Tellus (1993), 45A, 388–407
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

Computation of optimal unstable structures for a numerical weather prediction model

By R. BUIZZA[1,‡], J. TRIBBIA[2], F. MOLTENI[1] and T. PALMER[1], [1]*European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, England,* [2]*National Center for Atmospheric Research*, P.O. Box 3000, Boulder, Colorado 80307, USA*

(Manuscript received 30 November 1992; in final form 10 May 1993)

ABSTRACT

Numerical experiments have been performed to compute the fastest growing perturbations in a finite time interval for a complex numerical weather prediction model. The models used are the tangent forward and adjoint versions of the adiabatic primitive-equation model of the Integrated Forecasting System developed at the European Centre for Medium-Range Weather Forecasts and Météo France. These have been run with a horizontal truncation T21, with 19 vertical levels. The fastest growing perturbations are the singular vectors of the propagator of the forward tangent model with the largest singular values. An iterative Lanczos algorithm has been used for the numerical computation of the perturbations. Sensitivity of the calculations to different time intervals and to the norm used in the definition of the adjoint model have been analysed. The impact of normal mode initialization has also been studied. Two classes of fastest growing perturbations have been found; one is characterized by a maximum amplitude in the middle troposphere, while the other is confined to model layers close to the surface. It is shown that the latter is damped by the boundary layer physics in the full model. The linear evolution of the perturbations has been compared to the non-linear evolution when the perturbations are superimposed on a basic state in the T63, 19-level version of the ECMWF model.

1. Introduction

Ensemble prediction is an attempt to estimate the probability distribution of forecast states, given uncertain initial conditions, through a finite sample of nonlinear deterministic integrations of a numerical weather prediction (NWP) model (Epstein, 1969; Leith, 1974). A fundamental difficulty with ensemble prediction is that the number of degrees of freedom in such a model greatly exceeds the maximum practicable size of an ensemble. One sampling strategy being pursued at ECMWF is to first calculate those initial perturbations which have maximum growth over that initial part of the forecast in which error growth can be described by linearised dynamics. Such fields are then used to construct initial perturbations for the ensemble forecast.

Calculations of these finite-time instabilities, first suggested by Lorenz (1965) and subsequently investigated by Farrell (1982), have been performed by Borges and Hartmann (1992) using a barotropic model, and by Molteni and Palmer (1993) using T21 barotropic and T21, 3-level quasi-geostrophic (QG) models. These QG perturbations were then interpolated onto the grid of a primitive-equation model with terrain-following coordinate grid. Results of ensemble experiments with such interpolated perturbations are described in Mureau et al. (1992), where it was concluded that inconsistencies between the QG and primitive-equation model, particularly in respect of their orographic representation, made the initial perturbations less effective than expected on the basis of theory.

* NCAR is sponsored by the National Science Foundation.

‡ Corresponding author.

The principal reason for calculating fastest-growing perturbations using a QG model, was that it enabled conventional matrix algorithms to find the entire singular spectrum. With a primitive-equation model of the same resolution as the QG model, a matrix representation of the linearised dynamical equations was too large to allow the algorithmic operations to be carried out in computer memory.

On the other hand, sophisticated semi-direct methods now exist to find the dominant eigenvectors of a symmetric matrix. In the Lanczos algorithm (Strang, 1986), iteration of the operator associated with such a matrix can produce an estimate of a number of eigenvectors with extreme eigenvalues. In the problem posed here this symmetric operator is the product of the forward tangent model propagator multiplied by its adjoint (Lacarra and Talagrand, 1988).

This paper describes some initial calculations of the fastest growing perturbations using a T21, 19-level (hereafter T21L19) adiabatic primitive-equation system. Such a system has been developed at ECMWF primarily for the application of 4-D variational data assimilation (Courtier et al., 1991; Rabier and Courtier, 1992).

The theory for calculating the unstable subspace, and some of the practical details of the computations are given in Section 2. Results are shown in Section 3 for optimization times of 12, 24 and 36 h. In Section 4, we study the impact of non-linear normal-mode initialisation (NNMI) on perturbations computed over 12 and 24 h. In Section 5, we analyse the impact of the inner product definition. In Section 6 we show the nonlinear time evolution of the perturbations when they are interpolated into and integrated with a T63L19 model specification. Some conclusions are made in Section 7. The Lanczos method is described briefly in the Appendix.

2. Estimation of the fastest growing perturbations

In this section we will briefly describe the theoretical and numerical aspects of the computation of the fastest growing perturbations over a finite time interval. This technique makes use of the linearized version of a dynamical model of the atmospheric flow, and its adjoint. In this study, the dynamical model is a primitive-equation spectral model with 19 vertical levels, and triangular truncation of the horizontal fields to total wave number 21 (T21). The full non-linear model, together with its forward and adjoint tangent models are part of what is known as the Integrated Forecasting System (IFS), developed at ECMWF and Météo France. The model equations are described in Courtier et al. (1991). At the time of these experiments the forward and adjoint tangent versions of the adiabatic part of the IFS model were available (with a horizontal diffusion scheme).

The definition of the adjoint of the tangent version of the model equations depends on the inner product that characterizes the working space, and its associated norm. Most of the experiments have been performed with the norm of the state vector defined to be its total energy. The equations are linearized about the basic state trajectory, as defined in Section 2.3.

2.1. Theoretical aspects

Let χ be the state vector in the phase space of a system, whose evolution equations can be formally written as:

$$\frac{d\chi}{dt} = A(\chi). \tag{2.1}$$

The model equations can be linearised about a non-linear trajectory of (2.1). This gives the linearized equations for the perturbation x:

$$\frac{dx}{dt} = A_l x, \tag{2.2}$$

where $A_l = \delta A/\delta\chi|_{\chi(t)}$ is the tangent operator relative to the trajectory $\chi(t)$. The eq. (2.2) defines the tangent model.

Let $L = L(t_0, t)$ be the integral propagator, i.e., the mapping of $x(t_0)$ to $x(t)$ given by eq. (2.2). Since it is a linear operator, it can be represented by a matrix. We will use the same symbol for the operator and the matrix. The perturbation $x(t)$ at time t is given by:

$$x(t) = Lx(t_0). \tag{2.3}$$

Let us define an inner product on the linear vector space of perturbations:

$$(x; y) \equiv \langle x; Ey \rangle, \tag{2.4}$$

where E defines some weighting factors, and $\langle ...; ... \rangle$ is the Euclidean scalar product. We also define the associated optimisation norm as:

$$\|x\|^2 \equiv (x; x) = \langle x; Ex \rangle, \tag{2.5}$$

so that the norm of the perturbation at time t is given by:

$$\|x(t)\|^2 = (Lx(t_0); Lx(t_0)). \tag{2.6}$$

Let us define the adjoint of the operator L with respect to the inner product $(...; ...)$ defined by eq. (2.4):

$$(L^{*E}x; y) = (x; Ly) \tag{2.7}$$

Applying the adjoint definition (2.7), the norm (2.6) of the state vector $x(t)$ at time t can be evaluated as:

$$\begin{aligned}\|x(t)\|^2 &= (Lx_0; Lx_0)\\ &= (L^{*E} \cdot Lx_0; x_0).\end{aligned} \tag{2.8}$$

The square roots of the eigenvalues of $L^{*E} \cdot L$ are called the singular values of L, and the eigenvectors of $L^{*E} \cdot L$ are called the right singular vectors of L (hereafter SVs). Since in our case L is real, the SVs are orthogonal. Let σ_i^2 be the eigenvalues of $L^{*E} \cdot L$, and V the matrix whose columns are the normalized eigenvectors of $L^{*E} \cdot L$. Then, the following identity holds:

$$L^{*E} \cdot L = V \cdot \Sigma \cdot V^{\mathrm{T}}, \tag{2.9}$$

where Σ is a diagonal matrix with the values σ_i^2 on the diagonal. The norm of a SV $v_i(t)$ at time t is given by:

$$\begin{aligned}\|v_i(t)\|^2 &= (L^{*E} \cdot Lv_i(t_0); v_i(t_0))\\ &= \sigma_i^2 \, \|v_i(t_0)\|^2.\end{aligned} \tag{2.10}$$

The SVs v_i identify an orthogonal basis. They evolve into orthogonal patterns at time t, and their norm at time t can be computed from their initial norm using eq. (2.10). The singular values σ_i give the amplification factors between t_0 and t for each of them. The directions characterized by the largest singular values are the directions that grow fastest with respect to the norm (2.5).

In the numerical model used for this study, the adjoint of the tangent version of the model equations has been defined with respect to the Euclidean inner product $\langle ...; ... \rangle$. Let L^* identify the adjoint of L with respect to the Euclidean inner product $\langle ...; ... \rangle$, i.e.,

$$\langle L^*x; y \rangle = \langle x; Ly \rangle. \tag{2.11}$$

The adjoint with respect to the inner product $(...; ...)$ (2.4) can be deduced from L^* using (2.3) and (2.7) (the operator E is self-adjoint):

$$\begin{aligned}(x; Ly) &= \langle x; E \cdot Ly \rangle = \langle L^* \cdot Ex; y \rangle\\ &= \langle E^{-1} \cdot L^* \cdot Ex; Ey \rangle = (L^{*E}x; y).\end{aligned} \tag{2.12}$$

From (2.12) we have:

$$L^{*E} = E^{-1} \cdot L^* \cdot E. \tag{2.13}$$

Eq. (2.13) defines the adjoint L^{*E} of the linear operator L with respect to the inner product $(...; ...)$ (2.4), when the adjoint L^* of L with respect to the Euclidean inner product $\langle ...; ... \rangle$ is known. Using (2.13), the norm (2.8) of a vector at time t can be computed using the following expression:

$$\|x(t)\|^2 = (E^{-1} \cdot L^* \cdot E \cdot Lx_0; x_0). \tag{2.14}$$

Let us define the operator K:

$$K = L^{*E} \cdot L = E^{-1} \cdot L^* \cdot E \cdot L. \tag{2.15}$$

The fastest growing perturbations are the eigenvectors of the operator K with the largest eigenvalues. Fastest growing perturbations with respect to different norms can be computed by changing the definition of the matrix E.

2.2. The Lanczos algorithm

The Lanczos algorithm is very useful when only a few of the extreme eigenvectors are needed. It can be applied to large and sparse problems (see Lacarra and Talagrand, 1988, and Anderson, 1991, for meteorological applications). The algorithm does not access directly the matrix elements of the operator that define the problem, but it gives an estimate of the eigenvectors through successive applications of the operator. Appendix A gives an outline of the Lanczos theory; for further details see Strang (1986) and Simon (1984).

The Lanczos algorithm used here works only for symmetric problems. It should be pointed out that the matrix K in (2.15) is not symmetric: this is a consequence of the fact that its eigenvectors are orthogonal in the (...; ...) inner product rather than in the Euclidean one. For computational purposes, it is better to transform K into a symmetric matrix by a suitable coordinate transformation. This can be easily done by defining the vectors:

$$\tilde{x} = E^{1/2}x, \tag{2.16}$$

which satisfy the relationship:

$$\langle \tilde{x}; \tilde{y} \rangle = (x; y). \tag{2.17}$$

If:

$$Kx = \sigma^2 x \tag{2.18}$$

then:

$$\tilde{K}\tilde{x} = \sigma^2 \tilde{x}, \tag{2.19}$$

where:

$$\begin{aligned}\tilde{K} &= E^{1/2}KE^{-1/2} \\ &= E^{-1/2}L^*ELE^{-1/2}\end{aligned} \tag{2.20}$$

is a symmetric matrix.

The first step of the Lanczos algorithm is to do a partial transform of the matrix $\tilde{K}(n, n)$ into a tridiagonal matrix $T(j, j)$ with $j \ll n$ (see Appendix A for details):

$$\tilde{K} = Q \cdot T \cdot Q^{t}. \tag{2.21}$$

The matrix Q is defined by the Krylov sub-space generated by the operator $\tilde{K}$, i.e., the sub-space generated by the q_j vectors:

$$\text{span}\{q_1, \tilde{K}q_1, ..., \tilde{K}^j q_1, ...\}, \tag{2.22}$$

where q_1 is a randomly chosen starting vector. The Lanczos algorithm makes successive integrations of the forward and adjoint tangent version of the direct model to compute all the q_j vectors of the Krylov sub-space and the tridiagonal matrix T. From the q_j vectors and from the matrix T it evaluates an approximation of the eigenvectors of the operator $\tilde{K}$. These eigenvectors, after the inverse of the transformation (2.16), define the eigenvectors of the operator K defined in (2.15) (SVs of the operator L).

The number of iterations j determines the accuracy of the computations. As this number increases, more eigenvectors can be separated from all the others, independently from the choice of the starting vector q_1 of the Lanczos algorithm procedure. This separation starts from the boundaries of the spectrum interval of the eigenvalues. Generally, 20 iterations are enough to have an acceptable precision on the largest 5 eigenvalues. Increasing the number of iterations improves the accuracy of the following values. The accuracy of the eigenvectors is less than the accuracy of the eigenvalues, order $O(\varepsilon)$ when the precision of the eigenvalues is of the order $O(\varepsilon^2)$.

For each estimated eigenvalue θ, the algorithm also provides an uncertainty interval ρ:

$$|\sigma^2 - \theta| \leqslant \rho, \tag{2.23}$$

where σ^2 is a correct eigenvalue of the operator $\tilde{K}$. The bound ρ is a measure of the distance between an estimated eigenvalue and a true one, and it is a function of the number of iterations. We can define a measure of the accuracy of the estimated eigenvalue as:

$$\varepsilon = \left|\frac{\rho}{\theta}\right|.$$

After some tests it was decided to run 100 iterations for each experiment. In this way, the accuracy ε is less than 0.01 for all the eigenvalues up to the 30th–35th.

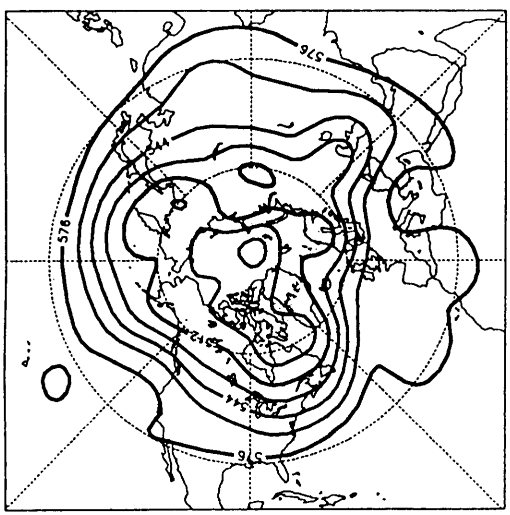


Fig. 1. 500 hPa geopotential height field at 12 UTC, 17.01.1987 (contour isolines every 160 m).

2.3. Basic state characteristics

All the experiments have starting date the 12 UTC of 17 January 1989. This initial state is characterized by a flow with a strong zonal character throughout the entire hemisphere (see Fig. 1). The T21L19 IFS tangent integration has been done by linearising around the time-evolving trajectory computed by a non-linear version of this model, except when explicitly mentioned. The time evolution of the basic state suffers from the lack of a planetary boundary layer (PBL) representation; as a result strong winds develop in the eastern part of the Himalayan region and over northern Africa. A test has been done to see if this influences the SVs structures: results from an experiment with constant basic state has been compared to results of an experiment performed following the trajectory. Small differences have been detected between the two.

3. 12 H, 24 H and 36 H singular vectors

In this section, we analyse the impact of the time interval over which the growth of the SVs structures has been maximized. Three time intervals have been considered: 12, 24 and 36 h. All the experiments have been performed with the NNMI procedure applied to the gravest 5 vertical modes: in Section 4 some results of the impact of NNMI on the SVs will be reported. A very weak horizontal diffusion has been used in these experiments: a stronger diffusion produces smaller amplification factors, but does not affect the structure of the SVs unless unrealistically large values (for a T21 truncation) are used. First we describe some of the fastest growing 24 h SVs, and then a comparison between the three unstable sub-spaces will be reported.

Fig. 2 shows the amplification factors of the first 100 SVs of the three experiments (singular values σ_i sorted in decreasing order):

$$\sigma_i = \frac{\|v_i(t)\|}{\|v_i(t_0)\|}. \tag{3.1}$$

The spectra of the 24 h and 36 h amplification factors have a similar trend: the first part of each spectrum is characterized by a steep decrease until a plateau is reached at about SV number 20. At about SV number 80 the SVs represent marginally

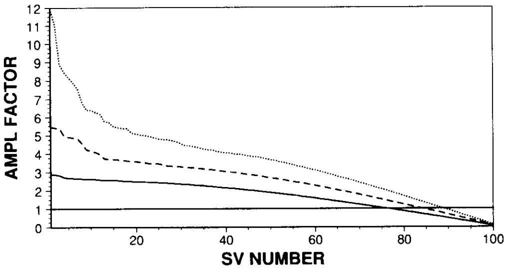


Fig. 2. Amplification factors for the SVs optimized over a 12-h, 24-h and 36-h time interval. The amplification factors are sorted in decreasing order.

growing or decaying modes. Conversely, the 12 h singular values are not easily separated in a group of fast growing perturbations and slowly growing perturbations.

Figs. 3 and 4 show the streamfunction of the first 6 24 h SVs, respectively at model level 11 (approximately 500 hPa over sea) and at model level 18 (approximately 1000 hPa). As a general consideration, all are very localized: only after the first 10–15 SVs one can find structures that cover half the Northern Hemisphere (NH). None of the 1st 20 SVs has significant amplitude in the Southern Hemisphere (SH). The 1st and the 3rd SVs are localized over the western Pacific, with maximum amplitude in the middle troposphere between 700 hPa and 500 hPa, where the mid-latitude jet has a region of maximum strength. They are characterized by baroclinically amplifying patterns, with a westward tilt with height. Horizontally they are characterized by a NW-SE elongated shape, also indicating barotropic amplification. The 2nd SV has maximum amplitude near the surface in the eastern part of the Himalayan region. The 4th SV has maximum amplitude near the surface, over northern Africa, which might be due to instability growth in a region of low level shear. The fifth SV has some structure at model level 11, but it is almost entirely confined to the lowest model levels. The sixth SV has maximum amplitude in the middle troposphere, in the north-western part of the Atlantic Ocean. It has westward tilt with height and NW–SE elongated shape. Some of these SVs are very similar to the SVs computed by Molteni and Palmer (1993) using a T21L3 QG model. As already shown in Molteni and Palmer (1993), the SVs have much more localized structures than the

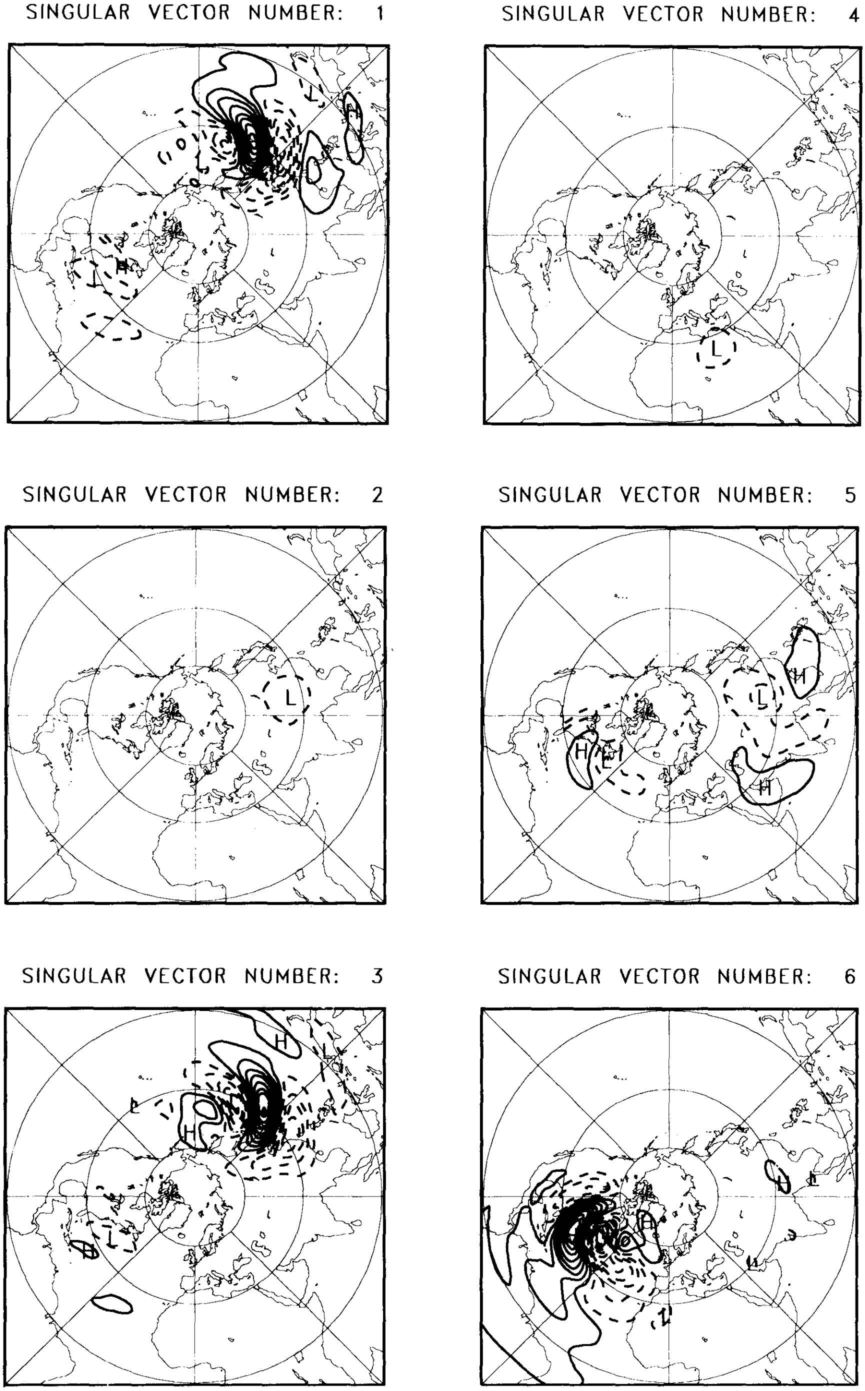


Fig. 3. Level-11 streamfunction of the first six SVs optimized over 24 h (increasing SV number from top-left to bottom-left, from top-right to bottom-right). The SVs are normalized to have unit total energy norm.

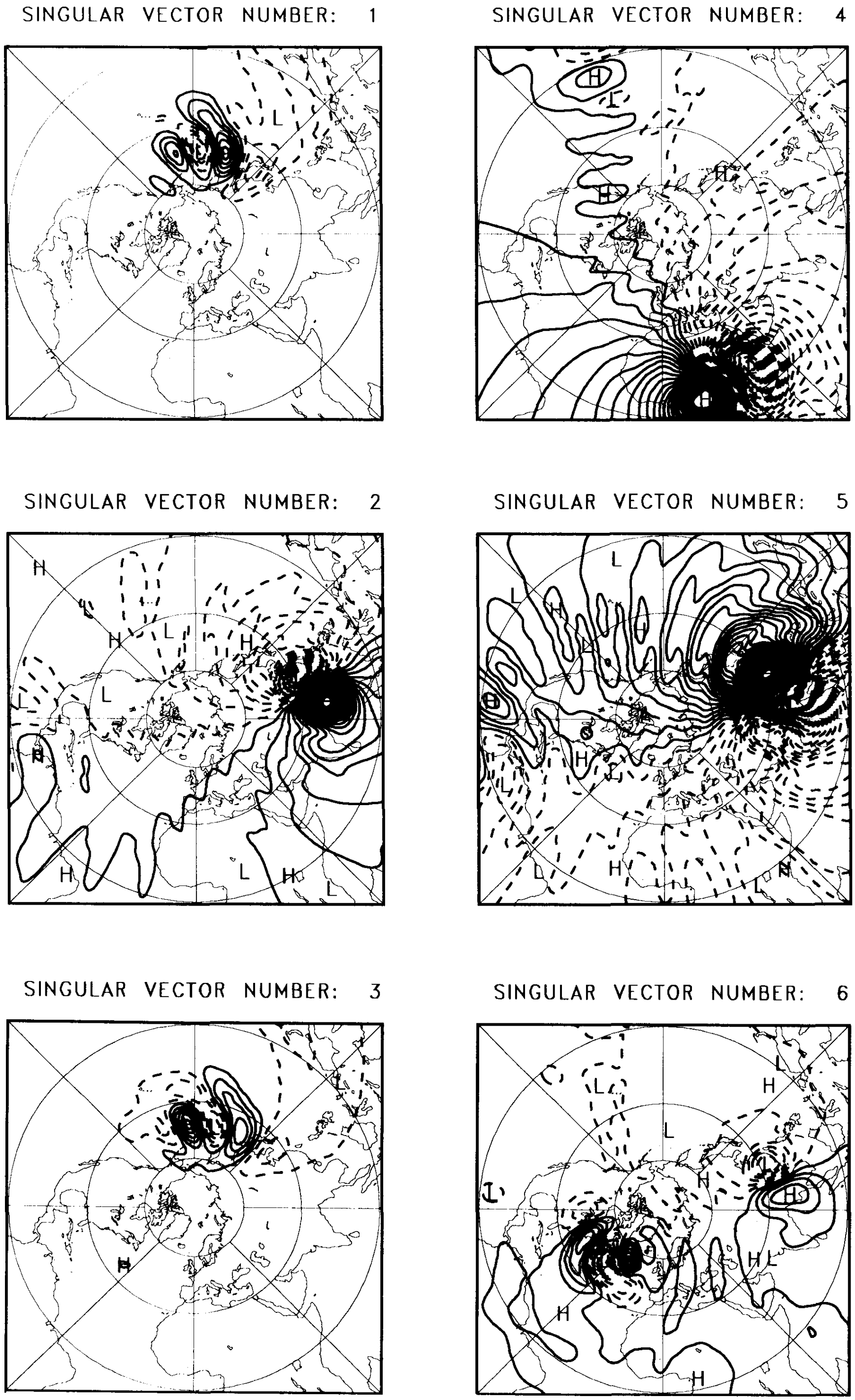


Fig. 4. As in Fig. 2 but for model level number 18.

leading normal modes of a time-average basic state.

Some SVs are almost completely confined to the lowest model levels and are "non-meteorological" in character (as it will be shown in Section 6, these structures do not grow in a model with a parameterization of boundary layer dissipation). They have a very strong baroclinic structure as can be detected from the temperature fields (not shown), with a positive perturbation at one level, negative at the following one, then positive again. The "meteorological" SVs have more realistic structures (in the sense that they have a structure not associated with the vertical discretization), with maximum amplitude in the region of the mid-latitude jet streams. The growth of the "non-meteorological" surface structures occurs probably because of the absence of a PBL physics in the version of the IFS used. These perturbations are kinematically structured in the vertical so as to extract energy from the low-level shear in the basic state (Farrell, 1982). In the real atmosphere the shear is balanced by boundary layer dissipation. The perturbations, computed without a PBL physics, can grow rapidly since they are not subject to the dissipation inherent in the boundary layer balance.

The 24-h and the 36-h SVs have very similar structures, although sorted in a different order. Table 1 gives the unstable sub-space projection matrix $M(i, j)$ of the 1st 20, 24-h SVs on the 1st 20, 36-h SVs, and vice-versa (recall that the space of the perturbations x has been defined using the inner product (2.4)). The 20 values in the jth column of Table 1 (elements $M(i, j)$ for $i = 1, 20$), are the squared scalar products between the jth 36 h SV, and the 20, 24-h SVs. The last value in each column is the sum of these squared scalar products, which represents how well each 36-h SV can be reconstructed from a linear combination of the 1st 20, 24-h SVs. The last column of the matrix gives the same information for the reconstruction of the 24-h SVs from a linear combination of the 36 h SVs.

Table 1. *Unstable sub-space projection matrix M*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	0	0	74	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83
2	82	0	0	0	1	0	0	0	0	0	0	0	6	0	0	2	0	3	0	0	95
3	0	0	7	69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77
4	0	89	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	92
5	1	0	0	0	72	0	2	9	0	0	0	0	0	0	0	0	0	0	0	0	85
6	0	0	0	0	1	59	27	0	0	0	0	0	0	0	0	0	0	0	0	1	88
7	0	0	0	0	2	27	56	0	0	0	0	0	0	0	0	1	1	1	0	0	88
8	0	0	0	0	14	0	0	64	3	0	1	1	0	0	0	1	0	3	0	0	86
9	0	0	0	0	0	0	0	7	58	1	4	0	6	0	0	1	0	0	0	0	78
10	0	0	0	2	0	0	0	0	4	0	64	8	0	0	0	0	0	0	0	0	79
11	0	0	1	0	0	0	0	2	1	0	7	68	0	0	0	0	0	0	0	0	79
12	1	0	0	0	0	0	0	0	15	0	1	0	48	1	0	2	0	5	0	0	73
13	0	0	0	0	1	0	0	2	0	0	0	1	0	0	0	0	46	9	5	1	66
14	0	3	0	0	0	0	0	0	1	48	0	0	0	9	1	0	0	0	3	0	66
15	0	0	0	0	0	0	0	1	0	8	0	0	0	1	0	1	1	0	15	0	27
16	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	34	2	19	0	11	70
17	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	9	0	29	9	53
18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	1	9	2	4	17	37
19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	5	1	1	1	12	25
20	0	1	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	10
	85	94	83	80	90	90	88	85	82	67	79	79	60	13	9	51	71	45	58	51	

The indices on the vertical axis refer to the 24-h 5-NMI SVs, the indices on the horizontal top axis to the 36-h 5-NMI experiment. The element $M(i, j)$ of the matrix M gives the scalar product between the ith 24-h SV and the jth 36-h SVs. For each line, the last column gives the % of norm of each of the 24-h SVs that is explained by the 36-h sub-space. For each column, the last line gives the analogous information for the 36-h SVs norms explained by the 24-h sub-space.

Table 2. *The same as in Table 1; vertical axis: 24-h 5-NMI SVs; horizontal axis: 12-h 5-NMI SVs*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	47	4	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	56
2	0	0	72	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	75
3	4	49	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	56
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	70	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	1	76
6	0	0	0	1	0	0	0	0	2	0	0	0	8	50	0	0	0	0	0	0	62
7	0	0	0	1	0	0	0	0	0	0	0	1	1	1	58	0	0	0	0	0	63
8	0	0	1	3	1	0	2	1	0	29	20	0	0	0	0	0	5	1	0	0	64
9	0	0	0	1	2	0	3	2	1	6	6	1	0	0	0	0	11	1	0	1	36
10	0	0	0	1	0	0	0	5	0	3	3	3	10	2	0	5	1	1	0	0	36
11	0	0	0	0	0	0	1	1	0	0	2	17	4	0	1	0	3	0	0	0	30
12	0	0	0	2	0	0	0	2	0	3	7	0	1	0	0	3	19	1	0	3	43
13	0	0	0	0	0	5	2	0	0	0	0	0	0	1	0	0	2	4	0	0	16
14	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5
15	0	0	0	0	0	25	4	0	0	0	0	0	1	0	0	2	0	0	0	1	33
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	6
17	0	0	0	0	4	7	0	0	2	0	0	0	0	1	0	0	0	0	0	1	17
18	0	0	0	0	0	0	2	2	11	0	0	0	1	3	1	0	0	0	0	1	22
19	0	0	0	0	1	1	6	2	22	0	0	0	0	0	0	0	0	0	0	0	34
20	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	52	54	74	79	10	42	21	16	39	41	42	26	29	60	66	12	44	11	1	9	

Table 1 shows that the 1st 36-h SV (a non-meteorological SV which has maximum amplitude near the surface in the Himalayan region), has an 82% squared-correlation with the 2nd 24-h SV. In other words, 82% of the energy of the 1st, 36-h SV is explained by its projection on the 2nd 24-h SV. The 2nd, 36-h SV (which has maximum amplitude near the surface over northern Africa) has an 89% squared-correlation with the 4th, 24-h SV. The 3rd and the 4th 36 h SVs have maximum amplitude in the middle troposphere, over the western Pacific Ocean, and project respectively on the 1st and the 3rd 24 h SVs. Analogous considerations can be made for the other SVs. Overall, the unstable subspace generated by the first 14 SVs of one experiment can explain at least 60% of the energy of the first 14 SVs of the other experiments. This indicates a rather strong parallelism between the two sub-spaces.

Table 2 gives the unstable sub-space relationship between the 12-h and the 24-h SVs. It clearly shows that the two sub-spaces are characterized by a very poor parallelism: the 20 SVs of one experiments can explain 60% of the energy of only 4 SVs of the other one. The 12-h SVs reveal the same non-meteorological structures of the 24-h and 36-h SVs. Moreover, some of them are characterized by large amplitude in the upper troposphere not detected in the 24th and 36-h SVs.

Looking at the 1st 8 SVs of each experiment, various types of perturbations with maximum amplitude in different part of the atmosphere can be identified: the Pacific (P) and the Atlantic (A) structures with maximum amplitude in the middle troposphere, the Himalayas (H) and the Tropical-African (T) structures with maximum amplitude near the surface. Moreover, some of the 1st, 12-h SVs have maximum amplitude in the upper troposphere, mainly over Europe, Asia and Africa: we will identify them as upper troposphere structures (U). Table 3 summarizes these considera-

Table 3. *Regions where the maximum amplitude of the first 8 SVs computed by the experiments run with a 12-h, 24-h and 36-h optimisation time interval: Pacific (P), Atlantic (A), Himalayas (H) and Tropical-African (T) regions*

SV number	1	2	3	4	5	6	7	8
12-h experiment	P	P	H	H	U	U	U	U
24-h experiment	P	H	P	T	H	A	A	H
36-h experiment	H	T	P	P	H	A	A	H

tions. The two P-SVs and the two H-SVs are present in all the experiments. However, a time interval longer than 12 h is necessary to separate the A-SVs and the T-SVs from the others.

From this section, the following conclusions can be drawn.

(i) The fastest growing meteorological perturbations have very localized structures similar to QG perturbations (Molteni and Palmer, 1993).

(ii) Some SVs are dominated by non-meteorological structures near the surface: this is likely due to the absence of PBL physics. All the SVs present traces of these non-meteorological structures near the surface.

(iii) Some differences exist between the 12-h unstable sub-space, and the 24-h and 36-h sub-spaces. The latter, however, show a very strong parallelism.

4. Impact of NNMI on the singular vectors

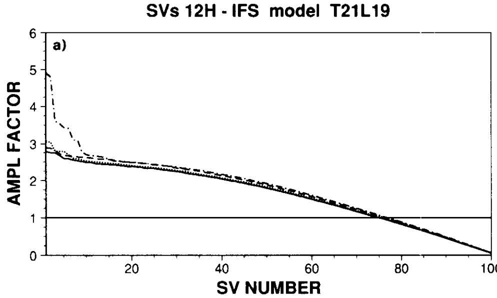
In order to study the impact of the (adiabatic) non-linear normal mode initialisation (NNMI) on the unstable sub-space definition, experiments have been performed with no initialisation, with the 1st 3, 5, 9 and 13 gravest vertical modes initialized (the experiments will be called NO-NNMI, 3-NNMI, 5-NNMI and so on). The experiments have been performed mainly over 2 time intervals, 12 and 24 h, with some additional experiments over 36 h. A linearized version of NNMI has been inserted before the time integration of the linear model, and formally can be considered as a part of the propagator operator L. The full non-linear version has been used to computed the basic state trajectory.

Fig. 5a shows the amplification factors of the 12-h experiments with NO-NNMI, 5-NNMI, 9-NNMI and 13-NNMI experiments. The 3-NNMI spectrum have been omitted for clarity. A large difference can be detected between the experiments with up to 9 initialized modes and the 13-NNMI experiment, in the first part of the spectrum. The 13-NNMI SVs are characterized by stronger growth. Looking at the 13-NNMI SVs structures, they show maximum amplitude in the upper troposphere, with a horizontal structure with high zonal wave number (approximately 16).

The difference in the structure of the SVs is also apparent through the partition of the total energy between the rotational and divergent part of the kinetic energy, and the potential energy. For the 5-NNMI SVs, approximately 70% of the total energy is due to potential energy, while in the 13-NNMI SVs with the upper troposphere, high wavenumber features the potential energy is less than 10% of the total energy.

Analysis of the unstable sub-spaces generated by the NO-NNMI and the 3-NNMI, 5-NNMI and 9-NNMI SVs indicates that the impact of NNMI is small when only the first gravest modes are initialized. Considering for example the NO-NNMI and the 5-NNMI SVs optimized over a 12-h time interval, the first 15 SVs of each experiment can explain at least 60% of the energy of the first 15 SVs of the other experiment.

The impact of NNMI can be detected also in the 24-h and 36-h SVs, but it is not so evident in the



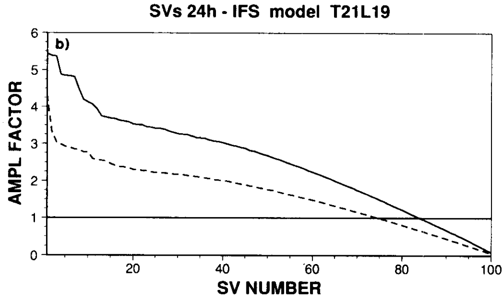


Fig. 5. (a) amplification factors for the SVs optimized over a 24-h time interval, not using the NMI procedure (solid line), applying it to the first 5 (dashed line), 9 (dotted line) and 13 vertical modes. (b) amplification factors of the SVs optimized over a 24-h time interval, computed with the total energy (solid line) and the kinetic energy norm (dashed line).

amplification factor curves as it is in the 12-h experiments. A longer integration period reduces the effect of the NNMI procedure on the fastest growing SV structures. Considering for example the NO-NNMI and the 5-NNMI SVs, the first 16 SVs of one experiment can explain at least 90% of the energy of the SVs of the other experiment.

It is worth pointing out one effect of the NNMI procedure on the norm of the state vector. The NNMI procedure projects a state vector onto the normal modes, decomposes the projection into the rotational and gravitational part, and then imposes a zero tendency on the gravitational part. This procedure can modify the energy of the state vector; this can increase or decrease, since the projection is made conserving the component of the state vector along the rotational directions, and modifying its component along the gravitational directions. For example, when only the first modes are initialized, the norm increases on the order of 1%, while initializing 13 modes can produce a norm increase of 80%. This large difference can be related to the divergence of the non-linear normal mode initialisation procedure when applied to the shallower vertical modes (e.g., Daley, 1981; Williamson and Temperton, 1981).

The small increase in the amplification factors when NNMI is applied to the first 3 gravest modes is due to the small increase of the norm produced by the NNMI procedure. This amplification effect in not enhanced when NNMI is applied to the first gravest 5–9 vertical modes, rather than to the first 3. On the contrary, a small decrease of the amplification factors when going from 3 to 5, and to 9 initialized vertical modes can be seen, associated with the increased number of constraints imposed by NNMI.

The following conclusions can be drawn.

(i) The impact of NNMI on the SVs is stronger when shorter optimisation time intervals are considered.

(ii) The unstable sub-spaces generated by the NO-NNMI, 3-NNMI, 5-NNMI and 9-NNMI experiments show strong similarity.

(iii) The NNMI procedure does not conserve the total energy: it can induce a variation in the energy norm of the state vector. This effect, however, is very small when only the first 5 modes are initialized. Not more than 9 modes should be initialized in a 19-level model, due to the divergence of the iterative procedure.

5. Impact of the norm definition on the singular vectors

Experiments have been performed to test the differences between the fastest growing perturbations evaluated when the norm has been defined to be the total energy (T-experiments) or the kinetic energy (K-experiments). In all these sensitivity studies the NNMI procedure has been applied to the first 5 modes. Two optimisation time intervals have been considered, 12 and 24 h.

The energy norm has been defined formally by eq. (2.2). Decomposing a state vector x into its rotational, divergent, temperature and surface pressure term, the weighting factors E that define the inner product (2.4) can be deduced from the following formula:

$$\|x\|^2 = \frac{1}{2}\int_0^1 \iint_\Sigma \bigg(\nabla\Delta^{-1}\zeta \cdot \nabla\Delta^{-1}\zeta + \nabla\Delta^{-1}D \cdot \nabla\Delta^{-1}D + R_\mathrm{a} T_\mathrm{r} (\ln \pi)^2 + \frac{C_\mathrm{p}}{T_\mathrm{r}} T^2 \bigg) \,\mathrm{d}\Sigma \left(\frac{\partial p}{\partial \eta}\right) \mathrm{d}\eta, \tag{5.1}$$

where ζ, D, T and π stand for vorticity, divergence, temperature and surface pressure, η is the vertical coordinate, T_r is a reference temperature and R_a, C_p are thermodynamic constants.

The first two terms are the rotational and the divergent part of the kinetic energy, the third is the surface pressure term, and the fourth is the potential energy term. When the norm has been defined as the kinetic energy, the third and the fourth terms have not been considered.

Fig. 5b shows the SVs' amplification factors of two experiments run with an optimisation time interval of 24 h. It shows that the amplification factors in the T-experiments are almost twice as large as those in the K-experiment. The reduced growth of the K-perturbations can be explained by the fact that at the initial time (and before NNMI), these perturbations are only wind structures, and have no temperature component. This is a consequence of temperature being an independent variable in a primitive-equation model, and if a zero-weight is given to the temperature field in the scalar product definition, then such a field cannot be used by the optimisation procedure which computes the SVs. A temperature component is inserted into the

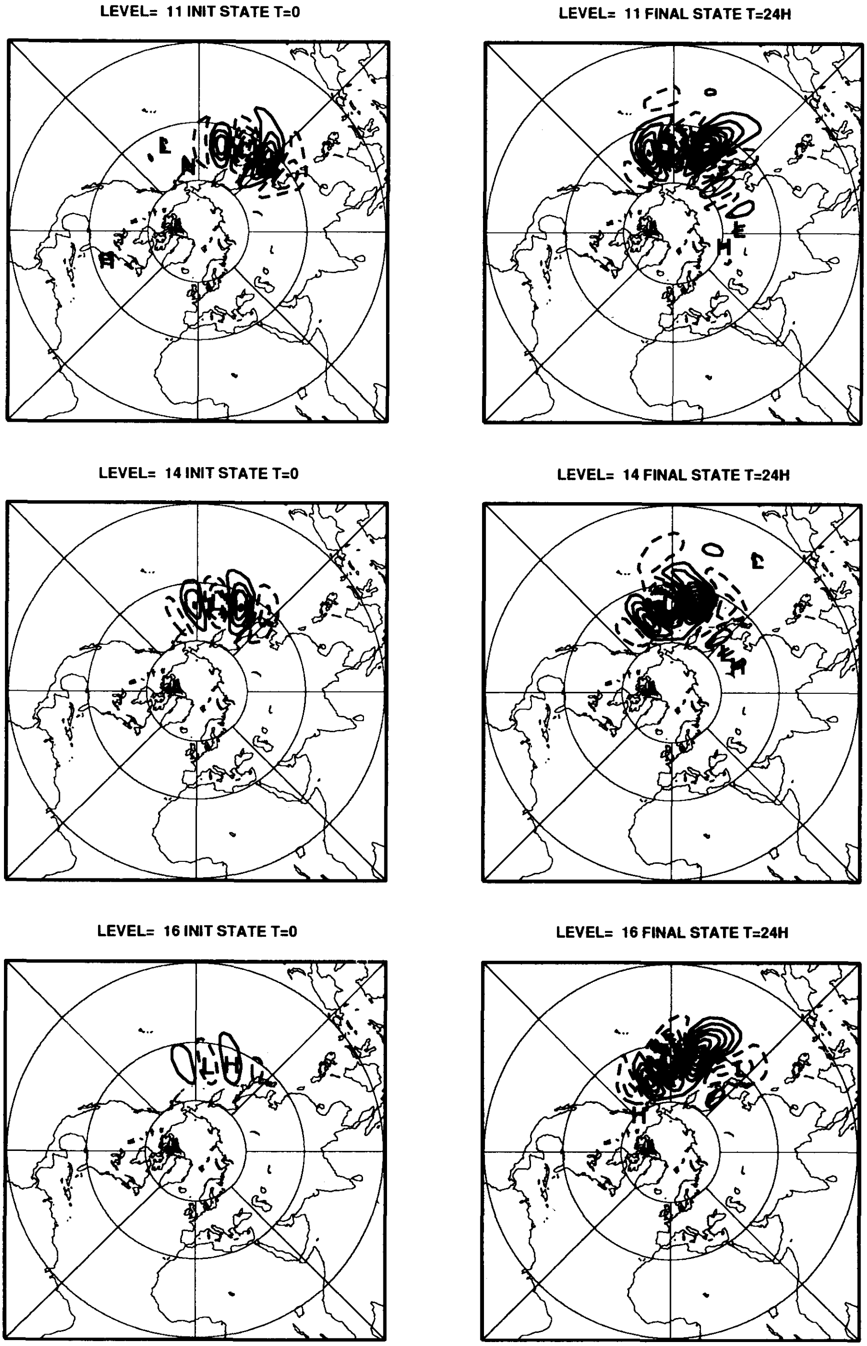


Fig. 6. Tangent linear time evolution of the fastest growing SV computed over a 24-h time interval, with NMI applied to the first 5 modes. The temperature field has been plotted at initial time $t = 0$, and after 24-h tangent linear integration, at three model levels (contour isolines every 1°): $L = 11$: approximately 500 hPa when surface pressure $p_s \approx 1000$ hPa. $L = 14$: approximately 800 hPa. $L = 16$: approximately 900 hPa.

K-perturbations by NNMI; however, the partial adjustment induced by NNMI cannot produce the same growth obtained with an independent optimisation of the temperature field.

Most 24-h K-SVs are characterized by patterns similar to the 24-h T-SVs: the 1st K-SV has a structure similar to the 5th T-exp SV. The 2nd and the 3rd K-SVs have a similar pattern. The 4th, the 5th and the 6th K-SVs can be seen as a combination of Pacific and Atlantic T-SVs. The 7th K-exp SV is similar to the 2nd T-SV. Generally speaking, the K-SVs show less localized structures than the T-SVs. They can be seen as combinations of T-SVs, although they have structures that extend into the upper troposphere. Some of them (e.g., the 9th, the 10th and the 11th), have very high amplitude in the upper troposphere, with structures never detected in the T-SVs.

We analysed the norm variation due to the NNMI procedure on the 24-h K-experiments state vector. NNMI always produces a norm decrease of at least 4%, with a maximum decrease of not more than 6% during the first iterations. This depletion was never detected in the T-experiments, which were always characterized by a slight increase (less than 1%).

The 12-h experiments support these results: moreover the impact of the norm definition on the SVs structures is more evident. This is related to the fact that norm sensitivity of the SVs decreases when longer time intervals are considered. The 1st 8 K-SVs have maximum amplitude in the upper troposphere, with a pattern never shown by the T-SVs. The NNMI procedure induces larger changes in the state vector norm: it produces a decrease of at least 4%, with a peak value of 9% at the 1st algorithm iteration. This NNMI effect has never been observed during the 12-h T-exp, where NNMI produced always a slight increase of about 1%.

6. Time evolution of the 24 h singular vectors

In this section, we study the time evolution of two of the SVs derived from an optimisation time period of 24 h, with the first 5 gravest vertical modes initialized, and norm defined as total energy.

Fig. 6 shows the time evolution of the 1st SV: its temperature field is plotted at three vertical levels at the starting time ($t = 0$, left hand side), and after 24 h using the T21L19 forward tangent model. Note that the meridional phase tilt is reversed from the initial to the final state, which indicates that barotropic energy conversion from the basic state to the perturbation changes from positive to negative within 24-h period. Fig. 7 shows the vertical cross section of the SV total energy at the initial state and after 24 h (the application of the NNMI procedure does not modify the energy profile in any appreciable way). From these figures it is clear how the SV grows. In particular Fig. 7 and the final state at model level 19 (last panel in Fig. 6), illustrates how the perturbation can grow close to the surface, due to the absence of PBL physics.

Fig. 8 shows the non-linear time evolution of the perturbation, when superimposed on the same initial condition in the T63L19 version of the full ECMWF model. The temperature at model level 11 (500 hPa) is shown for different integration time intervals. The growth of the perturbation during the 1st 24 h is very similar to the linear growth.

The Atlantic and Pacific IFS SVs have very similar structures to the SVs computed using a T21L3 version of a QG model (Molteni and Palmer, 1993). Fig. 9 shows the growth of a 24-h SV computed using the QG model that has a pattern similar to the 1st IFS SV. Although they have been computed with very different models, the two perturbations have the same horizontal scale and shape, and their time evolution in the 1st 24 h is very similar. In addition, both perturbations produce a positive deviation over western Canada and the north-eastern Pacific, at day 4. Table 4 gives the amplification factors of the T21L19 IFS

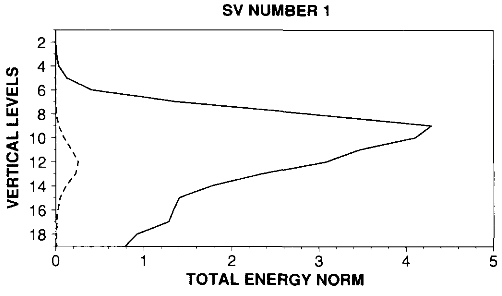


Fig. 7. Vertical cross section of the total energy norm of the 1st SV at the initial state (dashed line) and after 24-h tangent linear time integration (solid line).

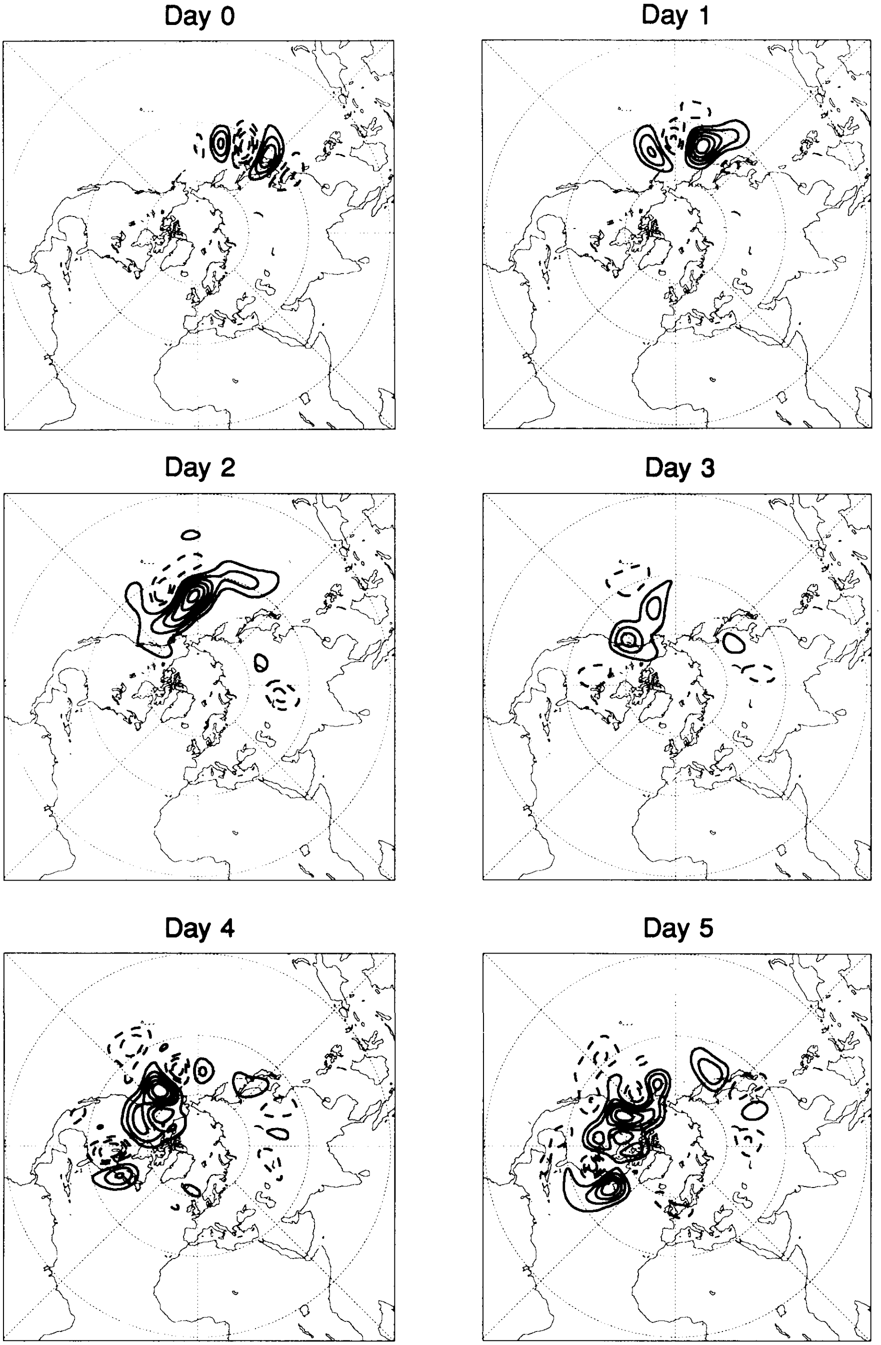


Fig. 8. Time evolution of the fastest growing SV generated by the IFS model over a 24-h time interval, computed using the T63L19 non-linear version of the ECMWF model. The level-11 temperature field at different forecast times are shown: starting from the top-left panel at day 0, 1, till day 5 (contour isolines every 1° at day 0, 1, 2, and every 2° afterwards).

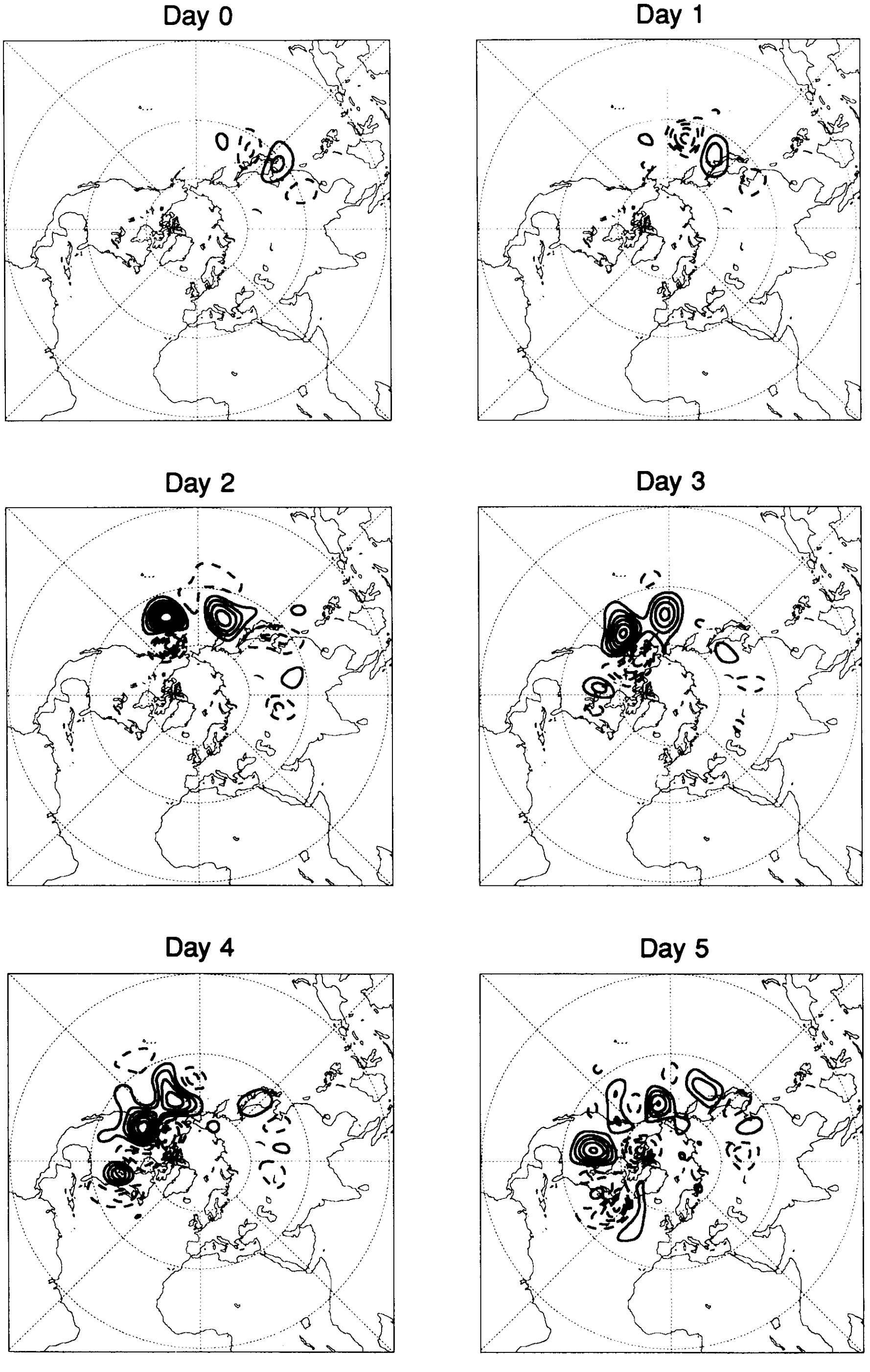


Fig. 9. As Fig. 8, but for the T63L19 time evolution of the SV computed by Molteni and Palmer (1993) using a T21L3 QG model (this QG-SV is very similar to the 1st IFS SV). Contour isolines are drawn every 1° at day 0, 1, 2, and every 2° afterwards.

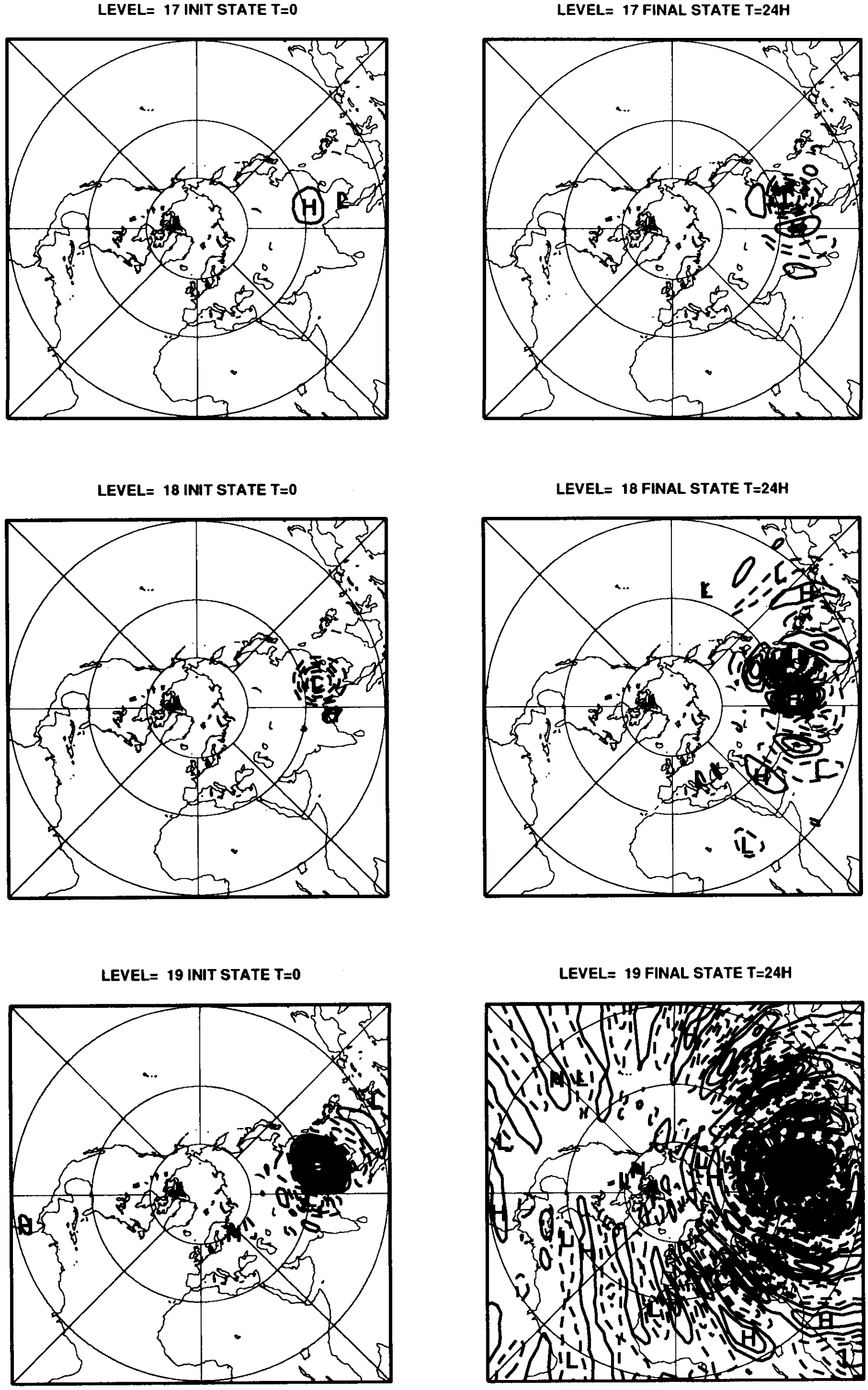


Fig. 10. As Fig. 6 but for the time evolution of the 2nd 24-h 5-NMI IFS SV: the temperature field has been plotted at the last 3 model levels (contour isolines every 2°).

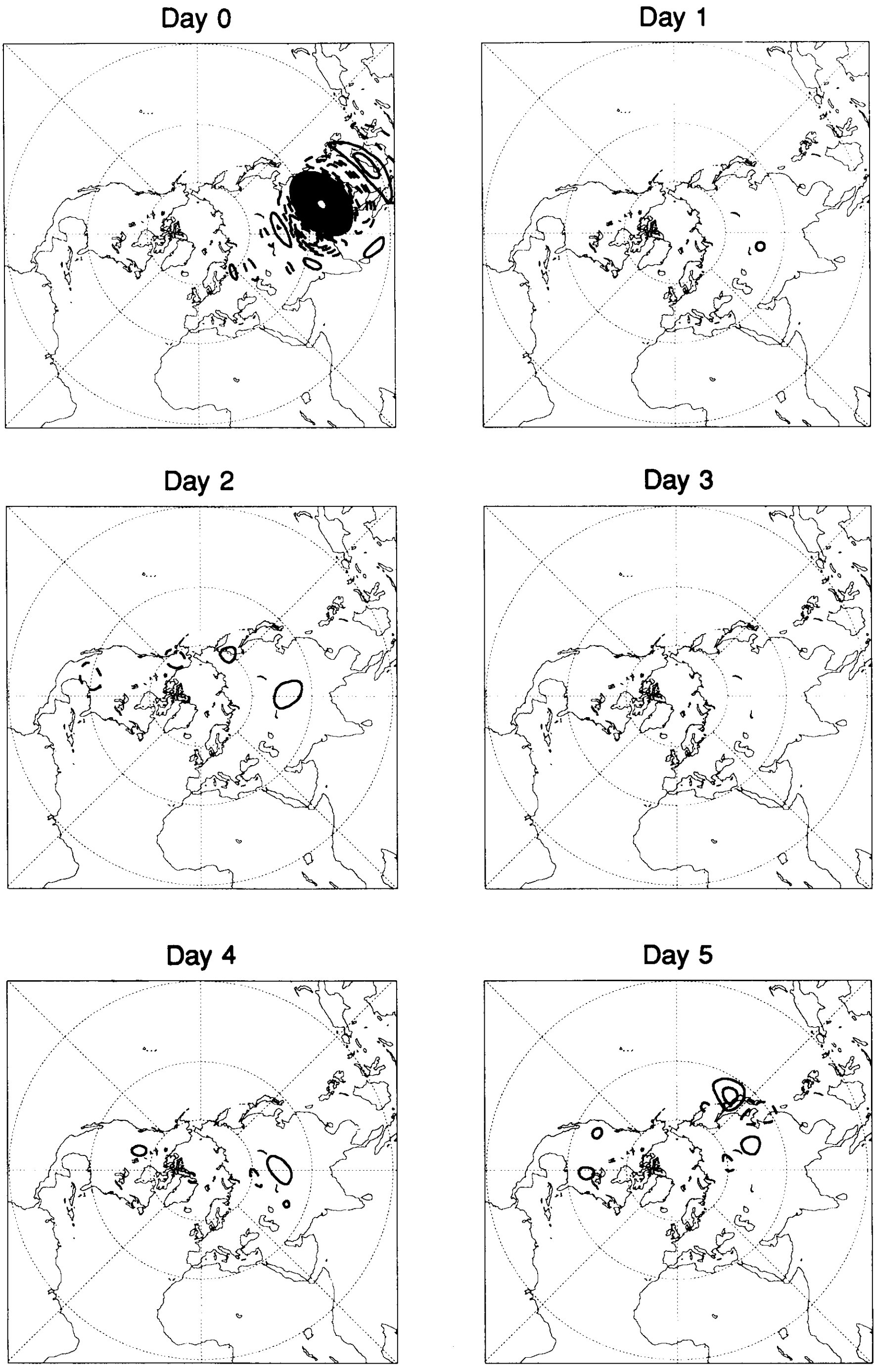


Fig. 11. As Fig. 8, but for the time evolution of the 2nd 24-h 5-NMI IFS SV. The level-18 temperature field at different forecast times are shown: starting from the top-left panel at day 0, 1, till day 5 (contour isolines every 1° for day 0, 1, 2, and every 2° afterwards).

Table 4. *Amplification factors of the IFS and QG SVs shown respectively in Figs. 8 and 9*

	Amplification			
	12-h	24-h	48-h	120-h
IFS	2.53	4.10	5.43	14.84
QG	1.83	2.91	5.10	11.09

and of the T21L3 QG SVs, computed as the root mean square (RMS) ratio between the 500 hPa geopotential height field at the considered time and at the initial time. The IFS SV seems to be slightly more efficient.

Fig. 10 shows the time evolution of the 2nd SV: since it has maximum amplitude near the surface, the temperature field has been plotted at the lowest 3 model levels. The left hand panels show the initial state at model levels 17, 18 and 19, the right hand panels the temperature after 24 h of tangent T21L19 IFS time evolution. Fig. 11 shows its evolution when the integration has been performed with the full version of the T63L19 ECMWF model: the temperature at 1000 hPa has been plotted. Note that the contour interval in Fig. 10 is 5 times the one of the analogous Fig. 6. During the linear evolution, the perturbation remains confined in the lowest model levels, and its growth is very remarkable. When the perturbation has been introduced in the full T63L19 model it has been drastically damped by the PBL diffusion in less than 24 h. This confirms the non-meteorological nature of the low-level adiabatic SVs.

7. Conclusions

In this paper, we have reported on some preliminary experiments performed at ECMWF on the computation of the unstable sub-space of a primitive equations model, by means of its tangent and adjoint versions. The fastest growing perturbations are the singular vectors (SVs) of the tangent propagator of the primitive-equation model, identified by the linear operator *L*. A Lanczos algorithm was applied to find the dominant SVs. Some of the sensitivity results presented above should be confirmed by further investigation, since all the experiments were characterized by the same initial date.

The 24-h and 36-h experiments performed with the first 5 vertical modes initialized have very similar unstable sub-spaces. Two classes of structures can be identified in the sub-spaces: one with maximum amplitude in the middle troposphere, and one with higher amplitude near the surface. An optimisation time interval of 12 h seems to be too short to give a clear identification of the fastest growing SVs.

The inclusion of the Non-linear Normal Mode Initialization procedure has a more detectable impact in the 12-h experiments than in experiments with longer optimisation time interval. Strong similarity was shown by the unstable sub-spaces generated without non-linear normal mode initialization, and NNMI applied to 3, 5 and 9 modes. Application of the Non-linear Normal Mode Initialization to more than 9 modes produces fast-growing non-meteorological structures, due to the non-convergence of the iterative procedure when applied to the shallower vertical modes.

The time evolution of the fastest growing SV computed using a tangent version of the T21L19 primitive equation model is very similar to its T63L19 full ECMWF model evolution over a period of one day. Moreover, the similarity between the full model evolution of the first primitive equation SV and the first T21L3 QG SV (Mureau et al., 1993) is remarkable. The primitive equation SVs with maximum amplitude near the surface are quickly damped when integrated using the T63L19 full diabatic ECMWF model. The absence of PBL physics in the version of the IFS model used for these experiments can explain why very strong surface perturbations can grow near the surface and do not grow in the full diabatic model where surface drag and vertical diffusion schemes are implemented.

Finally, although some of the adiabatic SVs have realistic structures which lead to strong amplification even in a diabatic model, a simple PBL parameterization is essential to eliminate non-meteorological low-level perturbations which arise in the adiabatic computation, but do not correspond to unstable modes of the real atmosphere. Experiments performed using a simple surface drag and vertical diffusion scheme are under way. Results of these experiments will be reported in a following paper.

8. Acknowledgments

We would like to thank R. Mureau for assistance with some of the integrations. Thanks also to C. Temperton for the helpful discussion on NNMI and to J. Du Croz of NAG, Oxford for the use of the Lanczos algorithm.

Appendix

The Lanczos algorithm was first proposed in 1950. Lanczos intended this algorithm to be used to compute a few of the extreme eigenvalues, and the corresponding eigenvectors, of a symmetric matrix. The Lanczos method is a technique applicable to large, sparse, symmetric eigenproblems. For details see Golub and Van Loan (1983).

Suppose the problem is to find the eigen-system of a symmetric matrix $A(n, n)$: the Lanczos algorithm does not access the elements of the matrix directly, so there is no need for conventional storage of the matrix. It is very useful when only a part of the spectrum is required (eigen-pairs less than $n/100$). The structure of the algorithm is such that the information about the matrix extremal eigenvalues tend to emerge after few iterations.

The main steps of the Lanczos procedure can be summarized in the following way.

(i) it computes a partial transformation of the matrix $A(n, n)$, that is to find two matrices $Q(n, j)$ (unitary matrix) and $T(j, j)$ such that:

$$Q^{t} \cdot A \cdot Q = T, \tag{A.1}$$

where the matrix T is tridiagonal:

$$\begin{matrix} \alpha_1 & \beta_2 & 0 & 0 & \cdot & \cdot & \cdot \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \cdot & \cdot & \cdot \\ 0 & \beta_3 & \alpha_3 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \beta_{j-1} & 0 \\ \cdot & \cdot & \cdot & 0 & \beta_{j-1} & \alpha_{j-1} & \beta_j \\ \cdot & \cdot & \cdot & 0 & 0 & \beta_j & \alpha_j \end{matrix} \tag{A.2}$$

and

$$Q = [q_1, q_2, ..., q_j], \tag{A.3}$$

where the vectors q_i of (A.3) are column vectors, and where $j \ll n$;

(ii) it computes the eigenvalues of $T(j, j)$, finding its diagonal decomposition:

$$T = S^{t} \cdot D \cdot S. \tag{A.4}$$

where the matrix D is diagonal

$$D = \mathrm{diag}[\theta_1, \theta_2, ..., \theta_j]. \tag{A.5}$$

(iii) the eigenvalues of D are an estimate of the eigenvalues of A, and an estimate of its eigenvectors is given by:

$$\begin{aligned} Y &= Q \cdot S, \\ Y &= [y_1, y_2, ..., y_j], \end{aligned} \tag{A.6}$$

where y_i is the eigenvector corresponding to the ith eigenvalue. More correctly, the eigen couples $[\theta_i, y_i]$ are the Ritz pairs of the matrix A for the sub-space $P(Q_j) = \mathrm{span}(q_1, q_2, ..., q_j)$.

REFERENCES

Borges, M. D. and Hartmann, D. L. 1992. Barotropic instability and optimal perturbations of observed non-zonal flows. *J. Atmos. Sci. 49*, 335–354.

Courtier, P., Freydier, C., Geleyn, J. F., Rabier, F. and Rochas, M. 1991. The Arpege project at Météo France. *Proceedings of ECMWF Seminar on Numerical methods in atmospheric models*, Shinfield Park, Reading RG2-9AX, UK, 9–13 September 1991, vol. 2, 193–231.

Daley, R. 1981. Normal mode initialization. *Rev. Geoph. and Space Phy. 19*, 450–468.

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

Farrel, B. F. 1982. The initial growth of disturbances in a baroclinic flow. *J. Atmos. Sci. 39*, 1663–1686.

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

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

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

Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus 17*, 321–333.

Molteni, F. and Palmer, T. N. 1993. Predictability and finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc. 119*, 269–298.

Mureau, R., Molteni, F. and Palmer, T. N. 1993. Ensemble prediction using dynamically-conditioned perturbations. *Q. J. R. Meteorol. Soc. 119*, 299–323.

Rabier, F. and Courtier, P. 1992. Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc. 118*, 649–672.

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

Strang, G. 1986. *Introduction to applied mathematics.* Wellesley-Cambridge Press, 758 pp.

Williamson, D. L. and Temperton, C. 1981. Normal mode initialisation for a multi-level grid point model. Part 2: Non-linear aspects. *Mon. Wea. Rev. 109*, 744–757.