Spatial Correlation of the 24-Hour ECMWF Forecast Error

JUHANI RINNE
Department of Meteorology, University of Helsinki, Helsinki, Finland

SIMO JÄRVENOJA
The HIRLAM Project, The Danish Meteorological Institute, Copenhagen, Denmark

(Manuscript received 4 August 1986, in final form 14 August 1987)

ABSTRACT

The spatial autocorrelation of the error of the ECMWF 24-h forecast of 500 mb geopotential height during 1980–83, hereafter referred to as the "forecast error," is studied. The leading EOF of the forecast error describes a "teleconnection" between the Himalayas and the Pacific/European sector. The third EOF resembles the analysis error. Approximations for the isotropic part of the autocorrelation are presented. They consist of two source terms having different spatial scales. A "random-like" term could be due to model errors in baroclinic processes or to random analysis errors. A "scale-dependent" term could be due to model errors in barotropic processes and over mountain areas, or to analysis errors over data-sparse areas or mountains. The terms do not contribute uniformly. The "scale-dependent" term is strongest over mountains. The relative contribution of that term decreased from forecasts of 1980–81 to those of 1982–83.

1. Introduction

Usually the spatial autocorrelation values of forecast errors are computed as correlations between errors at observation points. When the values are displayed as a function of distance, they form a point swarm. Approximations of the spatial autocorrelation functions are derived to fit that swarm, in what could be called the direct method because of the direct fitting to the autocorrelation. There are many ways to choose the function. For instance, Lönneberg and Hollingsworth (1986), starting from kinematic considerations, applied Bessel series.

The autoregressive method introduced by Thiebaux (1985) clearly differs from the previous classical approach. She first determines the stochastic autoregressive process that describes the generation of the forecast error. The autocorrelation function used in the fitting is the one corresponding to the autoregressive process. Thus, one first has to understand (guess) the process before determining the function.

Balgovind et al. (1983) determined the structure function of the forecast error from a forecasting equation of the error utilizing a simplified model. The model was based on the conservation of potential vorticity, and they had to assume that the spectrum of the accumulated forcing is that of white noise.

Thus there have been three different approaches in deriving autocorrelation functions: (i) the form of the function is fit directly, (ii) the form of the function is based on an autoregressive process, or (iii) it is based on a forecast equation. A fourth method was studied by Rinne and Järvenoja (1985); hereafter referred to as RJ, in which the form of the function is determined from the spectrum of the actual variable. The spectral functions are spherical harmonics, which are the "natural" functions of a random process on the sphere (cf. North et al., 1981). RJ applied this method to 500 mb geopotential height analyses.

It is rather easy to find a function to fit a point swarm. In the direct method the number of alternatives is therefore large. In other methods there are fewer degrees of freedom. For instance, in practice the model of an autoregressive process cannot be complicated. This automatically reduces the number of function candidates. On the other hand, there is no need to check the spectral properties of the function candidate as is done in the case of the direct method (Julian and Thiebaux, 1975; Thiebaux, 1975).

An easy-to-handle function yielded by the spectral approach may be so complicated in the autoregressive approach that it will not be chosen as a function candidate. Different approaches concentrate on different limited groups of functions. A new approach can discover new features.

There are good reasons to believe that the spectral approach is reasonable. It arises in a natural way in a study of autocorrelation. In RJ, with the aid of the spectral approach, we found a function that explained
the isotropic part of the autocorrelation of the 500 mb height. It seemed to perform as well as, or better than, the functions presented earlier in the literature. Moreover, there was only one parameter which varied geographically, thereby allowing information about the variation of the autocorrelation to be compressed.

In this paper the aim is to extend the spectral approach to the case of forecast errors. We are especially interested in the possibility of describing the geographical variation. This implies that the data must be given in a dense network, i.e., on a grid. Therefore, as an exception, we take for the forecast error values of the forecast minus analysis instead of forecast minus observation.

2. Data

The data used in the present study consist of ECMWF (European Centre for Medium Range Weather Forecasts) 500 mb height analyses at 1200 UTC, and 24-h forecasts for 1980–1983. The data have been collected at the Finnish Meteorological Institute via the WMO/CAS NWP (Numerical Weather Prediction) Data Study and Intercomparison Project. The analyses and forecasts have been archived on a $5^\circ \times 5^\circ$ latitude–longitude grid north of $20^\circ$N. For the computations, the data were converted onto a polar-stereographic grid (grid interval 381 km at $60^\circ$N) of 1404 points.

In deriving the form of the function, we only utilized the data for 1980–81. From these, the 24-h forecast error $e$ was determined for each day from

$$e(i, \tau) = zf(i, \tau) - za(i, \tau),$$

where $zf$ is the 24-h forecast valid at time $\tau$ and $za$ is the corresponding verification analysis. Index $i$ refers to a grid point.

Note that we use gridded analyses instead of observations from an irregular network. This makes the computations easy but, on the other hand, makes the definition of the forecast error nonconventional and dependent on the quality of the analyses.

Next, empirical orthogonal functions (EOF) were computed from the forecast error fields (for details, see Appendix). The mean of the fields, i.e., the systematic forecast error, was first removed. Two hundred EOFs were computed, which explain 90% of the total error variance. These space-dependent functions are normalized so that the mean square over the grid equals unity.

Data for 1982–83 are used only for comparison. All figures except Fig. 1b and Fig. 13 refer to the earlier period of 1980–81.

3. Method

a. Computation of the autocorrelation functions

The autocorrelations are computed between each pair of grid points. Thus there are 1404 autocorrelation fields, each of them having 1404 values. The autocorrelation fields are computed via EOFs, as detailed in the Appendix. The results are similar to those found by computing the autocorrelation values directly from forecast error fields. In the present case, the computer time required in both approaches is equal.

The autocorrelation fields are then represented with the aid of spherical harmonics (see Appendix) in which the pole is transferred to the base point, i.e., to the point where the autocorrelation value equals 1. This approach means that we will study the spectral representation of the autocorrelation fields rather than the fields themselves.

Only those terms which are independent of the direction of the great circle connecting the grid points are dealt with. Thus we will concentrate on the isotropic part of the autocorrelation fields, which is given by

$$r(i, j) = \sum_{n=0}^{27} C_n^i(1 + 2n)^{1/2} P_n(n) \cos \theta_{ij}$$

Here $P_n$ are the Legendre polynomials, and $\theta_{ij}$ is the angular distance along the great circle connecting points $i$ and $j$. The term $(1 + 2n)^{1/2}$ is due to the two-dimensional nature of the problem (cf. Appendix). For practical reasons, truncation is needed. In RJ, $n = 27$ was found to be large enough to make the residual negligible.

Formally, the spherical harmonics are computed over the entire globe. Therefore, outside the grid area we have assumed $r(i, j) = 0$. As a consequence, there will be a discontinuity at the grid boundary (a more detailed discussion is presented in RJ). The results thus become inaccurate when the base point is close to the boundary. For interior base points, due to the peaked or local character of the autocorrelation function, $r(i, j)$ is close to zero outside the grid anyway, and so the computations are more correct for interior points than for boundary points.

b. Approximations of the spectral presentation

The aim is to find an explicit form for $r(i, j)$. To do this, we follow RJ and try to find a general law for the coefficients in (2). More specifically, we try to approximate the coefficients $C_0(i), C_1(i), \ldots, C_n(i), \ldots$ as a function of $n$ and $i$. Here $n$ is the wavenumber (in fact, the ordinal number of a Legendre polynomial) and $i$ refers to the base point.

To approximate $C_n$, we choose

$$\hat{C}_n(i) = A(i)1.02[1 + 2a(i)]^{-1/2} \exp[-a(i)a'(i)/n!] + M(i)r''(i)$$

Here $A(i), a(i), M(i)$ and $r''(i)$ are parameters to be determined. They vary with $i$; i.e., they can vary geographically.

The form of (3) seems to be complicated. However, it is based on simple ideas. Much of the forecast error
is thought to be “randomlike” with a white spectrum. However, the $C_n$ coefficients are rather small when $n$ is large. Therefore we cannot use simply a pure white spectrum. This kind of speculation leads us to model a portion of the spectrum with $t^n$ to the last term in (3), where $t$ is expected to be rather close to 1 in order to have a nearly white spectrum. Thus, this term is expected to be nearly independent of the scale of the forecast error.

By definition, our forecast error is affected by the analyses because it is given as the difference between forecasts and the verification analyses. Thus, one can expect that the structure of the forecast error partly resembles that of the height field in the analyses. These were studied in RJ, from which the first right-hand term in (3) is directly taken. The scale (“wavenumber $n$”) of the component with the largest coefficient is given by $a(i)$ (not necessarily an integer).

The first and second right-hand terms in (3) will be denoted as “scale-dependent” and “randomlike” parts. The final approximation of the isotropic autocorrelation will be

$$\tilde{r}(i, j) = \sum_{n=0}^{27} C_n(i) (1 + 2n)^{1/2} P_n(\cos \theta),$$  \hspace{1cm} (4)

or, applying (3),

$$\tilde{r}(i, j) = \sum_{n=0}^{27} A(i) 1.02 [1 + 2a(i)]^{-1/2} \times \exp[-a(i)][a^n(i)/n!(1 + 2n)^{1/2} P_n(\cos \theta)]$$
$$+ \sum_{n=0}^{27} M(i)t^n(i)(1 + 2n)^{1/2} P_n(\cos \theta)$$  \hspace{1cm} (5)

Only three of the parameters $A, M, t$ and $a$ are independent, owing to the constraint that the autocorrelation must equal unity at the base point, i.e., $r(i, i) = 1$. If $M = 0$ and $i = j$, then, according to the normalization made in RJ, (5) gives $\tilde{r}(i, i) = A$. Thus the coefficient $A$ gives directly the contribution of the “scale-dependent” term in the autocorrelation function.

Values of $A, M, t$ and $a$ were determined to give the best fit between $C_n$ and $\tilde{C}_n$ so that $\tilde{r}(i, i) = 1$ in (5). This makes $A, M, t$ and $a$ slightly dependent on the truncation point, $n = 27$. In determining the values of $A, M, t$ and $a$, we used an iterative solution method which will not be presented here. The computations were rather tedious, and they did not converge at every grid point.

The right-hand side of (5) consists of two source terms. The second one is continuously decreasing with $n$ ($t < 1$), whereas the first one has a maximum at $n = a$. If in fact there are several maxima in $C_n$ as a function of $n$, then there are several candidates for $a$ and the fitting becomes difficult. This seems to be the reason for the nonconvergence of the computation at some grid points. The general success of the fitting, however, implies that overall the basic form (Eq. 3) is suitable.

4. Results

a. Autocorrelation fields

Figure 1 shows the autocorrelation functions for the forecast error (Fig. 1a, present data, 1980–81) and the height analysis (Fig. 1b, taken from RJ, data from 1946–70). The forecast error (Fig. 1a) shows a more peaked function, where the values rapidly decrease from 1. There is a sudden change in the rate of decrease at 1000 km, where a slow decrease towards zero begins. Thus the forecast errors at longer distances are practically independent of each other.

The autocorrelation of the 500 mb height in Fig. 1b is clearly different. It is more like a damped cosine oscillation with clear negative minima at longer distances. The smoothness of Fig. 1b, in comparison to Fig. 1a, is most probably due to the long data period (25 years) used in the former.

Figure 2 shows another typical case of the forecast error; here the base point is located at the North Pole. An exceptional case is given in Fig. 3, where the base point has been moved to the Himalayas. The field is flat, there are nonzero correlations at longer distances, and negative values are clearly present.

The negative values in Figs. 1a and 2 are local, i.e., they tend to belong to the anisotropic part of the autocorrelation field. Therefore they are not considered in the present study.

FIG. 1a. The autocorrelation field of the 24-h forecast error in 1980–81 for a base point close to Helsinki. Contour interval 0.1, negative values are indicated by broken lines.
b. EOFs of the forecast error

The first EOF of the forecast error shows teleconnectivity between the Himalayas and the Pacific/European sector (Fig. 4). If a large-scale cell (trough, ridge) at 140°W is in a wrong position in the forecasts, then it generates an error field like that in Fig. 5. The second EOF thus reflects phase errors in the forecasts. This EOF, like the first one, has no connection with the EOFs of the analysis error (Rinne and Järvenoja, 1984).

On the other hand, the maxima of the third EOF (Fig. 6) are located where the standard deviation of the analysis error is at its largest (cf. Figs. 1 and 2 in Rinne and Järvenoja, 1984). The EOFs thus seem to support the idea that a part of the forecast error is structurally like the analysis error.

c. The “scale-dependent” term

Values of \( A \) are shown in Fig. 7. They give the contribution of the error with a structure like that of the
analyses. A clear maximum is found over the Himalayas. The pattern in Fig. 7 resembles that in Fig. 4. The contribution of the "scale-dependent" term in (5) is thus large where the teleconnectivity in Fig. 4 is strong.

It was concluded in RJ that in the analyses $a \sim 7$ and that, as a consequence, the dominating scale is that of the seventh Legendre polynomial (the maximum of $\exp(-a)a^n/n!$ is found at $a = n$). In Fig. 8, three isolines for $a$ in the forecast error are given. The isolines of $a = 3$ delimit those areas where the scale is very large ($a < 3$). These areas are marginal boundary areas. At most grid points $3 \leq a \leq 6$. Thus the "scale-dependent" term is of a larger scale in the forecasts than in the height analyses.

FIG. 8. Values of the $a$ parameter as applied in the scale-dependent term. Contour interval 3. Within areas of $a > 9$ the convergence of the computations was weak, and the results are inaccurate there. At the boundaries the values are thought to be less reliable.
Isolines of $a = 9$ in Fig. 8 give additional information. They enclose rather well the areas where our fitting process did not converge satisfactorily. There the results are inaccurate.

d. The randomlike term

The contribution of the randomlike term to the correlation will not be shown because it is the complementary image of Fig. 7. The term dominates where the contribution of the scale-dependent error is small. In some areas the randomlike error is almost completely predominant. Over such areas, like the Atlantic and the northern USSR, our model [Eq. (5)] has essentially only one term and one dependent parameter.

The term describing the randomlike error is related to $t^n$. As previously noted, if $t = 1$ the spectrum is nearly white. This seems to be true. Values of $t$ are close to 1; they vary from 0.96 in the north to 0.88 at 40°N (Fig. 9). The values tend to decrease southward. Note that nothing can be said about the behavior at the boundary because of inaccuracies there.

Thus both terms indicate large-scale errors in the south. In the scale-dependent term the larger scale is indicated by a small wavenumber ($a$), in the random term by a smaller value of $t$ resulting in a somewhat red spectrum.

5. The representativeness of the model

The efficiency of the model expressed by (5) is interesting because it tells how successful the fit of the isotropic autocorrelation has been. One measure of efficiency is given by the coefficient variance explained. Our model is very efficient; over nearly the whole area the explained variance is close to 95% as measured in terms of the squared sum of the coefficients $(1 - \Sigma_{n=0}^\infty (C_n - \bar{C}_n)^2/\Sigma_{n=0}^\infty C_n^2)$. Exceptions are found, of course, at the boundaries and in areas of poor convergence in the fitting process. In addition, at the North Pole and over northern Greenland there is a small area where the efficiency is close to or less than 90%. It can thus be said that the present model describes the isotropic part of the autocorrelation field of the forecast error well almost everywhere.

Figure 10 illustrates in a single case (a grid point close to Helsinki, cf. Fig. 1a) the contribution of "scale-dependent" and randomlike terms to the isotropic part of the autocorrelation function. Here $A = 0.305$ and the former term equals 0.305 at zero distance. Thereafter it decreases very slowly. The randomlike term equals 0.695 at zero distance and then decreases very rapidly. Thus the random error generates the peaked form to which the scale-dependent term adds a long-distance tail. The sum of these terms thus has a complicated structure.

In Fig. 11 we show the true (anisotropic plus isotropic) autocorrelation function northeastward from the Helsinki grid point, the corresponding isotropic part of the autocorrelation function [Eq. (2)] and the present [Eq. (5)] approximation. The difference between the last two is quite small in comparison to the difference between the isotropic and observed curves. Note the kink in all the curves at 1000 km. This kink reflects the complicated form of the function.

It is seen that the theoretical curve approximates the isotropic curve well. In Fig. 12 a worse case is shown. The grid point, at the North Pole, is close to an area of nonconvergence in the computations.

6. An explicit formula for the autocorrelation function

As is shown, Eq. (5) gives a good representation of the isotropic field. Such a sum is impractical to use because coefficients and values of the Legendre polynomials must be determined 27 times at each point $i$. An explicit formula is required. The derivation of the formulae is not presented here because the algebra becomes lengthy and the result is only valid in a limited parameter range.

The main idea is to expand the right-hand terms of (5) into Taylor series and to combine coefficients of the same order to form one Taylor series. Then a function with a similar Taylor series is sought. Scale-dependent and randomlike parts are handled separately.

The task is to approximate $\sum_n (a^n/n!)(1 + 2n)^{1/2} \times P_n(\cos \theta)$ and $\sum_n t^n(1 + 2n)^{1/2} P_n(\cos \theta)$. The approximation of the former term is derived in RJ. The latter is approximated so that the final result is
\[
\hat{r}(i, j) \sim A \exp[a(\cos \theta_{ij} - 1)]J_0(a \sin \theta_{ij}) \\
+ B \left[ \exp(-49\theta_{ij}^2) - 45\theta_{ij}^2 \exp(-45\theta_{ij}^2) \right] \\
\times \left[ \Phi \left( \frac{\varphi - 0.5}{0.5} \right) - 0.5 \right],
\]
where \( \varphi \) is obtained from \( \sin \varphi = 10(t - 0.87) \). Here \( \theta_{ij} \) is the angular distance along the great circle connecting points \( i \) and \( j \). The \( J_0 \) and \( \Phi \) are the zero-order Bessel function and the cumulative normal distribution, respectively. Parameters \( A, a \) and \( t \) can be taken from Figs. 8–10 and \( B \) is computed from \( B = 1 - A \). The approximation of the random part is relatively simple except for the last multiplier, which describes the dependence on \( t \). This dependence was easily approximated by \( \Phi \) when a new variable, \( \varphi \), was introduced. We had no other reason to make use of the normal

---

**Fig. 10.** The isotropic autocorrelation function (continuous line) as the sum of two source terms. These are the scale-dependent term (broken line) and the randomlike term (dotted line). Values of \( A, a \) and \( t \) are 0.305, 4.8 and 0.96, respectively. These correspond to a point close to Helsinki (autocorrelation field in Fig. 1a).

**Fig. 11.** The autocorrelation function for the Helsinki grid point taken northeastward (toward 80°N, 60°E). The broken line shows the observed function. Its isotropic representation by a Legendre polynomial (Eq. 2) is given by the dotted line. The continuous line shows the corresponding isotropic approximation given by (5) in the text.
distribution. Note that one has to take $t \leq 0.97$. Excluding this limitation, (6) approximates (5) well at distances less than 3000 km. Parameters can vary with the base grid point ($i$), which means that $\hat{r}(i, j)$ may not be equal to $\hat{r}(j, i)$. If one uses constant parameters throughout the area, then this small inconsistency can be avoided.

It is obvious that for $A$ and $a$ no closed formula can be given. Their geographical variation seems to be too irregular. However, for $t$ we find (Fig. 9) that approximately $\sin \theta \approx 10(t - 0.87)$, where $\theta$ is the latitude of the base point. Thus $\theta$ in (6) can just be taken as that latitude.

At shorter distances ($\leq 1200$ km; cf. Fig. 10) a simplified version of (6) is

$$\hat{r}(i, j) \sim A + (1 - A) \exp(-4900t^2)$$  \hspace{1cm} (7)

In this formula the scale-dependent term is approximated by a constant because this term is usually smaller and changes slowly at shorter distances. Everything except $A$ is given explicitly. This parameter must be taken from a presentation such as that in Fig. 7.

As a summary of the different approximations, we present the following list:

1) The autocorrelation fields are computed via the EOFs. This procedure includes a truncation, the effect of which is negligible.

2) The autocorrelation fields are expanded in a series of spherical harmonics. Once again, a truncation is necessary. Further, the direction-dependent terms are removed, i.e., only the isotropic part remains.

3) The $\hat{C}_n$ coefficients of the spherical harmonics are estimated as the sum of two terms, where the kernel functions are $t^n$ and $a^n/n!$ [Eq. (3)]. The autocorrelation function can then be given as a sum of spherical harmonics [Eq. (4)]. The estimated $\hat{C}_n$ coefficients of the series are known when three independent parameters, i.e., $t$, $a$ and $A$, are given.

4) The series in (5) is approximated by a simplified form given in (6). A heavily truncated version is given in (7).

It has thus been shown that the spatial autocorrelation function can be generated as the sum of two source terms, the scales of which vary geographically.

7. Effect of a forecast system change

The effect of changes in the ECMWF system was studied by taking another set of forecasts from the years 1982–83. Our fitting procedure still failed to converge at some grid points around Greenland. Obviously there is a systematic feature in the $C_n$ coefficients that our fitting procedures fail to find. No clear changes are observed in $a$, thus there are no changes in the scale of the scale-dependent error. The same is true for $t$ and the randomlike error. The efficiency in the explanation of the isotropic field does not change essentially. The distribution of $A$ does not change; there are still maxima over the eastern and western Pacific, the Himalayas, Europe and the Rocky Mountains. However, there is a decrease in the actual values. Practically everywhere the values in the $A$ field of 1980–81 should be multiplied by 0.8 in order to obtain the field of 1982–83 (Fig. 13). Hence, in comparison to 1980–81, the relative contribution of the scale-dependent and randomlike errors in 1982–83 are smaller and larger, respectively. The only clear exception is the Pacific cell, which has not changed.
The largest absolute changes in $A$ are found over the Himalayas and Europe. When autocorrelation fields for 1980–81 and 1982–83 are compared directly, the largest differences are found over just those areas. Changes are observed over Greenland and the Aleutians, too. (These are areas of nonconvergence in Fig. 7 or 13 and the present method fails there.)

Thus it is apparent that the autocorrelation function of the forecast error can be sensitive to system changes. In the present case the change is mainly due to the decrease in the relative contribution of the scale-dependent error. The distribution patterns and values of the parameters remain much the same, however. The only exception is that the relative contribution of the two source terms has changed similarly over the whole of the Northern Hemisphere. This indicates that the separation into two terms is reasonable.

8. Discussion

The two source terms were introduced into (3) on a heuristic basis. The quality of the results shows that there obviously exist two such terms. In the following we try to characterize these terms and present tentative explanations for them. Some comparison with the results of other authors will also be presented.

The randomlike term was introduced to describe the random errors. This seems to be justified because $t \approx 1$, i.e., the corresponding spectrum is nearly white. It is not precisely white, because $t < 1$ and because our computations include a truncation at $n = 27$ in (4). If the spectrum were white, then the autocorrelation function would equal zero except at the origin. The fitted function is a smoothed version of the sharp peak.

The autocorrelation function falls rapidly with distance and the effective radius of the randomlike term is short. An error having such a short-scale autocorrelation function could be generated by baroclinic processes. On the other hand, the analysis system only uses a certain number of adjacent stations at a time and can thus produce randomlike small-scale errors.

The scale-dependent term is different. It has a spectral maximum at (meridional) wavenumber $a$. The effective radius varies with $a$ but is longer than in the previous term (for instance, Fig. 10). Such large-scale errors are produced in the analysis process over data-sparse areas through erroneous observations or through the first-guess field. As previously discussed, the third EOF (Fig. 6) shows that the large-scale forecast error can be similar to the analysis error. The contribution of the scale-dependent term is dominant over mountain areas (Fig. 7). Thus the term could be due to analysis errors over mountains. Large-scale errors can also be generated by the model especially at the scale of the most unstable barotropic waves. Thus this scale-dependent term can be due to analysis errors in data-sparse areas and over mountains, and/or model errors over mountains and in barotropically unstable waves. Lönnberg and Hollingsworth (1986) have found that, in the vertical direction, the forecast error shows both a barotropic and a baroclinic structure.

We next compare the scale-dependent term with the corresponding term in Thiebaux et al. (1986). They concluded that “it seems plausible that we could obtain better fits by adding a constant parameter . . . $R_{00}$” and studied the form of $R_{00} + R \exp(-b\theta^2)$. This is similar in form to our (7) and thus $R_{00}$ corresponds to a simplified version of our “scale-dependent” term.

Hence $R_{00}/R_0$ in Thiebaux et al. should correspond to $A/(1-A)$. It follows that in their results $A \sim 0.15$–0.20. These values are smaller than ours (Fig. 7). This is natural because the definitions of the forecast error differ and, as a consequence, the effect of the analysis error is obviously smaller in their study. Further, in their study $R_{00}/R_0$ or $A/(1-A)$ decreases poleward in agreement with our Fig. 7.

We continue by assuming that $R_{00}$ in Thiebaux et al. corresponds to our “scale-dependent” term. Then $\exp(-b\theta^2)$ in Thiebaux et al. corresponds to our randomlike term, which in (6) is

$$\exp(-49\theta^2) - 45\theta^2 \exp(-45\theta^2) \Phi.$$  

To compare these terms, we present them in Taylor series, which are

$$\exp(-b\theta^2) = 1 - 118.96\theta^2 + 7076\theta^4 - \cdots \quad (8)$$  

(Thiebaux et al., 40°–65°N, their Table 2)

and

$$\exp(-49\theta^2) - 45\theta^2 \exp(-45\theta^2) \Phi$$

$$= 1 - 69\theta^2 + 1903\theta^4 - \cdots \quad (9a)$$

[our Eq. (6), $\varphi = 60^\circ$N]
or, in an alternative form,

\[
\sum_{n=0}^{\infty} t^n (1 + 2n)^{1/2} P_n (\cos \theta) = 1 - 65 \theta^2 + 1845 \theta^4 + \ldots
\]

(9b)

[our Eq. (5), \( \varphi = 60^\circ N \).]

Comparing (9a) and (9b), it is seen that the left-hand side in (9b) is well approximated by that in (9a). As a consequence, the comparison with Thiebaux et al. can be based as well on Eq. (6) (9a) as on Eq. (5) (9b). The series in (8) and (9) clearly differ. However, the functional forms are rather similar. This is seen by multiplying \( b \) in (8) by 0.55. The right-hand sides in (8) and (9) now become similar. Corresponding multiplication factors in zones 0–30°N, 25–50°N, 40–65°N, 60–90°N are 1.4, 1.05, 0.9, respectively. Here we have applied \( \varphi = 0^\circ, 30^\circ, 60^\circ, 90^\circ N \) in our (6). Note here that the case of \( \varphi = 0^\circ \) is an extrapolation because this case lies outside our computing area. The comparison at \( \varphi = 0^\circ \) is thus less realistic. In the zones of 25°–50° and 60°–90°N the functions are, up to the fourth order, close to each other. The only exception is for 40°–65°N, where the functions have similar analytical forms but the values differ because of different parameters. The difference is not negligible; our function is damping much more slowly.

It is not easy to explain the difference because so many factors can contribute. The data sources differ; previously we have seen that data from the very same center but from different periods can give different results. Thiebaux et al. combine observations whereas we study separately each pair of grid points. The definitions of the forecast error differ. The fitting procedures differ: we fit with the spectral values whereas Thiebaux et al. fit conventionally with the observations (the autoregressive technique is applied only in finding function candidates). In our case, the autocorrelation equals unity at zero separation. In Thiebaux et al., the corresponding autocorrelation is not given because of the unknown effect of the observation error. This makes the fitting more difficult because at shorter distances there are no observations. Finally we have dealt with the scale-dependent term in detail whereas Thiebaux et al. present it simply as a constant. This, of course, affects the determination of the randomlike term.

Thus we can claim that scale-dependent and randomlike terms can be found in Thiebaux et al. The former is represented by a constant, the latter one is given by \( \exp(-b \theta^2) \). This fits rather well with our function except at approx. 55°N, where there is a clear difference in parameter values. Our randomlike term is more complicated than \( \exp(-b \theta^2) \). This is natural because our function gives both the autocorrelation function and its latitudinal variation.

Other functions studied by Thiebaux et al. differ from ours because in their Taylor series the first and third power terms (\( \theta, \theta^3 \)) are present. This means that our results do not correspond to those of the autoregressive method. The function of \( \exp(-b \theta^2) \) in Thiebaux et al. originates from Gandin (1963).

Our results agree with the discussion presented in Balgovind et al. (1983). They observe that the Southern Hemisphere functions do not possess such long-distance tails as those of the Northern Hemisphere. Their explanation lies in the effect of thermal (land–sea contrast) and orographic forcing in the forecasts. We only make it more precise by adding that, in the analyses, the land–sea contrast is a consequence of a nonuniform observational network, and that mountains can cause difficulties in the analysis process. Further, Balgovind et al. observe that "the external, barotropic motions do not dominate the horizontal forecast error variances". We differ slightly in stating that the "scale-dependent" error can nevertheless be locally dominant.

We also experimented with the function given by Balgovind et al. (1983). We tried to find such a combination of our parameters (\( A, a, t \)) that the corresponding function could be fitted with theirs. No such combination was found. Balgovind et al.'s Bessel representation does not fit well with the present function type.

Our model of the two source terms only applies, in a strict sense, to 24-h forecast errors as determined by the difference between the forecast and the analysis. We believe that the model can be generalized to the usual case of 6-h errors given as the difference between the forecast and observations. In this case the analysis errors are introduced into the forecast error via the initial field of the forecast process.

9. Summary

The autocorrelation function of the nonsystematic and isotropic error in ECMWF 24-h 500 mb height forecasts has been studied separately at each grid point. In the preliminary computations, the EOFs of the forecast error have been determined. The first EOF shows a teleconnection between the Himalayas and other parts of the Northern Hemisphere. The third EOF shows patterns which resemble those of the analysis error.

The autocorrelation fields are represented with the aid of spherical harmonics. The coefficients of the harmonics are approximated by a function that consists of two parts. The first part has a structure taken from studies of analyses of geopotential height (Rinne and Järvenoja, 1985). The second part is taken to resemble that which would be generated by a random process.

A large-scale effective radius is characteristic of the first part. The values of the autocorrelation function are clearly nonzero at several thousand kilometers. It is suggested that this long distance part can be generated by analysis errors over mountains or data-sparse areas.
Model errors at barotropically unstable wave lengths and over mountains are another potential source of the term. It is also suggested that this term explains the existence of the constant \( R_{00} \) introduced by Thiebaux et al. (1986).

The second part of the autocorrelation function typically vanishes at distances greater than 1000 km. This randomlike part dominates over mountains like the Himalayas, Alps and Rocky Mountains. Two maxima are also found over the Pacific.

Though the structure of the autocorrelation function becomes rather complicated, there are nevertheless only three free parameters. These determine the dominant scale of the scale-dependent part, the redness of the nearly white spectrum of the randomlike part, and the relative contributions of the two parts. These contributions seem to be connected to the large-scale patterns shown by the first EOF of the forecast error. The parameters vary geographically so that they cannot be given explicitly. The only exception is "the redness of the randomlike part". This parameter can approximately be given in terms of the latitude.

Finally, an attempt has been made to simplify the complicated form of the autocorrelation function found. The most heavily truncated approximation turns out to be similar to the classical Gaussian shape or negative squared exponential function. It is, however, necessary to note that this really is a truncated version. Nevertheless, even in that form the two principal features can be seen: first, the more or less irregular geographical variation of the parameters, second, the two source terms that make the autocorrelation function.

The results are applied by making a comparison between ECMWF forecasts for 1981–82 and 1982–83. There has been a clear change only in one term: the relative contribution of the scale-dependent error has decreased nearly uniformly over the whole of the Northern Hemisphere.

Acknowledgments. This study has been financially supported by the Academy of Finland. Dr. H. Savijärvi kindly read one version of the manuscript.

APPENDIX

The EOF Representation of the Forecast Error: Computation and Spherical Harmonic Representation of Autocorrelation Fields

Following Rinne and Järvenoja (1985), the forecast error fields \( e \) are expanded in empirical orthogonal functions (EOF)

\[
e(i, \tau) = \text{em}(i) + \sum_{\tau=1}^{200} E_{\tau}(\tau)h_{\tau}(i) + \text{residual (i, } \tau) \tag{A1}
\]

where \( \text{em} \) is the mean (systematic) forecast error, \( h_{\tau} \) are spatial functions and \( E_{\tau} \) are time-dependent coefficients of EOF component \( \tau \). The point of truncation used is 200. Indices \( i \) and \( \tau \) refer to grid point and time, respectively. We shall now consider the nonsystematic forecast error \( \delta e \):

\[
\delta e(i, \tau) = e(i, \tau) - \text{em}(i) \approx \sum_{\tau=1}^{200} E_{\tau}(\tau)h_{\tau}(i). \tag{A2}
\]

In (A2) we have neglected the small residual term. Using the orthogonality property of EOFs, we simply obtain for the autocovariance between points \( i \) and \( j \)

\[
\overline{\delta e(i, \tau)\delta e(j, \tau)} = \sum_{\tau=1}^{200} \overline{E_{\tau}^2(\tau)h_{\tau}(i)h_{\tau}(j)} \tag{A3}
\]

where the overbar \( \overline{\cdot} \) refers to the time mean, and \( E_{\tau}^2(\tau) \) terms are eigenvalues associated with the \( h_{\tau} \) functions. Now the autocovariances are transformed to the autocorrelations \( r \):

\[
r(i, j) = \sum_{\tau=1}^{200} \overline{E_{\tau}^2(\tau)h_{\tau}(i)h_{\tau}(j)} / [\sigma(i)\sigma(j)], \tag{A4}
\]

where the gridpoint variances \( \sigma^2(i) \) are given by

\[
\sigma^2(i) = \sum_{\tau=1}^{200} \overline{E_{\tau}^2(\tau)h_{\tau}^2(i)}.
\]

Next, the autocorrelation fields \( r \) are expanded in spherical harmonics. The pole of the harmonics is located at the base point \( i \) for which the autocorrelation field is determined. The zero longitude falls on the meridian of point \( i \). The longitude \( \lambda \) and the colatitude \( \theta \) are determined with respect to the zero longitude and point \( i \) (see RJ).

\[
r(i, j) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} [S_n^m(i) \sin \lambda + C_n^m(i) \cos \lambda]Y_n^m(\cos \theta), \tag{A5}
\]

where

\[
\sin (m\lambda)Y_n^m(\cos \theta)
\]

are the spherical harmonics. Index \( i \) refers to the base point. The separation between points \( i \) and \( j \) \( (j = 1, \ldots, 1404 \text{ in our case}) \) is taken along the connecting great circle and is given by \( \theta \).

Coefficients \( S_n^m \) and \( C_n^m \) are computed in the usual way from

\[
S_n^m = \int_{-\pi}^{\pi} \int_{0}^{\pi} r(i, j) \sin (m\lambda)Y_n^m(\cos \theta) d\theta d\lambda. \tag{A6}
\]

In the case of a discrete grid, this is approximated by
\[
S_n^m = \sum_j r(i, j) \sin(m\lambda_j)Y_n^m(\cos\theta_{ij})dA(j), \quad (A7)
\]

where \(\lambda_j\) and \(\theta_{ij}\) are the latitude of point \(j\) and the angular distance from point \(j\) to point \(i\), respectively, and \(dA(j)\) is the areal element associated with grid point \(j\) \([\sum_j dA(j) = 1]\).

It is assumed that \(r(i, j) = 0\) if \(j\) is outside the grid area. When the base point is on the boundary, the assumption is clearly invalid and there will be a discontinuity across the boundary. Hence the computations are not realistic for these points.

In the present case, we are studying only the isotropic part of the autocorrelation. Then \(m = 0\) in (A5), and taking into account that \(Y_n^0(\cos\theta_{ij}) = (1 + 2n)^{1/2} \times P_n(\cos\theta_{ij})\), we obtain a simpler formula [Eq. (2) in the main text] for the isotropic part of the autocorrelation.

REFERENCES


