Tellus (1991), *43A*, 177–190

Variation of the predictability in a low-order spectral model of the atmospheric circulation

By P. L. HOUTEKAMER, *KNMI, P.O. Box 201, 3730 AE De Bilt, The Netherlands*

(Manuscript received 18 July 1990; in final form 15 January 1991)

ABSTRACT

Temporal variations in the predictability of a simple atmospheric model are studied. The model is a spectral two-layer quasi-geostrophic hemispheric model, truncated at T5. It has chaotic properties, illustrated by 4 positive Lyapunov exponents and a smooth power spectrum of the first empirical orthogonal function (EOF). The chaotic behavior is responsible for the eventual separation of initially nearby phase points. We use the adjoint of the tangent linear equations to obtain the few directions in which errors grow most rapidly. With the growth rates in these directions, we can estimate the distribution function for the errors. Comparison with a Monte Carlo method shows satisfactory agreement. With a 100-day run, we show how the predictability varies over the attractor. We discuss how, and under what conditions, the method can be used in a global circulation model (GCM). From a dimension derived from an EOF analysis, we obtain an estimate for the dimension of error growth. This estimate implies that information on the probability distribution of errors can be obtained at about 150 times the computational cost of a single model run.

1. Introduction

Weather forecasts tend to go wrong after some time. There is a consensus that detailed deterministic predictions are limited to a range of about 10 days. One reason is the presence of systematic errors that are introduced by imperfections of the forecast model. Another more fundamental reason is the intrinsic deterministic growth of errors. The growth of systematic and intrinsic errors makes the predicted weather gradually diverge from the subsequently observed weather. Systematic errors can be reduced either by improving the physical description which is used for the model or by post-processing of model output (Thiébaux and Morone, 1990). Intrinsic errors are caused by the unavoidable sensitivity of the model to small errors in the initial conditions. Their possible importance was first discussed by Lorenz (1963). In this paper, we will concentrate on the intrinsic errors.

Some predictions have a higher quality than others. To make effective use of high-quality forecasts and to be able to warn against low-quality forecasts, a measure for the temporal predictability is needed. Stochastic dynamic predictions (Epstein, 1969) require the integration of the mean values of the model variables and of the corresponding covariance matrix. Because in addition to the N equations for the mean field, there are $N(N+1)/2$ equations for the evolution of the second moments, this method may not be feasible in practice. Because only the 1st and 2nd moments are considered, the method can only be applied to the short-range evolution of initially small errors. A probably less expensive method to study variations in the predictability of the atmosphere would be to take a sufficiently large ensemble with perturbed initial conditions (Leith, 1974). The divergence of the ensemble is a measure of the quality of the forecast. This method should work, even in the regime of non-linear error growth. A review of methods to estimate the predictability of forecasts in the 5 to 10 days range, where non-linearity is important, is given by Palmer and Tibaldi (1988).

It is shown in a study by Lacarra and Talagrand (1988) with an f-plane shallow-water model that the linear approximation for error growth can be valid for ranges up to about 48 h in meteorologi-

cally realistic situations. The linear approximation greatly simplifies the analysis. If the evolution of the errors is linear, it is possible to make use of the adjoint of the tangent linear equations. Le Dimet and Talagrand (1986) have given a rather general presentation of the use of adjoint equations in the context of data assimilation. An example of a minimization of a distance function using the adjoint is given by Courtier and Talagrand (1987). Their model contains 231 independent degrees of freedom and is integrated for 24 h. Their distance function is a measure for the agreement between a model-run and a set of verifying observations. They find strong indications that their distance function varies quadratically with respect to the initial conditions, implying that the tangent linear equations of their model are indeed sufficient to describe the 24-h evolution of forecast errors. Consequently, the minimization leads to the trajectory which shows the best agreement with the available observations. Thus, the adjoint equations can be used both for data assimilation and for error growth studies.

Our goal is to determine confidence limits for short-range numerical forecasts. We study the probability distribution of the errors in an ensemble in a 30-dimensional baroclinic model as obtained with a Monte Carlo method, and try to reproduce this distribution with less time consuming methods. The error growth is described with the tangent linear equations. Because our model is small, we can compute all growth rates rigorously. The growth rates have a very steep spectrum, allowing us to approximate the probability distribution for the errors using only the few dominating growth rates. The dominating growth rates correspond to the largest eigenvalues of a symmetric matrix. We use a Lanczos algorithm (see e.g., Parlett 1980) to estimate these eigenvalues. A Lanczos algorithm requires only a few matrix vector multiplications. Individual multiplications are done using both the tangent linear equations and the adjoint of the tangent linear equations. The applicability of the Lanczos algorithm depends on the number of relevant eigenvalues or, putting it in other words, on the dimensionality of the errors. We give a relationship between the dimensionality of the instantaneous circulation and the dimensionality of the errors. If the same relationship holds for a GCM, then the algorithm can be applied to it.

2. Description of the model

Our model is a highly truncated 2-layer hemispheric spectral model. Both layers have 3 zonal and 6 wave modes. It is comparable with the highly truncated 2-layer spectral channel model of Reinhold and Pierrehumbert (1982) and the standard 2-layer model discussed by Holton (1979). We will use the notation introduced by Lorenz (1960) where ψ_1 is the 250-mb streamfunction, ψ_3 is the 750-mb streamfunction, $\psi = \frac{1}{2}(\psi_1 + \psi_3)$ is the interpolated 500-mb streamfunction and $\tau = \frac{1}{2}(\psi_1 - \psi_3)$ is the 250–750 mb thickness. The set of equations is:

$$\frac{\partial}{\partial t}\Delta\psi_1 = -J(\psi_1, \Delta\psi_1 + f) + \frac{f_0}{\Delta p}\omega_2, \tag{1}$$

$$\frac{\partial}{\partial t}\Delta\psi_3 = -J(\psi_3, \Delta\psi_3 + f) - \frac{f_0}{\Delta p}\omega_2 - f_0 J(\psi_3, h) - C\,\Delta(\psi_3 - \psi_3^*), \tag{2}$$

$$\frac{\partial}{\partial t}2\tau = -J(\psi, 2\tau) + \frac{\sigma\,\Delta p}{f_0}\omega_2 + 2Q(\tau^* - \tau), \tag{3}$$

where $J(A, B)$ denotes the jacobian operator:

$$J(A, B) = \frac{1}{a^2}\left(\frac{\partial A}{\partial \lambda}\frac{\partial B}{\partial \mu} - \frac{\partial A}{\partial \mu}\frac{\partial B}{\partial \lambda}\right). \tag{4}$$

Here, λ is the geographic longitude, μ is the sine of the geographic latitude and a is the radius of the earth. We non-dimensionalize eqs. (1)–(4) using the radius a as unit of length and the inverse of the angular speed of rotation of the earth as unit of time. f is the Coriolis parameter and f_0 is the value of the Coriolis parameter at a latitude of 45°. C is the Ekman damping coefficient. The Ekman damping is towards the 750 mb-streamfunction ψ_3^*, which corresponds to a zonal westerly wind in the lower layer. We enforce a 250–750 mb thickness with the function τ^*. The function τ^* corresponds to a meridional temperature gradient, and Q is the coefficient for the cooling towards this gradient. h is the mountain height and σ is the static stability parameter. The pressure difference Δp between the two layers is 500 mb. The vertical motion $\mathrm{d}p/\mathrm{d}t$ at the 500 mb-level is ω_2. Eqs. (1) and (2) are the vorticity equations for the 250-mb and the 750-mb layer, respectively. Eq. (3) is the thermodynamic energy equation.

We define the radius of deformation Λ^{-1} by:

$$\Lambda^2 = \frac{f_0^2}{\sigma(\Delta p)^2}. \tag{5}$$

Eliminating ω_2 from the eqs. (1), (2) and (3), we get:

$$\frac{\partial}{\partial t}\Delta\psi = -J(\psi, \Delta\psi) - J(\tau, \Delta\tau) - J(\psi, f) + \frac{f_0}{2}J(\tau - \psi, h) + \frac{C}{2}\Delta(\tau - \psi + \psi_3^*),$$

$$\frac{\partial}{\partial t}(\Delta - 2\Lambda^2)\tau = -J(\psi, \Delta\tau) - J(\tau, \Delta\psi) - J(\tau, f) + \frac{f_0}{2}J(\psi - \tau, h) + \frac{C}{2}\Delta(\psi - \tau - \psi_3^*) + 2\Lambda^2 J(\psi, \tau) - 2\Lambda^2 Q(\tau^* - \tau). \tag{6}$$

The streamfunctions are projected on a basis of eigenfunctions of the Laplace operator Δ. These eigenfunctions are:

$$Y_{m,n}(\lambda, \mu) = P_{m,n}(\mu)\, e^{im\lambda}. \tag{7}$$

Here, $P_{m,n}(\mu)$ denote associated Legendre functions of the first kind, which are defined by:

$$P_{m,n}(\mu) = \left(\frac{(2n+1)(n-m)!}{(n+m)!}\right)^{1/2} \times \frac{(1-\mu^2)^{m/2}}{2^n n!}\frac{\mathrm{d}^{m+n}}{\mathrm{d}\mu^{m+n}}(\mu^2 - 1)^n. \tag{8}$$

The streamfunction is now approximated as follows:

$$\psi(\lambda, \mu, t) = \sum_{n=1}^{5}\sum_{m=-n}^{+n}\psi_{m,n}(t)\, Y_{m,n}(\lambda, \mu), \quad m + n = \text{odd}. \tag{9}$$

The restriction to modes with $m+n$ odd excludes currents across the equator. It makes the model hemispheric with a T5 triangular truncation. The 3 modes with m equal to zero are the real zonal modes. They are a function of the geographic latitude only. The remaining 6 wave-modes have a real and an imaginary part. Consequently, 15 coefficients are needed to describe the 500-mb streamfunction and another 15 coefficients to describe the 250–750 mb shear streamfunction, making the total system 30-dimensional. The moderate number of modes is a compromise between a completely unrealistic low-dimensional model, which can be handled by any computational method, and a higher-dimensional model in which rigorous testing becomes computationally expensive.

A Galerkin projection of eq. (6) on these eigenfunctions gives a system of 30 coupled non-linear ordinary differential equations. We evaluate the right-hand side of this equation using the interaction coefficient method (Platzman, 1960). The interaction coefficients are computed with the analytical expressions given by Silberman (1954). An alternative method is the transform method (Orszag, 1970). This method becomes more efficient if the truncation is made less severe (Bourke et al., 1977).

The Ekman damping coefficient C and the cooling coefficient Q are given a value of $2\pi/100$ a day. Thus, the e-folding time for both effects is about 16 days. These unrealistically low values of C and Q make the circulation attractor high-dimensional. If they are chosen equal to zero, the circulation is no longer attracted towards realistic patterns and spreads out in the 30-dimensional phase space.

The forcings are in the $Y_{0,1}$ mode only. The strength of τ^* corresponds to a westerly thermal wind of 28 m/s and ψ_3^* corresponds to a lower layer westerly wind of 5 m/s. ψ_3^* can be considered to reproduce the effect of surface baroclinicity. Without this forcing, the 750-mb westerlies are very weak. Due to the effects of the topography and the baroclinicity, the enforced westerly winds become unstable.

The radius of deformation Λ^{-1} is 667 km. The model is not very sensitive to the exact value of this parameter.

The topography is described with the $Y_{2,3}$ mode, standing for the influence of the Rocky Mountains and the Tibetan plateau. The topography serves both to destabilize the enforced zonal flow and to locate preferential circulation patterns at fixed geographical positions. The amplitude of the mountain height is 1.56 km. In Fig. 1 we show both the topography and the barotropic component of the climate. The climate is computed with a 30-year run. Since the thermal forcing is in the $Y_{0,1}$ mode and the topographic forcing is in the

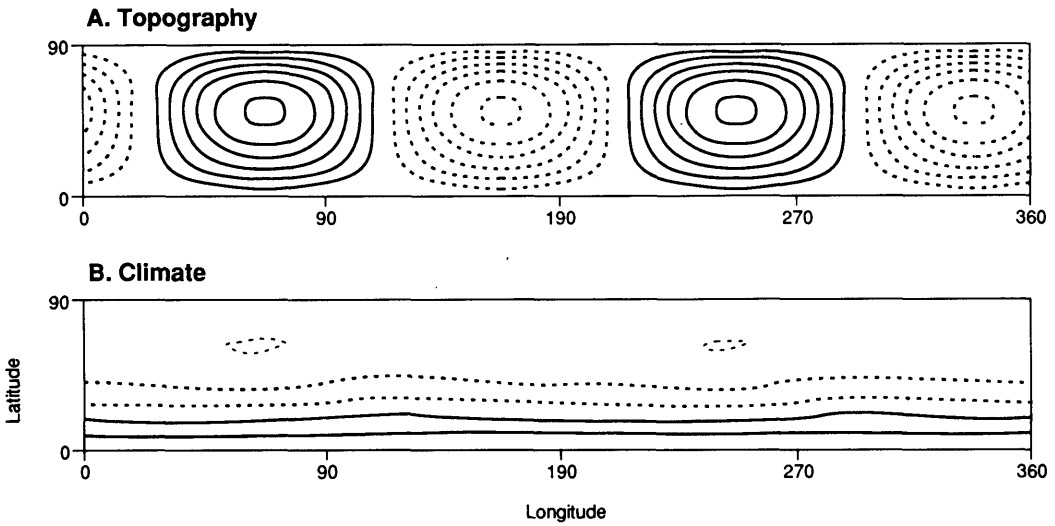

Fig. 1. (a) Contours of the topography. The projection is area preserving. The vertical coordinate has been scaled with the sine of the geographic latitude. The horizontal coordinate is the geographic longitude. Negative heights are indicated with dashed lines. The contour interval is 250 m. (b) Contours of the geopotential height at 500 mb for the 30-year climate. The contour interval is 300 m. Projection as in (a).

$Y_{2,3}$ mode, it follows, as we observe in Fig. 1, that the model climate is π-periodic in longitude. This means that the climate is restricted to the $2 \times (3 + 2 \times 3) = 18$ dimensional subspace (m even, $m + n$ odd) of the $2 \times (3 + 2 \times 6) = 30$ dimensional space ($m + n$, odd) of the model. Because the symmetry is broken in the specification of the initial state, the dynamics themselves are not confined to the 18-dimensional subspace.

The most significant component of the model circulation is an eastwards moving $Y_{4,5}$ wave. The wave is accelerated and decelerated by the varying intensity of the zonal wind. During short-time intervals, it may move westwards. All other waves continuously move westwards. The influence of the zonal wind is insufficient to keep them at fixed geographical longitude.

Fig. 2 shows the power spectrum of the first EOF. We will explain how we determine the EOFs in Section 5. A power spectrum shows how the energy of a signal is distributed as a function of frequency. A periodic signal shows only isolated peaks in its power spectrum. A chaotic signal shows a much smoother spectrum (Schuster, 1988). Fig. 2 was obtained from four 2048-day integrations. The power spectrum contains most energy on periods between 9 and 30 days. This is a first indication that our model shows chaotic behaviour. The second EOF has an almost identical power spectrum. The 1st and 2nd EOF approximately describe the travelling $Y_{4,5}$ wave.

A chaotic trajectory is, by definition, not periodic. We verify numerically that solutions do not return negligibly close to previously visited phase points within a short time. This behaviour suggests that the asymptotic orbit defines a strange attractor. A strange attractor is characterized by a number of growth directions for errors, at least one neutral direction in which errors neither grow nor shrink and a number of directions with negative growth.

We compute all 30 Lyapunov exponents with the algorithm given by Wolf et al. (1985). A basis of 30 orthonormal vectors is integrated forward with the tangent linear equations. After every 1.6 days, a Gram-Schmidt orthonormalisation is applied. The change in length of the vectors fixes the Lyapunov exponents. After integrating the model over 3.5 years, we find 4 positive Lyapunov exponents, 1 exponent zero and 25 negative exponents.

Kaplan and Yorke (1979) conjectured the

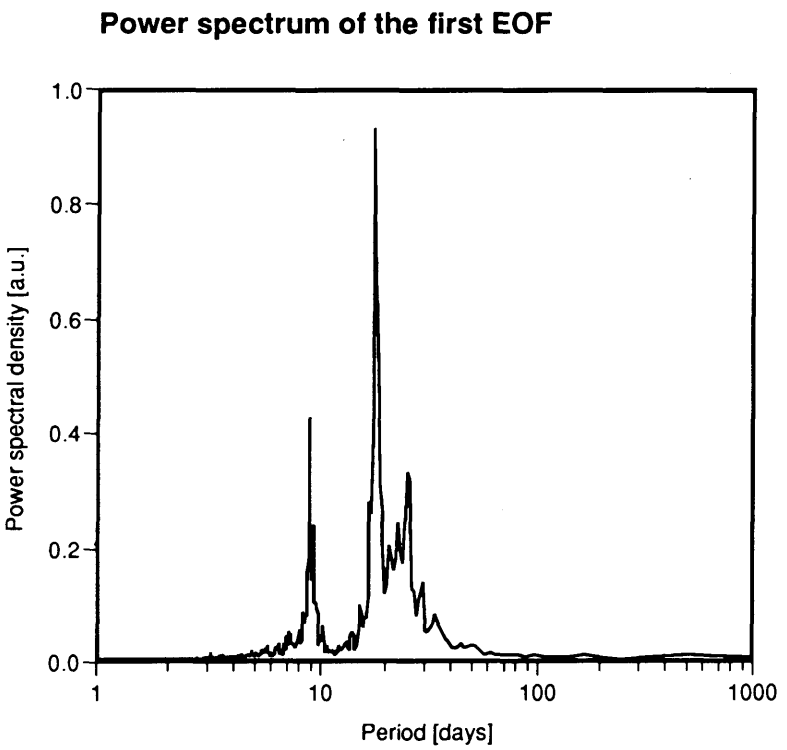

Fig. 2. Powerspectrum of the first EOF. Vertical units are arbitrary. The area under the curve is a measure of the amount of energy contained in that part of the spectrum.

following general formula for the dimension of arbitrary strange attractors:

$$D_{\mathrm{KY}} = j + \frac{\sum_{i=1}^{j} \alpha_j}{|\alpha_{j+1}|} \tag{10}$$

Here, the Lyapunov exponents are ordered $\alpha_1 > \alpha_2 > \cdots > \alpha_{30}$, j is the largest integer for which $\sum_{i=1}^{j} \alpha_j > 0$ and D_{KY} is, in almost all cases, equal to the information dimension. In some cases, D_{KY} is indeed shown to be equal to the information dimension within the precision of the estimate. In other non-generic cases, the conjecture does not hold. We will call D_{KY} the Kaplan-Yorke dimension. We find a Kaplan-Yorke dimension of 10.7 ± 0.2 for our model. Due to these observations, we assume that the asymptotic dynamics of the system are characterized by a strange attractor, and consequently we expect complex unpredictable behaviour.

3. Determination of growth rates

We assume that errors in the initial conditions are small. As a consequence, we can describe the initial evolution of errors by a linearization of eq. (6) with respect to the unperturbed solution. We give two methods for the computation of the growth rates of the errors. The first method uses the tangent linear equations and gives us all 30 growth rates. The second method uses both the tangent linear equations and their adjoint. It gives us only the largest growth rates, but requires less computations. The latter method can also be applied to a larger model.

The tangent linear equations can be written as:

$$\frac{\partial}{\partial t}\begin{pmatrix} \delta\,\Delta\psi \\ \delta\,\Delta\tau \end{pmatrix} = S^{-1} L \begin{pmatrix} \delta\,\Delta\psi \\ \delta\,\Delta\tau \end{pmatrix}. \tag{11}$$

The operators S^{-1} and L are:

$$S^{-1} = \begin{pmatrix} I & 0 \\ 0 & \Delta(\Delta - 2\Lambda^2)^{-1} \end{pmatrix}, \tag{12}$$

$$L \begin{pmatrix} \delta\,\Delta\psi \\ \delta\,\Delta\tau \end{pmatrix} = \begin{pmatrix} -J(\delta\psi, \Delta\psi) - J(\psi, \delta\,\Delta\psi) \\ \quad -J(\delta\tau, \Delta\tau) - J(\tau, \delta\,\Delta\tau) \\ -J(\delta\psi, \Delta\tau) - J(\psi, \delta\,\Delta\tau) \\ \quad -J(\delta\tau, \Delta\psi) - J(\tau, \delta\,\Delta\psi) \end{pmatrix}$$

$$+ \begin{pmatrix} -\frac{f_0}{2} J(\delta\psi - \delta\tau, h) - J(\delta\psi, f) \\ \quad + \frac{C}{2}(\delta\,\Delta\tau - \delta\,\Delta\psi) \\ \frac{f_0}{2} J(\delta\psi - \delta\tau, h) - J(\delta\tau, f) \\ \quad - \frac{C}{2}(\delta\,\Delta\tau - \delta\,\Delta\psi) \end{pmatrix}$$

$$+ 2\Lambda^2 \begin{pmatrix} 0 \\ J(\delta\psi, \tau) - J(\delta\tau, \psi) + Q\delta\tau \end{pmatrix}. \tag{13}$$

For the development of the adjoint model distances are computed in a kinetic energy norm. The kinetic energy of a field is given by:

$$K = \int_\Sigma (\nabla\psi \cdot \nabla\psi + \nabla\tau \cdot \nabla\tau)\, d\Sigma$$

$$= \sum_{n=1}^{5} \sum_{m=-n}^{+n} n(n+1) \times \{|\psi_{m,n}(t)|^2 + |\tau_{m,n}(t)|^2\}, \qquad m+n = \text{odd}.$$

The factor $n(n+1)$ in this expression is removed with the rescaling:

$$\tilde{\psi}_{m,n}(t) = (n(n+1))^{1/2}\, \psi_{m,n}(t), \tag{14}$$

$$\tilde{\tau}_{m,n}(t) = (n(n+1))^{1/2}\, \tau_{m,n}(t). \tag{15}$$

We put the real and imaginary parts of $\tilde{\psi}_{m,n}$ and $\tilde{\tau}_{m,n}$ is one single real valued vector q (of length 30). The vector q contains the coefficients q_i of the state vector with respect to a 30-dimensional basis $\{e_1, e_2, ..., e_{30}\}$ which is orthonormal with respect to the kinetic energy inner product:

$$(e_i, e_j) = \delta_{ij}. \tag{16}$$

The kinetic energy or length squared associated with the vector q becomes:

$$R^2 = \sum_{i=1}^{30} q_i^2 = (q, q). \tag{17}$$

In the rest of our paper, we use the kinetic energy inner product to define lengths or error-magnitudes.

An initial error vector $u(0)$ is integrated forward with the tangent linear equations to give a vector $u(t)$. If an ensemble with more than 30 initial vectors is to be integrated, then it is more efficient to compute the 30×30 matrix A defined by:

$$u(t) = Au(0). \tag{18}$$

The matrix A is time dependent, but for convenience of notation, we write A rather than $A(t)$. The columns of A follow from the forward integration of the unit-vectors of the basis. Thus, we first take a unit vector e_i, we transform this vector to streamfunction coefficients using eqs. (14), (15), (16) and (17), then we integrate eq. (11) and transform the vector back to the basis $\{e_1, e_2, ..., e_{30}\}$. The kinetic energy of the error at time t is given by:

$$(u(t), u(t)) = (Au(0), Au(0)) = (A^*Au(0), u(0)). \tag{19}$$

A^* is the adjoint matrix of A, which is just the transpose matrix:

$$A^* = A^{\mathrm{T}}. \tag{20}$$

The eigenvalues of A^*A are all positive. In order of decreasing magnitude, we call them: $\lambda_1^2, \lambda_2^2, ..., \lambda_{30}^2$. The eigenvectors of A^*A are orthogonal. Once we have the eigenvalues and eigenvectors of A^*A, we can derive statistics for the error at time t. If for instance, we start with an ensemble of random initial error vectors of unit length, the average error kinetic energy after time t is given by (Lorenz, 1985):

$$\overline{(u(t), u(t))} = \sum_{i=1}^{30} \frac{\lambda_i^2}{30}. \tag{21}$$

The model analysed in the present paper is sufficiently small so that the above formulation can be applied; this is not the case for models with a larger

number of degrees of freedom. In the following paragraphs, we sketch an alternative method, which does not require the computation of the entire matrix A and makes use of the adjoint equations. Basic facts about adjoint operators and a specific example with the adjoint of the vorticity equation can be found in Talagrand and Courtier (1987). Because we use the same inner product, our adjoint model is comparable to theirs.

The adjoint of the tangent linear equation (eq. (11)) can be written as:

$$\frac{\partial}{\partial t}\begin{pmatrix}\delta\,\Delta\psi'\\ \delta\,\Delta\tau'\end{pmatrix} = -L^*(S^{-1})^*\begin{pmatrix}\delta\,\Delta\psi'\\ \delta\,\Delta\tau'\end{pmatrix}. \tag{22}$$

$(S^{-1})^*$ and L^* are the adjoints of the operators S^{-1} and L given in eqs. (12) and (13). The adjoint operators read:

$$\begin{pmatrix}\delta\,\Delta\psi''\\ \delta\,\Delta\tau''\end{pmatrix} = (S^{-1})^*\begin{pmatrix}\delta\,\Delta\psi'\\ \delta\,\Delta\tau'\end{pmatrix} = \begin{pmatrix}I & 0\\ 0 & (\Delta - 2\Lambda^2)^{-1}\,\Delta\end{pmatrix}\begin{pmatrix}\delta\,\Delta\psi'\\ \delta\,\Delta\tau'\end{pmatrix}, \tag{23}$$

$$L^*\begin{pmatrix}\delta\,\Delta\psi''\\ \delta\,\Delta\tau''\end{pmatrix} = \begin{pmatrix}-J(\Delta\psi,\,\delta\psi'') + \Delta J(\psi,\,\delta\psi'')\\ \quad -J(\Delta\tau,\,\delta\tau'') + \Delta J(\tau,\,\delta\tau'')\\ -J(\Delta\tau,\,\delta\psi'') + \Delta J(\tau,\,\delta\psi'')\\ \quad -J(\Delta\psi,\,\delta\tau'') + \Delta J(\psi,\,\delta\tau'')\end{pmatrix} + \begin{pmatrix}-\frac{f_0}{2}J(h,\,\delta\psi''-\delta\tau'') - J(f,\,\delta\psi'')\\ \quad +\frac{C}{2}(\delta\,\Delta\tau''-\delta\,\Delta\psi'')\\ \frac{f_0}{2}J(h,\,\delta\psi''-\delta\tau'') - J(f,\,\delta\tau'')\\ \quad -\frac{C}{2}(\delta\,\Delta\tau''-\delta\,\Delta\psi'')\end{pmatrix} + 2\Lambda^2\begin{pmatrix}J(\tau,\,\delta\tau'')\\ -J(\delta\psi,\,\tau'') + Q\delta\tau''\end{pmatrix}. \tag{24}$$

We can see from eqs. (23) and (24) that the complexity of the adjoint model is comparable to the complexity of the tangent linear model (eqs. (12) and (13)). Backward integration with the adjoint model replaces the multiplication with the matrix A^* in the same way as forward integration with the tangent linear equations replaces the multiplication with the matrix A. For a large model, this provides us with a practical tool for obtaining the largest eigenvalues of the matrix A^*A.

We use a Lanczos algorithm with orthogonalization (see e.g., Parlett, 1980) to find the largest eigenvalues of A^*A. Below, we give a description of the Lanczos algorithm. In a series of j iterative steps we produce a basis $\{\Delta_1, \Delta_2, ..., \Delta_j\}$, which approximates the basis of eigenvectors $\{v_1, v_2, ..., v_j\}$. We start with a random direction vector Δ_1 of unit length:

$$\Delta_1 = \sum_{k=1}^{30} b_k v_k = \sum_{k=1}^{30} a_k e_k,$$
$$(\Delta_1, \Delta_1) = 1. \tag{25}$$

Notice that at this point, we still do not know the coefficients b_k in eq. (25). The coefficients a_k give the random direction. If we apply A^*A to the vector Δ_1, the coefficient for the kth vector increases with λ_k^2:

$$A^*A\,\Delta_1 = \sum_{k=1}^{30} b_k \lambda_k^2 v_k. \tag{26}$$

This property suggests using the matrix A^*A for getting the new direction Δ_2 and subsequent new directions Δ_{i+1}. We use a Gram-Schmidt orthonormalisation to make a new direction Δ_{i+1} orthonormal to the previous directions $\Delta_1, \Delta_2, ..., \Delta_i$:

$$\Delta'_{i+1} = A^*A\,\Delta_i - \sum_{k=1}^{i} \frac{(A^*A\,\Delta_i, \Delta_k)}{(\Delta_k, \Delta_k)}\Delta_k, \tag{27}$$

$$\Delta_{i+1} = \frac{\Delta'_{i+1}}{(\Delta'_{i+1}, \Delta'_{i+1})^{1/2}}. \tag{28}$$

After every iteration step, we can estimate the eigenvalues of A^*A with the eigenvalues of the tridiagonal $j \times j$-matrix $E_j^{\mathrm{T}}E_j$:

$$E_j = [A\,\Delta_1,\, A\,\Delta_2,\, ...,\, A\,\Delta_j]. \tag{29}$$

The estimate is exact if the eigenvalues $\lambda_j^2, \lambda_{j+1}^2, ..., \lambda_{30}^2$ are equal to zero. The largest eigenvalue is estimated most accurately. In a typical case, we need 4 iterations to find the largest eigenvalue, 3 additional iterations for the second

largest eigenvalue, successive iterations give one more accurate estimate per iteration. The maximum number of 30 iteration steps gives us all 30 eigenvalues.

The only time consuming step in the iteration procedure is the computation of $A^*A\,\Delta_i$. The computation of $A\,\Delta_i$ is just as expensive as one model integration. The subsequent computation of $A^*(A\,\Delta_i)$ is roughly $2\times$ as expensive as the one model integration, because differentiation doubles the number of quadratic terms. This makes one iteration step $3\times$ as expensive as a single model run. It follows that the explicit computation of A, which requires 30 model integrations, is as expensive as 10 iteration steps. Hence, the computation of the matrices $E_{10}, E_{11}, ..., E_{30}$ is no longer an efficient strategy for the computation of the largest eigenvalues. If such a large number of eigenvalues is needed, it is cheaper to compute A explicitly and to compute the full set of 30 eigenvalues.

The method using the adjoint may be applied to a model with many degrees of freedom. In the remaining sections, we explain how the growth rates can be used to compute the probability distribution for the error and why a limited number of iterations should also be sufficient in a realistic model.

4. Probability distribution of errors

In this section, we use the above iterative procedure to estimate the largest eigenvalues and demonstrate how these eigenvalues can be used to establish an estimate of temporal predictability. We take an integration time of 4 days rather than 2 days as suggested in Section 1, because the error growth in our model is slower than the error growth in the atmosphere. In our model, errors grow with a factor of 2.5 in 4 days.

To determine the temporal probability distribution of the errors, we use a Monte Carlo method. We integrate an ensemble of 40,000 random errors with the tangent linear equations. All errors initially have length one. All errors are added to the same starting point on the attractor. After 4 days, their length R is given by:

$$R = (u(t=4), u(t=4))^{1/2}. \tag{30}$$

We consider the probability distribution of this radius to be a perfect measure of the temporal predictability. We will now show how this distribution can be approximated using the largest eigenvalues of A^*A.

The matrix $E_5^{\mathrm{T}}E_5$ is computed in order to estimate the 5 dominant eigenvalues. In the following paragraphs, we first give our method to derive an error distribution function on the basis of a number of eigenvalues. We then apply our method to the example with the 5 eigenvalues. Comparison with the Monte Carlo result indicates that we need to do 3 more iteration steps.

First, we need a reference distribution function f_{ref} for the square root of the fraction of the kinetic energy of a unit-vector with random direction which projects onto a coordinate axis. This distribution is identical for every possible coordinate axis. We perform a Monte Carlo experiment to obtain this distribution function. We can now use f_{ref} for the projection of an initial error vector of unit length on vector v_i. The distribution function f_i for the square root of the error energy, which after an integration of 4 days, projects onto the coordinate axis spanned by Av_i is:

$$f_i(r) = f_{\mathrm{ref}}\left(\frac{r}{\lambda_i}\right)\frac{1}{\lambda_i}. \tag{31}$$

In other words, the reference distribution function f_{ref} is stretched by a factor λ_i. Consequently, it must also be divided by λ_i, as the integrated probability must remain 1. Taking successive convolutions of the individual functions f_i with $i = 1, 2, ..., j$ we get the distribution function for the error:

$$f_{1,\Sigma}(r) = f_1(r), \tag{32}$$

$$f_{i+1,\Sigma}(r) = \int_{\sigma=0}^{r} f_{i,\Sigma}((r^2-\sigma^2)^{1/2})\, f_{i+1}(\sigma) \times \left(\frac{r^2}{r^2-\sigma^2}\right)^{1/2} \mathrm{d}\sigma, \tag{33}$$

$$f(R=r) \approx f_{30,\Sigma}(r) \approx f_{j,\Sigma}(r). \tag{34}$$

The functions $f_{i,\Sigma}$ give the distribution for the contribution of the first i eigenvalues to the error. Thus, eq. (32) is satisfied by definition. We compute $f_{i+1,\Sigma}$ with the approximation that the distribution of the energy on axis Av_{i+1} is independent of the distribution of the energy on the

axes $\{Av_1, Av_2, ..., Av_i\}$. This approximation allows us to use the product of $f_{i,\Sigma}$ and f_{i+1} in eq. (33). The integral is over all possible states with energy r^2 on the first $i+1$ axes. Differentiation of the argument of $f_{i,\Sigma}$ gives the geometry factor in eq. (33):

$$\frac{\mathrm{d}}{\mathrm{d}r}(r^2-\sigma^2)^{1/2}=\left(\frac{r^2}{r^2-\sigma^2}\right)^{1/2}. \tag{35}$$

The real error distribution is given by $f(R)$. The first approximation sign in eq. (34) is due to the assumed independence of the functions f_i. This approximation is reasonable for the first few convolutions because we use a rather large number of 30 dimensions. At the last possible convolution to obtain $f_{30,\Sigma}$ from $f_{29,\Sigma}$ and f_{30}, the approximation would break down completely. However, because λ_{30}^2 is negligibly small, we do not need to perform it; in fact it will appear that we do not need to go beyond the convolution of $f_{7,\Sigma}$ and f_8. As long as j is much smaller than 30, the most significant error is the one indicated by the 2nd approximation sign, which is caused by taking only j distribution functions.

Fig. 3a compares the distribution $f_{5,\Sigma}$ which is based on 5 eigenvalues with the Monte Carlo result. The tail of the distribution is estimated accurately. At lower lengths, there is a systematic error. Here, the error growth is significantly underestimated because we assumed 25 eigenvalues to be zero. One might correct for this statistically. An alternative is to perform a few more iterations. Fig. 3b compares the distribution $f_{8,\Sigma}$ which is based on 8 eigenvalues with the Monte Carlo result. We conclude that we are able to reproduce the temporal distribution function for the forecast error reasonably accurately on the basis of 8 eigenvalues. This does not mean that we can predict whether one particular forecast will have a large error or not. Indeed, from the Monte Carlo result, we observe that the error in the forecast depends heavily on the stochastic direction of the initial perturbation. It does mean that we can say how a large ensemble of equally likely initial perturbations will spread out in space.

To analyse the variability of the temporal predictability over the attractor, we compute a series of error distribution functions for 4-day integrations starting at 100 consecutive days. Thus consecutive integrations have 3 days in common. For every distribution, we compute the probabilities $P(R<2)$, $P(R<3)$ and $P(R<4)$.

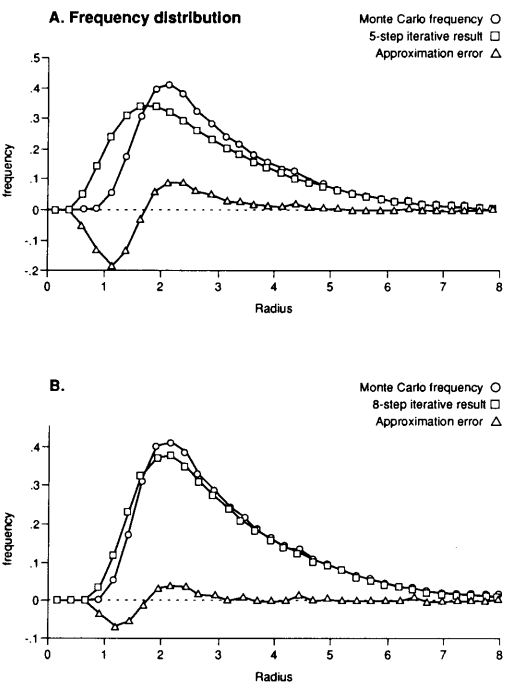

Fig. 3. (a) Comparison of the error size distribution obtained from 40,000 Monte Carlo integrations and the profile based on E_5. The iteration result is marked with squares; the Monte Carlo curve with circles; the difference between the two curves with triangles. (b) Comparison of the error size distribution with the profile based on E_8 instead of E_5. Otherwise it is the same as (a).

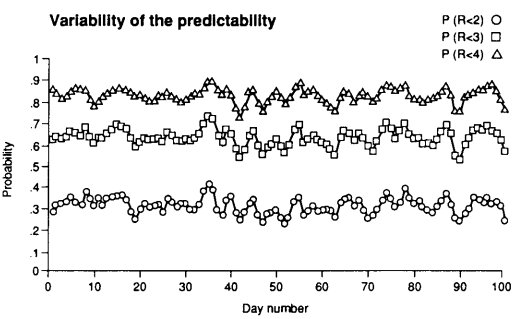

Fig. 4. Variations in time of the probabilities $P(R<2)$, $P(R<3)$ and $P(R<4)$ during a 100-day integration. $P(R<2)$ is indicated with circles; $P(R<3)$ with squares; $P(R<4)$ with triangles.

The first one gives the probability of a high-quality forecast. The second one gives the probability that a forecast is at least of moderate quality. $1-P(R<4)$ is the probability of a low-quality forecast. The 3 probabilities are plotted in Fig. 4. We observe that the probability that a forecast is of low quality may vary between 10 and 25%. We conclude that there is a considerable variation in the predictability as a function of the position on the attractor.

5. Dimensionality of the error growth

In this section, we compare the dimensionality derived from the error growth analysis with the dimensionality which can be derived from an EOF analysis. EOF analysis is a well-known technique (see, e.g., Preisendorfer, 1988) that is often applied to meteorological data. We will first apply it to a 30-year model data set to estimate the dimension of the model attractor. We will then see to what extent this information on the dimension of the model can be used to estimate the dimension of the error growth. The assumption is that the dimension of error growth will be of the order of the dimension of the model attractor. If this appears to be true, we can estimate the dimension of error growth in the real atmosphere with the help of an EOF analysis of the atmospheric circulation. This then gives us information on the number of eigenvalues of A^*A that have to be computed in a realistic atmospheric model with many degrees of freedom. The argument for making the assumption on the dimension of error growth is that a model with a perturbed initial condition will rapidly converge to the model attractor in the first few hours of the integration. Thus, after some time, both the perturbed and the reference run can be described with the same EOF basis. As time progresses, the EOF basis will then also become an appropriate basis for the errors.

We compute the 30×30 symmetric scatter matrix F defined by:

$$F=\sum_{t=1}^{n} q(t)\, q^{\mathrm{T}}(t). \tag{36}$$

Here the vector $q(t)$ is the state vector at time t. It is defined by eqs. (14), (15), (16) and (17). The summation is over a 30-year run. Successive values of t are 1 day apart. The matrix F has a set of 30 orthonormal eigenvectors (i.e., orthonormal with respect to the kinetic energy inner product). These eigenvectors are the "Empirical Orthogonal Functions (EOFs)". We put the eigenvalues l_i of F in descending order and scale them such that their sum equals 100:

$$l_1 > l_2 > \cdots > l_{30} \geqslant 0,$$

$$\sum_{i=1}^{30} l_i = 100.$$

The individual l_i give the % of the variance described by their corresponding eigenvector. If the l_i decrease sharply towards 0, as a function of index number i, then we need only a few EOFs to describe most of the variance of the system. We then call the system low-dimensional. We plotted the logarithms of the eigenvalues l_i against index number in Fig. 5. With the eigenvectors 1, 2, ..., 9, we explain 97% of the variance. To explain the entire 100% of the variance, we need the maximum number of 30 EOFs.

We can do something similar for the error variance with the eigenvalues of A^*A. The ith eigenvalue λ_i^2 gives the average error energy projecting on Av_i after 4 days of integration. For a total of 100 different initial positions on the attractor, we compute $\lambda_1^2, \lambda_2^2, \ldots, \lambda_{30}^2$. After averaging

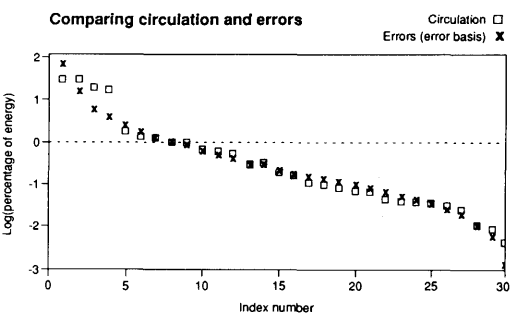

Fig. 5. Logarithmic eigenvalue graph for the instantaneous circulation and for the errors. The squares give the variance of the instantaneous circulation described with the EOFs. The crosses give the variance of the errors described with the basis-vectors for the errors.

and scaling, we get a set of numbers m_i comparable to the l_i for the EOF analysis:

$$m_i = 100 \frac{\overline{\lambda_i^2}}{\sum_{j=1}^{30} \overline{\lambda_j^2}}$$

$$m_1 > m_2 > \cdots > m_{30} \geqslant 0;$$

$$\sum_{i=1}^{30} m_i = 100.$$

The logarithms of the m_i are shown in Fig. 5. The vectors 1, 2, ..., 9 again explain 97% of the variance of the errors.

Comparing the eigenvalues of the scatter matrix with the eigenvalues of A^*A, we observe that the errors project a higher % of their energy on the 1st vector. This difference is compensated with lower %s at the next 3 vectors. The remaining 26 eigenvalues have comparable %s. We conclude that the dimensionality of the error vectors is only slightly lower than the dimensionality of the instantaneous circulation patterns. Thus, we have now validated our a priori assumption for our 30-dimensional model.

To verify the argument for our assumption, we take the same set of 100 initial positions on the attractor and from each point we do 100 integrations with a random initial error. According to our argument, the EOF basis will become an appropriate basis for these errors at time progresses. In Fig. 6 we project errors after 4 days of integration on the EOF basis. As a reference, we have copied the information on the logarithms of the l_i from the previous figure. It is clear that the EOF basis is more efficient for the description of the state vectors than for the error vectors. However, considering the random initial error direction, we observe that the tangent linear equations are quite efficient in pushing a perturbed initial point towards the attractor. Combining this with the fact that the basis of A^*A is optimum for the errors, we conclude that in general, we may expect the dimensionality of the errors to be comparable to the dimensionality of instantaneous circulation patterns. Thus, we are now ready to apply the same reasoning to a GCM.

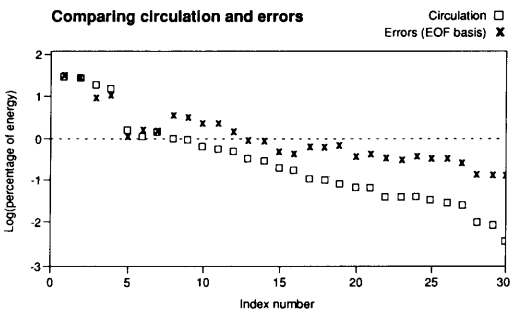

Fig. 6. Logarithmic eigenvalue graph for the instantaneous circulation and for the errors. The squares give the variance of the instantaneous circulation described with the EOFs. The crosses give the variance of the errors described with the EOFs.

Rinne and Karhila (1979) and Rinne and Järvenoja (1979) give EOFs of the 500 mb instantaneous height field in the northern hemisphere based on a large data sample. They show that 25 EOFs suffice to describe 86% of the variance. Doubling the number of EOFs to 50 increases the described variance to 94%. We now assume that a 2-day error projects as efficiently on the basis of EOFs as an instantaneous field. This means that a 50-dimensional EOF basis suffices to describe 94% of the mean square error field. As we have just seen, this assumption is optimistic. We define a projection matrix D to project the final error field on the EOF basis and get for the truncated mean square error R_D^2:

$$R_D^2 = (DAu(t=0), DAu(t=0))$$
$$= (A^*D^*DAu(t=0), u(t=0)). \quad (37)$$

This equation is almost identical to eq. (19). The matrix D^*D has 50 eigenvalues with value 1.0. Its other eigenvalues are zero. The matrix A^*D^*DA has 50 non-zero eigenvalues. We can find all of them with 51 iterations. Thus, if we are satisfied with 94% accuracy we need to do 51 iterations. Having noticed this, we can remove the projection D, apply the iterative procedure, which determines its own optimum basis for the errors, and estimate the eigenvalues of the complete system. Using a more relevant basis, we expect to need less than 51 iterations to arrive at the same % of described variance. This may compensate for the optimistic assumption at the beginning, as it did in our 30-dimensional model. The total computational cost of 51 iterations is roughly the same as the cost of 153 model runs.

The above numbers can be improved if we are interested in geographically localized rather than global predictability. Peagle and Haslam (1982) give an EOF analysis of 7 years of winter data over a portion of the northern hemisphere centered over the western United States. They find that only 18 EOFs are needed to describe 94% of the variance. This reduces the dimension and thus also the number of iterations by a factor of 3.

We started our analyses with a homogeneous random initial error field. This simplified the exposition of the problem, but is not realistic. We can describe the errors in the observations by means of a vector w such that:

$$u(0) = Mw, \tag{38}$$

$$\overline{(w_i, w_j)} = \frac{\delta_{ij}}{N_{\text{obs}}}, \tag{39}$$

$$(w, w) = 1. \tag{40}$$

Eq. (38) relates the vector w of observation errors with an initial error field $u(0)$. If the measurement errors are small, M is a linear operator. In general, because of the complexity of assimilation methods, the operator M may be cumbersome to evaluate. Eq. (39) states that all N_{obs} observations are independent. Eq. (40) assigns unit length to the error vector. The adapted equation for the error energy at time t now becomes:

$$R^2_{\text{DM}} = (DAMw, DAMw)$$
$$= (M^*A^*D^*DAMw, w). \tag{41}$$

Because the dimension of our problem is already constrained by the matrix D^*D we can make our initial error basis as large as the basis of $u(0)$ without causing significant computational problems. In practice, though, the number of observations may be much smaller than the number of model variables. Thus, the model matrix A, the projection matrix D, the assimilation matrix M and the maximization procedure in general all have the effect of rapidly eliminating irrelevant initialization directions.

6. Discussion and conclusions

In this paper, we have studied variations in the deterministic error growth in a low-order spectral model. We studied short-range error growth only, which allowed us to use the tangent linear equations for evolution of the errors. With these equations and their adjoints, we computed the directions with large error growth. From the growth rates in these directions, we could establish a probability distribution for the errors. From this distribution, we obtained parameters for the skill of a forecast.

It would be of value to know whether the above method can be applied to a large model. We made the hypothesis that the dimension of the error growth is of the order of the dimension of the instantaneous circulation patterns. We have tested the hypothesis in our low-order model, where indeed the number of relevant error growth directions happened to be almost identical to the number of relevant EOFs. This is a plausible result because a run with small initial error will converge to the model attractor. From an analysis of the real atmospheric circulation, we obtained a number of 50 relevant EOFs. According to the hypothesis, this is equivalent to 50 relevant error growth directions. Computing these directions is as expensive as about 150 model runs. With these directions and the corresponding growth rates, one can get parameters for the skill.

The above methods may still be judged prohibitively expensive. A reduction in the cost may be obtained by choosing a regional instead of a global error norm and by relaxed requirements on the numerical precision of the skill forecast. All of this depends on the yet-to-be-determined spectrum of eigenvalues which might be different from its hypothesized shape. To get a useful estimate of the skill, one should use an inhomogeneous initial error field which is consistent with errors in the observations. Formally, this can be done with the operator M, which transforms a vector with measurement errors into an initial error field. Nonlinear error growth is not discussed in this paper; it limits the applicability of the above method to short-range predictions.

7. Acknowledgements

We wish to thank Professor J. Grasman and Dr. J. D. Opsteegh for many stimulating discussions during the course of this work. We thank Drs. F.

Selten for his help with the EOF analysis and Dr. J. Barkmeijer for his help with the derivation of the adjoint equations. We thank Dr. R. Pasmanter for his careful reading of the manuscript. This investigation is supported by the working group on Meteorology and Physical Oceanography (MFO) with financial aid from the Netherlands Organization for the advancement of Research (NWO). Computing facilities are provided by the Royal Dutch Meteorological Institute.

8. Appendix

List of symbols

a	radius of the earth
A	integration operator for the tangent linear equations
C	Ekman coefficient
D	matrix for the projection on an EOF-basis
D_{KY}	Kaplan-Yorke dimension
e_i	basis vector for rescaled vectors
E_j	matrix which contains the most relevant information of A
F	scatter matrix
f	Coriolis parameter
$f(R)$	distribution function for the total length of an error vector
f_0	Coriolis parameter at 45° latitude
f_{ref}	a reference distribution function
f_i	distribution function for the length of Av_i
$f_{i,\Sigma}$	cumulative distribution function for the error
h	mountain height
I	identity operator
J	Jacobian operator
K	kinetic energy
l_i	scaled eigenvalue of F
L	linear operator used to describe the tangent linear equations
m_i	scaled eigenvalue of A^*A
M	operator to couple observating errors to initial errors
N	the number of model variables
N_{obs}	the number of observations
$P_{m,n}$	associated Legendre function
q	rescaled state vector
Q	cooling coefficient
r	partial length of a vector
R	length of a vector
R_D	length of a vector in a truncated system
R_{DM}	length of a vector in a truncated system with data assimilation
S	linear operator used to describe the tangent linear equations
t	time
u	error vector in rescaled coordinates
v_j	eigenvector of A^*A
w	vector with observation errors
$Y_{m,n}$	eigenfunction of the Laplace operator
α_j	Lyapunov exponent
Δ	Laplacian operator
Δ_j	approximates the eigenvector v_j
Δp	a pressure difference of 500 mb
λ	geographic longitude
λ_k	singular value of A
Λ	the inverse of the radius of deformation
μ	geographic latitude
σ	the static stability parameter
τ	250–750 mb thickness
$\tilde{\tau}$	rescaled 250–750 mb thickness
τ^*	enforced 250–750 mb thickness
ψ	interpolated 500-mb streamfunction
$\tilde{\psi}$	rescaled 500-mb streamfunction
ψ_1	250-mb streamfunction
ψ_3	750-mb streamfunction
ψ_3^*	enforced 750-mb streamfunction
ω_2	vertical wind at 500 mb

REFERENCES

Bourke, W. B., Mc Avaney, B., Puri, K. and Thurling, R. 1977. Global modeling of atmospheric flow by spectral methods. *Methods in Computational Physics 17*, 267–334.

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

Epstein, E. S. 1969. Stochastic dynamic prediction. *Tellus 21*, 739–759.

Holton, J. R. 1979. *An introduction to dynamic meteorology* (ed. W. L. Donn). New York, Academic Press, 391 pp.

Kaplan, J. L. and Yorke, J. A. 1979. Chaotic behavior of multidimensional difference equations. In: *Functional differential equations and approximation of fixed*

points, (eds. H. O. Peitgen and H. O. Walther). Heidelberg-New York, Springer, 228–237.

Lacarra, J. and Talagrand, O. 1988. Short-range evolution of small perturbations in a barotropic model. *Tellus 40A*, 81–95.

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

Leith, C. E. 1974. Theoretical Skill of Monte Carlo Forecasts. *Mon. Wea. Rev. 102*, 409–418.

Lorenz, E. N. 1960. Energy and numerical weather prediction. *Tellus 12*, 364–373.

Lorenz, E. N. 1963. Deterministic Nonperiodic Flow. *J. Atmos. Sci. 20*, 130.

Lorenz, E. N. 1985. The growth of errors in prediction. In: *Turbulence and predictability in geophysical fluid dynamics and climate dynamics* (eds. M. Ghil, R. Benzi and G. Parisi). Amsterdam, North-Holland, 243–265.

Orszag, S. A. 1970. Transform method for the calculation of vector-coupled sums: Application to the spectral form of the vorticity equation. *J. Atmos. Sci. 27*, 890–895.

Palmer, T. N. and Tibaldi, S. 1988. On the prediction of forecast skill. *Mon. Wea. Rev. 116*, 2453–2480.

Paegle, J. N. and Haslam, R. B. 1982. Empirical orthogonal function estimates of local predictability. *J. Appl. Meteor. 20*, 117–226.

Parlett, P. 1980. *The symmetric eigenvalue problem.* Series in computational mathematics, Englewood Cliffs New Jersey, Prentice Hall, 348 pp.

Platzman, G. W. 1960. The spectral form of the vorticity equation. *Journal of Meteorology 17*, 635–644.

Preisendorfer, R. W. 1988. *Principal component analysis in meteorology and oceanography* (ed. C. D. Mobley). Amsterdam, Elsevier, 425 pp.

Reinhold, B. B. and Pierrehumbert, R. T. 1982. Dynamics of weather regimes: Quasi stationary waves and blocking. *Mon. Wea. Rev. 110*, 1106–1145.

Rinne, J. and Järvenoja, S. 1979. Truncation of the EOF series representing 500 mb heights. *Quart. J. R. Meteorol. Soc. 105*, 885–897.

Rinne, J. and Karhila, V. 1979. Empirical orthogonal functions of 500 mb height in the northern hemisphere determined from a large data sample. *Quart. J. R. Meteorol. Soc. 105*, 873–884.

Silberman, I. 1954. Planetary waves in the atmosphere. *Journal of Meteorology 11*, 27–34.

Schuster, H. G. 1988. *Deterministic Chaos An Introduction* (ed. W. Greulich). Weinheim, VCH, 270 pp.

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

Thiébaux, H. J. and Morone, L. L. 1990. Short-term systematic errors in global forecasts: their estimation and removal. *Tellus 42A*, 209–229.

Wolf, A., Swift, J. B., Swinney, H. L. and Vastano, J. A. 1985. Determining Lyapunov exponents from a time series. *Physica D 16*, 285–317.