearly in log pressure ($\ln p$) from mandatory and significant level data. By interpolating the model 6-h forecast to these new levels (in addition to the 11 original analysis levels), it thus became possible to calculate observed residuals at these levels as well. The operational analysis has continued to use only mandatory level data but, in preparation for a possible increase in the number of analysis levels, the statistics and structure of the observed residuals at seven of these additional levels were also examined as part of this study. Thus the full set of levels to be considered below is 1000, 950, 900, 850, 800, 750, 700, 600, 500, 400, 350, 300, 250, 200, 150, 100, 70, and 50 mb. Since, during the period under study, CMC was receiving significant level data only over the North American region, we focus our attention on that region and in particular on the midlatitudes (30°–60°), although more northerly and southerly regions are also considered. As in TMS only those stations having reported at least 100 times during the given season and only station pairs having more than 100 common reports are retained in the calculations. As in TMS and HL and LH the mean of the observed residuals is removed separately for each station prior to the calculation of the covariances.

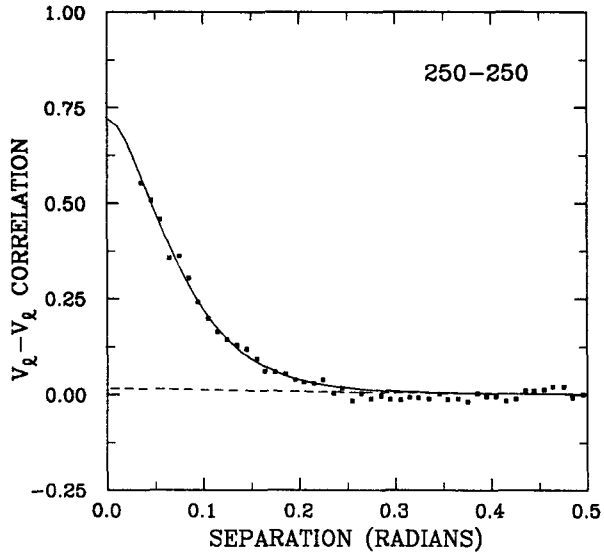

FIG. 1a. The dots show the interval-averaged correlations versus separation (in radians) of the longitudinal component of the wind residuals at 250 mb. The solid curve, which is the plot of a fitted correlation model, and the dashed curve are described in the text. Note that 0.5 radians ≈ 3200 km.

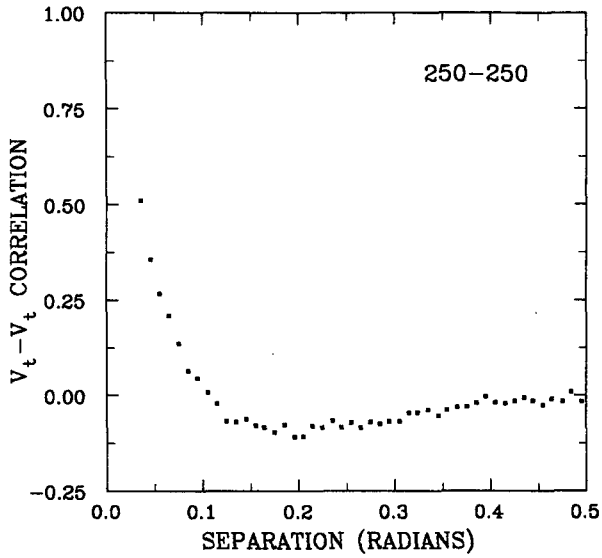

FIG. 1b. Interval-averaged correlations of the 250 mb transverse wind component residuals plotted against separation (in radians).

b. *Interval averaging of correlations*

Since it has been found (e.g., TMS) that the anisotropic variation in correlation structure is small, attention is confined to isotropic representations. We therefore find it convenient to reduce the full correlation arrays to arrays of great-circle distance interval-averaged values. These are computed as in TMS except that (i) the interval over which correlations are averaged is 0.01 radians (≈64 km), (ii) Fisher's z-transform is used in this calculation as in HL/LH, and (iii) interval-averaged correlations are not produced for intervals containing less than three station pairs.

c. *Correlation functions*

As discussed by Buell (1972) and HL/LH, the study of wind correlation structure is facilitated by transforming to longitudinal (l) and transverse (t) components of velocity, along and perpendicular to, the line joining two observations. Interval averaged correlations for each of these components are presented in Fig. 1 for the 250 mb level. (The plotted curves in Fig. 1a are discussed below.) As discussed by HL, the positive longitudinal correlation and the change in sign of the transverse correlation are indicative of nondivergence. In fact, only nondivergent correlation models are considered in this study. Thus we are basically seeking a set of geostrophically consistent correlation functions for height and wind. Once a height representation has been chosen, the corresponding wind representation is implied by geostrophy. To determine if this representation is appropriate, fits to wind correlations, such as those in Fig. 1, can be examined. The relatively narrow structure of transverse–transverse (tt) correlations made the zero-intercepts obtained by ex-

trapolating fits of this quantity to zero separation very sensitive to small variations in the values of the few interval-averaged correlations at small separations. Therefore we choose to focus on the fits to the longitudinal–longitudinal (*ll*) correlations. Given a correlation function for height, the corresponding function for *ll* correlations is obtained assuming isotropy, homogeneity, and geostrophy by applying the operator

$$R = (1/r)(d/dr) \tag{1}$$

and multiplying by $-L^2$, where r is radial distance and L is a scale parameter having dimensions of length (Buell 1972; Hollingsworth and Lönnberg 1986).

The interval averaged height correlations for the 700 mb level are presented in Fig. 2. (The plotted curves are discussed below.) Note that unlike the *ll* correlations which rapidly approach a zero asymptote, the height correlations, like those presented in TMS, fall to zero rather slowly, due to the presence of a component having a large horizontal scale.

The function which TMS found gave the best fits to observed height residuals was the degenerate second-order autoregressive (SOAR) function

$$(1 + cr)\exp(-cr) \tag{2a}$$

with an additional additive constant to account for the large-scale horizontal component. Both the scale parameter c and the additive constant were determined by the fitting procedure. Adoption of the correlation function (2a) for height would imply via (1) an exponential form for *ll* correlations; i.e.,

$$\exp(-cr). \tag{2b}$$

Attempts to fit this exponential function to *ll* correlations yielded very large "zero-intercepts," at times exceeding 1. This could be related to deficiencies in our assumptions of isotropy, homogeneity, and geostrophy or, as discussed by Buell (1972), to the fact that (2b) does not have a horizontal tangent at $r = 0$. Consequently it was decided to adopt a different set of height and wind correlation functions.

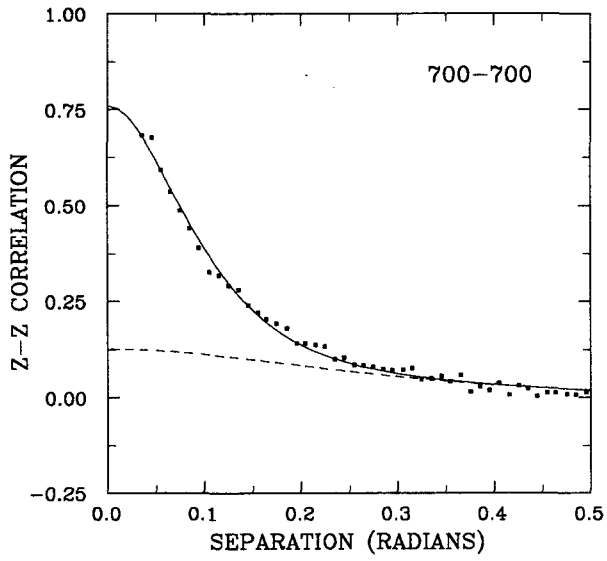

FIG. 2. As in Fig. 1a but for the 700 mb height residuals.

Initially we chose the function

$$f(c, r) = (1 + cr + c^2r^2/3)\exp(-cr), \tag{3a}$$

which is a degenerate form of the third-order autoregressive (TOAR) function (take $a = c$ and let $b \to 0$ in Eq. (2.3) of TMS and note that there should be no division by d in the definitions of α, β, and γ). To this functional representation we added a constant term as before. The corresponding *ll* correlation function is then the degenerate SOAR function

$$f'(c, r) = (1 + cr)\exp(-cr). \tag{3b}$$

Note that here, as in the rest of the paper, a prime indicates the normalized logarithmic derivative.

We observe that the additive constant, which we added to the height correlation representations (2a) and (3a), is annihilated upon application of the operator R. This agrees well with the presence of the large-scale horizontal component in the observed height correlations and the absence of such a component in the observed wind correlations. We note also that both sets of height and wind correlation functions, Eqs. (2) and (3), have been considered by Buell (1972), although without the additional additive constant in the height correlation representation.

In certain ways the constant term in the height correlation representation is analogous to the large-scale term of HL and LH. One troubling aspect of these terms is that their magnitude depends on the distance, D, over which the correlation functions are fit. A second undesirable aspect is that the resulting correlation functions do not go asymptotically to zero. This is unrealistic and effectively limits the data search radius to $D/2$. A correlation representation dropping asymptotically to zero would not impose a limitation on the data selection of the analysis algorithm and was therefore deemed preferable. Rather than including an additive constant to represent the large-scale horizontal component, we finally decided to represent this component by a second, broader TOAR function. The height correlation representation thus becomes a sum of two TOAR functions

$$F(c, r) = (1 + \alpha)^{-1}[f(c, r) + \alpha f(c/N, r)] \tag{4a}$$

where α is a constant, N is a positive integer greater than 1, and f is as defined in (3a). The corresponding *ll* correlation function is then a sum of two SOAR functions

$$F'(c, r) = \left(1 + \frac{\alpha}{N^2}\right)^{-1}[f'(c, r) + (\alpha/N^2)f'(c/N, r)] \tag{4b}$$

where f' is as defined in (3b).

After some experimentation, we chose $N = 3$ and $\alpha = 0.2$, as these values gave good fits. We note that although the large-scale term is one-fifth as large as the remaining term in the height correlation representation (4a), the corresponding term in the wind correlation representation (4b) is only one-forty-fifth as large as the other term, due to the factor $1/N^2$ which is produced by applying the operator R. This difference in relative magnitudes agrees with the observed structure of height and wind correlations discussed above.

d. Fitting procedure

The fitting procedure used to determine scale parameters and zero-intercepts is as described in TMS. Fitting the pair of correlation functions given by (4) to wind and height correlations gives the solid curves in Figs. 1a and 2. In each figure the dashed curve represents the contribution of the term with scale parameter c/N; i.e., the large-scale component. We note that although the figures extend from 0 to 0.5 radians, the fitting procedure only considers those points lying between 0 and 0.4 radians ($\approx$2500 km), as in TMS. As discussed by Rutherford (1972), HL and LH and TMS, the value of the fitted function at zero separation, referred to as the zero-intercept, gives an estimate of the ratio of the forecast error variance to the sum of the forecast error variance plus the observation error variance. This ratio determines the relative weight given to the observations versus the forecast in an ensuing analysis. Knowledge of this ratio allows the total error variance to be separated into forecast-error and observation-error components. It should be noted that the "observation error" calculated in this way includes error due to unresolved scales.

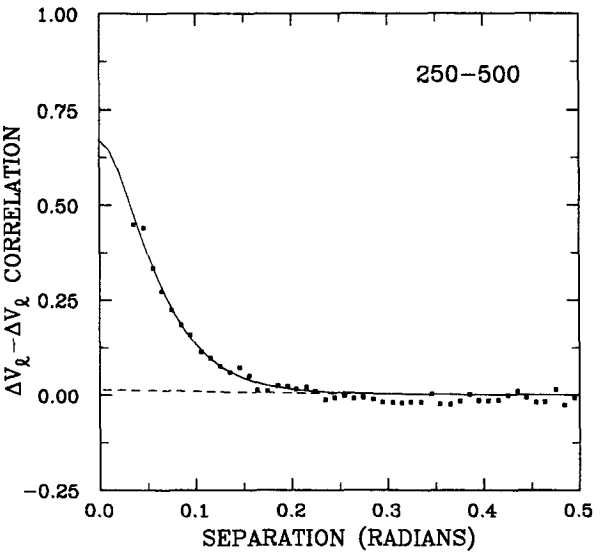

FIG. 3. As in Fig. 1a but for the 250–500 mb shear of the longitudinal component of the wind residuals.

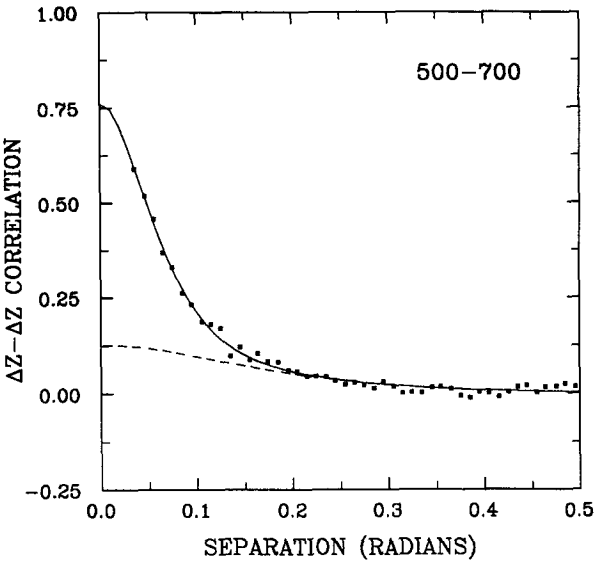

FIG. 4. As in Fig. 1a but for the 500–700 mb thickness residuals.

The vertical (interlevel) correlations were determined using the procedure of HL and LH. Defining angle brackets to be an ensemble average, we have the following relation for the observed residuals of any variable g at any pair of levels A and B

$$\langle (g_A - g_B)^2 \rangle = \langle g_A{}^2 \rangle - 2\langle g_A g_B \rangle + \langle g_B{}^2 \rangle. \quad (5a)$$

The corresponding prediction and observation errors satisfy

$$\langle (g_A{}^p - g_B{}^p)^2 \rangle = \langle (g_A{}^p)^2 \rangle - 2\langle g_A{}^p g_B{}^p \rangle + \langle (g_B{}^p)^2 \rangle \quad (5b)$$

and

$$\langle (g_A{}^o - g_B{}^o)^2 \rangle = \langle (g_A{}^o)^2 \rangle - 2\langle g_A{}^o g_B{}^o \rangle + \langle (g_B{}^o)^2 \rangle. \quad (5c)$$

Knowledge of the variances in (5a) and determination of the zero-intercepts by applying the fitting procedure separately to g_A, g_B and $g_A - g_B$ allows the determination of all variances in (5b) and (5c) and hence for the indirect determination of $\langle g_A{}^p g_B{}^p \rangle$ and $\langle g_A{}^o g_B{}^o \rangle$.

We apply this procedure to the longitudinal component of the wind and to the height. Figure 3 shows the results of fitting (4b) to the correlations of the 250 to 500 mb shear of the longitudinal component of the wind, while Fig. 4 shows the fit of (4a) to the 500 to 700 mb thickness correlations. We note that the thickness correlation has a significantly narrower structure than the 700 mb height correlation of Fig. 2. This is consistent with the difference between height and temperature correlation function widths reported by TMS.

3. The statistical structure of the observed residuals

a. Vertical structure of the wind residuals

The structure function (4b) was fit to the correlations of the observed *ll* wind component residuals, or errors, at all isobaric levels. Figure 5 shows the rms of this wind residual for the midlatitudes of the North American region, together with the partition of this total into observation and prediction error obtained from the fitting procedure. The sum of the squares of the observation and prediction errors equals the square of the total residual. All quantities in Fig. 5 have been multiplied by $\sqrt{2}$ so that they can be easily interpreted as vector-wind quantities. The total error increases from about 4.5 m s^{-1} in the lower troposphere to reach a maximum of just over 8 m s^{-1} at 250 mb and then falls to about 6.5 m s^{-1} at the uppermost levels. The observation error has little vertical variation, ranging from about 3 to 4.5 m s^{-1}, and is generally smaller than the prediction error especially from 200–400 mb.

A comparison of Fig. 5 with the corresponding figure (Fig. 3) of HL shows generally similar vertical structures. One difference is that our results exhibit larger observed residuals in the upper levels. We ascribe this difference partly to the fact that, during the period studied here, the CMC analysis scheme had only one level above 100 mb, and that level (at 50 mb) was not directly coupled in the vertical to the other analysis levels, whereas the ECMWF analysis scheme studied by HL had five analysis levels above 100 mb. Our assimilation model might also be less accurate in the stratosphere for various reasons, including the fact that it has less vertical resolution there and a lower topmost level. We will return to this point in section 3d below. A second difference between our results and those of HL is the somewhat greater prediction errors versus observation errors that we obtain, indicative of somewhat higher zero-intercepts.

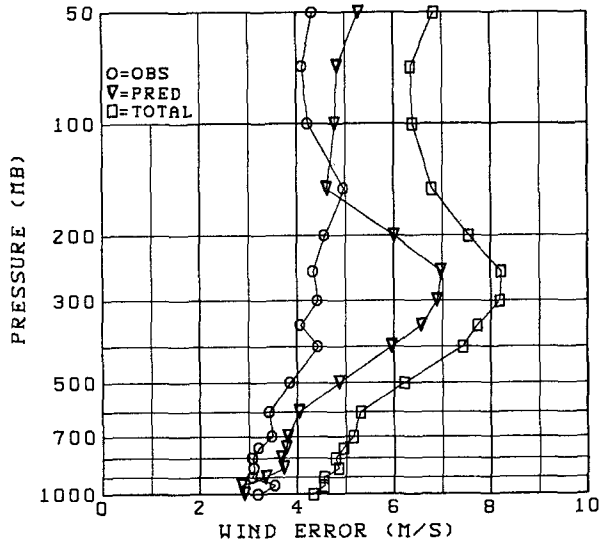

FIG. 5. Vertical profile of the observed wind residual (meters per second) (i.e., total perceived forecast error) denoted TOTAL, and the corresponding profiles of prediction and observation error.

In Fig. 6 we show the radiosonde wind observation error correlations at selected levels. Hollingsworth and Lönnberg (1986), given the following argument due to P. Julian: "Since most of the North American rawinsondes use a radio-theodolite wind-finding system, there should be little vertical correlation of instrument error between successive levels." They then examine the magnitude of the vertical correlation of their wind observation error as a measure of the reasonableness of their error separation. This seems like a useful check despite the fact that the error due to unresolved scales that is included in the observation error could be vertically correlated. This test can be applied to the correlations of Fig. 6 although strictly speaking correlations involving the seven nonmandatory levels included in the calculations should not be considered as they have been computed by interpolation from neighboring mandatory and significant levels. Thus, for example, the wind observation error correlation between 850 mb and both 800 and 900 mb is approximately 0.55. However, correlations between mandatory levels are small, except for several large correlations involving the uppermost levels, indicative of previously discussed problems at these levels. The results indicate that the error separation procedure is functioning properly.

The corresponding correlations for wind prediction error are presented in Fig. 7 as solid lines. (The dashed lines are discussed below.) A comparison indicates that these correlations have vertical structures similar to those presented by HL (Fig. 16).

b. Vertical structure of height residuals

The structure function (4a) was fit to the height correlations. In Fig. 8 we show the vertical profile of the observed height residuals for the North American midlatitudes and the corresponding vertical profiles of observation and prediction error. The total error, which is slightly greater than 10 m in the lower troposphere, increases to reach 26 m at 250 mb and then after decreasing slightly at 200 mb increases with height to reach 46 m at 50 mb. The overall vertical structure is very similar to that presented by LH in their Fig. 2. However, as in the case of the winds, our observed residuals are larger in the upper levels. The prediction error has a vertical structure much like that of the total error and is larger at all levels than the observation error. The latter increases from approximately 5 m in the lower troposphere to 23 m at 50 mb. Lönnberg and Hollingsworth (1986) (see their Fig. 2) obtained larger observation errors than prediction errors above 200 mb.

Figures 9 and 10 show the radiosonde height observation and prediction error correlations at selected levels. Comparison with the corresponding figures from

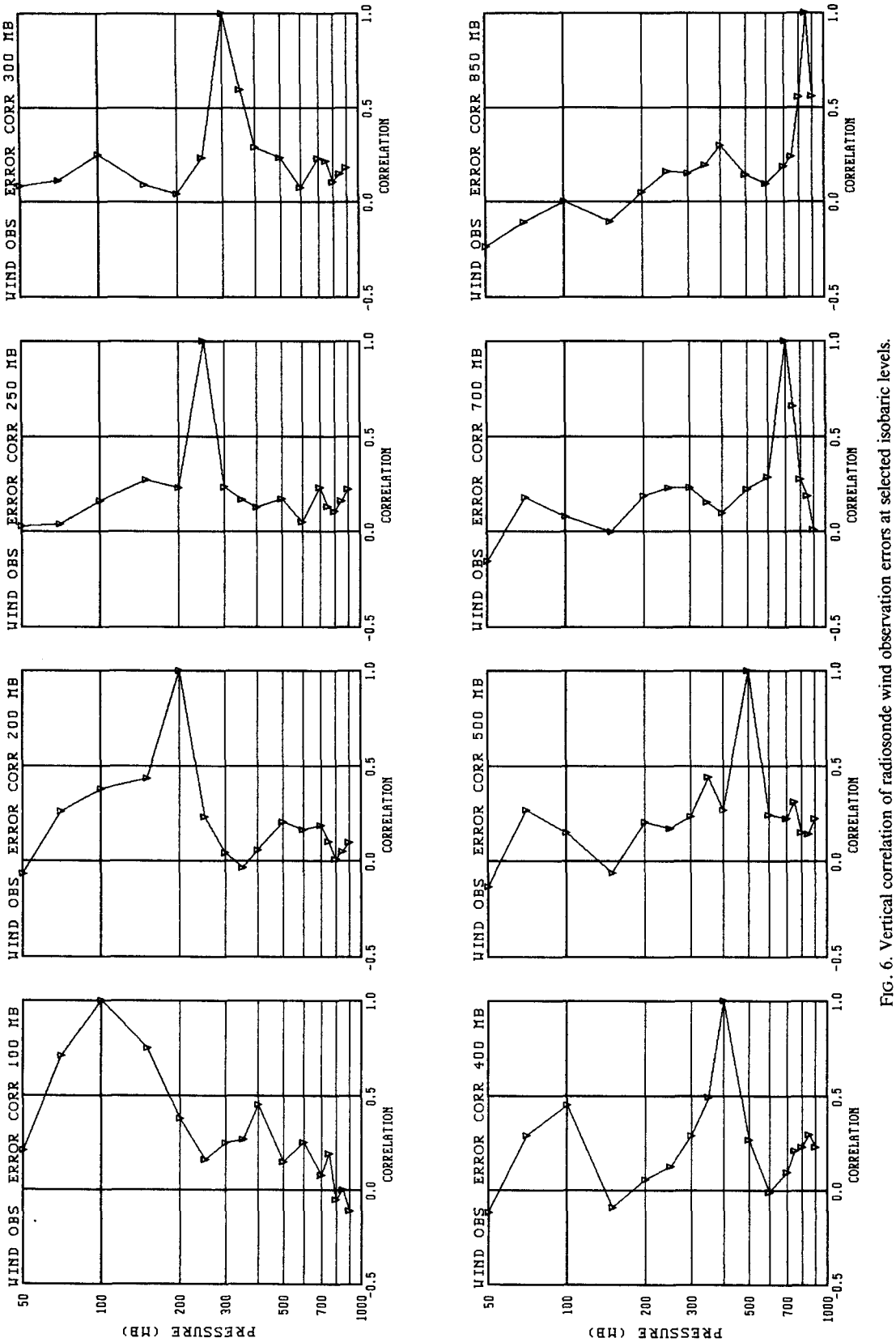

FIG. 6. Vertical correlation of radiosonde wind observation errors at selected isobaric levels.

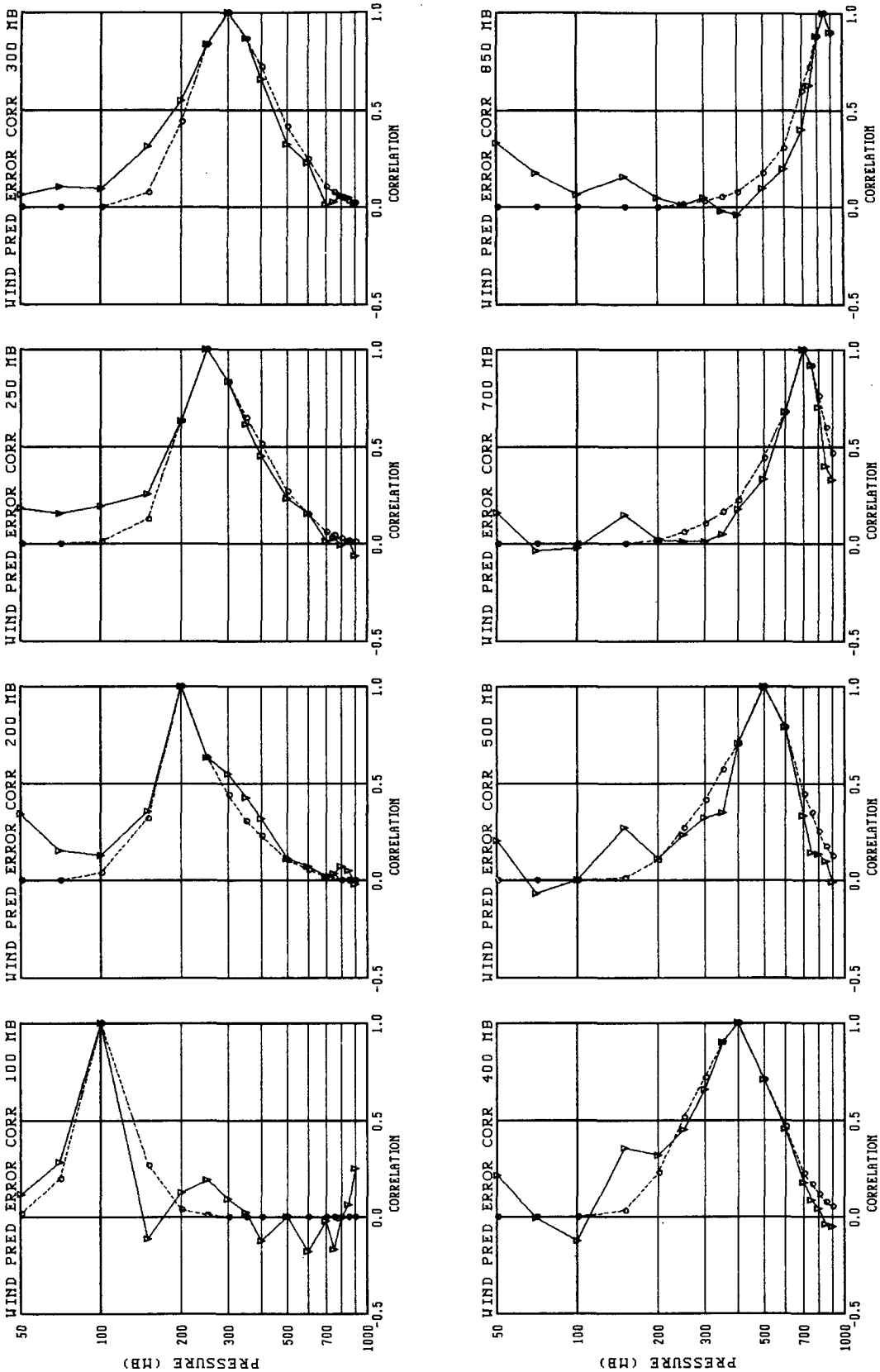

FIG. 7. Vertical correlation of wind prediction errors at selected isobaric levels. The raw correlations obtained from the fitting procedure are plotted using solid lines, while the correlations obtained by applying Cats' procedure are plotted using dashed lines.

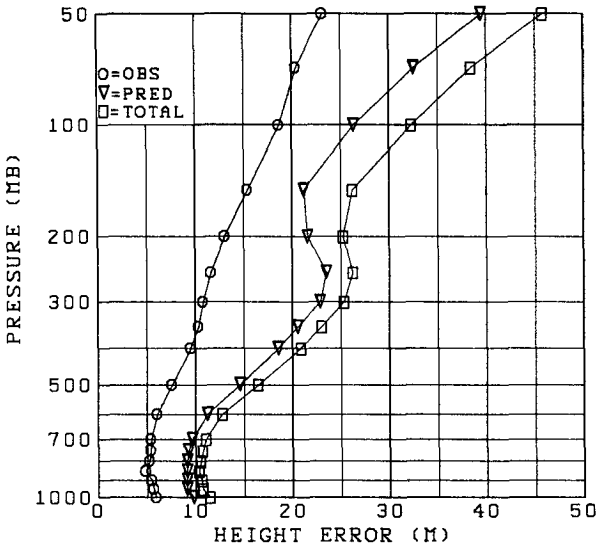

FIG. 8. Vertical profile of the observed height residual (m) (i.e., total perceived forecast error) denoted TOTAL, and the corresponding profiles of prediction and observation error.

LH (Figs. 3 and 10) indicates overall similarity with regard to vertical structure. Two curious features of our correlations are the relative maxima (i) in observation error correlation between the upper levels and 600 mb, and (ii) in prediction error correlation between 100 mb and the levels in the lower troposphere (upper left frame of Fig. 10).

To check the separation of error, we calculated the observation error correlations for thickness from the height observation error covariance matrix, as in LH. The layers were defined as those between adjacent levels and the results are presented in Fig. 11. The interlayer correlations are quite small, which reinforces confidence in the error separation procedure.

c. *Horizontal structure of wind and height residuals*

Having adopted the set of structure functions (4a) and (4b), the horizontal structures of height and wind residual correlations are given by the scale parameter, c, determined by the fitting procedure. As the height and wind fits have been performed independently, each has yielded a set of scale parameters. These have been plotted, as a function of isobaric level, in Fig. 12.

Since (4a) and (4b) are geostrophically consistent, the extent to which the two sets of values agree reflects the degree to which the height and wind residuals are geostrophically related. In fact, the two sets of values agree quite well, which is further confirmation of the method and results. The largest difference in scale parameters occurs in the lower troposphere where the larger scale parameters obtained from the wind fits indicate narrower wind structures than height structures. A comparison of the height scale parameters of Fig. 12 with those presented in Table 4 of TMS, which were based on observed height residuals for complete latitude bands from the winter of 1981/82, indicates very similar vertical profiles. The overall structure is also seen to be similar to that of HL (Fig. 13) and LH (Figs. 6 and 18) who have plotted length scale as a function of pressure (see also Fig. 18 below).

d. *Latitudinal variation*

The results presented up to this point have been calculated at North American radiosonde stations located between 30° and 60°N. In Figs. 13a and 13b we present vertical profiles of the total observed residuals of wind and height for North American radiosonde stations in three overlapping latitude bands: 25°–45°, 30°–60° and 45°–90°. The profiles for the 30°–60° latitude band have appeared above in Figs. 5 and 8.

The latitudinal variation in the total observed wind residual is small up to 400 mb. Above that level the variation becomes large, with relatively small residuals poleward of 45°, intermediate values in midlatitudes, and relatively large residuals equatorward of 45°. By contrast, the latitudinal variation in the total observed height residual, shown in Fig. 13b, is relatively small, even in the upper levels.

The latitudinal variation of the observed wind residuals in Fig. 13a might be explained by the fact that in winter the tropospheric jet is located at relatively low latitudes. One might then expect a different latitudinal variation in summer, with the largest wind residuals occurring further north. We investigate this by examining Fig. 14, the corresponding figure for the summer season (1 June–31 August 1988). At 300 mb and below, the summer wind residuals are consistent with this hypothesis. However, above 250 mb, although the variation is smaller than before, the magnitude of the observed wind residuals still increases equatorward. This suggests that the large observed wind residuals in the 25°–45° latitude band are due, in part, to the equatorial wall in the hemispheric data assimilation scheme under study here. The fact that this effect is observed only at the upper levels points to possible problems in the tropical stratosphere either in the forecast model or in the assimilation scheme. Similarities between the curves for low-latitude and middle-latitude bands in both Figs. 13a and 14 suggest that these problems may be affecting the results for the 30°–60° latitude band as well.

In order to examine how the width of the horizontal structure functions varies with latitude, we present in Fig. 15 vertical profiles of the scale parameter, c, obtained from the height fits for the three overlapping latitude bands considered above. The profile for the 30°–60° latitude band has appeared above in Fig. 12. The profiles from all three latitude bands have similar vertical structures with relative minima at 800 mb and relative maxima near 400 mb. Nevertheless c is seen to vary substantially in the troposphere. Since c varies

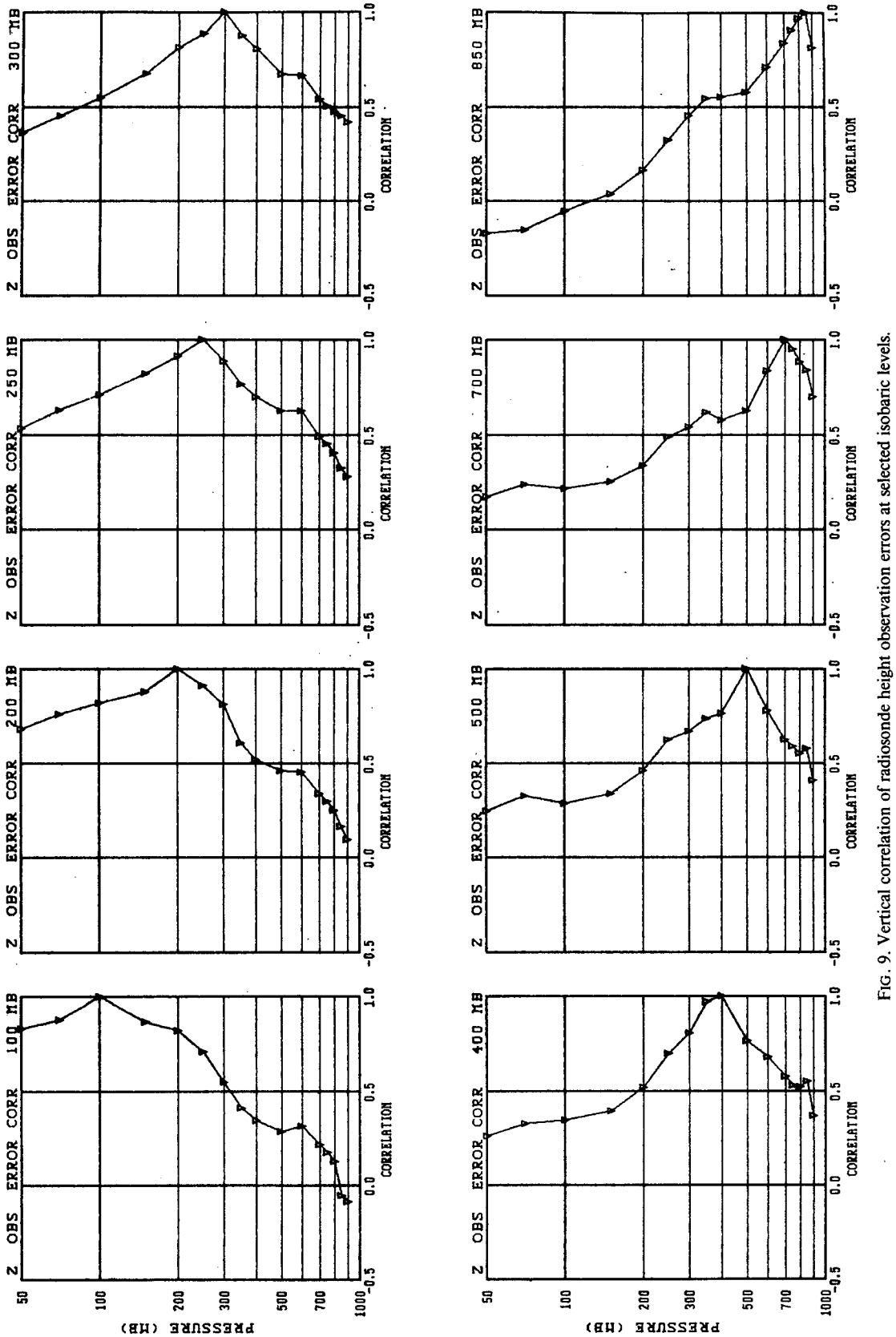

FIG. 9. Vertical correlation of radiosonde height observation errors at selected isobaric levels.

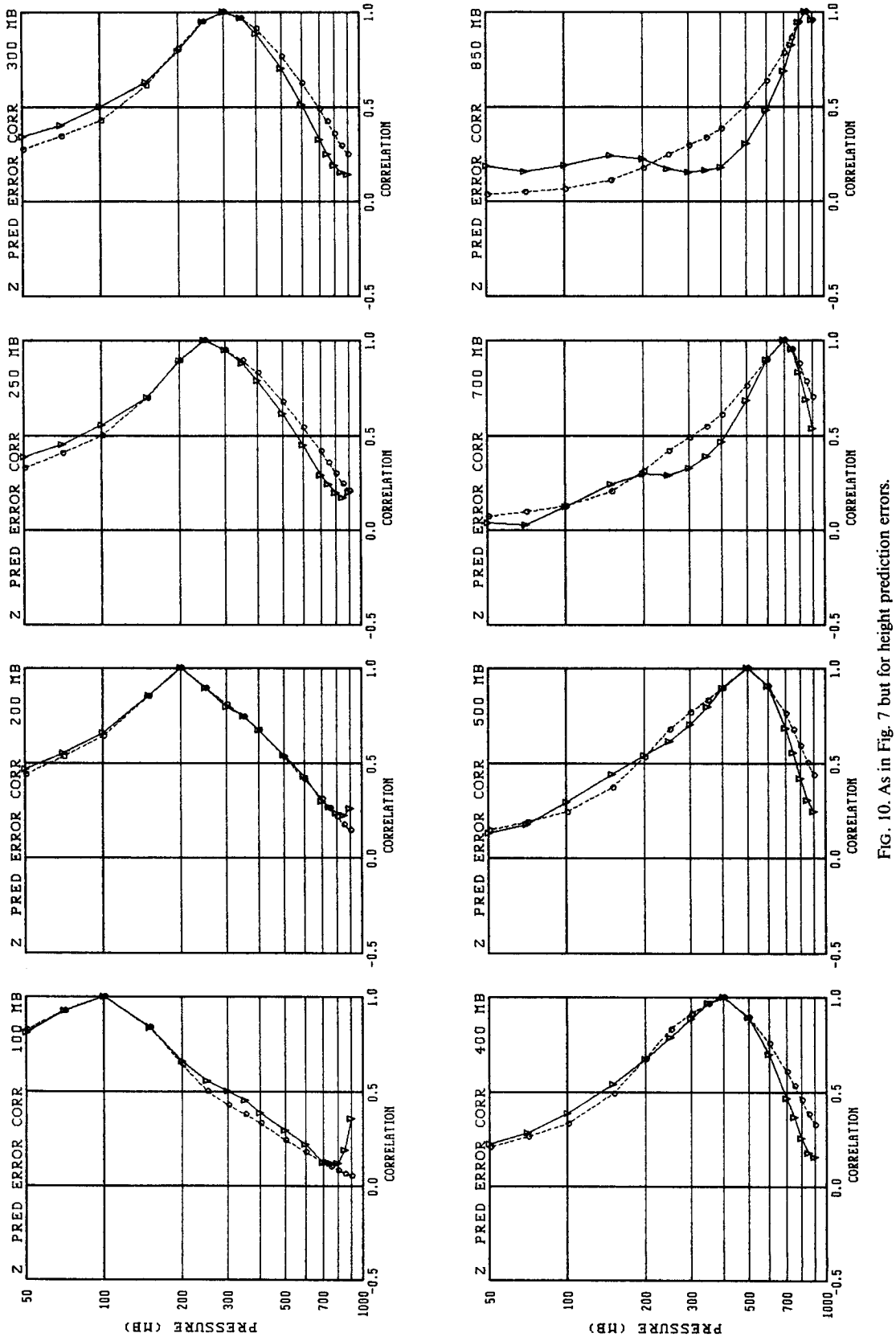

FIG. 10. As in Fig. 7 but for height prediction errors.

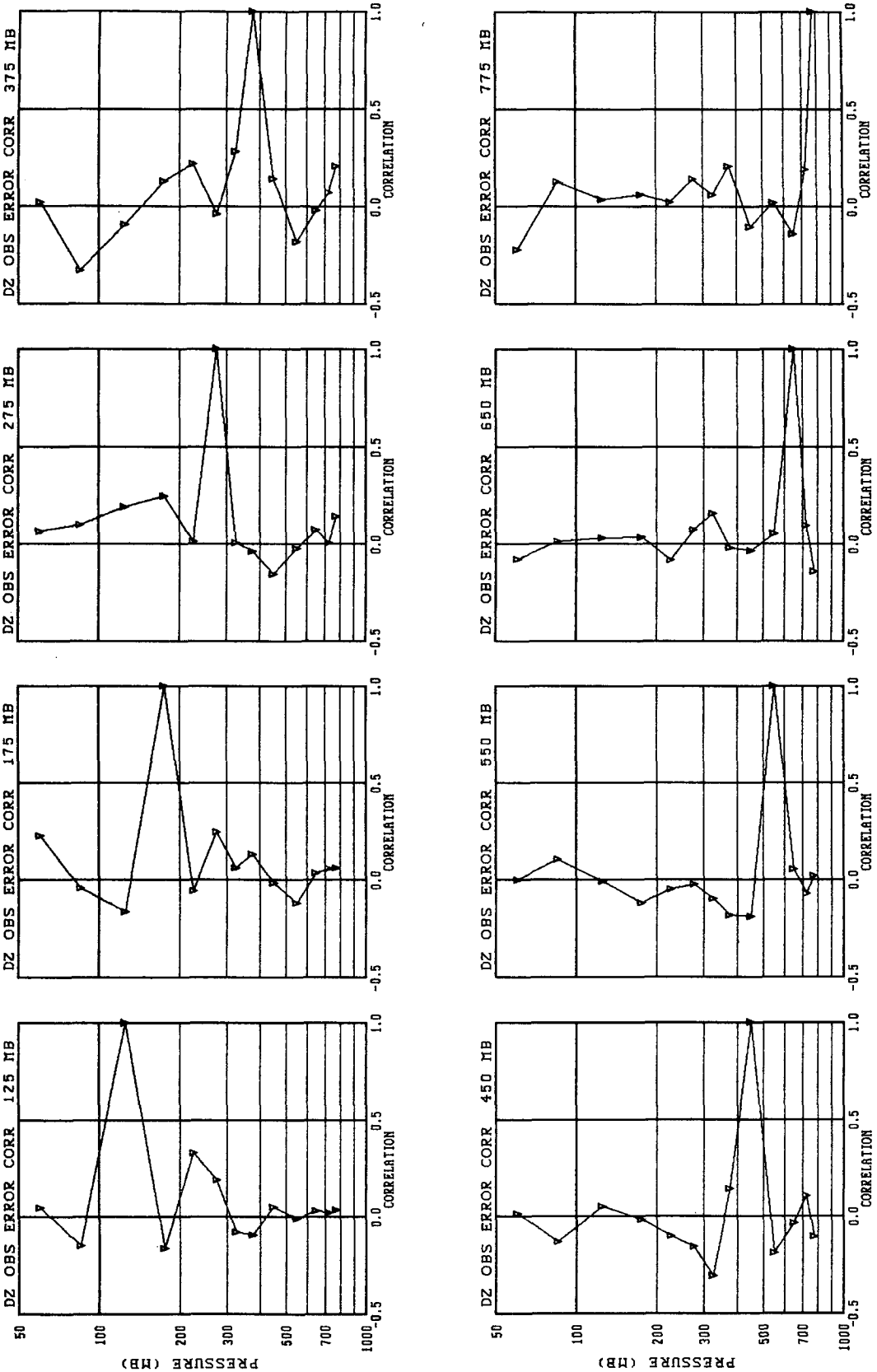

FIG. 11. Vertical correlation of radiosonde thickness observation errors for selected layers. The layers are those between adjacent levels; e.g., 125 mb represents the 100 to 150 mb thickness and so on until 775 mb, which represents the 750 to 800 mb thickness.

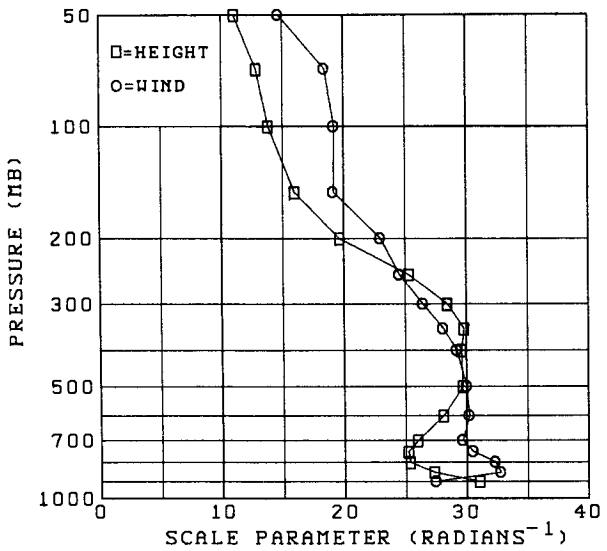

FIG. 12. Vertical profiles of the scale parameter, c (radians^{-1}), determined by applying the fitting procedure to the correlations of observed height and wind residuals.

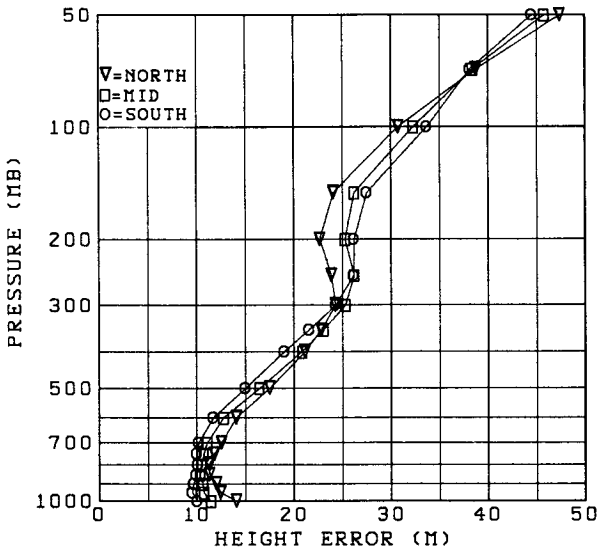

FIG. 13b. As in Fig. 13a but for total observed height residual (m).

inversely as the correlation function width, the increase of c with latitude indicates a narrowing of correlation structure poleward. From 100–250 mb there is very little latitudinal variation in correlation function width. Above 100 mb c decreases with latitude, an effect that may be due to deficiencies in the 50 mb analyses referred to earlier. A comparison of Fig. 15 with Table 4 of TMS indicates similar latitudinal variation of structure function width.

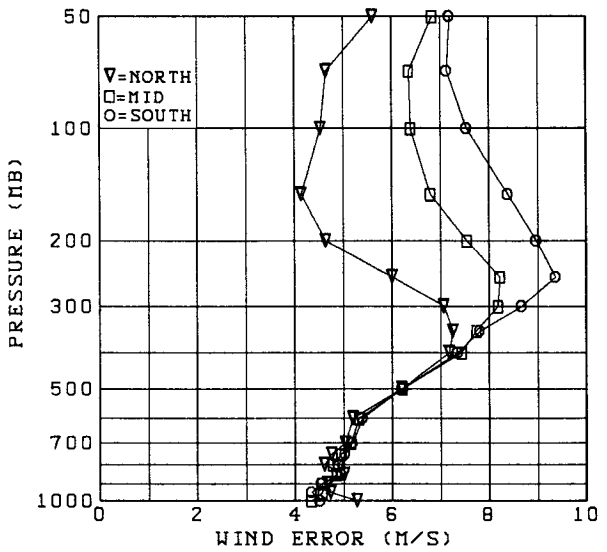

FIG. 13a. Vertical profiles of the total observed wind residual (meters per second) at North American radiosonde stations are shown for each of three overlapping latitude bands: 25°–45° (denoted SOUTH), 30°–60° (denoted MID), and 45°–90° (denoted NORTH).

The latitudinal variation in the structure function widths resulting from the wind fits (not shown) was smaller and less consistent than that shown in Fig. 15 for the height residuals.

4. Specifying the interpolation statistics

We wish to use the information about the statistical structure of the observed residuals to specify the interpolation statistics to be used in the analysis procedure. It must be recognized that the simplifications and modeling assumptions that have been made to this point (e.g., homogeneity) and those about to be made (e.g., geostrophy) mean that, as noted by Rutherford

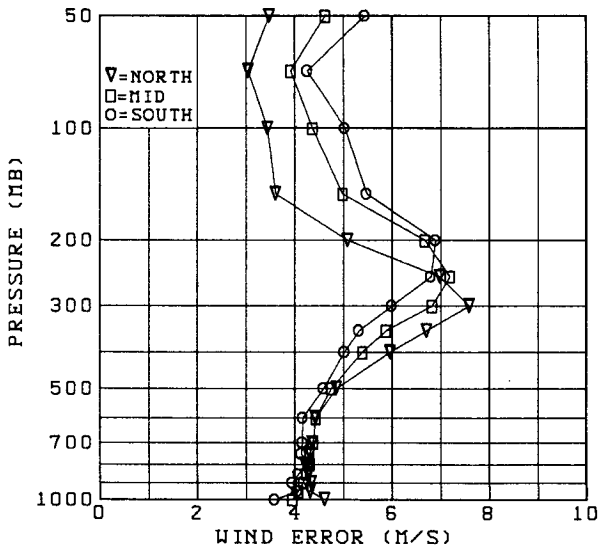

FIG. 14. As in Fig. 13a but for summer.

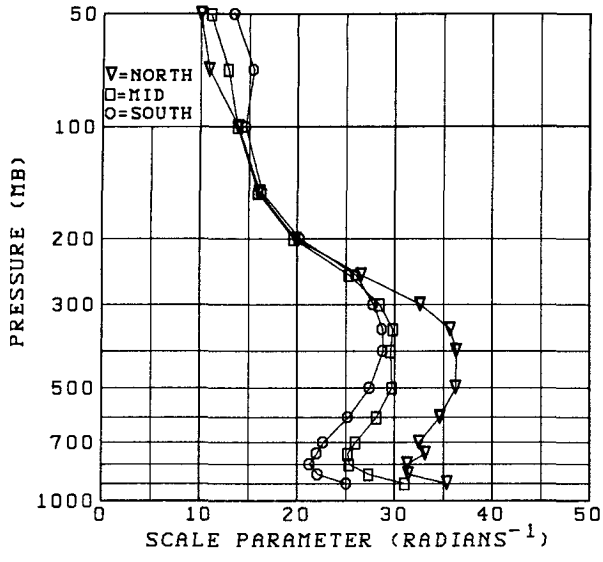

FIG. 15. Vertical profiles of the scale parameter, c (radians^{-1}), obtained from the height fits for each of three overlapping latitude bands: 25°–45° (denoted SOUTH), 30°–60° (denoted MID), and 45°–90° (denoted NORTH).

(1976), "all claims to 'optimality' must be relinquished."

a. Three-dimensional prediction error covariance model

Following Rutherford (1976) three-dimensional prediction error covariances, when required, are calculated assuming horizontal/vertical separability. As mentioned in the Introduction, the full three-dimensional analysis problem is dimensionally split into a vertical analysis step followed by a set of horizontal analyses. Three-dimensional covariances are needed in the CMC o.a. only in the cases of single-level data and to fill in missing levels in sounding data. In the case of complete radiosonde soundings, data is available at all analysis levels, so no vertical correlations are used. In this sense the scheme relies as much as possible on the data to provide the vertical coupling and less heavily on the assumption of horizontal/vertical separability of prediction error covariances than other schemes (e.g., Lorenc 1981).

1) HORIZONTAL PREDICTION ERROR CORRELATION MODEL

The model is presented using the formalism of Daley (1985). Assuming nondivergence the covariances involving streamfunction and height can be written as

$$\left.\begin{aligned} \langle \psi_i \psi_j \rangle &= E_\psi{}^2 F(r) \\ \langle z_i \psi_j \rangle = \langle \psi_i z_j \rangle &= E_z E_\psi \mu F(r) \\ \langle z_i z_j \rangle &= E_z{}^2 F(r) \end{aligned}\right\} \quad (6)$$

where F is a nondimensional correlation function, E_ψ and E_z are the standard deviations of the error of streamfunction, ψ, and height, z, respectively, and r is the distance between arbitrary points i and j:

$$r^2 = (x_i - x_j)^2 + (y_i - y_j)^2,$$

where x_i, x_j and y_i, y_j are the Cartesian coordinates of the points; and μ, the geostrophic coupling coefficient, is a dimensionless constant with magnitude close to 1 outside of the tropics and equal to zero at the equator.

The horizontal prediction error correlation model is based on (4) which can be rewritten as

$$F(r) = (1+\alpha)^{-1}\{[1 + cr + (c^2r^2/3)] \times \exp(-cr) + \alpha[1 + (cr/N) + (c^2r^2/3N^2)] \times \exp(-cr/N)\} \quad (7a)$$

$$F'(r) = \left(1 + \frac{\alpha}{N^2}\right)^{-1}\{[1 + cr]\exp(-cr) + (\alpha/N^2)[1 + (cr/N)]\exp(-cr/N)\}. \quad (7b)$$

Using Helmholtz's theorem and the definition of $F(r)$, we obtain expressions for the horizontal zz, uu and vv correlations and the related cross-correlations in Cartesian coordinates. Details are given in the Appendix and the set of equations (A3) gives the expressions for the complete set of horizontal correlations.

Having set $N = 3$ and $\alpha = 0.2$, it remains to specify c, the scale parameter. To be consistent with the assumption of horizontal/vertical separability, c should not vary in the vertical. However, it is evident from Fig. 12 that c does vary substantially with pressure, especially above 300 mb. We therefore choose to follow Rutherford (1976) by allowing vertical variation. At each isobaric level c is set equal to the average of the value obtained from the height fit and that yielded by the wind fit.

As mentioned in the Introduction, the correlation model previously used in the CMC o.a. was based on a negative squared exponential function. The implied set of correlations at 500 mb for this latter model has been plotted in Fig. 16 with $\mu = 1$. A comparison with the complete set of new correlation functions as derived in the Appendix, and plotted in Fig. 17 with the scale parameter for 500 mb and $\mu = 1$, reveals two significant differences. The first difference relates to the strength of the cross-correlations and negative lobes of the uu and vv correlations. These correlations are weaker in Fig. 17 than the corresponding correlations obtained from the negative squared exponential. For example, in the case of the latter function, the maximum negative uu, zu and uv correlations are approximately −0.4, −0.6 and −0.35; while the corresponding values from Fig. 17 are about −0.2, −0.45 and −0.25. For (2a), the degenerate second-order autoregressive (or Markov) function, Daley (1983; Fig. 2) had observed even weaker correlations, with corresponding values of about −0.1, −0.3 and −0.1.

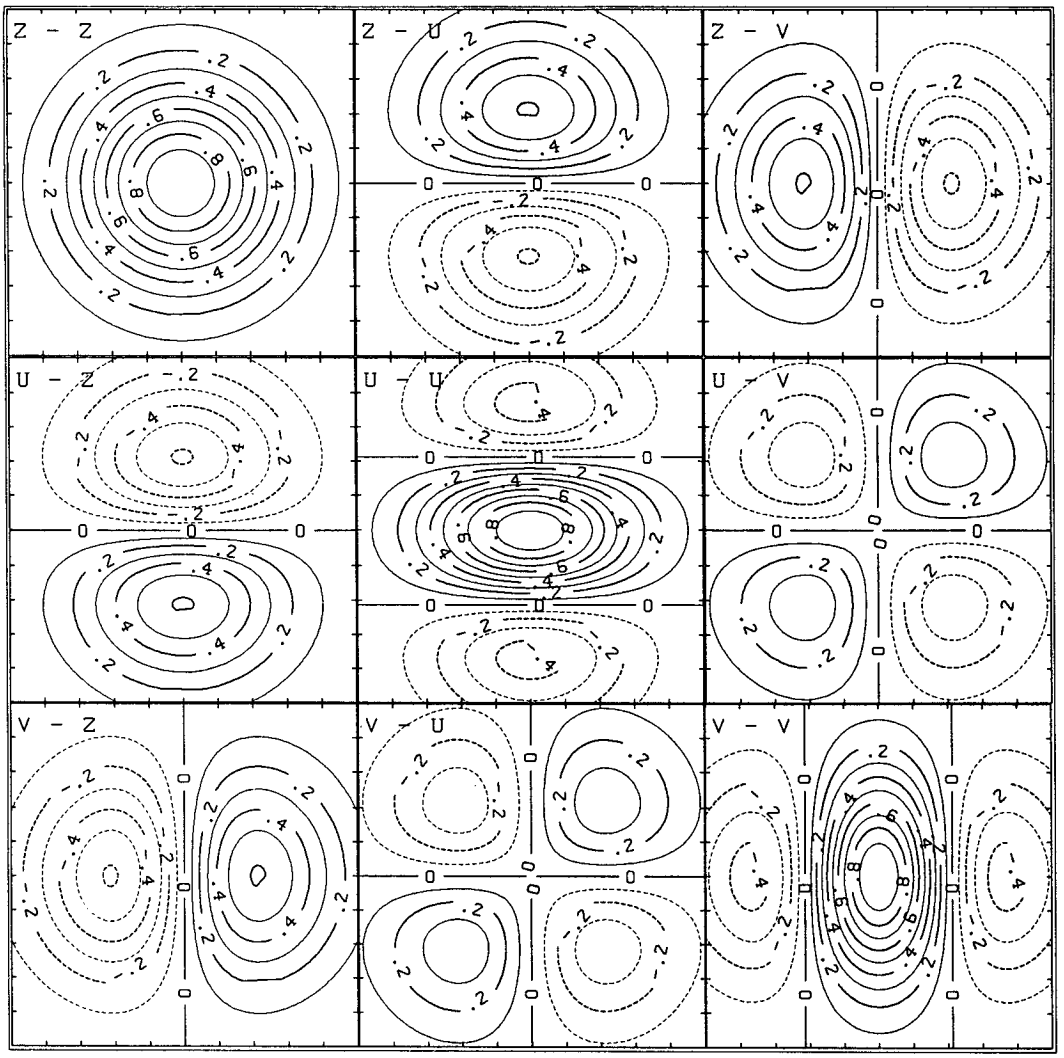

FIG. 16. The complete set of horizontal prediction error correlations for the variables z, u and v based on the negative squared exponential function. The horizontal scale parameter is that which had previously been used at 500 mb in the CMC operational o.a. Tic marks along margins are 250 km apart.

The second difference relates to the horizontal scale of the two sets of functions. The correlations plotted in Fig. 17 have significantly smaller scale than those appearing in Fig. 16, except in the case of zz. To extend this comparison to all isobaric levels, we transform to length scales as in HL/LH. For a correlation function F, the length scale, L_f, is defined as

$$L_f^2 = -\left(\frac{F}{\nabla^2 F}\right)_{r=0}. \tag{8}$$

Substituting (7a) into (8) gives

$$L_f = \left[\frac{3(1+\alpha)}{2\left(1+\frac{\alpha}{N^2}\right)}\right]^{1/2} c^{-1}. \tag{9}$$

For F, a negative squared exponential, expressed as $\exp(-br^2)$, substitution into (8) yields a corresponding expression for L_f as a function of b, $L_f = 1/(2\sqrt{b})$. In Fig. 18 we have plotted, for the 11 CMC analysis levels, the length scales for both the previously used negative squared exponential function and the newly derived correlation function. The length scale of the latter has been obtained from (9), using for c the average of the height and wind values plotted in Fig. 12.

Figure 18 shows that in the troposphere the horizontal scale of the newly derived correlation functions is significantly smaller than that of the functions used previously. This reflects improvements in the resolving capacity of the CMC data assimilation system over the years. Above 200 mb, the length scales in the older scheme seem to have been arbitrarily fixed to the value at 200 mb.

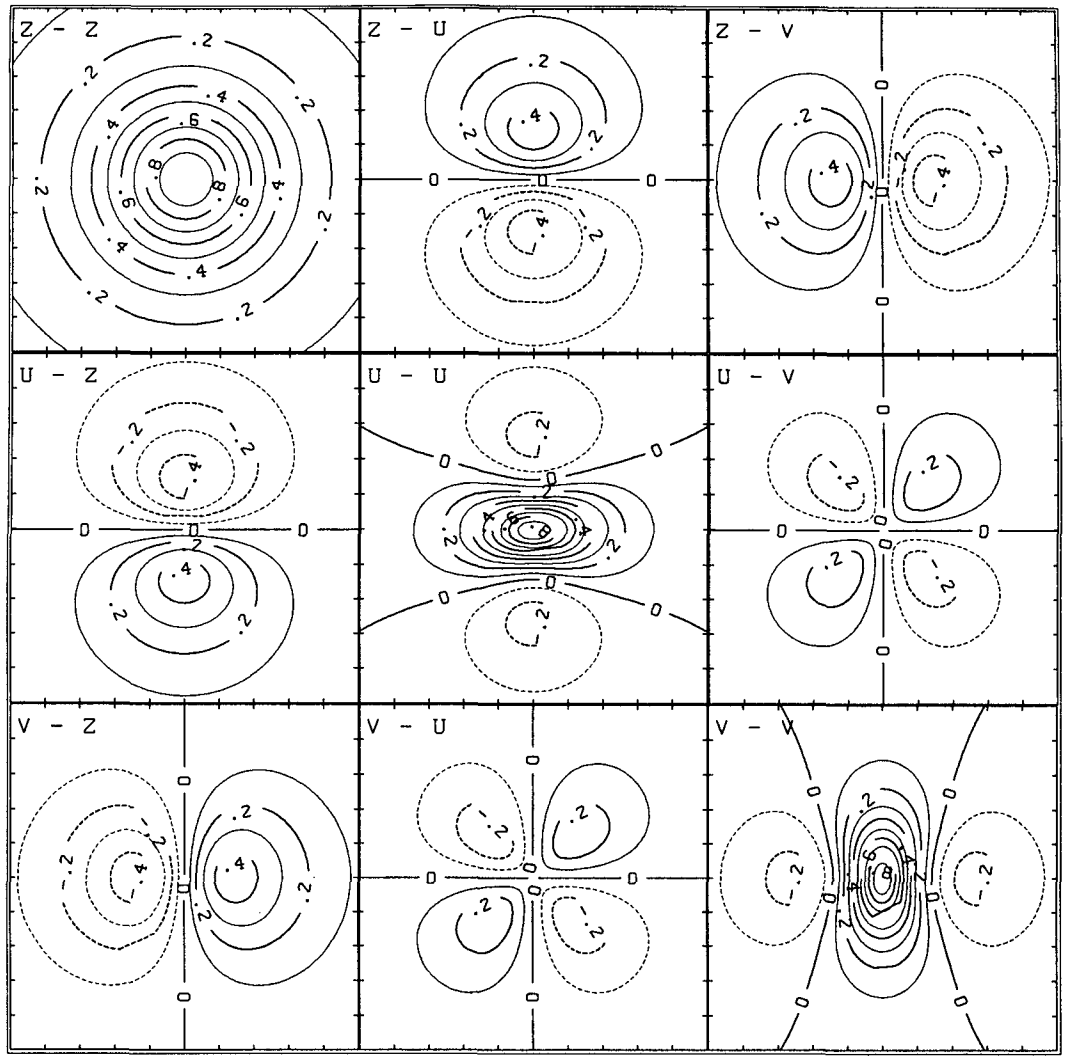

FIG. 17. The complete set of horizontal prediction error correlations for the variables z, u and v based on (7a) with $N = 3$ and $\alpha = 0.2$. The scale parameter, c, is that obtained at 500 mb. Tic marks along margins are 250 km apart.

2) VERTICAL PREDICTION ERROR COVARIANCES

Using the notation of Lorenc (1981), we denote the streamfunction and height vertical prediction error correlations by $V_{\psi kl}$ and V_{zkl}, where k and l denote level indices. Since both the observed and analyzed residuals, which are the input and output of the analysis, are specified at a fixed set of predetermined isobaric levels, $V_{\psi kl}$ and V_{zkl} can be specified as matrices of interlevel correlations. In section 3 we obtained estimates of these correlations by using the error separation procedure and plotted many of the results as the solid lines in Figs. 7 and 10. However, for use in the analysis we would prefer values that vary more smoothly and more monotonically. Such values are obtained by applying a procedure due to Cats (see Lönnberg and Shaw 1987, p. 2.21) to the raw $V_{\psi kl}$ and V_{zkl} matrices. The procedure generates a full correlation matrix using only the correlations between adjacent levels. Before applying the procedure, adjustments were made to some of these correlations; e.g., the negative wind prediction error correlation between 100 and 150 mb was changed to a positive value. The results of applying the procedure (with Cats' parameter $\sigma = 1.4$) have been plotted as the dashed lines in Figs. 7 and 10. As can be seen, application of the procedure has the desired effect.

Examination of Figs. 7 and 10 indicates that the vertical wind prediction error correlations are substantially narrower than the corresponding height prediction error correlations. A similar difference in vertical structure has been noted by LH in their examination of ECMWF observed residuals. The geostrophic assumption implies that outside of the tropics, $V_{\psi kl}$ and V_{zkl} should be specified to be equal (Rutherford 1976; Lorenc 1981). On the other hand, since about 1980

the CMC analysis has been using covariances for all analysis variables that have been empirically and independently determined, using a method based on empirical orthogonal functions for the vertical structure (Hollett 1975). In line with this approach, we choose to use the $V_{\psi kl}$ and V_{zkl} determined above when analyzing wind and height respectively, in order to faithfully represent the differences in vertical correlation width between wind and height.

The prediction errors $E_v{}^P$, for wind, and $E_z{}^P$, for height, must also be specified for each level. In general, the values adopted are those presented in section 3, which were produced by the error separation procedure. Adjustments were made when the partition of total error at a given level seemed inconsistent with those at neighboring levels; e.g., partition of wind error at 150 mb (see Fig. 5). Vertical temperature covariances are determined as described by Rutherford (1976); i.e., by differencing the rows and columns of the height covariance matrix using a cubic spline technique.

b. Observation error covariances

As described by Rutherford (1976), it is assumed in the CMC analysis program "that all observation errors are random and uncorrelated." As noted in section 3 (Figs. 6, 9 and 11), this assumption is satisfied by radiosonde wind and thickness (or temperature) errors, but not by height errors. Observation errors for radiosonde data were initially specified to be those obtained from the error separation procedure. However, earlier observation error estimates from a different dataset produced smaller wind observation errors and preliminary tests with these produced wind analyses of better quality. To maintain these benefits we artificially reduced the wind observation error estimates of Fig. 5 by 18%. No such modifications were made to the height observation errors. We note that Lorenc et al. (1988) and Shaw et al. (1987) have also found that reducing the observation errors (i.e., giving more weight to the data) is beneficial.

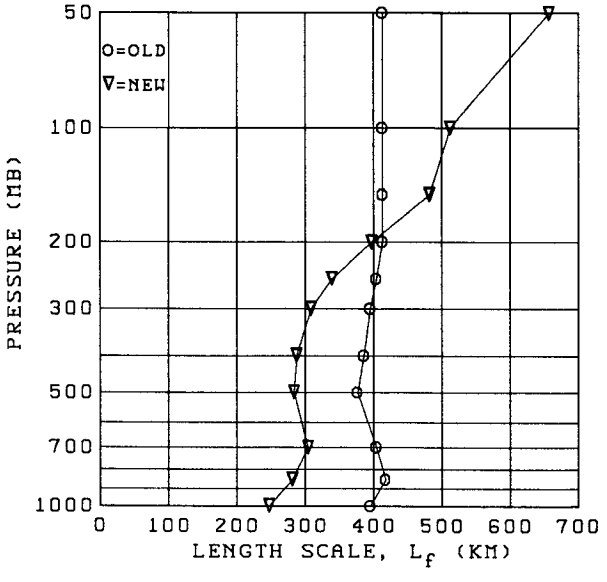

FIG. 18. Vertical profiles of length scale, L_f (km), for both the previously used negative squared exponential function (denoted OLD) and the newly derived correlation function (denoted NEW).

TABLE 1. Standard deviation of observation error for surface, aircraft, and satellite wind data. Wind errors are for wind components.

Surface (land, ship, buoy)	
Wind	3.6 m s^{-1}
Height	14 m*
Temperature	1.5°C
AIREP	
Wind (below 500 mb)	3.8 m s^{-1}
Wind (above 500 mb)	5.1 m s^{-1}
SATWIND	
Wind (below 500 mb)	3.8 m s^{-1}
Wind (above 500 mb)	7.7 m s^{-1}

* Later changed to 8 m as discussed in section 5.

Observation errors for satellite-derived temperature profiles (SATEMs) are specified separately and have not been changed. Surface, aircraft reports (AIREPs) and satellite wind reports (SATWINDs) have been assigned the observation errors given in Table 1. These are based on the values given by Shaw et al. (1987) and Lönnberg (1988).

c. Latitudinal variation

In order to account for the observed latitudinal variation in interpolation statistics, the complete error separation procedure has been carried out using North American radiosonde data for the three overlapping latitude bands: 25°–45°N, 30°–60°N and 45°–90°N. The results were used to specify the interpolation statistics for the three latitude bands: 0°–30°, 30°–60° and 60°–90°. Relatively large bands were chosen to ensure that each contained enough radiosonde data for statistically meaningful results to be obtained. In order that the statistics used in the o.a. program vary smoothly with latitude, the values at any given analysis point are obtained by interpolating the statistics linearly to the given latitude.

d. Humidity

We briefly describe the specification of the interpolation statistics for humidity. As discussed by Rutherford (1976), the humidity variable in the CMC o.a. is dewpoint depression, $T - T_d$. Horizontal correlations of observed residuals of this variable were found to be quite narrow, consistent with the small-scale variability of humidity. To represent the horizontal prediction error correlation, we chose to use (2a), the degenerate SOAR function, $(1 + cr)\exp(-cr)$. Using this function, the aforementioned error separation procedure

was carried out in order to determine the interpolation statistics to be used in the o.a. procedure.

5. Impact of the new statistics on the assimilation cycle

An 11-level version of the CMC o.a., running with the new interpolation statistics, was tested in a three-week data assimilation cycle. The cycle was driven by the operational model, which by late 1988 was a T89 version of the CMC spectral model. To serve as a basis of comparison, a control assimilation cycle was run using the operational o.a. Although both o.a. programs were producing analyses at 11 levels, the 50 mb analysis, vertically decoupled in the operational o.a., was vertically coupled to the levels below by the new interpolation statistics. Attention will be focused on the third week of the assimilation cycle, after the system had completely adjusted to this major change in interpolation statistics.

One method we used to evaluate the data assimilation systems was to calculate observed minus analysis ($O - A$) and observed minus 6-h forecast ($O - F$) rms statistics. These have been used by Hollingsworth et al. (1986) in their evaluation of the ECMWF data assimilation system during a 3-month period in 1984. The magnitude of the $O - A$ statistics indicates how well the analyses fit the observations, while the magnitude of the $O - F$ statistics is a measure of the accuracy of the trial field. Ideally we would like to reduce both of these statistics; however, a closer fit to the data does not necessarily lead to better 6-h forecasts. This will only occur if the analysis step has correctly filtered out the observational error (i.e., instrument error plus unresolved scales). One could choose to try to fit the observations exactly; however, in principle this would lead to worse 6-h forecasts; i.e., larger $O - F$ statistics.

Figure 19 shows $O - A$ and $O - F$ rms differences for the one-week period, 1200 UTC 03 December 1987

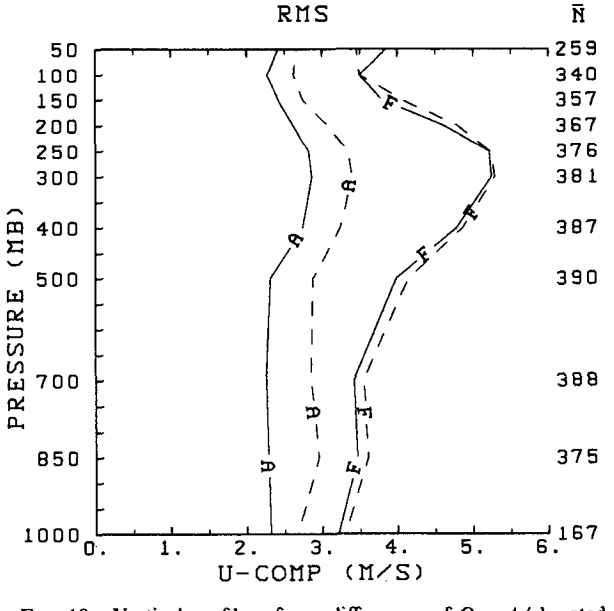

FIG. 19a. Vertical profiles of rms differences of $O - A$ (denoted A) and $O - F$ (denoted F) of the zonal wind component u (meters per second) averaged over all radiosonde observations in the Northern Hemisphere north of 30°N for the period 1200 UTC 03 December 1987–0000 UTC 10 December 1987. Dashed curves are from the assimilation cycle with the previously used interpolation statistics and solid curves are from the cycle with the new interpolation statistics. For each level, $\bar{N}$, the average number of radiosonde observations over this region per main synoptic hour is indicated at the right.

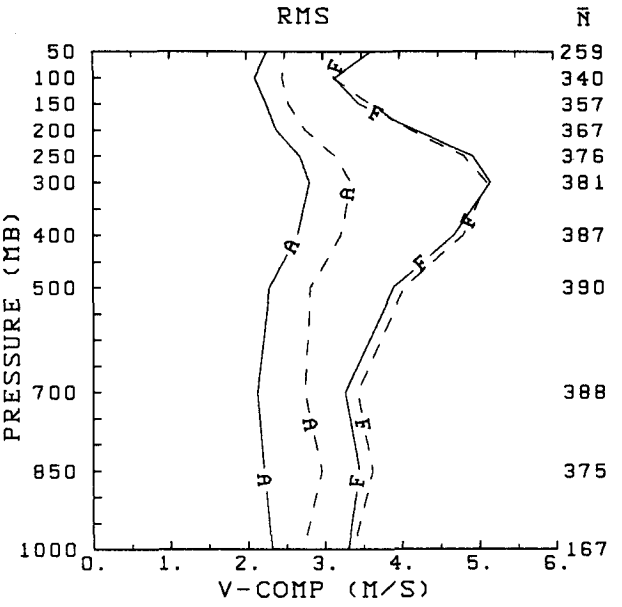

FIG. 19b. As in Fig. 19a but for the meridional wind component v (meters per second).

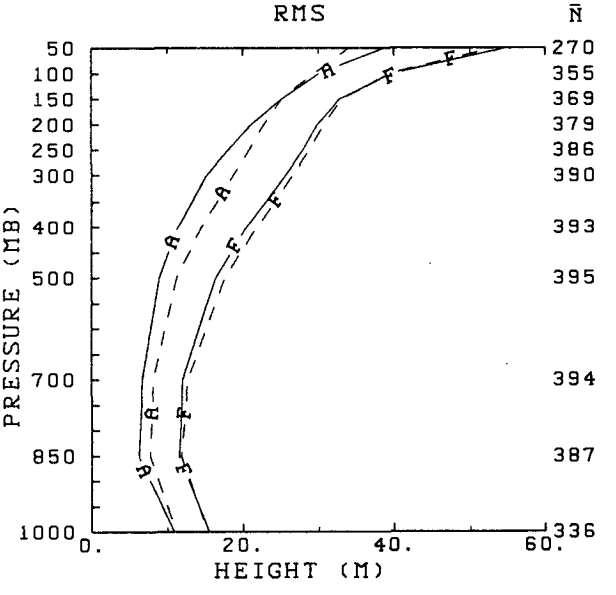

FIG. 19c. As in Fig. 19a but for z (m).

to 0000 UTC 10 December 1987. These statistics, averaged over all radiosonde observations in the Northern Hemisphere extratropics, are presented for the three variables u, v and z. Use of the modified o.a. results in a substantially closer fit of the analyses to the observations. This is especially so in the case of the wind components, for which the $O - A$ statistics drop from values generally near 3.0 m s^{-1} to values generally below 2.5 m s^{-1}. There is also a reduction in the $O - F$ statistics. Thus the changes to the analyses brought about by the new statistics have indeed led to improvements in the 6-h forecast. Although the improvements in the $O - F$ statistics are relatively smaller than those of the $O - A$ statistics for all variables, in an absolute sense the reduction in $O - A$ and $O - F$ variances is about the same. An examination of the observed minus initialized analysis statistics (not shown) also revealed a comparable reduction due to the use of the new interpolation statistics.

Consistent with these results, an examination of the synoptic charts indicated that the impact of the new interpolation statistics was greatest on the wind analyses: observations were drawn in better, jets tended to be stronger and vorticity fields more detailed with more intense centers. To illustrate this, Fig. 20 shows the 300 mb height and isotach analyses for 1200 UTC 4 December 1987 over western Canada. It can be seen that the assimilation cycle that uses the new interpolation statistics produces a stronger jet maximum over central Alberta. Observed wind speeds are generally better respected, an effect particularly noticeable in the case of the 62 m s^{-1} wind reported at Prince George

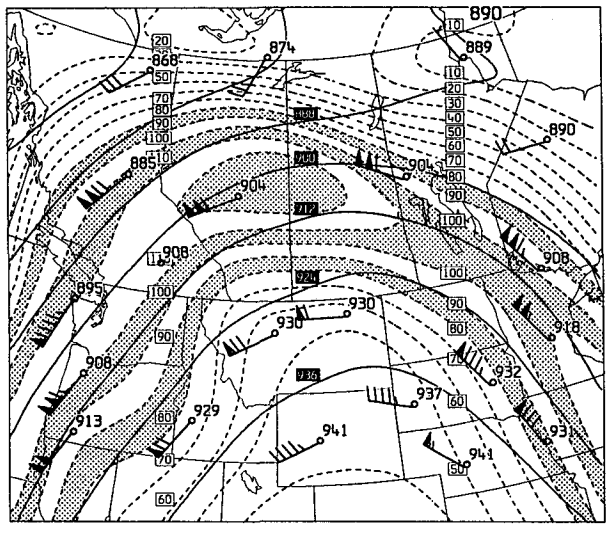

FIG. 20a. Analyses of height (solid contours) and wind speed (dashed contours) at 300 mb for 1200 UTC 04 December 1987 from the control assimilation cycle. Conventional units are used; i.e., the heights are given in dam (contour interval 12 dam), the winds are in knots (contour interval 10 kt; 1 kt $\approx$ 0.514 m s^{-1}). The winds are plotted as follows: a half barb is 5 kt, a full barb is 10 kt, and a solid flag is 50 kt.

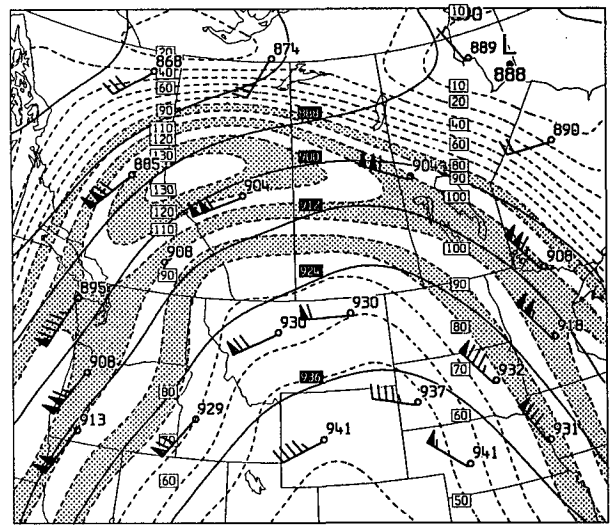

FIG. 20b. As in Fig. 20a but from the assimilation cycle that uses the new interpolation statistics.

in central British Columbia and the 57 m s^{-1} wind at Edmonton in central Alberta. Many of the differences between these analyses are evident in the 6-h forecasts, which served as trial fields (not shown). This is true in particular in the vicinity of the jet maximum over central Alberta; e.g., the predicted wind speeds at Prince George, British Columbia, and Edmonton, Alberta were 50 and 65 m s^{-1}, respectively, in the control cycle and 59 and 60 m s^{-1}, respectively, in the cycle that used the new interpolation statistics.

To illustrate the impact on the vorticity we present in Fig. 21 the 500 mb height and vorticity analyses for

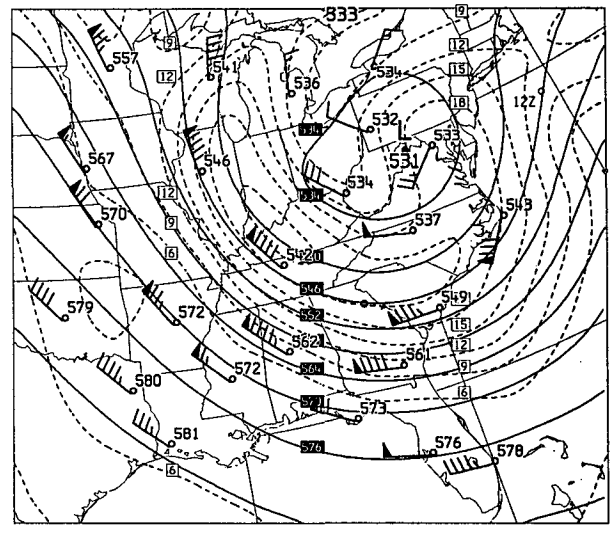

FIG. 21a. Height (dam) and absolute vorticity (10^{-5} s^{-1}) analysis at 500 mb for 1200 UTC 04 December 1987 from the control assimilation cycle. Contour interval for height is 6, for vorticity is 3. Plotting convention is as in Fig. 20a.

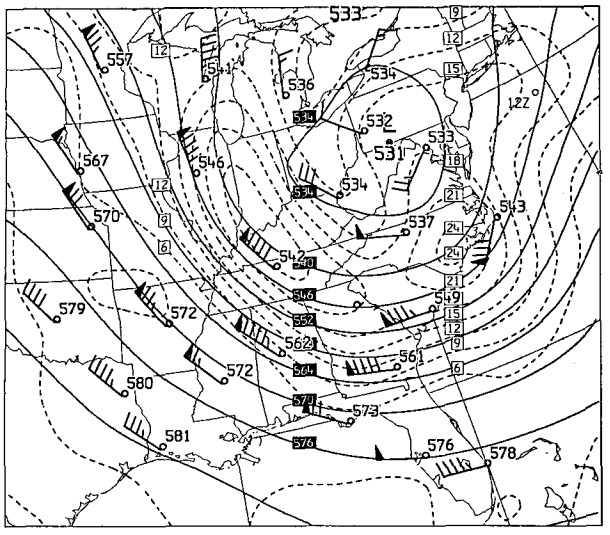

FIG. 21b. As in Fig. 21a but from the assimilation cycle that uses the new interpolation statistics.

this same time over the eastern United States. A series of short-wave troughs is moving around the 531 dam low located south of Lake Erie. The analysis produced by the cycle using the new interpolation statistics (Fig. 21b) clearly identifies two vorticity centers, one south of Lake Michigan and the other over North Carolina. The associated short-wave troughs can also be inferred from the shape of the height contours. By contrast, in the analysis produced by the control cycle (Fig. 21a) only the short-wave trough south of Lake Michigan is clearly evident.

In view of these results a parallel data assimilation cycle was implemented in December 1988 at the CMC in order to examine the suitability of the modified o.a. for operational implementation. For this parallel cycle and the other experiments discussed below, the observation error assigned to surface (1000 mb height) observations was reduced to 8 m as indicated in Table 1. This was done to give more weight to these observations, particularly in data sparse areas where a closer fit to the available surface data was desirable. Once per day a parallel early analysis was run from the parallel cycle and a parallel 48-h forecast was produced with the Canadian regional finite-element model to assess the impact of the modified o.a. on the model forecasts. Various aspects of this forecast model have been recently described by Staniforth and Benoit (1985), Benoit et al. (1989) and Tanguay et al. (1989).

In Table 2 we present objective verification scores for the 34 forecasts that were done. These scores were computed over the CMC verification grid, which covers Canada, most of the United States, and adjacent waters. Both sets of forecasts were verified against the operational analyses. The modest but consistent reductions in gradient scores (S1), standard deviations, and biases indicate that use of the new interpolation statistics results in better model forecasts. Thus in addition to improving the 6-h forecasts, we have also improved the 24 and 48-h forecasts of an independent model.

We have also developed a scheme, similar to that described by Lönnberg and Shaw (1987, p. 2.21), to change the variances and vertical correlation matrices as a function of season. For any given time of year, appropriate values of these quantities are calculated by interpolation using the values derived above for winter and a corresponding set of values for summer.

To complete the evaluation of the new statistics another three-week assimilation cycle was performed, this time for a period in May 1988. As before a control cycle was performed using the old interpolation statistics. Examination of the third week of the cycles, 1200 UTC 16 May to 0000 UTC 23 May 1988, revealed the benefits of the new statistics as in the December case. In fact, the improvements in the rms $O - F$ statistics for the wind components were even larger than before, as illustrated in Fig. 22.

Based on results such as these and on a separate subjective evaluation by CMC meteorologists, it was decided to proceed with operational implementation. This took place at 0000 UTC 26 January 1989 using the interpolation statistics for winter. The scheme to allow the interpolation statistics to vary with the time of year was implemented operationally in July 1989.

6. Summary and concluding discussion

The method used to obtain interpolation statistics for the data assimilation procedure in operation at the Canadian Meteorological Centre has been presented. New interpolation statistics have been derived using the method and the impact of using them in the analysis step of the data assimilation procedure has been evaluated by performing extended assimilation cycles. A comparison with the scheme previously in operation

TABLE 2. S1 scores, standard deviation (std. dev.) of the error, and biases over the CMC verification grid for 24- and 48-h forecasts for 34 winter cases. Verification scores of forecasts produced from the control analysis cycle are denoted "Old"; those from the cycle using the new interpolation statistics are denoted "New." Mean-sea-level is denoted MSL.

	24 h		48 h	
	Old	New	Old	New
		MSL		
S1	33.9	33.0	46.5	45.5
Std. dev. (mb)	3.2	3.2	5.1	5.0
Bias (mb)	−0.6	−0.5	−1.8	−1.8
		500 mb		
S1	17.1	16.3	27.2	26.3
Std. dev. (dam)	2.4	2.3	4.2	4.1
Bias (dam)	−2.1	−1.9	−3.1	−3.0

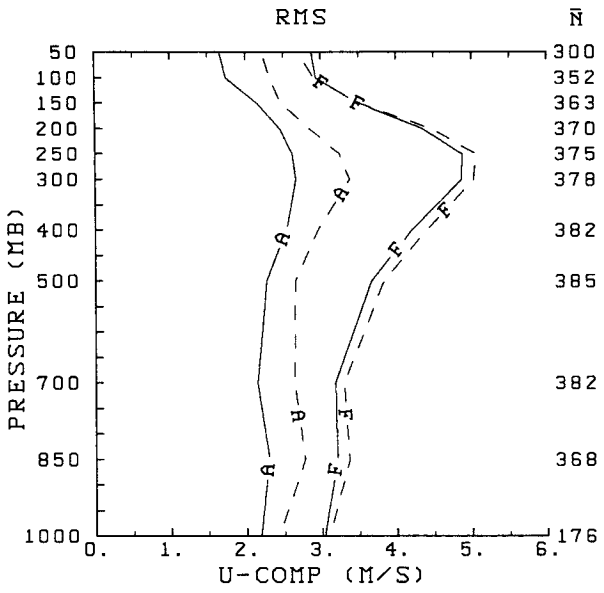

FIG. 22a. As in Fig. 19a but for the period 1200 UTC 16 May 1988–0000 UTC 23 May 1988.

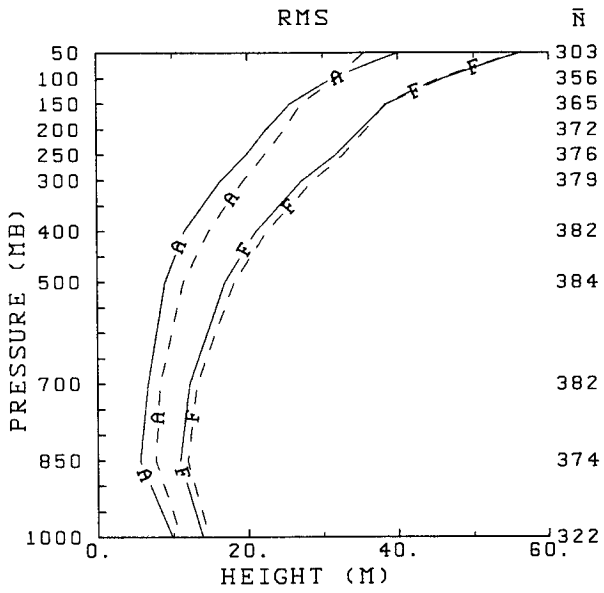

FIG. 22c. As in Fig. 19c but for the period 1200 UTC 16 May 1988–0000 UTC 23 May 1988.

indicated that use of the new interpolation statistics results in a significant improvement in the fit of the analyses to the observations. This is especially true in the case of the winds, for example, rms values of $O - A$ dropped from values generally near 3.0 m s^{-1} to values generally below 2.5 m s^{-1} during a one-week period in December. An examination of synoptic charts revealed stronger jets and more intense vorticity centers. There was also better agreement between the 6-h forecasts, used as trial fields, and the radiosonde observations. An evaluation of 24 and 48-h forecasts produced from the new analyses by an independent model revealed measurable improvements in verification scores.

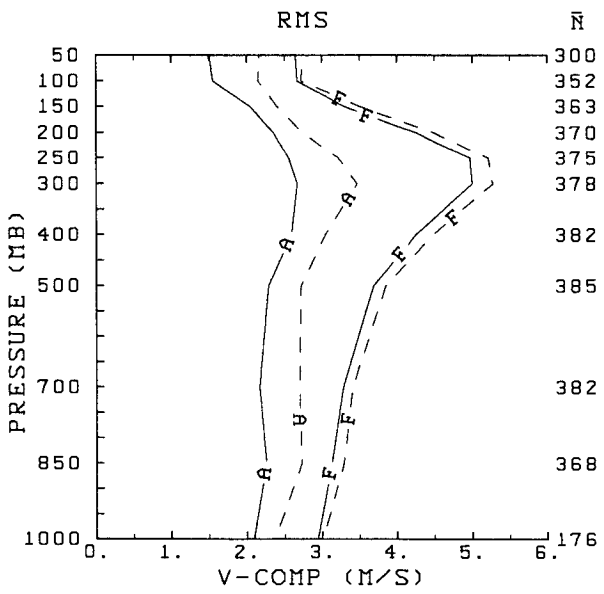

FIG. 22b. As in Fig. 19b but for the period 1200 UTC 16 May 1988–0000 UTC 23 May 1988.

As in earlier work, our approach in deriving the interpolation statistics has been based on a functional representation of the horizontal structure of correlations of observed residuals as determined from radiosonde data. In a previous study (TMS) the sum of a degenerate second-order autoregressive (SOAR) function $(1 + cr)\exp(-cr)$ and an additive constant, to account for longwave error, was adopted as the "function of choice" for height residuals. In extending that study to include the horizontal structure of wind correlations, we have found it desirable to modify this choice in two ways. The degenerate SOAR function has been replaced by a degenerate third-order autoregressive (TOAR) function $(1 + cr + c^2r^2/3)\exp(-cr)$ and the additive constant has been replaced by a second, broader TOAR function. To determine the vertical (interlevel) prediction and observation error covariance matrices, we adopted the procedure of HL and LH. This involves performing fits to wind shear error correlations and thickness error correlations.

As a prerequisite to specifying new interpolation statistics, the procedure has been used to study CMC observed residuals of wind and height at North American radiosonde stations during the winter of 1987–88. The results relating to the magnitude of the observed residuals, separation of error and vertical and horizontal correlation structure have been presented in section 3. These have been compared with the results of HL and LH, keeping in mind that there are differences between the CMC and ECMWF data assimila-

tion systems, especially the assimilating models and analysis procedures, and the hemispheric versus global domains. In general our results are similar to theirs except that the CMC observed residuals of both height and wind are larger at levels above the tropospheric jet. This may be due, in part, to the fact that, during the period studied here, the CMC analysis scheme had only a single (vertically decoupled) level above 100 mb, whereas the ECMWF analysis scheme studied by HL and LH had 5 levels above 100 mb. Possibly relevant here was the finding that observed wind residuals varied substantially with latitude above 250 mb in both winter and summer. In winter observed residuals at 150 mb were twice as large from 25°–45°N as from 45°–90°N. These large observed wind residuals point to possible problems in the tropical stratosphere of the CMC hemispheric data assimilation scheme.

A further difference between our results and those of HL and LH is that our error separation procedure yields prediction errors that are generally larger than the observation errors. The prediction and observation errors obtained by HL/LH were about equal, indicating that their fitting procedure, which uses 10 terms of a Fourier–Bessel expansion, yields lower zero-intercepts. In fact, as shown by HL (Fig. 1), retention of more terms in the Fourier–Bessel expansion yields higher zero-intercepts that would agree more closely with ours. We also found that the variation of horizontal scale of prediction error correlation with isobaric level and latitude was very much as described in TMS; i.e., an overall increase in horizontal scale upward, and, in the troposphere, equatorward.

The study of CMC observed residuals of wind and height forms the basis for specifying the new interpolation statistics. By stratifying the North American radiosonde data into three overlapping latitude bands and deriving interpolation statistics for each, we have tried to represent the observed latitudinal variation in the structure of the observed residuals. The results of Hollingsworth et al. (1986) and TMS show that there are also large regional differences in the structure of observed residuals at the same latitude; however, the scheme we have implemented does not account for these differences. As discussed in section 4, we have chosen to make several modifications to the raw interpolation statistics obtained from the fitting procedure, rather than use them directly. In general, the modifications are necessitated by inhomogeneities in our set of observed residuals. For example, the fact that a significant number of radiosonde reports do not include 1000 mb values and/or terminate below 50 mb, leads to difficulties in calculating accurate vertical correlations involving these levels.

For this reason, we decided not to proceed with the introduction of an automatic interpolation statistics updating procedure as envisioned in TMS. Instead we have implemented a scheme that produces appropriate statistics for any given time of the year by interpolation from a fixed set of winter and summer statistics. The winter and summer statistics can be recomputed every few years, or more frequently when major changes are made to some component of the data assimilation procedure.

This study and those of HL and LH, Hollingsworth et al. (1986), TMS, and Williamson et al. (1981) indicate important deficiencies in many of the simplifying assumptions (namely, horizontal/vertical separability, homogeneity, geostrophy) normally made in applying statistical interpolation operationally for NWP. As pointed out by Rutherford (1976), these assumptions are not basic to Gandin's optimum interpolation method but have been adopted to facilitate operational implementation. Daley (1985) has developed a more general formalism to relax the geostrophic constraint and account for synoptic-scale divergence. Work aimed at easing the other constraints could also lead to improvements in operational data assimilation.

Acknowledgments. We would like to thank André Robert and Donald Shantz for helpful suggestions regarding the choice of the correlation model, Roger Daley for reviewing the manuscript, Claudette Thibeault for her invaluable help with the statistics programs and the graphics, and Diane Lespérance for expertly typing the manuscript.

APPENDIX

Correlations for Multivariate Analysis

Using Helmholtz's theorem

$$u = -\frac{\partial \psi}{\partial y}; \quad v = \frac{\partial \psi}{\partial x},$$

ignoring horizontal derivatives of E_ψ, and using the notation of Daley (1985), the prediction error covariances are given by

$$\left.\begin{aligned}
\langle u_i u_j \rangle &= E_\psi{}^2 \Gamma[F(r)] \\
\langle v_i v_j \rangle &= E_\psi{}^2 \Delta[F(r)] \\
\langle v_i u_j \rangle = \langle u_i v_j \rangle &= E_\psi{}^2 \Theta[F(r)] \\
\langle z_i u_j \rangle = -\langle u_i z_j \rangle &= E_z E_\psi \mu \Xi[F(r)] \\
\langle z_i v_j \rangle = -\langle v_i z_j \rangle &= -E_z E_\psi \mu \Pi[F(r)] \\
\langle z_i z_j \rangle &= E_z{}^2 F(r)
\end{aligned}\right\} \quad \text{(A1)}$$

where

$$\left.\begin{aligned}
\Gamma &= -[R + (y_i - y_j)^2 R^2] \\
\Delta &= -[R + (x_i - x_j)^2 R^2] \\
\Theta &= [(y_i - y_j)(x_i - x_j) R^2] \\
\Xi &= [(y_i - y_j) R] \\
\Pi &= [(x_i - x_j) R]
\end{aligned}\right\}$$

and R has been defined in (1).

Defining $F(r)$ as in (7a) and substituting into (A1) yields expressions for each of the prediction error covariances in terms of the parameters E_ψ, E_z, c, N, α and μ.

Expressions for the correlations can then be derived using

$$\text{corr}(a, b) = \frac{\langle a, b \rangle}{[\langle a, a \rangle \langle b, b \rangle]^{1/2}}. \tag{A2}$$

This yields:

$$\begin{aligned}
\text{corr}(u_i, u_j) &= F'(r) - (y_i - y_j)^2 c^2 \frac{\left(1 + \frac{\alpha}{N^4}\right)}{\left(1 + \frac{\alpha}{N^2}\right)} F''(r) \\
\text{corr}(v_i, v_j) &= F'(r) - (x_i - x_j)^2 c^2 \frac{\left(1 + \frac{\alpha}{N^4}\right)}{\left(1 + \frac{\alpha}{N^2}\right)} F''(r) \\
\text{corr}(u_i, v_j) = \text{corr}(v_i, u_j) &= c^2 (y_i - y_j)(x_i - x_j) \frac{\left(1 + \frac{\alpha}{N^4}\right)}{\left(1 + \frac{\alpha}{N^2}\right)} F''(r) \\
\text{corr}(z_i, u_j) = -\text{corr}(u_i, z_j) &= -\mu \left[\frac{\left(1 + \frac{\alpha}{N^2}\right)}{3(1 + \alpha)}\right]^{1/2} c(y_i - y_j) F'(r) \\
\text{corr}(z_i, v_j) = -\text{corr}(v_i, z_j) &= \mu \left[\frac{\left(1 + \frac{\alpha}{N^2}\right)}{3(1 + \alpha)}\right]^{1/2} c(x_i - x_j) F'(r) \\
\text{corr}(z_i, z_j) &= F(r)
\end{aligned} \tag{A3}$$

where $F(r)$ and $F'(r)$ have been defined in (7) and

$$F''(r) = \left(1 + \frac{\alpha}{N^4}\right)^{-1} \times [\exp(-cr) + (\alpha/N^4) \exp(-cr/N)].$$

With

$$E_v^2 = \langle u_i u_i \rangle = \langle v_i v_i \rangle$$

$$E_z^2 = \langle z_i z_i \rangle$$

the set of equations (A3) can be used to express the covariances $\langle u_i u_j \rangle$, $\langle v_i v_j \rangle$, $\langle u_i v_j \rangle$, $\langle v_i u_j \rangle$, $\langle z_i u_j \rangle$, $\langle u_i z_j \rangle$, $\langle v_i z_j \rangle$, $\langle z_i v_j \rangle$ and $\langle z_i z_j \rangle$ in terms of the parameters E_v, E_z, c, N, α and μ.

REFERENCES

Balgovind, R., A. Dalcher, M. Ghil and E. Kalnay, 1983: A stochastic-dynamic model for the spatial structure of forecast error statistics. *Mon. Wea. Rev.*, **111**, 701–722.

Benoit, R., J. Côté and J. Mailhot, 1989: Inclusion of a TKE boundary layer parameterization in the Canadian regional finite-element model. *Mon. Wea. Rev.*, **117**, 1726–1750.

Buell, C., 1972: Correlation functions for wind and geopotential on isobaric surfaces. *J. Appl. Meteor.*, **11**, 51–59.

Daley, R., 1983: Spectral characteristics of the ECMWF objective analysis system. ECMWF Tech. Rep. No. 40.

——, 1985: The analysis of synoptic scale divergence by a statistical interpolation procedure. *Mon. Wea. Rev.*, **113**, 1066–1079.

——, C. Girard, J. Henderson and I. Simmonds, 1976: Short-term forecasting with a multi-level spectral primitive equations model. Parts I, II, *Atmosphere*, **14**, 98–134.

Hollett, S. R., 1975: Three dimensional spatial correlations of PE forecast errors. M.S. thesis, Dept. Meteor., McGill University, 73 pp.

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*, **38**, 111–136.

——, D. B. Shaw, P. Lönnberg, L. Illari, K. Arpe and A. J. Simmons, 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, **114**, 861–879.

Lönnberg, P., 1988: High resolution analysis experimentation at ECMWF. Preprints, *Eighth Conf. on NWP*, Baltimore, Amer. Meteor. Soc., 165–171.

——, 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*, **38**, 137–161.

——, and D. Shaw, Eds., 1987: *ECMWF Data Assimilation Scientific Documentation*. 96 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.]

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109,** 701–721.

——, R. S. Bell, T. Davies and G. J. Shutts, 1988: Numerical forecast studies of the October 1987 storm over southern England. *Meteor. Mag.*, **117,** 118–130.

Mitchell, H., C. Charette, M. Valin, B. Brasnett and C. Thibeault, 1988: Improving the objective analysis, CMC Tech. Document No. 32, 39 pp. [Available from CMC, 2121, Route Transcanadienne, Dorval, Québec, Canada, H9P 1J3.]

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

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29,** 809–815.

——, 1976: An operational three-dimensional multivariate statistical objective analysis scheme. *Proc. JOC Study Group Conf. on Four-Dimensional Data Assimilation,* GARP Rep. No. 11, WMO, Geneva, 98–121. [Available as No. 1 of the series of Notes Scientifiques et Techniques de DRPN, Dorval, Québec, Canada, H9P 1J3.]

Shaw, D. B., P. Lönnberg, A. Hollingsworth and P. Undén, 1987: Data assimilation: The 1984/85 revisions of the ECMWF mass and wind analysis. *Quart. J. Roy. Meteor. Soc.*, **113,** 533–566.

Staniforth, A., and R. Benoit, 1985: Numerical weather forecasting research in the Canadian weather service. *Telematics and Informatics,* **2,** 279–287.

Tanguay, M., A. Simard and A. Staniforth, 1989: A three-dimensional semi-Lagrangian scheme for the Canadian regional finite-element forecast model. *Mon. Wea. Rev.*, **117,** 1861–1871.

Thiébaux, H. J., H. L. Mitchell and D. W. Shantz, 1986: Horizontal structure of hemispheric forecast error correlations for geopotential and temperature. *Mon. Wea. Rev.*, **114,** 1048–1066.

Williamson, D. L., R. Daley and T. W. Schlatter, 1981: The balance between mass and wind fields resulting from multivariate optimal interpolation. *Mon. Wea. Rev.*, **109,** 2357–2376.