IFS DOCUMENTATION

PART V: THE ENSEMBLE PREDICTION SYSTEM (CY25R1)

(Operational implementation 9 April 2002)

Edited by Peter W. White

(Text written and updated by members of the ECMWF Research Department)

Table of contents

Copyright

Part V: THE ENSEMBLE PREDICTION SYSTEM

CHAPTER 1 Theory

Table of contents

1.1 INTRODUCTION

The ensemble prediction system (EPS) is a technique to predict the probability distribution of forecast states, given a probability distribution of random analysis error and model error. More specifically, the operational EPS is a set of (50) integrations of a lower resolution (T_L255) version of the operational model from initial conditions which are constructed by adding small dynamically active perturbations to the operational analysis for the day. Random model errors due to physical processes and subgrid-scale effects are represented by stochastically perturbing the tendencies of parameterized physical processes. In situations where the ensemble is tightly distributed around the operational integration, this forecast can be thought of as very likely to occur. More generally, the EPS gives a probability forecast of some given flow type, or some given category based on temperature, precipitation or other forecast variable.

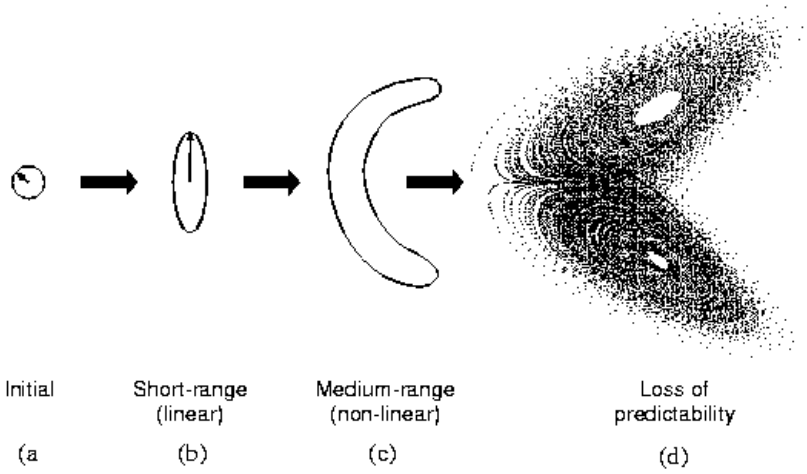

Figure 1.1 A schematic illustration of the growth of an isopleth of the forecast error probability distribution function, from (a) initial phase, to (b) linear growth phase, to (c) nonlinear growth phase, to (d) loss of predictability. See text for further details.

Fig. 1.1 shows a schematic illustration of the phase-space evolution of the probability distribution function (PDF) of analysis error throughout the forecast range. A specific isopleth (e.g. the 1 standard deviation isopleth) is illustrated. It is assumed that at initial time the distribution is normal along each phase-space direction. At initial time (Fig. 1.1 (a)), the isopleth is shown as isotropic, i.e. bounding an n-sphere, where n is the dimension of phase space ($O(10^7)$ for the ECMWF operational forecast model). In general, this error will not be isotropic—analysis error is likely to be larger along directions which are less well observed, and vice versa. However, it is straightforward to define an inner product on phase space, with respect to which the initial probability density function (PDF) is isotropic. This inner product, defined from the analysis error covariance matrix, is fundamental in the theory of singular vectors below.

In the early part of the forecast, error growth is governed by linear dynamics. During this period an initially spherical isopleth of the PDF will evolve to bound an n-dimensional ellipsoidal volume (Fig. 1.1 (b)). The major axis of the ellipsoid corresponds to a phase-space direction which defines the dominant finite-time instability of that part of phase space (relative to the analysis error covariance metric). The arrow shown in Fig. 1.1 (b) points along the major axis of the ellipsoid. It can be thought of as evolving from the arrow shown in Fig. 1.1 (a). Note that the arrows in Figs. 1.1 (a) and (b) are not parallel to one another. This illustrates the non-modal nature of linear perturbation growth.

The arrows at initial and forecast time define the dominant singular vector at initial and final time (with respect to the analysis error covariance metric). At forecast time, the dominant singular vector defines the dominant eigenvector of the forecast error covariance matrix. See Section 1.2.1 for more details.

The growth of the (isopleth of the) PDF between Figs. 1.1 (b) and (c) describes a nonl inear evolution of the PDF. In Figs. 1.1 (c) the PDF has deformed from its ellipsoidal shape in Figs. 1.1 (b). The nonlinear deformation will cause the PDF to evolve away from a normal distribution. Put another way, in the nonlinear phase, the PDF in any give direction is partially determined by perturbations which, in the linear phase, were orthogonal to that direction. Finally, Fig. 1.1 (d) shows (schematically) the situation where the evolved PDF has effectively become indistinguishable from the system's attractor, so that all predictability has been lost.

The number of degrees of freedom of the operational ECMWF model is (very) much larger than the largest practicable ensemble size. This raises the question of whether any particular strategy is desirable in sampling the initial PDF. If initial errors can occur independently in all the phase-space directions, then a strategy of random-under-sampling could lead to an EPS whose reliability was poor, especially for cases of small ensemble spread. In particular, if the spread from a randomly under-sampled ensemble was found to be small on a particular occasion, this could either be because the flow was especially predictable, or because the ensemble perturbations poorly sample the unstable subspace in which the analysis error lay. From a credibility perspective, it is important to try to minimize the latter type of occurrence.

An alternative strategy is to base the perturbations on the singular vectors. Clearly, by focusing on the unstable subspace, the cases of small spread being associated with large forecast error should be minimized, at least in the linear and weakly nonlinear range. In addition to this, there are five related reasons why the initial perturbations for the ECMWF EPS are based on the dominant singular vectors.

Firstly, as shown by *Rabier et al.* (1996), the sensitivity of day-2 forecast error to perturbations in the initial state projects well into the space of dominant singular vectors. *Rabier et al*. have shown that cases of severe forecast failure can be dramatically improved if the analysis is modified using the sensitivity perturbations.

Secondly, provided the metric or inner product for the singular vectors is an accurate reflection of the analysis error covariance matrix then, as mentioned, the evolved singular vectors point along the largest eigenvectors of the forecast error covariance matrix. As such (see *Ehrendorfer and Tribbia*, 1997), perturbations constructed from the dominant singular vectors represent the most efficient means for predicting the forecast error covariance matrix,

given a pre-specified number of allowable tangent model integrations.

Thirdly, singular vector perturbations may provide a relatively efficient means of sampling the forecast error PDF in the nonlinear range, particularly during transitions in weather regimes. For example, *Mureau et al.* (1993) have shown a case where the singular vector perturbations were successful in capturing a major transition to blocking, where random perturbations were inadequate. *Gelaro et al.* (1998) have documented further such cases. A more systematic study of the ability of ensembles to describe the probability of regime transitions in the weakly nonlinear forecast range has been made by *Trevisan et al.* (1998) using an intermediate-complexity model of the extratropical circulation. Relatively small ensembles initialised using firstly singular vectors, and secondly local Lyapunov vectos, were compared with a large Monte-Carlo ensemble based on random perturbations. It was found that small ensemble spread from the singular-vector ensemble was a reliable indicator of small ensemble spread from the Monte-Carlo ensemble. By contrast, small ensemble spread from the Lyapunov-vector ensemble was a much less reliable indicator of small spread from the Monte-Carlo ensemble.

Fourthly, in practice the initial PDF is only poorly known. Hence it is difficult to even define a truly random initial sampling.

Fifthly, from a purely pragmatic point of view, it would seem to be wasteful to integrate explicitly those perturbations that are likely to grow slowly, and thus resemble the control forecast. Such perturbations can be implicitly taken into account in constructing a forecast PDF, by increasing the weight given to the control forecast relative to the perturbed forecasts.

At present, the EPS is based on the notion that forecast uncertainty is dominated by error or uncertainty in the initial conditions. This is consistent with studies that show that, when two operational forecasts differ, it is usually differences in the analyses rather than differences in model formulation that are critical to explaining this difference. In addition, random model errors due to physical parameterizations and the effects of subgrid-scale processes are represented by stochastic perturbations of the physical tendencies of the model, see *Buizza et al.* (1999). The physical tendencies are multiplied with a random number between 0.5 and 1.5. Each ensemble member uses a different realization of this random number. The same perturbation is applied in a tile of 10 degrees latitude and 10 degrees longitude over a period of 6 hours. The ratio of spread associated with initial error and model error can be tuned to the ratio of the influence of initial and model error obtained from studies of divergent operational forecasts.

Each EPS perturbation is a linear combination of the computed singular vectors. This is done so that a given perturbation covers as much of the Northern and Southern hemisphere as possible. The amplitude of the perturbation is then defined after comparison with the statistics of analysis error. These processes are described in Section 2.2.

The EPS was first implemented operationally in 1992 (*Palmer et al.*, 1993). A general description of the ECMWF EPS is described in *Molteni et al.* (1996). The singular vector computations are described in *Buizza and Palmer* (1995), and *Barkmeijer et al.* (1998, 2001).

1.2 SINGULAR VECTORS

1.2.1 Formulation of the singular vector computation

One way to define singular vectors is by means of a maximization problem. The scalar which has to be maximized can be written as:

$$\frac{[\mathbf{EMx},\mathbf{Mx}]}{[\mathbf{Dx},\mathbf{x}]} \tag{1.1}$$

where $[,]$ denotes the Euclidean inner product, $[\mathbf{x}, \mathbf{y}] = \sum_i (x_i y_i)$, and $\mathbf{D}$ and $\mathbf{E}$ are positive definite operators. The operator $\mathbf{M}$ is the propagator of the tangent model. It assigns to a particular vector $\mathbf{x}$ the linearly evolved vector $\mathbf{Mx}$ for a given forecast time and with respect to a reference trajectory. Hence, the scalar defined by (1.1) is the ratio between the $\mathbf{E}$-norm of the evolved vector $\mathbf{x}$ and the $\mathbf{D}$-norm of $\mathbf{x}$ at initial time. Notice that the norm at initial and final time may differ. The leading singular vector has the property that it maximizes the scalar, the second singular vector maxizes the scalar in the space $\mathbf{D}$-orthogonal to the leading singular vector, and so forth. In this way, one obtains a set of singular vectors which are $\mathbf{D}$-orthogonal at initial time and $\mathbf{E}$-orthogonal at final time. The actual computation of the singular vectors in the IFS is done by solving an equivalent eigenvalue problem. Observe that the solutions of the maximization problem (1.1) also satisfy the following generalized eigenvalue problem (1.2).

$$\mathbf{M}^{*}\mathbf{EMx} = \lambda \mathbf{Dx} \tag{1.2}$$

where $\mathbf{M}^{*}$ is the adjoint of $\mathbf{M}$. The defining equation (1.2) can be generalized by activating operators in the singular vector computation, see Section 1.2. It is, for instance, possible to set the state vector to zero outside a prescribed area at optimization time, by using a projection operator $\mathbf{P}$. Consequently, the growth of singular vectors outside the target area is not taken into account in the actual computation. In using this projection operator, the eigenvalue problem (1.2) becomes $\mathbf{P}^{*}\mathbf{M}^{*}\mathbf{EPMx} = \lambda \mathbf{Dx}$. To keep the notation as simple as possible, these additional operators will be left from the basic eigenvalue problem (1.2).

The operator $\mathbf{D}$ determines the properties by which the singular vectors are constrained at initial time. As such, it can be interpreted as an approximation of the inverse of the analysis error covariance matrix $\mathbf{P}^{a}$. Currently, there are two methods to compute singular vectors, depending on the form of $\mathbf{D}$. Both methods will be discussed in Sections 1.2.2 and 1.2.3.

1.2.2 Use of the total energy norm

When using the total energy norm, or any other simple operator, at initial time, the generalized eigenvalue problem (1.2) can be simplified to an ordinary eigenvalue problem. In this case the $\mathbf{D}$-norm of $\mathbf{x}$ reads as

$$\begin{aligned}[\mathbf{x}, \mathbf{Dx}] = \frac{1}{2}\int_0^1 \int_\Sigma \left\{ \Delta^{-1}\zeta_{\mathbf{x}}.\zeta_{\mathbf{x}} + \Delta^{-1}D_{\mathbf{x}}.D_{\mathbf{x}} + \frac{c_p}{T_{\mathrm{ref}}}T_{\mathbf{x}}^2 + w_q \frac{L_{\mathrm{cond}}^2}{c_p T_{\mathrm{ref}}} q_{\mathbf{x}}^2 \right\} \mathrm{d}\Sigma \, \frac{\partial p}{\partial \eta} \, \mathrm{d}\eta \\ + \frac{1}{2}\int_\Sigma R_{\mathrm{dry}} T_{\mathrm{ref}} P_{\mathrm{ref}} \ln \pi^2 \, d\Sigma\end{aligned} \tag{1.3}$$

where $\zeta_{\mathbf{x}}$, $D_{\mathbf{x}}$, $T_{\mathbf{x}}$ and $\ln \pi_{\mathbf{x}}$ stands for the vorticity, divergence, temperature, specific humidity and logarithm of the surface pressure component of the state vector $\mathbf{x}$, and c_p is the specific heat of dry air at constant pressure, L_{cond} is the latent heat of condensation at $0°\mathrm{C}$, R_{dry} is the gas constant for dry air, $T_{\mathrm{ref}} = 300\,K$ is a reference temperature and P_{ref} = 800 hPa is a reference pressure. The parameter w_q defines the relative weight given to the specific humidity term.

Since the operator $\mathbf{D}$ is a diagonal matrix, one can easily define a matrix $\mathbf{C}$ so that $\mathbf{C}^2 = \mathbf{D}^{-1}$. Multiplying both sides of (1.2) to the left and right with $\mathbf{C}$, yields the equation

$$\mathbf{CM}^{*}\mathbf{EMCx} = \lambda \mathbf{X} \tag{1.4}$$

which can be solved using the Lanczos algorithm, see Appendix Section A.1. The energy metric is believed to be a first-order approximation to the analysis error covariance metric (*Palmer et al*. 1998).

1.2.3 Use of the Hessian of the 3D-Var objective function

In the incremental formulation of 3D-Var, the Hessian of the objective function $\Im$ can be used as an approximation of the inverse of the analysis error covariance matrix. The objective function has the form

$$\Im(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d}) \tag{1.5}$$

and the increment $\delta\mathbf{x}^{\mathrm{a}}$ where $\Im$ attains its minimum, provides the analysis $\mathbf{x}^{\mathrm{a}}$ which is defined by adding $\delta\mathbf{x}^{\mathrm{a}}$ to the background $\mathbf{x}^{\mathrm{b}}$

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}^{\mathrm{a}} \tag{1.6}$$

The operators $\mathbf{B}$ and $\mathbf{R}$ are covariance matrices of the background and observation error respectively and $\mathbf{d}$ is the innovation vector

$$\mathbf{d} = \mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}} \tag{1.7}$$

where $\mathbf{y}^{\mathrm{o}}$ is the observation vector and $\mathbf{H}$ is a linear approximation of the observation operator in the vicinity of $\mathbf{x}^{\mathrm{b}}$.

The Hessian $\nabla\nabla\Im$ of the objective function is given by

$$\nabla\nabla\Im = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} \tag{1.8}$$

Provided that the background error $(\mathbf{x}^{\mathrm{b}}-\mathbf{x}^{\mathrm{t}})$ and the observation error $(\mathbf{y}^{\mathrm{o}}-\mathbf{x}^{\mathrm{t}})$ are uncorrelated, with $\mathbf{x}^{\mathrm{t}}$ the true state of the atmosphere, the Hessian $\nabla\nabla\Im$ is equal to the inverse of the analysis error covariance matrix $\mathbf{P}^{\mathrm{a}}$. This follows by noting that the objective function is quadratic and attains its unique minimum at $\delta\mathbf{x}^{a}$ and consequently, by using (1.6),

$$\mathbf{B}^{-1}[\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{t}}-(\mathbf{x}^{\mathrm{b}}-\mathbf{x}^{\mathrm{t}})] + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{H}(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{t}}) + \mathbf{H}\mathbf{x}^{\mathrm{t}} - \mathbf{y}] = 0 \tag{1.9}$$

Rewriting (1.9) gives

$$(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{t}}) = \mathbf{B}^{-1}(\mathbf{x}^{\mathrm{b}}-\mathbf{x}^{\mathrm{t}}) - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^{\mathrm{t}}-\mathbf{y}) \tag{1.10}$$

Using the assumption that the background and observation error are uncorrelated the above equation implies that

$$(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})\mathbf{P}^{\mathrm{a}}(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{\mathrm{T}} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{\mathrm{T}} \tag{1.11}$$

By now multiplying each side of (1.11) to the right with its transpose, the desired result follows.

The defining eigenvalue problem for the singular vectors becomes

$$\mathbf{M}^{*}\mathbf{E}\mathbf{M}\mathbf{x} = \lambda\nabla\nabla\Im\mathbf{x} \tag{1.12}$$

Since the objective function is quadratic in the incremental formulation, the Hessian $\nabla\nabla\mathfrak{I}\mathbf{x}$ in (1.12) can be evaluated by computing the difference between two gradients: $\nabla\nabla\mathfrak{I}\mathbf{x} = \nabla\mathfrak{I}(\mathbf{x}+\mathbf{x}^b) - \nabla\mathfrak{I}\mathbf{x}^b$. The generalized eigenvalue problem (1.12) is solved by using the Jacobi–Davidson algorithm, see Appendix Section A.2.

APPENDIX A EIGENVALUE ALGORITHMS

A.1 THE LANCZOS ALGORITHM

Algorithms based on Lanczos theory are very useful to solve an eigenvalue problem when only a few of the extreme eigenvectors are needed. It can be applied to large and sparse problems. The algorithm does not access directly the matrix elements of the operator that defines the problem, but it gives an estimate of the eigenvectors through successive application of the operator.

Consider the eigenvalue problem

$$\mathbf{A}\mathbf{x} = \sigma_i^2\mathbf{x} \tag{A.1}$$

where the matrix $\mathbf{A}$ is $N \times N$ dimensional, and symmetric. Without loss of generality, we can also suppose that it is real.

If $\mathbf{A}$ is a real, symmetric matrix, then there exists an orthogonal real matrix $\mathbf{Q}$ such that

$$\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{A}\mathbf{Q} = \mathbf{D}((\lambda_1), \ldots, \lambda_N), \tag{A.2}$$

where $\mathbf{D}(\lambda_1, \ldots, \lambda_2)$ is a diagonal matrix, and $\mathbf{Q}^{\mathrm{T}}$ denotes the transpose of $\mathbf{Q}$ (Schur decomposition theorem).

The Lanczos algorithm does not directly compute the diagonal matrix $\mathbf{D}$, but it first computes a partial transformation of the matrix $\mathbf{A}$ using a tridiagonal matrix $\mathbf{T}$

$$\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q} = \mathbf{T}, \tag{A.3}$$

with

$$\mathbf{T} = \begin{bmatrix} \alpha_1 & \beta_2 & 0 & . & . & . & . \\ \beta_2 & \alpha_2 & \beta_3 & 0 & . & . & . \\ 0 & \beta_3 & . & . & . & . & . \\ . & 0 & . & . & . & 0 & . \\ . & . & . & . & . & \beta_{J-1} & 0 \\ . & . & . & 0 & \beta_{J-1} & \alpha_{J-1} & \beta_J \\ . & . & . & . & 0 & \beta_J & \alpha_J \end{bmatrix}, \tag{A.4}$$

and with

$$\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_j] \tag{A.5}$$

where the vectors $\mathbf{q}_j$ are column vectors, and where the number of iterations J is much smaller than the dimen-

sionality of the problem, $J \ll N$. Then, the Lanczos algorithm finds the diagonal decomposition of $\mathbf{T}$

$$\mathbf{T} = \mathbf{S}^{\mathrm{T}}\mathbf{D}\mathbf{S}. \tag{A.6}$$

The elements of the diagonal matrix $\mathbf{D}$ are an estimate of the eigenvalues of $\mathbf{A}$, and an estimate of the eigenvectors are given by $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_j]$, with

$$\mathbf{Y} = \mathbf{Q}\mathbf{S}, \tag{A.7}$$

The actual computation is performed by writing Eq. (A.3) as

$$\mathbf{A}\mathbf{Q} = \mathbf{Q}\mathbf{T}. \tag{A.8}$$

Equating columns of Eq. (A.8), it follows that

$$\mathbf{A}\mathbf{q}_j = \beta_{j-1}\mathbf{q}_{j-1} + \alpha_j\mathbf{q}_j + \beta_j\mathbf{q}_{j+1} \tag{A.9}$$

for $j = 1, J$. The orthogonality of the vectors $\mathbf{q}_j$ implies that

$$\alpha_j = \mathbf{q}_j^{\mathrm{T}}\mathbf{A}\mathbf{q}_j. \tag{A.10}$$

Moreover, if

$$\mathbf{r}_j = (\mathbf{A} - \alpha_j\mathbf{I})\mathbf{q}_j - \beta_{j-1}\mathbf{q}_{j-1} \tag{A.11}$$

is non-zero, then

$$\mathbf{q}_{j+1} = \frac{\mathbf{r}_j}{\beta_j}, \tag{A.12}$$

where $\beta_j = \pm\sqrt{\langle \mathbf{r}_j;\mathbf{r}_j \rangle}$. An iterative application of these equations, with a randomly chosen starting vector $\mathbf{q}_1$, defines the Lanczos iterative procedure. The total number of iterations J determines the accuracy of the computation. As this number increases, more eigenvalues/eigenvectors can be separated from the others, independently from the choice of the starting vector $\mathbf{q}^{\mathrm{T}}$. This separation starts from the boundaries of the eigenvalue spectrum. The accuracy of the eigenvectors is less than the accuracy of the singular values, say to order ε when the precision of the singular values is of order ε^2.

The reader is referred to *Golub and van Loan* (1983) for a theoretical description of the Lanczos algorithm. The Lanczos code is available in NAG issue 17.

A.2 THE JACOBI–DAVIDSON ALGORITHM

The generalized eigenproblem

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x} \tag{A.13}$$

is usually handled by bringing it back to a standard eigenproblem

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{A.14}$$

The matrix $\mathbf{B}^{-1}\mathbf{A}$ is in general nonsymmetric, even if both $\mathbf{A}$ and $\mathbf{B}$ are symmetric. However, if $\mathbf{B}$ is symmetric and positive definite, the $\mathbf{B}$-inner product is well defined. The matrix $\mathbf{B}^{-1}\mathbf{A}$ is symmetric in this inner product if $\mathbf{A}$ is symmetric:

$$[\mathbf{w}, \mathbf{B}^{-1}\mathbf{A}\mathbf{v}]_{\mathbf{B}} = [\mathbf{B}\mathbf{w}, \mathbf{B}^{-1}\mathbf{A}\mathbf{v}] = [\mathbf{w}, \mathbf{A}\mathbf{v}] = [\mathbf{A}\mathbf{w}, \mathbf{v}] = [\mathbf{B}^{-1}\mathbf{A}\mathbf{w}, \mathbf{v}]_{\mathbf{B}}. \tag{A.15}$$

The proposed method to solve (A.14) constructs a set of basis vectors $\mathbf{V}$ of a search space ν, c.f. the Lanczos method. The approximate eigenvectors are linear combinations of the vectors $\mathbf{V}$. The classical and most natural choice for the search space ν, for instance utilized in the Lanczos method, is the so-called Krylov subspace, the space spanned by the vectors

$$\mathbf{v}, \mathbf{B}^{-1}\mathbf{A}\mathbf{v}, (\mathbf{B}^{-1}\mathbf{A})^2\mathbf{v}, \ldots (\mathbf{B}^{-1}\mathbf{A})^{i-1}\mathbf{v} \tag{A.16}$$

This i-dimensional subspace is denoted by $\mathbf{K}^i(\mathbf{v}, \mathbf{B}^{-1}\mathbf{A})$. The vector $\mathbf{v}$ is a starting vector that has to be chosen. The Krylov subspace is well suited for computing dominant eigenpairs since the vector $(\mathbf{B}^{-1}\mathbf{A})^{i-1}\mathbf{v}$ points more and more in the direction of $\mathbf{B}^{-1}\mathbf{A}$ the dominant eigenvector of for increasing i.

Given a search space ν, the approximate eigenpair $(\theta, \mathbf{u})$ of (A.15) is a linear combination of the basis vectors of ν:

$$\mathbf{u} = \mathbf{V}\mathbf{y}. \tag{A.17}$$

A suitable criterion for finding an optimal pair is the Galerkin condition that the residual

$$\mathbf{r} = \mathbf{B}^{-1}\mathbf{A}\mathbf{u} - \theta\mathbf{u} = \mathbf{B}^{-1}\mathbf{A}\mathbf{V}\mathbf{y} - \theta\mathbf{V}\mathbf{y} \tag{A.18}$$

is $\mathbf{B}$-orthogonal to the search space to ν. Hence

$$\mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{r} = 0. \tag{A.19}$$

and consequently, using (A.19),

$$\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V}\mathbf{y} - \theta\mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{V}\mathbf{y} = 0. \tag{A.20}$$

Note that the resulting eigenproblem is of the dimension of the search space, which is generally much smaller than of the original problem. The basis vectors are usually orthogonalized so that $\mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{V} = 1$. Approximate eigenpairs that adhere to the Galerkin condition are called Ritz pairs.

It can be shown that the residuals $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_i$ form a $\mathbf{B}$-orthogonal basis for $\mathbf{K}^i(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A})$ when the approximate eigenpairs are computed according to (1.37).

A.3 SOLUTION METHOD

As was stated before, the natural search space for the generalized eigenvalue problem is the Krylov subspace $\mathbf{K}^i(\mathbf{v}, \mathbf{B}^{-1}\mathbf{A})$. This basis can be generated by expanding the basis by new residual vectors. The problem in the

construction of a basis for this space is that operations with $\mathbf{B}^{-1}$ are needed. Since in our application the inverse of $\mathbf{B}$ is not known explicitly, its action is approximated by the Conjugate Residual method (CR). To compute the vector

$$\mathbf{r} = \mathbf{B}^{-1}\tilde{\mathbf{r}} \tag{A.21}$$

one iteratively solves the system

$$\mathbf{B}\mathbf{r} = \tilde{\mathbf{r}} \tag{A.22}$$

Iterative solution methods require, apart from vector operations, only multiplications with $\mathbf{B}$. The vector $\mathbf{x}$ can in principle be determined to high accuracy. This, however, may require many multiplications with $\mathbf{B}$ and hence may be very expensive. Therefore, the action of $\mathbf{B}^{-1}$ is approximated to low accuracy, by performing only a few steps with an iterative solution method. The number of iterations is controlled by NINNER. The subspace generated in this way is not a Krylov subspace and the basis vectors are not the residuals (A.19) but only approximations to it. As a consequence they are not perfectly $\mathbf{B}$-orthogonal. This has to be done explicitly.

The complete algorithm can be summarized as follows.

- Choose a starting vector $\mathbf{v}$
- Compute $\mathbf{Bv}$, $\mathbf{B}$-normalize $\mathbf{v}$
- Repeat the following 7 steps NITERL times

1) compute $\mathbf{AV}$, $\mathbf{V}^T\mathbf{AV}$
2) Solve small eigenproblem $\mathbf{V}^T\mathbf{AVy} = \theta\mathbf{y}$
3) Select Ritz value θ and $\mathbf{y}$
4) Compute Ritz vector $\mathbf{u} = \mathbf{Vy}$ and residual $\tilde{\mathbf{r}} = \mathbf{AVy} - \theta\mathbf{BVy}$
5) Compute approximately $\mathbf{v} = \mathbf{B}^{-1}\tilde{\mathbf{r}}$ with the CR-method
6) $\mathbf{B}$-orthonormalize new $\mathbf{v}$ against $\mathbf{V}$
7) Expand $\mathbf{V}$ with the resulting vector

The matrix $\mathbf{V}$ contains the basis for the search space, the vector $\mathbf{v}$ contains the new basis vector and $(\mathbf{u}, \theta)$ is an approximate eigenpair. In step 3 the pair with the largest θ is selected because the dominant part of the spectrum is only of interest. However, if the eigenpair approximation reaches a certain accuracy ε (XKAPA in namelist NAMLCZ), i.e., $\tilde{\mathbf{r}} = \mathbf{Au} - \theta\mathbf{Bu} \leq \varepsilon$, a smaller θ is selected. In step 5 a few CR steps are performed to approximate the action of $\mathbf{B}^{-1}$. In step 6 the Modified Gram–Schmidt procedure is used for reasons of numerical stability. For a more detailed descripton of the Jacobi–Davidson algorithm the reader is referred to *Sleijpen and van der Vorst* (1996)

Part V: THE ENSEMBLE PREDICTION SYSTEM

CHAPTER 2 Computational details

Table of contents

2.1 THE SINGULAR VECTOR CODE

The singular vector computation is called from CUN3 in CNT0 or from CUN2 in CUN1, depending on whether the Hessian of the objective function is used or not. General routines related to the singular vector computation can be found in the directory 'sinvect'. The technical routines needed for the Lanczos or Jacobi–Davidson algorithm are contained in the directory 'lanczos' of IFSAUX.

2.1.1 Set-up routines

Details of the set-up routines are given in Table 2.1.

TABLE 2.1 CONSTANTS IN NAMELIST NAMLCZ CONTROLLING THE SINGULAR VECTOR COMPUTATION.

NAME	TYPE	PURPOSE	DEFAULT
LANCZOS	LOGICAL	Activates singular vector computation with Lanczos algorithm (NCONF=601)	TRUE
LJACDAV	LOGICAL	Activates singular vector computations with Jacobi–Davidson algorithm (NCONF=131)	FALSE
LOCNORM	LOGICAL	Switch to localize norm computation in grid space	TRUE
ALAT1	REAL	NW corner of local area (defined by point 1)	30
ALON1	REAL		359.5

TABLE 2.1 CONSTANTS IN NAMELIST NAMLCZ CONTROLLING THE SINGULAR VECTOR COMPUTATION.

NAME	TYPE	PURPOSE	DEFAULT
ALAT3	REAL	SE corner of local area (defined by point 3)	90
ALON3	REAL		0
NLEVMIN	INTEGER	Minimum level of local area	1
NLEVMAX	INTEGER	Maximum level of local area	NLEVG
LSPTRLC0	LOGICAL	Switch to truncate in spectral space at initial (0) time	FALSE
LSPTRLC1	LOGICAL	Switch to truncate in spectral space at final (1) time	FALSE
NWTRMIN0(1)	INTEGER	Spectral coefficients with total wavenumber outside window are set to zero	0
NWTRMAX0(1)	INTEGER		NXMAX
LNEWNORMT0	LOGICAL	Switch to re-define the norm at initial time	FALSE
NEWNORMT0	INTEGER	Re-defines initial norm 1: total energy 2: kinetic energy 3: vorticity squared 4: stream function squared 5: rotational kinetic energy	1
NITERL	INTEGER	Maximum number of Lanczos or Jacob–Davidson inner iterations	70
NINNER	INTEGER	Number of Jacobi–Davidson inner iterations	2
NJDSTOP	INTEGER	Value of NSTOP when LJACDAV=TRUE	144
LEVOLC	LOGICAL	Switch to evolve singular vectors	TRUE
NEIGEVO	INTEGER	Number of singular vectors to evolve	35
NLANTYPE	INTEGER	Determines type of singular vectors 1: energy type norms are used 2 and 3: obsolete 4: same as 1 but now NCONF = 131 5: Hessian is used at initial time 6: for computing eigensystem of the Hessian	1

The routine SULCZ defines the constants as listed in Table 2.1 and determines the configuration of the singular vector computation.

2.1.2 Main routines

The defining equation (1.2) for the singular vectors takes the form

$$\mathrm{OPK}(\mathbf{x}) = \lambda \mathrm{OPM}(\mathbf{x}) \tag{2.1}$$

The routine OPM is the identity in case a simple energy norm is used at initial time, or it is equal to the Hessian of the 3D-Var objective function. In the routine OPK, the propagators of the tangent and adjoint model are evaluated. Depending on the setting of the constants in NAMLCZ, additional operators may be active, such as SPTRLCZTL when LSPTRLC0=TRUE. It truncates the state vector in spectral space by setting all harmonic coefficients with total wavenumber smaller than NWTRMIN0, or larger than NWTRMAX0, to zero at initial time.

2.2 CALCULATION OF INITIAL PERTURBATIONS

For a description of the evolution of the ECMWF Ensemble Prediction System (EPS) the reader is referred to *Molteni et al.* (1996) and *Buizza et al.* (1998, 1999, 2002). As mentioned in the previous Sections, for each initial date, an ECMWF ensemble comprises one 'control' forecast, which is a forecast started from the operational analysis, and $N_{\rm ens}$ perturbed forecasts. The initial conditions for the perturbed integrations are constructed by adding and subtracting to the operational analysis $N_{\rm ens}/2$ orthogonal perturbations defined as linear combinations of extratropical singular vectors. In addition, tropical perturbations are added, which are constructed by Gaussian sampling of the leading diabatic singular vectors; see *Barkmeijer et al.* (2001). The latter perturbations are designed to perturb the tracks of tropical cyclones (*Puri et al.* 2001). At the time of writing (December 2002), $N_{\rm ens} = 50.$.

The methodology used in the Ensemble Prediction System to define these linear combinations in the extratropics is a modification of the procedure described in *Palmer et al.* (1993). Its aim is to create perturbations which cover most of the Northern and Southern Hemisphere, and have an amplitude comparable (in any region) to the estimates of root-mean-square analysis error provided by the ECMWF data assimilation procedure.

The generation of the ensemble initial perturbations is compounded of 5 steps:

- *(i)* retrieval from the ECMWF MARS archive of the unperturbed analysis (mc/sv/inidata);
- *(ii)* retrieval from the ECMWF MARS archive of the estimate of analysis error (mc/sv/getea);
- *(iii)* generation of initial perturbations: selection of extratropical singular vectors, phase-space rotation, and scaling of perturbations to analysis error; Gaussian sampling of diabatic singular vectors in the tropics (mc/sv/rot);
- *(iv)* *g*eneration of the perturbed initial conditions: spectral expansion and vertical interpolation of initial perturbations, and addition of the initial perturbations to the unperturbed analysis (mc/sv/pertinic);
- *(v)* non-linear integrations starting from unperturbed (mc/fc/cf/control) and perturbed initial conditions (mc/fc/pf/nn);

Hereafter a brief description of each step is reported. The singular vectors' resolution is T42L40, and the model's resolution is $T_{\rm L}255$L40, where the subscript $_{\rm L}$ denotes a linear grid; this is the operational configuration at the time of writing (December 2002), see *Buizza et al.* (2002) for a description of the benefits of increasing the spatial resolution for the EPS.

2.3 UNPERTURBED ANALYSIS RETRIEVAL

This task (mc/sv/inidata) retrieves from the ECMWF MARS archive the unperturbed analysis at full resolution

2.4 ANALYSIS ERROR ESTIMATE RETRIEVAL

This task (mc/sv/getea) retrieves the estimate of analysis error from the ECMWF MARS archive . At the time of writing (December 2002), three variables (temperature and the two horizontal wind components) are retrieved on five model levels (30, 34, 39, 44, 49) on a $6^\circ \times 6^\circ$ latitude/longitude grid.

2.5 GENERATION OF INITIAL PERTURBATIONS

This task (mc/sv/rot) generates the initial perturbations.

2.5.1 Extratropical perturbations

The first step is the selection of $N_{ens}/2$ singular vectors among the ones computed (see Section 1.2). The selection (comp_rotmat) proceeds as follows:

(i) The first 4 singular vectors are always selected;

(ii) For each singular vector a localization function is defined in three-dimensional grid-point space, equal to 1 wherever the local energy (per unit mass) of the singular vector field is greater than 1% of its maximum value over the grid and 0 elsewhere;

(iii) An overlap function is defined at each point as the sum of localization functions of the first four singular vectors. In general, the overlap function gives the number of selected singular vectors which 'cover' any grid point;

(iv) Each subsequent singular vector (from the 5th onwards) is examined in turn, and is selected only if more than half of its total energy lies in regions where the current overlap function is less than 4. If this is the case, the localization function for the new singular vector is used to update the overlap function.

Step *(iv)* is repeated until $N_{ens}/2$ singular vectors are selected. The final overlap function gives the number of singular vectors with at least 1% of their maximum local energy (i.e. 10% of their maximum amplitude) at any location.

Once $N_{ens}/2$ singular vectors have been selected, an orthogonal rotation in phase space and a final re-scaling are performed to generate the ensemble perturbations. If $\mathbf{V}$ is the matrix whose columns are the selected singular vectors, $\mathbf{R}$ is an $(N_{ens}/2)\times(N_{ens}/2)$ orthogonal matrix and $\mathbf{D}$ is a diagonal matrix of scaling factors. The matrix $\mathbf{P}$ containing the ensemble initial perturbations is computed by first defining

$$\mathbf{P}' = \mathbf{VR} \tag{2.2}$$

and then:

$$\mathbf{P} = \mathbf{P}'\mathbf{D} \ . \tag{2.3}$$

Let $\mathbf{p}'_i = (u_i', v'_i, T'_i)$ be one of the orthonormal perturbations defined by Eq. (2.2), in terms of the zonal and meridional wind and temperature, respectively. Moreover, let e_u, e_v, e_T be the estimates of root-mean-square analysis error for these variables. The continuous function

$$f_i = \overline{[(u_i'/e_u)^8 + (v_i'/e_v)^8 + (T'_i/e_T)^8]}^{1/8} \ , \tag{2.4}$$

where the overbar represents a mean over the grid-point space, gives an estimate of the maximum local ratio between the perturbation amplitude and the estimated analysis error.

The rotation matrix $\mathbf{R}$ is defined in such a way to minimise the cost function

$$CF = \sum_{i=1}^{N_{ens}} f_i^2 \ . \tag{2.5}$$

Since *CF* is not a simple quadratic function of the independent variables, the minimization cannot be reduced to the solution of a linear problem. Instead, we perform the minimization iteratively by constructing $\mathbf{R}$ as the product of a series of 2×2 elementary rotation matrices.

In practice, the purpose of the phase-space rotation is to generate perturbations which have the same globally-averaged energy as the 'original' singular vectors, but have a smaller local maximum and a more uniform spatial distribution. The iterative algorithm has proved effective in performing this task, despite the fact that the highly nonlinear nature of the cost function may generate more than one minimum in phase space.

Once the rotation has been performed, the perturbations are re-scaled in order to have a realistic local amplitude. The non-null elements of the diagonal matrix **D** in Eq. (2.3) are given by:

$$d_i = \alpha / f_i , \tag{2.6}$$

where α is a constant factor which represents the maximum acceptable ratio between perturbation amplitude and analysis error. On the basis of the comparison of the average divergence of a generic ensemble member from the control forecast (ensemble spread), and of the average divergence of the control forecast from the verifying analysis (control forecast error) the current value α is now $\alpha = 2.0$.

From the first quarter of 1998 onwards, also linearly evolved singular vectors are used in the generation of initial perturbations. First, two sets $\{\mathbf{p_i}\}$ and $\{\mathbf{ep_i}\}$, each consisting of $N_{ens}/2$ perturbations, are determined based on initial singular vectors and linearly evolved singular vectors of two days before respectively. Both sets are constructed using the above described method with $\alpha = 2.0$. The actual extratropical ensemble perturbations $\{\mathbf{x_i}\}$ are defined by adding perturbations from each set:

$$\mathbf{x}_i = \mathbf{p}_i + \mathbf{ep}_i, i = 1,\ldots,\frac{N_{ens}}{2}$$

The perturbations of the initial conditions in the extratropics consist of symmetric plus/minus pairs. Considering the j th initial perturbation, with $j = 1, N_{\mathbf{ens}}/2$, the initial condition of ensemble forecast $(2j-1)$ is constructed by *adding* the j th initial perturbation to the analysis, whilst the initial condition of ensemble forecast $2j$ is constructed by *subtracting* the j th initial perturbation to the analysis.

2.5.2 Tropical perturbations

Initial EPS perturbations in the tropics are included since cycle 24R3 (introduced operationally in January 2002). The perturbations are generated for $M_{\mathrm{tg}} \leq 4$ target areas by Gaussian sampling of $N_{\mathrm{basis}} = 5$ leading diabatic singular vectors for each area. The Caribbean (0˚-25˚N and 100˚-60˚W) is always target area, as is every tropical storm of category larger than 1 between 25˚N and 25˚S. In the event that these criteria produce more than four target areas, the closest areas are merged. Target areas are determined in task (mc/sv/targets). The diabatic singular vectors are computed in the tasks (mc/sv/sv1 ... sv4).

The coefficients for the Gaussian sampling of the diabatic singular vectors are determined in task (mc/sv/rot). The standard deviation β for the Gaussian sampling of the diabatic singular vectors is proportional to the root mean square value of the scaling coefficients d_i, which are determined as described in the previous section:

$$\beta = \left(\sum_{k=1}^{N_{\mathrm{basis}}} d_k{}^2 / N^2{}_{\mathrm{basis}} \right)^{\frac{1}{2}}$$

Then $M_{tg} N_{\mathrm{basis}} N_{\mathrm{ens}}$ independent random numbers $q_{jk}^{(m)}$ are drawn from a normal distribution with standard deviation β and zero mean. To avoid perturbations with very large amplitudes the distribution is truncated at $\pm 3\beta$. Let $\mathbf{v}_k^{(m)}$ denote the kth singular vector of target area m. Then the tropical perturbations for the j th perturbed

forecast are constructed as the linear combination

$$\mathbf{x}_j^{\text{tropics}} = \sum_{m=1}^{M_{tg}} \sum_{k=1}^{N_{\text{basis}}} q_{jk}^{(m)} \ \mathbf{v}_j^{(m)} \ , \ j = 1,\ldots,N_{ens}.$$

2.6 GENERATION OF THE PERTURBED INITIAL CONDITIONS

This task (mc/sv/pertinic) generates the perturbed initial conditions at the resolution of the nonlinear integrations.

2.7 NONLINEAR INTEGRATIONS OF THE CONTROL AND OF THE PERTURBED FORECASTS

This task controls the time integration of the unperturbed (mc/fc/cf/control) and perturbed ensemble members (mc/fc/pf/nn). The seed for the random number generator used to perturb the tendencies is set to a number that is determined from the number of the perturbed forecast (NENSFNB) and the initial date and time of the forecast.

Part V: THE ENSEMBLE PREDICTION SYSTEM

REFERENCES

Barkmeijer J., van Gijzen, M. and Bouttier, F. 1998: Singular vectors and estimates of the anlaysis error covariance metric. *Quart. .J. Roy. Meteor. Soc.*,**124**, 1695-1713.

Barkmeijer J., Buizza, R., Palmer T. N., Puri, K. and Mahfouf, J.-F. 2001: Tropical singular vectors computed with linearized diabatic physics. *Quart. .J. Roy. Meteor. Soc.*,**127**, 685-708.

Buizza, R. and T.N. Palmer, 1995: The singular vector structure of the atmosphere global circulation. *J. Atmos. Sci.*, **52**, 1434-1456.

Buizza, R., Petroliagis, T., Palmer, T. N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A. and Wedi, N., 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*,**124**, 1935-1960.

Buizza, R., Barkmeijer, J., Palmer, T. N., Richardson, D. S.., 1999: Current status and future developments of the ECMWF Ensemble Prediction System. *Meteorol. Appl.*,**6**, 1-14.

Buizza, R., Miller, M. and Palmer T. N. 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. .J. Roy. Meteor. Soc.*,**125**, 2887-2908.

Buizza, R., Richardson, D. S.. and Palmer T. N. 2002: Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *ECMWF Tech. Memo* **389**.

Ehrendorfer, M. and Tribbia, J. J., 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **53**, 286-313.

Gelaro, R., Buizza, R., Palmer, T. N. and Klinker, E., 1998: Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *J. Atmos. Sci.*, **55**, 1012-1037.

Golub, G. H., and van Loan, C. F., 1983: *Matrix computation*. North Oxford Academic Publ. Co. Ltd., pp 476.

Molteni, F., Buizza, R. , Palmer, T. N. and Petroliagis, T., 1996: The ECMWF Ensemble Prediction System: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Mureau, F., Molteni, F. and Palmer, T.M., 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **199**, 299-298.

Palmer, R.N., Gelaro, R., Barkmeijer, J. and Buizza, R., 1998: Singular vectors, metrics and adaptive observations. *J. Atmos. Sci*. **55**, 633–653.

Palmer, T.N., Molteni, F., Mureau, R. Buizza, R., Chapelet, P. and Tribbia, J., 1993: 'Ensemble Prediction'. Proceedings of the ECMWF Seminar on Validation of models over Europe*: Vol 1*, ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Puri, K., Barkmeijer J. and Palmer T. N. 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. .J. Roy. Meteor. Soc.*,**127**, 709-734.

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A., 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121-150.

Simon, H. D., 1984: The Lanczos algorithm with partial re-orthogonalization. *Math. Comp.*, **42**, **165**, 115-142.

Sleijpen, G. L. G., and van der Vorst, H. A., 1996: A Jacobi - Davidson iteration method for linear eigenvalue prob-

lems. SIAM J. *Matrix Anal. Appl.*, **17**, 401 - 425

Trevisan, A, Pancotti, F. and Molteni, F., 2001: Ensemble prediction in a model witxh flow regimes. *Quart. J. Roy. Meteor. Soc.* **127**, 343-358.