TECHNICAL MEMORANDUM

189

Unstable perturbations computed using the adjoint technique

R. Buizza

Research Department

October, 1992

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

ABSTRACT

The generation of effective perturbations is one of the major problems in ensemble forecasting. Perturbing the initial conditions along the most unstable directions of the phase space of the system ia a technique which may be most effective for determining the principal weather types consistent with given initial data.

Numerical experiments have been performed to compute the most unstable perturbations (perturbations growing fastest in a finite time interval) in a primitive equation model. The models used are forward and adjoint tangent versions 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 perturbations are the singular vectors of the propagator of the forward tangent model. A Lanczos algorithm has been used for the numerical computation of the perturbations.

This Technical Memorandum presents some preliminary results. The computations have been done for only one date. Although sensitivity studies require more significant data sets to be complete, we are confident that some of the results will be confirmed by further experimentation.

Sensitivity of the calculations to different time intervals has been studied. Sensitivity to the norm used in the definition of the adjoint of the tangent linear version of the full model and to the horizontal diffusion coefficients have been analysed. The impact of the normal mode initialization has also been studied.

Four classes of fastest growing perturbations have been found; two are characterized by a maximum amplitude in the middle troposphere, while the other two have most amplitude close to the surface. It is shown that the latter are damped by the boundary layer physics in the full model.

The T21L19 tangent time evolution of the perturbations has been compared to the non-linear evolution when the perturbations are superimposed on a basic state in the T63L19 version of the ECMWF model.

1. INTRODUCTION

Ensemble forecast prediction (*Leith*, 1974) is based on the integration of the model equations from different initial conditions. Perturbations are superimposed on a basic state to provide a fair representation of the uncertainty of the initial state of the atmospheric flow.

Our aim is to compute the unstable sub-space for a T21L19 primitive equation model, and to study its linear and non-linear time evolution. The model used for this study is the Integrated Forecasting System (IFS), a result of a collaboration between Météo-France and ECMWF. The model equations are described in *Courtier et al.* (1991). The version of the IFS model used during this study is adiabatic, plus a horizontal diffusion scheme.

Lacarra and Talagrand (1988) showed that "the short time evolution of the forecast error can be approximated to a high degree of accuracy by a linear differential system". Their results were obtained with a rather simplified model. *Rabier and Courtier* (1991), applying the adjoint technique to four dimensional assimilation, studied the validity of the linear approximation of the IFS model. They found that the linear

approximation describes to a good degree of approximation the time evolution of small perturbations for about two days. In our study we will compute the fastest growing perturbations over time intervals no longer than 36 hours.

Molteni and Palmer (1992) calculated in both a barotropic and a 3-level quasi-geostrophic model the fastest growing perturbations using a linear approximation of the model equations. *Mureau et al.* (1992) used these perturbations for predictability experiments with a T63L19 version of the European Centre for Medium-Range Weather Forecasts (ECMWF) model.

In our work, the perturbations are computed using the forward and adjoint models. After this introduction, section 2 describes the technique for calculating the unstable subspace, and the characteristics of the IFS model. The basic state characteristics are described in sub-section 2.3. In section 3 we compare the results of experiments with optimization times of 12, 24 and 36 hours. In section 4 we study the impact of the normal mode initialization procedure (NMI) on the perturbations computed over two time intervals, 12 and 24 hours. In sections 5 and 6 we analyse, respectively, the impact of the inner product definition and the sensitivity to the coefficients of the horizontal diffusion scheme.

In section 7 we present the time evolution of the perturbations, when they are used to perturb the initial conditions when a T63L19 version of the ECMWF full model is integrated. The growth is compared to the tangent T21L19 IFS model integration. In the last section some conclusions are drawn.

Appendix A briefly describes the Lanczos method used during the experiments, while Appendix B reports the actual configuration for an IFS-Lanczos computation inside the ECMWF IFS environment.

2. THE ADJOINT TECHNIQUE APPLIED TO THE IFS SYSTEM

In the first sub-section 2.1 we will briefly describe the adjoint technique applied to the computation of the fastest growing perturbations over a finite time interval. Some details of the primitive equations model used in the computation will be given in sub-section 2.2. The definition of the adjoint of the tangent version of the model equations depends on the inner product that characterizes the working space. Once an inner product has been defined, it is possible to define 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 following the basic state trajectory.

2.1 The adjoint technique

Let χ be the state vector in the phase space of the system. The evolution equations can be formally written in the following way:

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

To find the fastest growing perturbations we have to identify, in the phase space of the system, the directions that guarantee the fastest growth of the norm of the perturbations. These directions define the unstable subspace of the system. For small time intervals and small initial amplitude the growth is almost linear, and the problem can be studied in the linear approximation. The model equations 2.1 can be linearized around a basic state. This gives the linearized equations for the perturbation x:

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

where A_l is the tangent operator that corresponds to the model operator A.

Let us identify with $L=L(t_0,t)$ the propagator computed from 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:

$$x(t) = L\, x(t_0) \tag{2.3}$$

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

$$(x;y) = < x;E\, y > \tag{2.4}$$

where E defines some weighting factors, where <..;..> is the usual Euclidean scalar product. We can also define the associated norm of a generic vector:

$$\| x \|^2 = (x;x) = < x;E\, x > \tag{2.5}$$

The norm of the perturbation at time t is given by:

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

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

$$\left(L^{*_E}x;y\right) = (x;L\, y) \tag{2.7}$$

Using the definition 2.7 in 2.6, the norm of the state vector $x(t)$ at time t can be computed in the following way:

$$\|x(t)\|^2 = (L\, x_0;L\, x_0) = \left(L^{*_E}\cdot L\, x_0;x_0\right) \tag{2.8}$$

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

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

where the matrix 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:

$$\| v_i(t) \|^2 = \left(L^{*E} \cdot L v_i(t_0); v_i(t_0) \right) = \sigma_i^2 \| v_i(t_0) \|^2 \tag{2.10}$$

The computation of the perturbations with the fastest growth of the norm 2.5 of the vector $x(t)$ is so reduced to an eigenvalue problem.

In the IFS code, the adjoint of the tangent version of the model equations has been defined with respect to the Euclidean inner product <..;..>. Let us identify with L^* the adjoint of L with respect to the Euclidean inner product <..;..>:

$$< L^* x; y > = < x; Ly > \tag{2.11}$$

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):

$$(x; Ly) = < x; E \cdot Ly > = < L^* \cdot E\, x; y > = < E^{-1} \cdot L^* \cdot Ex; Ey > = \left(L^{*E}\, x; y \right) \tag{2.12}$$

From 2.12 we have:

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

Equation 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 <..;..> is known. Using 2.13 in 2.8, the norm of a vector at time t can be computed using the following expression:

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

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.

The method implemented to solve this eigenvalue problem is based on a Lanczos algorithm developed by the Numerical Algorithm Group (NAG, Oxford, see Appendices A and B, and *Simon* for references). The Lanczos algorithm is a very useful technique when only 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 define the problem, but throughout successive applications of the operator it gives an estimate of the eigenvectors. Appendix A gives some detail about the Lanczos theory. Appendix B reports the implementation of the Lanczos algorithm in the IFS environment.

The NAG-Lanczos algorithm solves only symmetric problems, and it uses as an inner product to orthogonalize and normalize the working vectors, the Euclidean inner product <..;..>. A vector $\tilde{x}$ that has unity norm in the NAG-Lanczos code is normalized in the Euclidean space Ξ:

$$\| x \|^2_{LAN} = < \tilde{x};\tilde{x} > = 1 \tag{2.16}$$

while in the perturbation space a normalized vector is characterized by:

$$\| x \|^2 = (x;x) = < x;E\,x > = 1 \tag{2.17}$$

We can define a coordinate transformation:

$$\tilde{x} = E^{1/2}\, x \tag{2.18}$$

so that the vector $\tilde{x}$ is normalized with respect to the Euclidean inner product, and its corresponding vector x defined by the transformation 2.18 is normalized with respect to the inner product 2.4. Using 2.18 in 2.14, the norm at time t of a perturbation x is given by:

$$\begin{aligned} \| x(t) \|^2 &= \left(L^{*E} \cdot L\, x_0;x_0\right) \\ &= < E^{-1} \cdot L^{*} \cdot E \cdot L x_0;Ex_o > = < E^{-1} \cdot L^{*} \cdot E \cdot L \cdot E^{-1/2}\, \tilde{x}_0;\, E^{1/2} \cdot \tilde{x}_0 > \\ &= < E^{-1/2} \cdot L^{*} \cdot E \cdot L \cdot E^{-1/2}\, \tilde{x}_o;\tilde{x}_o > \end{aligned} \tag{2.19}$$

Starting from an initial vector $\tilde{x}_0$ in the Euclidean space Ξ, the norm at time t is determined by the operator K_{LAN}:

$$K_{LAN} = E^{-1/2} \cdot L^{*} \cdot E \cdot L \cdot E^{-1/2} \tag{2.20}$$

The eigenvalue problem of the operator K_{LAN} defined in 2.20 in the Euclidean space Ξ, is the analog of the eigenvalue problem of the operator K in the perturbation space of the system.

The NAG-Lanczos algorithm solves the eigenvalue problem for the operator K_{LAN} defined in 2.18. Note that this operator is symmetric. Each iteration of the NAG-Lanczos algorithm is defined by the action of the operator K_{LAN} on the initial vector. The initial vector $\tilde{x}_0$ is defined at the beginning of each iteration (see Appendix A for further details). Starting from $\tilde{x}_0$ in the Euclidean space Ξ, the action of the operator K_{LAN} implies: a coordinate transformation from the Euclidean space Ξ to the perturbation space, followed by the IFS tangent model integration, by its adjoint integration, and by the inverse of the coordinate transformation back to the Euclidean space Ξ.

2.2 The IFS model and the Lanczos algorithm

The IFS model is the result of a collaboration between Météo France and ECMWF. 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 with a horizontal diffusion scheme of the IFS model were available.

The first step of the Lanczos algorithm is to partially transform the matrix $K_{LAN}(n,n)$ into a tridiagonal matrix $T(j,j)$ with $j<n$ (see Appendix A for details):

$$K_{LAN} = Q \cdot T \cdot Q^t \tag{2.21}$$

The matrix Q is defined by the Krilov sub-space generated by the operator K_{LAN} (see Appendix A2), that is the sub-space generated by the q_j vectors:

$$span\left\{q_1, K_{LAN}q_1, \ldots, K_{LAN}^j q_1, \ldots\right\} \tag{2.22}$$

where q_1 is a randomly chosen starting vector. Lanczos makes successive integrations of the forward and adjoint tangent version of the direct model to compute all the q_j vectors of the Krilov sub-space and the tridiagonal matrix T. From the q_j vectors and from the matrix T it evaluates an approximation of the singular values and SVs of the operator L (see Theorem A4.2, in Appendix A). These eigenvectors, after the inverse of the transformation 2.18, define the eigenvectors of the operator K defined in 2.15 (SVs of the operator L). The square root of a singular value gives the growth rate of the norm of the corresponding SV. The fastest growing perturbations are the SVs with the highest singular values.

During all the tangent integrations, the linearization is around a basic state trajectory in phase space. The number of iterations determines the accuracy of the computations. As this number increases, more SVs can be separated from all the others, indipendently 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 singular values. Generally, 20 iterations are enough to have an acceptable precision on the largest 5 singular values. Increasing the number of iterations improves the accuracy of the following values. (The accuracy of the SVs is less than the accuracy of the singular values, say to order ϵ when the precision of the singular values is of the square order ϵ^2).

After some tests we decided to run 100 iterations for each experiment. This guarantees that 25-35 singular values are characterized by a relative error of less then 1%. Since we are interested in the definition of an unstable sub-space with a dimension of less then 20, this number of iterations is sufficient. For each singular value θ computed by the algorithm, we can define the uncertainty interval ρ:

$$|\sigma - \theta| \leq \rho \tag{2.23}$$

where σ is a correct singular value of L (eigenvalue of the K_{LAN} operator). The bound ρ is a measure of the distance between an estimated singular value and a true one. The accuracy of the first 20 values is very high. The relative error:

$$\epsilon = \left|\frac{\rho}{\theta}\right|$$

is less than our acceptable precision $\epsilon_{acc} = 0.01$ for all the singular values till the 30th-35th.

2.3 The basic state characteristics

All the experiments have starting date 17 January 1989, at 12UTC. The initial state is characterized by a zonal flow throughout the entire hemisphere. The T21L19 IFS tangent integration has been done following the trajectory computed by the same model, except when explicitly mentioned. The time evolution of the basic state suffers from the lack of the planetary boundary layer (PBL) representation: strong winds develop in the eastern part of the Himalayan region and over northern Africa. A test has been done to see if this can influence the SVs structures: results from an experiment with constant basic state has been compared to results of an experiment performed following the trajectory. Very small differences have been detected between the two.

3. 12H, 24H AND 36H SINGULAR VECTORS

In this part we will analyse the impact of the time interval over which the perturbations growth has been maximized on the SVs structures. Three time intervals have been considered: 12, 24 and 36 hours. All the

experiments have been performed with the NMI procedure applied to 5 modes: in section 3 some considerations on the impact of NMI on the SVs will be reported. First we will describe some of the fastest growing 24H SVs, and then a comparison between the three unstable sub-spaces will be reported.

Fig. 1 shows the amplification factors λ_i of the first 100 SVs of the three experiments (singular values σ_i):

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

The 24h and 36h curves are very similar to each other: the first part of them with the highest values are characterized by a steep derivative. The values decrease till a plateau is reached at about SV number 20. At about SV number 80 the SVs do not grow any more (values less then 1). The 12h singular values can not easily be separated in a group of fast growing perturbations and slowly growing perturbations.

Figures 2 and 3 show the streamfunction of the first 6 24h SVs, respectively at model level 11 (approximately 500 hPa) and at model level 18 (approximately 1000 hPa). As a general consideration we can say that all are very localized: only after the first 10-15 SVs structures that cover half the Northern Hemisphere (NH) start to be present. None of the first 20 SVs propagates into the Southern Hemisphere (SH). The first and the third SVs are localized over the eastern Pacific, with maximum amplitude in the middle troposphere between 700 hPa and 500 hPa, where the mid-latitude jet has a region of maximum amplitude. They are characterized by a barotropically and baroclinically unstable patterns, with a westward tilt with height. Horizontally they are characterized by a NW-SE elongated shape. The second SV has maximum amplitude near the surface in the eastern part of the Himalayan region. The fourth SV has maximum amplitude near the surface, over northern Africa: it might be due to instability growth in a low level shear region. The fifth SV presents some structures at model level 11, but it is almost all confined in the lowest model levels. The sixth SV presents 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* (1992) using a T21L3 QG model. As already shown in *Molteni and Palmer* (1992), the SVs have much more localized structures than the leading normal modes of a time-average basic state.

Some SVs are almost completely confined to the lowest model levels. They have a very strong baroclinic structure as can be detected from the temperature fields (not shown), with a positive perturbation at one level, a negative at the following one, then a positive again. The other SVs have more realistic structures (in the sense that they have a structure that is more similar to an error field), with maximum amplitude in the region of the mid-latitude jet streams. The growth of the "spurious" surface structures occors probably because of the absence of a PBL physics in the version of the IFS used.

The 24h and the 36h SVs present very similar structures, although sorted in a different order. Table 1 gives the unstable sub-space projection matrix M(i,j) of the first 20 24h SVs on the first 20 36h SVs, and vice-versa (remember that the Hilbert space H of the perturbations x has been defined using the inner product 2.2). The 20 values of the first column of Table 1 (elements M(1,j) for j=1,20), are the squared scalar products between the first 36h SV, and the 20 24h SVs. The last value is the sum of these squared scalar products, which represents how well the first 36h SV can be reconstructed from a linear combination of the 24h unstable sub-space generated by the first 20 SVs. The last column of the matrix gives the same information for the reconstruction of the 24h SVs from a linear combination of the 36h SVs.

Looking at Table 1, the 1st 36h SV (which has maximum amplitude near the surface in the Himalayan region) projects 82% of its norm on the 2nd 24h SV although some differences are present. The 2nd 36h SV (which has maximum amplitude near the surface over northern Africa) projects 89% of its norm on the 4th 24h SV. The 3rd and the 4th 36h SVs have maximum amplitude in the middle troposphere, over the western Pacific Ocean, and project respectively on the 1st and the 3rd 24h SVs. Analogous considerations can be made for the other SVs. It is remarkable that the unstable sub-space generated by the first 14 SVs of one experiment can explain at least 60% of the norm of the first 14 SVs of the other experiments. This indicates a strong parallelism between the two sub-spaces.

Table 2 gives the unstable sub-space relationship between the 12h and the 24h 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 norm of only 4 SVs of the other. The 12h SVs reveal the same spurious structures of the 24h and 36h SVs. Moreover, some of them are characterized by large amplitude in the upper troposphere never detected in the 24th and 36h SVs.

Looking at the first 8 SVs of each experiment, four classes 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, the first 12h SVs have maximum amplitude in the upper troposphere mainly over Europe, Asia and Africa: we will identify them as upper troposphere structures (U). The following table summarizes this:

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

The two P-SVs and the two H-SVs are present in all the experiments. A time interval longer than 12 hours is instead necessary to separate the A-SVs and the T-SV from the others.

From this section the following conclusions can be drawn:

i) The fastest growing perturbations have very localized structures similar to QG perturbations (*Molteni and Palmer*, 1992).

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

iii) Remarkable differences exist between the 12h unstable sub-space, and the 24h and 36h sub-spaces. These latter show instead a very strong parallelism.

4. NMI IMPACT ON THE SINGULAR VECTORS

Experiments have been performed with no initialization, and with the first 3, 5, 9 and 13 gravest vertical modes initialized, and with NMI applied to all the vertical modes (NO-NMI, 3-NMI, 5-NMI and so on), to study the impact of NMI on the unstable sub-space definition. The experiments have been performed over two time intervals, 12 and 24 hours. A few experiments have also been done over 36 hours to confirm the results.

Fig. 4 shows the amplification factor curves of the 12h experiments with NO-NMI, 5-NMI, 13-NMI and 19-NMI experiments. The 3-NMI and 9-NMI curves have been omitted for clearness. A large difference can be detected between the experiments with up to 9 initialized modes and the other two experiments, in the first part of the curves. The 13-NMI and the 19-NMI SVs are characterized by stronger growth. Looking at the structures of the corresponding 19-NMI SVs, they show maximum amplitude in the upper troposphere, with a horizontal structure with high wave number (approximately 16). Generally speaking, all the first 20 SVs of the 13-NMI and of the 19-NMI experiments retain this small scale pattern.

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. The 5-NMI SVs have approximately 70% of the norm projected on the potential energy, while the 19-NMI SVs with the upper troposphere high wave number features have less than 10% projected on it, and 40-50% projected on the rotational and divergent part of the kinetic energy.

The analysis of the unstable sub-spaces generated by the NO-NMI and the 3-NMI, 5-NMI and 9-NMI SVs confirm that the NMI impact is small when only the first modes are initialized. Considering for example the NO-NMI and the 5-NMI SVs optimized over a 36h time interval, the first 15 SVs of each experiment can explain at least 60% of the norm of the first 15 SVs of the other experiment.

The impact of the NMI procedure can be detected also in the 24h and 36h SVs, but it is not so evident in the amplification factor curves as it is in the 12h experiments. A longer time period integration reduces the effect of the NMI procedure on the fastest growing perturbation structures. Considering for example the NO-NMI and the 5-NMI SVs, the first 16 SVs of one experiment can explain at least 90% of the norm of the other experiment SVs.

It is worthwhile to point out one effect of the NMI procedure on the norm of the state vector. The NMI procedure projects a state vector on 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 direction, and modifying its component along the gravitational direction. We verified the existence of this effect: when only the first modes are initialized this can induce a norm increase of the order of 1%, while initializing 13 modes can produce a norm increase of 80%.

This large difference can be related to the non-convergence of the non-linear normal mode initialization procedure when applied to the shallower modes. Due to the divergence of the solving iterative algorithm, a large variation of the norm of the state vector can occur.

The following conclusions can be drawn:

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

ii) Not more than 9 modes should be initialized. The unstable sub-spaces generated by the NO-NMI, 3-NMI, 5-NMI and 9-NMI experiments show strong similarity.

iii) The NMI 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.

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 NMI procedure has been applied to the first 5 modes. Two optimisation time intervals have been considered, 12 and 24 hours.

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

$$\| x \|^2 = \frac{1}{2} \int_0^1 \iint_\Sigma (\nabla\Delta^{-1}\zeta \cdot \nabla\Delta^{-1}\zeta$$

$$+ \nabla\Delta^{-1}D \cdot \nabla\Delta^{-1}D + R_a T_r (\ln\pi)^2 + \frac{C_p}{T_r} T^2) \, d\Sigma \left(\frac{\partial p}{\partial \eta}\right) d\eta \qquad (5.1)$$

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 3rd and the 4th terms have not been considered.

Fig. 5 shows the SVs' amplification factors of two experiments run with an optimisation time interval of 24h. It shows that the amplification factors in the T-experiment are almost twice as large as those in the K-experiment.

The 24h K-SVs are characterized by patterns similar to the 24h T-SVs: the first 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 to be combinations of T-SVs, although they present structures that extend into the upper troposphere. Some of them (e.g. the 9th, the 10th and the 11th), present very high amplitude in the upper troposphere, with structures never detected in the T-SVs.

We analysed the norm variation due to the NMI procedure on the 24h K-experiments state vector. NMI 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-exp, that were always characterized by a slight increase (less then 1%).

The 12h experiments confirm these results: moreover the impact of the norm definition on the SVs structures is more evident. The first 8 K-SVs have maximum amplitude in the upper troposphere, with a pattern never shown by the T-SVs. The NMI procedure induces higher changes in the state vector norm: it produces a decrease of at least 4%, with a peak value of 9% at the first algorithm iteration. This NMI effect has never been observed during the 12h T-exp, where NMI produced always a slight increase of almost 1%.

6. DIFFUSION COEFFICIENTS SENSITIVITY

The horizontal diffusion term in the IFS model, for a general variable y, has been defined in the following way:

$$\frac{\partial y}{\partial t} + H\, y = 0 \tag{6.1}$$

where the diffusion operator H, in the spectral space, is:

$$h\,(n) = \frac{1}{\tau} f\left(\frac{n}{N}\right) \tag{6.2}$$

where the constant N defines the truncation, and τ and is the characteristic damping time for wave number N. The function $f(x)$ is given by:

$$f(x) = 0,\ \forall x\ 0 \le x < 0.5$$
$$f(x) = (2x - 1)^2,\ \forall x\ 0.5 \le x \le 1. \tag{6.3}$$

The horizontal diffusion can be applied to vorticity, divergence and temperature.

All the results reported in the previous sections have been obtained with a diffusion coefficient $\tau = 48h$ constant for all the variables. The wave number range on which horizontal diffusion operates has been kept equal to half the wave spectrum ($11 \le n \le 21$).

Apart from the control experiment with a damping time of 48 hours, two more experiments have been run, with damping times of 24 and 4 hours. The SVs of these experiments have been optimized over a time interval of 24 hours (the same as the control). We will call them D24-exp and D4-exp, while the control D48-exp.

The D24 and the D48 amplification factor curves have very similar values, while the D4 curve has values up to 30% smaller than the control. Differences can also be detected comparing the D4-SVs and the D48-SVs: the D4-SVs are less localized in space: from the 6th SV their structures cover almost all the NH. The growth of the D4-SVs near the surface is reduced, although it is still present. The comparison of the unstable sub-space of the three experiments confirm these considerations, with the D4 and the D48 sub-spaces showing a poor parallelism.

The same conclusions can be drawn from the analysis of the impact of the horizontal diffusion coefficient when an optimisation time interval of 12h is considered.

7. 24H SINGULAR VECTORS TIME EVOLUTION

The results of the experiments reported in the previous sections identify the configuration with an optimisation time interval of 24 hours, the first 5 vertical modes initialized, and with the norm defined to be the total energy as the one giving the most acceptable results. In this section we will study the time evolution of two of the SVs derived from this configuration.

Fig. 6 shows the time evolution of the 1st SV: the temperature field is plotted at three vertical levels at the starting time (t=0, left hand side), and after 24 hours tangent time integration using the IFS T21L19 model. Note that the meridional phase tilt is reversed from the initial to the finale state, which indicates that barotropic energy conversion from the basic state to the perturbation changes from positive to negative within 24 hours period. Fig. 7 shows the vertical cross section of the SV total energy at the initial state, after the application of the NMI procedure and after 24 hours. 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), point out how the perturbation can grow close to the surface.

Fig. 8 shows the time evolution of the perturbation, when superimposed to the same initial condition in the T63L19 version of the full ECMWF model. The temperature at model level 11 (500 hPa) has been reported for different integration time intervals. The growth of the perturbation during the first 24 hours is very similar to the linear growth.

The IFS SVs have very similar structures to the SVs computed using a T21L3 version of a QG model (*Molteni and Palmer*, 1992). Fig. 9 shows the growth of a perturbation computed using the QG model that has a pattern similar to the 1st IFS SV. It is remarkable how similar the initial structures and their time evolutions are, although they have been computed with very different models. The following table gives the amplification factors of the T21L19 IFS 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:

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

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 side panels show the initial

state at model levels 17, 18 and 19, the right hand side panels the temperature after 24 hours 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 in less than 24 hours.

8. CONCLUSIONS

This study shows how the adjoint technique suggested by *Lorenz* (1965) can be applied to the computation of the unstable sub-space of a primitive equations model. The fastest growing perturbations are the singular vectors (SVs) of the tangent linear version of the primitive equation model, identified by the linear operator L. A Lanczos algorithm was applied since only a few of the extreme eigenvalues and eigenvectors were required. The sensitivity results presented above should be considered as preliminary results, since all the experiments were characterized by the same initial date.

The 24h and 36h experiments performed with the first five vertical modes initialized present very similar results. The two unstable sub-spaces generated by the first 20 SVs are very similar. Four classes of structures can be identified in the sub-spaces: two with maximum amplitude in the middle troposphere, and two with higher amplitude near the surface. A time interval of 12 hours seems to be too short to let a clear identification of the fastest growing SVs.

The NMI procedure has a large impact on the SVs when more than 9 modes are initialized, and this confirms that only the first modes should be initialized (*Williamson and Temperton*, 1981). The impact was more detectable in the 12h experiments. Strong similarity was shown by the unstable sub-spaces generated by the NO-NMI, 3-NMI, 5-NMI and 9-NMI experiments.

An increase of the horizontal diffusion coefficients of an order of magnitude has a remarkable effect on the unstable sub-space definition. This effect is stronger if a 12 hours optimisation time period is considered.

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.

The time evolution of the fastest growing IFS SV computed using the tangent linear version of the T21L19 IFS model is very similar to its T63L19 full ECMWF model evolution. Moreover, the similarity between

the full model evolution of the 1st IFS SV and the 1st T21L3 QG SV (*Mureau et al.* 1992) is remarkable. The IFS SVs with maximum amplitude near the surface are quickly damped when integrated using the T63L19 full ECMWF model.

Generally speaking, the results seem to point out that a 24h optimisation time interval is better than a 12h interval to separate the SVs. Moreover, when NMI is applied to the computation, only the first modes should be initialized to avoid spurious effect to appear in the results. After the implementation of a simple PBL physics, the next step of our study will be to compare the perturbations with forecast error patterns. Moreover we will seek to confine the perturbation to the Northern Hemisphere. A more detailed comparison will be done with the quasi-geostrophic model perturbations. The goal of the project is to realize ensemble forecasts perturbing the initial condition along the unstable directions computed using the IFS model.

ACKNOWLEDGMENTS

The prelimary work implementing the Lanczos code into the IFS model was performed by J Tribbia. I would like to thank him, and F Molteni and T Palmer for useful technical discussions, and R Mureau who also provided me with the T63L19 time evolutions. Thanks also to C Temperton for the helpful discussion on NMI and to J Du Croz of NAG, Oxford for the discussion on the Lanczos algorithm theory.

APPENDIX A1: EIGENVALUE PROBLEM: THE LANCZOS METHOD

The Lanczos algorithm was first proposed in 1950. Lanczos intended his 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 eigen-problems. The following notes have been based on chapter 9 of *Golub and Van Loan, "Matrix computations"*, (1983). The aim is to give a basic idea of the method.

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 tranformation of the matrix $A(n,n)$, that is to find two matrices $Q(n,j)$ (unitary matrix) and $T(j,j)$ such that:

$$Q^t \cdot A \cdot Q = T \tag{A1.1}$$

where the matrix T is tridiagonal:

$$\begin{matrix} \alpha_1 & \beta_2 & 0 & 0 & . & . & . \\ \beta_2 & \alpha_2 & \beta_3 & 0 & . & . & . \\ 0 & \beta_3 & \alpha_3 & . & . & . & . \\ 0 & 0 & . & . & . & 0 & 0 \\ . & . & . & . & . & \beta_{j-1} & 0 \\ . & . & . & 0 & \beta_{j-1} & \alpha_{j-1} & \beta_j \\ . & . & . & 0 & 0 & \beta_j & \alpha_j \end{matrix} \tag{A1.2}$$

and

$$Q = [q_1, q_2, \ldots, q_j] \tag{A1.3}$$

where the vectors q_i of A1.3 are column vectors, and where $j<n$.

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

$$T = S^t \cdot D \cdot S \tag{A1.4}$$

where the matrix D is diagonal

$$D = diag\ [\theta_1,\theta_2,...,\theta_j] \tag{A1.5}$$

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

$$Y = Q \cdot S$$

$$Y = [y_1,y_2,..,y_j] \tag{A1.6}$$

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

APPENDIX A2: THE LANCZOS THEORY

In this section the theory of the Lanczos algorithm and a review of the essential mathematical theorems on which it is based, will be summarized.

Hypothesis A2.1: suppose

$$A \in R^{nXn} \tag{A2.1}$$

is large, sparse and symmetric, and that few of its largest and/or smaller eigenvalues are desired.

Definition A2.1: the Rayleigh quotient is defined by the following expression

$$r(x) = \frac{x^t \cdot A \cdot x}{x^t \cdot x} \tag{A2.2}$$

where x is a column vector.

Definition A2.2: the eigenvalues of a matrix A will be identified by

$$\lambda_n(A) \le \ldots \le \lambda_j(A) \le \ldots \le \lambda_1(A) \tag{A2.3}$$

The following relation holds between the eigenvalues of A and the Rayleigh quotient:

$$\lambda_n(A) \le r(x) \le \lambda_1(A) \tag{A2.4}$$

The relation A2.4 is very useful and can be used for the evaluation of the eigenvalues. Suppose that

$$\{q_i\} \in R^{nXn} \tag{A2.5}$$

is a sequence of orthogonal vectors, and define:

$$M_j = \lambda_1\left(Q_j^t \cdot A \cdot Q_j\right) = \max_y \frac{y^t \cdot (Q_j^t \cdot A \cdot Q_j) \cdot y}{y^t \cdot y} = \max_y r(Q_j \cdot y) \le \lambda_1(A) \tag{A2.6a}$$

$$m_j = \lambda_j\left(Q_j \cdot A \cdot Q_j\right) = \min_y \ldots = \min_y r(Q_j \cdot y) \ge \lambda_n(A) \tag{A2.6b}$$

where

$$Q_j = [q_1,\ldots,q_j] \tag{A2.7}$$

The Lanczos procedure can be derived by considering how to generate the vectors defined by A2.5 so that the scalars defined by A2.6 are increasingly better estimates of the extreme eigenvalues of A defined by A2.3.

This request, that the scalars A2.6 would be increasingly better estimates of the eigenvalues A2.3 of **A**, forces the choice of the vector

$$q_{j+1} \tag{A2.8}$$

computed at step j+1. It must be chosen in the direction of the gradient of the Rayleigh quotient A2.2:

$$\nabla_x r(x) \tag{A2.9}$$

Since the following relations hold

$$\nabla_x r(x) \in span\ \{A \cdot x, x\}$$

$$\nabla_x r(x) \in span\ \{q_1,...,q_{j+1}\} \tag{A2.10}$$

to satisfy A2.10 we must have

$$span\ \{q_1,...,q_{j+1}\} = span\ \{q_1, A \cdot q_1,...,A^{j-1} \cdot q_1\} = \mathrm{X}(A,q_1,j) \tag{A2.11}$$

where

$$X(A,q_1,j) \tag{A2.12}$$

is the Krilov sub-space generated by the matrix A.

Equation A2.11 answers the question of the choice of the vectors A2.5 from a theoretical point of view. These vectors A2.5 are the columns of the Krilov matrix

$$K_j = \left[q_1, A \cdot q_1,...,A^{j-1} \cdot q_1\right]$$

Practically, since the matrix A is supposed to be sparse, the necessity is to write a procedure numerically not too expensive, that evaluates the vectors A2.5. This algorithm can be built up writing explicitly the relation A1.1, where the matrix Q is the one defined by A2.7. First, is section 3, we will prove that for every matrix defined as in Hypothesis A2.1, it is always possible to find a decomposition A1.1. Then, in section 4, we will deduce the algorithm that computes the decomposition matrices.

APPENDIX A3: THE SCHUR DECOMPOSITION

The following theorems deal with matrix decomposition.

Theorem A3.1 (Schur decomposition): if

$$A \in C^{n \times n}$$

then there exists a unitary matrix

$$Q \in C^{n \times n}$$

such that

$$Q^t \cdot A \cdot Q = T = D + N \tag{A3.1}$$

where D is strictly diagonal, and N is strictly upper triangular. The decomposition A3.1 is also called Schur decomposition.

Definition A3.1: a matrix A is normal if

$$A \cdot A^t = A^t \cdot A \tag{A3.2}$$

Corollary A3.1: a matrix $A(n,n)$ is normal if and only if there exists a unitary matrix $Q(n,n)$ such that

$$Q^t \cdot A \cdot Q = D = diag\,(\lambda_1,\ldots,\lambda_n) \tag{A3.3}$$

Theorem A3.2 (Jordan decomposition): if

$$A \in C^{n \times n}$$

then there exists a non-singular matrix

$$X \in C^{n \times n}$$

such that

$$X^{-1} \cdot A \cdot X = diag\,(J_1,\ldots,J_k)$$

where the matrices J_j are in the Jordan form

$$\begin{matrix} \lambda_1 & 1 & 0 & 0 & . & . & 0 \\ 0 & \lambda_2 & 1 & 0 & . & . & 0 \\ 0 & 0 & \lambda_3 & 1 & 0 & . & 0 \\ 0 & . & . & . & . & . & 0 \\ 0 & . & . & 0 & \lambda_{n-2} & 1 & 0 \\ 0 & . & . & . & 0 & \lambda_{n-1} & 1 \\ 0 & . & . & . & . & 0 & \lambda_n \end{matrix} \tag{A3.4}$$

In case of a real symmetric matrix A, the Schur decomposition theorem gives the following corollary:

Corollary A3.2: if $A(n,n)$ is a real symmetric matrix, then there exists an orthogonal real matrix Q such that

$$Q^t \cdot A \cdot Q = diag\ (\lambda_1,...,\lambda_n) \tag{A3.5}$$

Theorem A3.1 gives the Schur decomposition of a complex matrix. If the complex matrix is normal, corollary A3.1 proves that it is possible to find a diagonal matrix such that equation A3.3 holds. Theorem A3.2 gives the Jordan decomposition for a complex matrix. Corollary A3.2 gives the Schur decomposition for real symmetric matrices.

The practical implementation of an algorithm capable of giving this decomposition will be described in the following section A4. From now on we will deal with real symmetric matrices: the possibility of writing a resolvent algorithm for our problem is so based on corollary A3.2.

APPENDIX A4: THE ALGORITHM PROCEDURE

The Schur decomposition theory applied to a real symmetric matrix leads to the formula A3.5 of corollary A3.2, which involves the diagonalization of the starting matrix A. Instead of the direct computation of this diagonal matrix, the practical implementation of an algorithm capable to compute the Schur decomposition first makes use of a tridiagonal matrix T, as defined by A1.2. Then it finds the diagonal decomposition of this tridiagonal matrix.

Suppose the matrix A is as in Hypothesis A2.1, that the matrix Q is defined by A2.7, and that T has the tridiagonal form A1.2, and that these matrices are related one each other by equation A1.1. Equation A1.1 can be written as

$$A \cdot Q = Q \cdot T \tag{A4.1}$$

Equating columns of A4.1 we have

$$A \cdot q_i = \beta_{i-1}\, q_{i-1} + \alpha_i\, q_i + \beta_i\, q_{i+1} \tag{A4.2}$$

for $i = 1,...,j-1$.

The orthogonality of the vectors q_i implies that

$$\alpha_i = q_i^t \cdot A \cdot q_i \tag{A4.3}$$

Moreover, if

$$r_i = (A - \alpha_i I) \cdot q_i - \beta_{i-1}\, q_{i-1} \tag{A4.4}$$

is non zero, then

$$q_{i+1} = \frac{r_i}{\beta_i} \tag{A4.5}$$

where

$$\beta_i = \pm \| r_i \|_2 \tag{A4.6}$$

An iterative application of these formulae defines the Lanczos iteration procedure (the starting vector is randomly chosen):

$r_0 = q_1,\ \beta_0 = 1,\ q_0 = 0,\ j = 0$

do while $(\beta_j \neq 0)$

$$q_{j+1} = \frac{r_j}{\beta_j}$$

$$j = j + 1$$

$$\beta_j = \| r_j \|_2$$

$$\alpha_j = q_j^t \cdot A \cdot q_j$$

$$r_j = (A - \alpha_j I) \cdot q_j - \beta_{j-1}\, q_{j-1} \qquad \text{(A4.7)}$$

The vectors q_j are called Lanczos vectors.

The algorithm A4.7, after the j-th iteration, gives the elements of the matrices

$$Q_j = [q_1,...,q_j],\ Q_j \in R^{nXj}$$

$$T_j = \{\alpha_n,\ n=1,j;\ \beta_n,\ n=1,j-1\},\ T_j \in R^{jXj} \qquad \text{(A4.8)}$$

The following two theorems are very important: the first proofs that the algorithm is well written, the second relates the eigensystem of A with the computed matrices Q, T, and the diagonal decomposition of T.

Theorem A4.1: let

$$A \in R^{nXn} \qquad \text{(A4.9)}$$

be a symmetric matrix, and let the vector

$$q_1 \in R^n,\ with\ \| q_1 \|_2 = 1$$

The Lanczos algorithm A4.7 runs until iteration j=m, where

$$m = rank\,[q_1, A \cdot q_1, ..., A^{n-1} \cdot q_1] \qquad \text{(A4.10)}$$

Moreover we have that

$$A \cdot Q_j = Q_j \cdot T_j + r_j\, e_j^t,\ j=1,...,m \qquad \text{(A4.11)}$$

and Q has orthonormal columns.

Theorem A4.1 guarantees that if the rank of the Krilov space associated to the matrix A defined by A4.9 is smaller than n, the algorithm stops.

Theorem A4.2: suppose that j-steps of the algorithm A4.7 have been performed, and that

$$S_j^t \cdot T_j \cdot S_j = diag(\theta_1,...,\theta_j) \tag{A4.12}$$

is the Schur decomposition of the tridiagonal matrix T_j. If we define

$$Y_j = [y_1,...,y_j] = Q_j \cdot S_j \qquad Y_j \in R^{nXj} \tag{A4.13}$$

then we have

$$\| A \cdot y_i - \theta_i y_i \|_2 = |\beta_j||s_{ij}| \tag{A4.14}$$

where

$$S_j = (s_{pq}) \tag{A4.15}$$

Theorem A4.2 solves the second point of section 1, in particular guarantees the validity of the equation A1.2 and that the elements of the diagonal matrix D are an approximation of the eigenvalues of the matrix A. The theorem provides a computable estimate of the approximation of the computed eigenvalues: from A4.14 it is easy to deduce that

$$\min_{\mu\in\lambda(A)} |\theta_i-\mu| \le |\beta_j||s_{ij}| \qquad i = 1,...,j \tag{A4.16}$$

The couples

$$(\theta_i,y_i) \tag{A4.17}$$

are called the Ritz pairs of A relative to the sub-space $P(Q_j) = \{y \in R^n \mid y=A \cdot x \text{ for some } x\in R^j\}$ where the Ritz pairs are defined by the following theorem.

Theorem A4.3: suppose A as in A4.9 is symmetric, and that

$$Q_1 \in R^{nXk},\ Q_1^t \cdot Q_1 = I_k$$

If

$$Z^t \cdot \left(Q_1^t \cdot A \cdot Q_1\right) \cdot Z = D = diag(\theta_1,...,\theta_k) \tag{A4.18}$$

is the Schur decomposition of

$$Q_1^t \cdot A \cdot Q_1 \tag{A4.19}$$

and

$$Q_1 \cdot Z = Y = [y_1,...,y_k] \tag{A4.20}$$

then the matrices D, Y define the Ritz pairs, and the following relation holds

$$\| A \cdot y_i - \theta_i y_i \|_2 = \| \left(I - Q_1 \cdot Q_1^t\right) \cdot A \cdot Q_1 \cdot Z \cdot e_i\|_2 \le \|\left(I - Q_1 \cdot Q_1^t\right) \cdot A \cdot Q_1 \|_2 \tag{A4.21}$$

for i=1,...,k.

An indication of the rate of convergence of the Lanczos algorithm can be deduced from the following theorem and corollary (Kaniel-Paige Theory):

Theorem A4.4: let A be as in A4.9, symmetric, with eigenvalues and corresponding eigenvectors

$$\lambda_1 \geq \ldots \geq \lambda_n,\ z_1,\ldots,z_n \tag{A4.22}$$

If

$$\theta_1 \geq \ldots \geq \theta_j$$

are the eigenvalues of T_j obtained after j-steps of the Lanczos algorithm, then

$$\lambda_1 \geq \theta_1 \geq \lambda_1 - \frac{(\lambda_1 - \lambda_n)\tan(\phi_1)^2}{\left[c_{j-1}(1 + 2\rho_1)\right]^2} \tag{A4.23}$$

where

$$\cos\phi_1 = |q_1^t z_1| \qquad \rho_1 = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n} \tag{A4.24}$$

and

$$c_{j-1}(x)$$

is the j-1 Chebychev polynomial.

Corollary A4.1: using the same notation as in theorem A4.4,

$$\lambda_n \leq \theta_j \leq \lambda_n + \frac{(\lambda_1 - \lambda_n)\tan(\phi_n)^2}{\left[c_{j-1}(1 + 2\rho_n)\right]^2} \tag{A4.25}$$

where

$$\rho_n = \frac{\lambda_{n-1} - \lambda_n}{\lambda_1 - \lambda_{n-1}} \qquad \cos(\phi_n) = |q_1^t z_n|$$

Estimates can be done also for the interior eigenvalues: the bounds are not so satisfactory as A4.24 and A4.25. For details, see *Kaniel* (1966), *Paige* (1971), or *Scott* (1978).

In conclusion, the Lanczos algorithm defined by A4.7 gives the decomposition A1.1 of a real symmetric matrix A into a tridiagonal matrix T. The diagonal decomposition of this matrix T given by theorem A4.2 gives the Ritz pairs A4.17 that provides an estimate of the eigensystem of the matrix A. Theorem A4.4 and Corollary A4.1 gives the rate of convergence of the procedure.

APPENDIX B1: THE NAG LANCZOS CODE IN THE IFS ENVIRONMENT

The NAG Group (Numerical Algorithm Group Ltd, Oxford) developed a code based on the Lanczos theory that computes the eigensystem of a real symmetric matrix. A pre-release of this code has been passed to ECMWF. The code is called LANSO for Lanczos Algorithm with Selective Orthogonalization. The code finds selected eigenvalues, and optionally the corresponding eigenvectors, for a real symmetric generalized eigenvalue problem

$$K \cdot x = \lambda M \cdot x \qquad \text{(B1.1)}$$

where the matrices K, M are symmetric, and M is positive semi-definite.

The two matrices need not be explicitly stored, but are defined by the action of two user-supplied subroutines OPK and OPM.

Let us continue using the matrix language instead of an operator language (since there is a 1:1 correspondence between linear operators and matrices). Let the matrix L identify the tangent version of the operational model, available in the IFS environment. Our problem is to compute the eigensystem for the matrix

$$L^t \cdot L \qquad \text{(B1.2)}$$

where the transpose of L identifies the adjoint of the tangent version of the operational model with respect to the chosen norm.

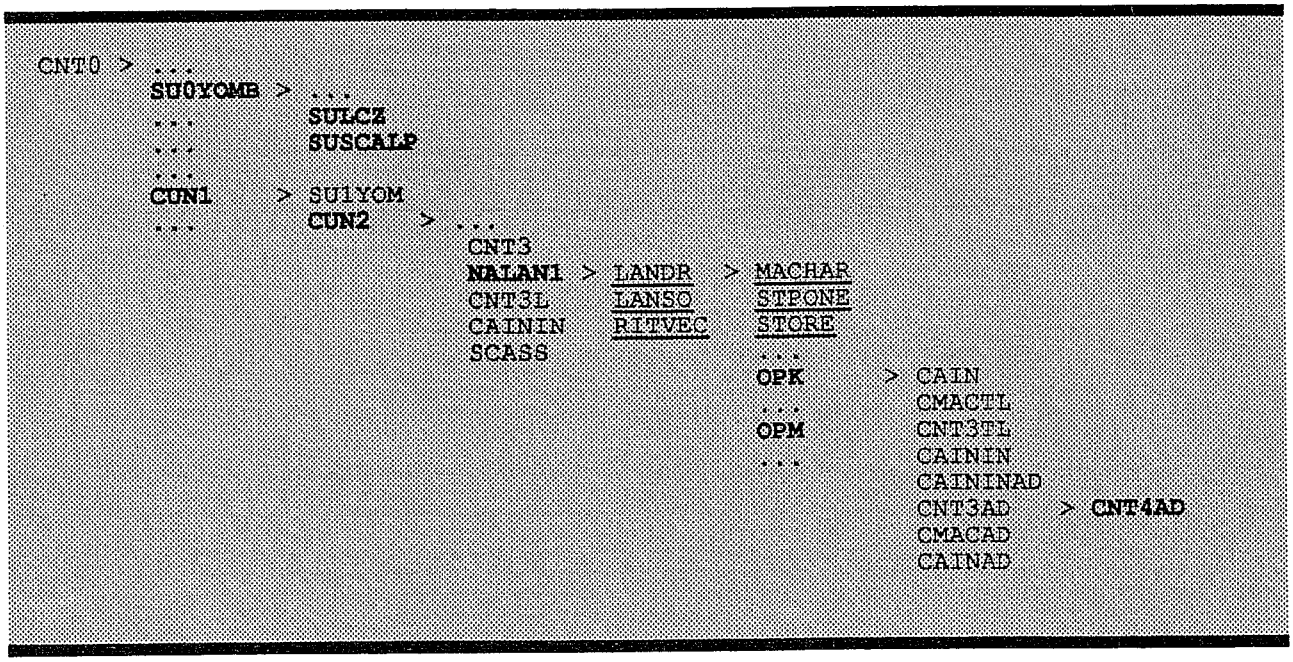
The matrix K in B1.1 in the Lanczos code must be equal to the matrix defined by the product B1.2. In our case the matrix M in B1.1 is the identity. The main steps of the eigensystem computation using the LANSO code in the IFS environment can be represented in the following way:

- set up the IFS variables;
- evaluate the trajectory integrating the full model equations;
- call the LANSO code, specifying the number of iteration and the number of required Ritz pairs. For each iteration, the LANSO code calls (among others NAG routines) the routine OPK which:
  - integrates the forward tangent model;
  - integrates the adjoint version of the tangent model;
- output the computed Ritz pairs.

The output of the LANSO code, after n iterations, consists of:

- n Ritz values, each characterized by a different bound, with higher precision at the ends of the spectrum;
- corresponding Ritz vectors that give an estimate of the eigenvectors.

The flowtree of the IFS-Lanczos evaluation of the eigensystem is the following:



The routines written in bold characters are either modifications of existing IFS routines, modified for a better use in the eigensystem evaluation with the Lanczos algorithm (IFS configuration 601), or routines set up for the computation. The underlined routines are part of the NAG code.

The more important routines are briefly described below:

SULCZ: (set up Lanczos) defines the Lanczos control variables and flags: it reads the IFS input namelist and update the default values;

SUSCALP: (set up scalar product) defines the norm;

CUN1: controls the job at level 1;

CUN2: controls the job at level 2; it initializes the Lanczos computation, it computes the trajectory, it calls the Lanczos routine at level 1, and it checks the growth rate of the first 5 unstable modes;

NALAN1: controls the Lanczos computation at level 1: it specifies the constants to control the Lanczos algorithm, it calls the first NAG routine LANDR, and it writes the computed eigensystem;

LANDR: is the first routine of the NAG code (see NAG documentation for further details), it calls the routine LANSO that does the Lanczos computation;

LANSO: makes the computation of the eigensystem: for each iteration it calls the two routines OPK and OPM that must be supplied by the user;

OPK: defines the operator K of the Lanczos algorithm (see text): in our case it calls the tangent linear and the adjoint model integration;

OPM: defines the operator M of the Lanczos algorithm (see text): in our case M is the identity;

Some of the variables that control the Lanczos computation can be changed for each integration: a value different from the default set by SULCZ can be specified in the IFS control namelist file, in the **NAMLCZ** namelist. These variables are:

NITERL: (integer) number of maximum Lanczos iterations (default = 100);

NWEIG: (integer) maximum number of wanted eigenpairs (default = 50);

NLANTYPE: (integer) it controls the type of Lanczos problem, defining the operator OPK that calls the IFS integration:

- 1 for only tangent linear integration;
- 2 for only adjoint integration;
- 3 for linear and adjoint integration (default);

NLANNORM: (integer) defines the norm which is set up by the routine SCALP:

- 1 for total energy (default);
- 2 for only kinetic energy;

XKAPA: (real) defines the acceptable precision for the eigenvalue computation (default = 0.01);

LANNMI: (logical) controls the use of NNMI:

- .T. to call the NNMI (default);
- .F. not to call it;

APPENDIX C1: TABLES AND FIGURES

Eigensystem projection

Horizontal axis index: SV optimized over 36 hours
Vertical axis index: SV optimized over 24 hours
M(i,j)=NINT(100*((SV24(i)*SV36(j))**2))

	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	

Table 1. Unstable sub-space projection matrix M. The indices on the vertical axis refer to the 24h 5-NMI SVs, the indices on the horizontal top axis to the 36h 5-NMI experiment. The element M(i,j) of the matrix M gives the scalar product between the i-th 24h SV and the j-th 36h SVs. For each line, the last column gives the % of norm of each of the 24h SVs that is explained by the 36h sub-space. For each column, the last line gives the analogous information for the 36h SVs norms explained by the 24h sub-space.

Eigensystem projection

Horizontal axis index: SV optimized over 12 hours
Vertical axis index: SV optimized over 24 hours
M(i,j)=NINT(100*((SV24(i)*SV12(j))**2))

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

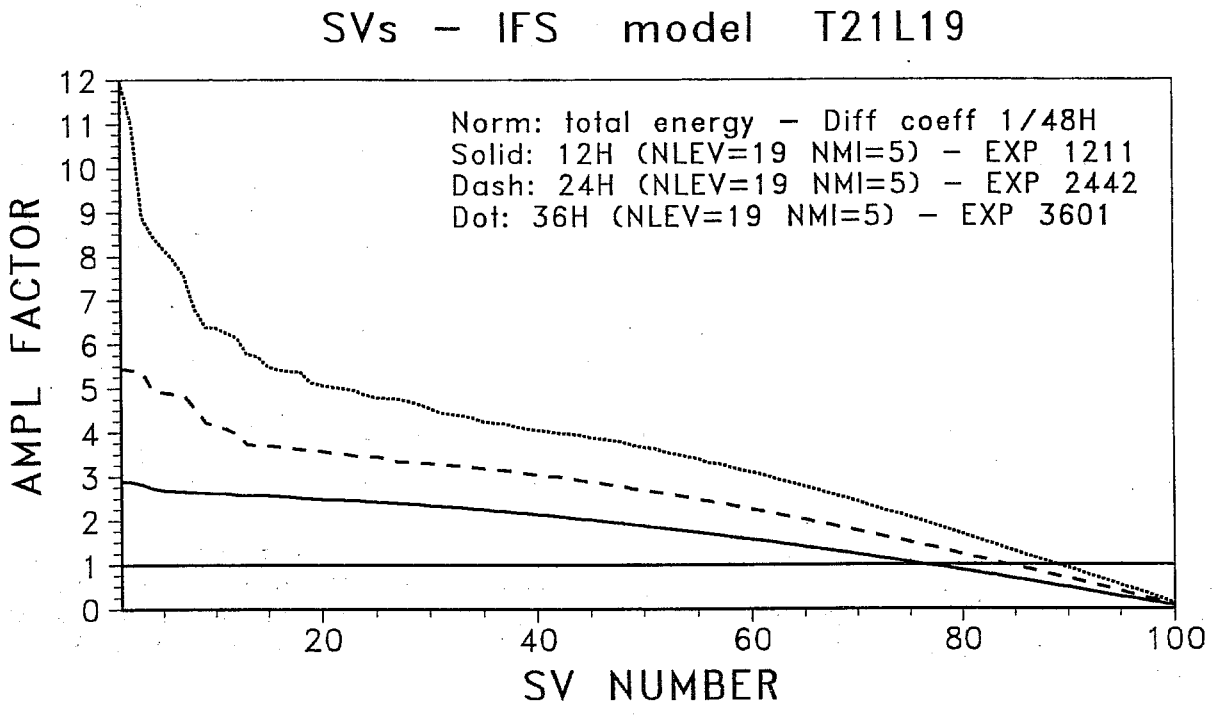


Figure 1. Amplification factors for the SVs optimized over a 12h, 24h and 36h time interval.

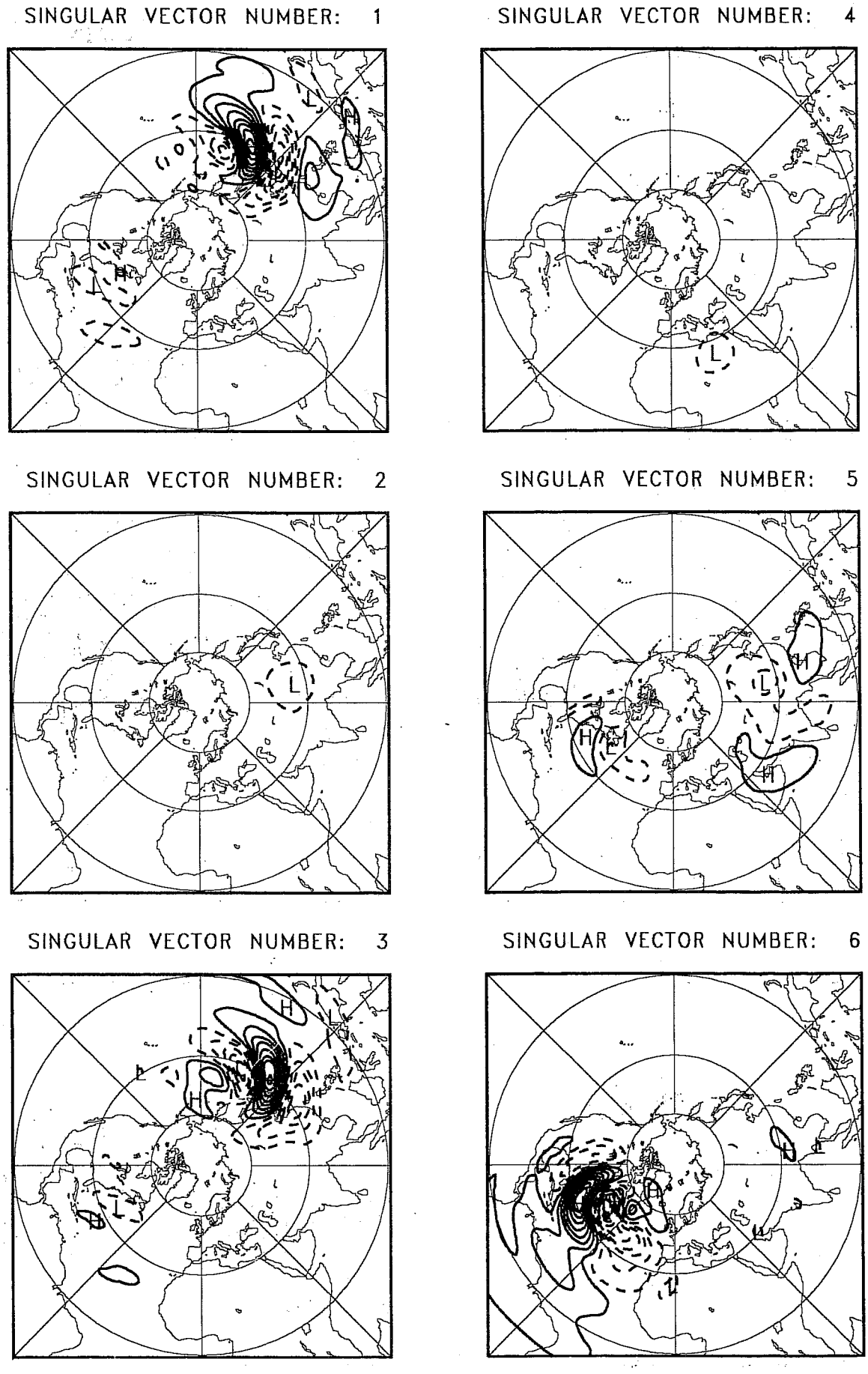


Figure 2. 500hPa streamfunction of the first six SVs optimized over 24h (increasing SV number from top-left to bottom-left, from top-right to bottom-right).

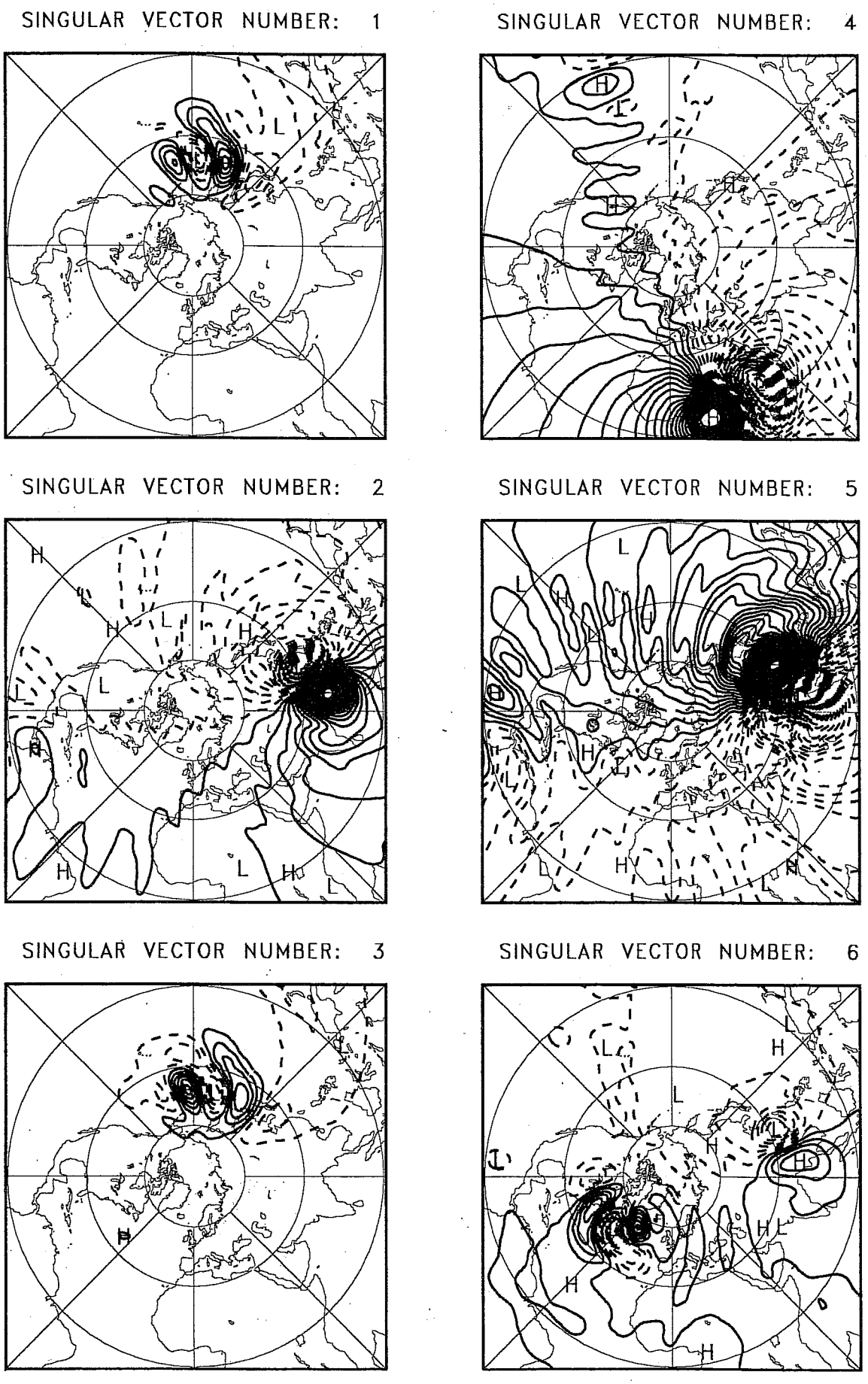


Figure 3. As in Fig. 2 but for model level number 18 (approximately 1000hPa).

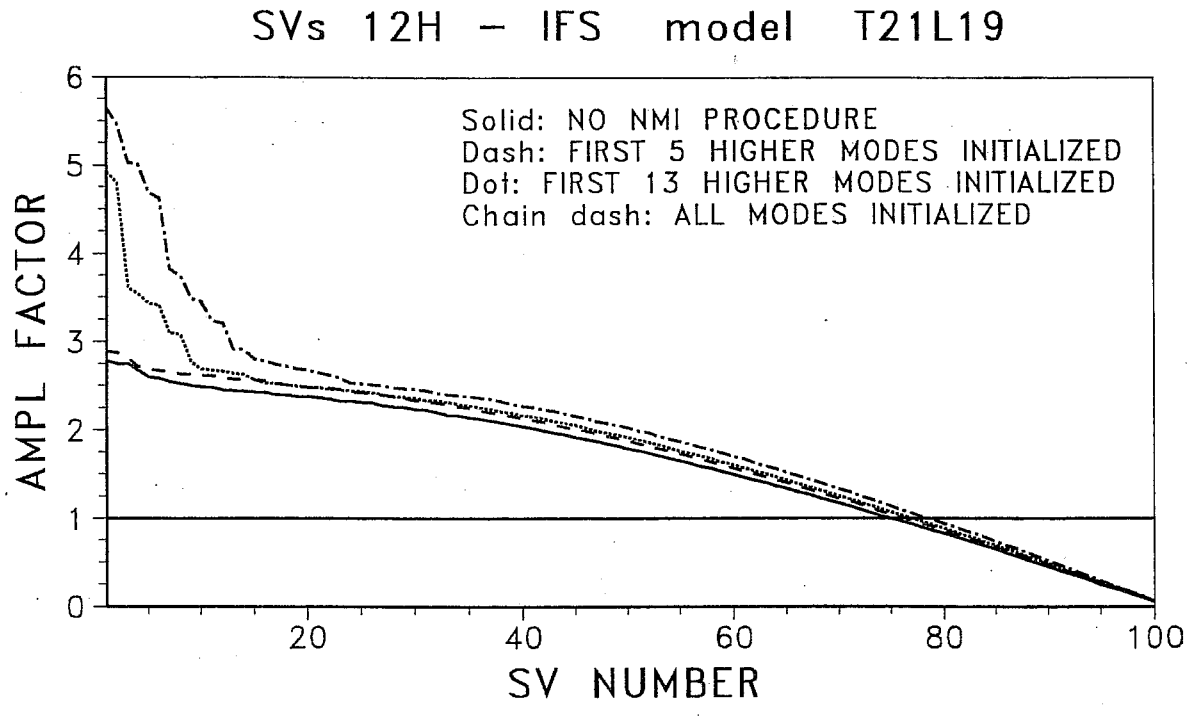


Figure 4. Amplification factors for the SVs optimized over a 24h time interval, not using the NMI procedure, applying it to the first 5, 13 vertical modes, and initializing all the vertical modes.

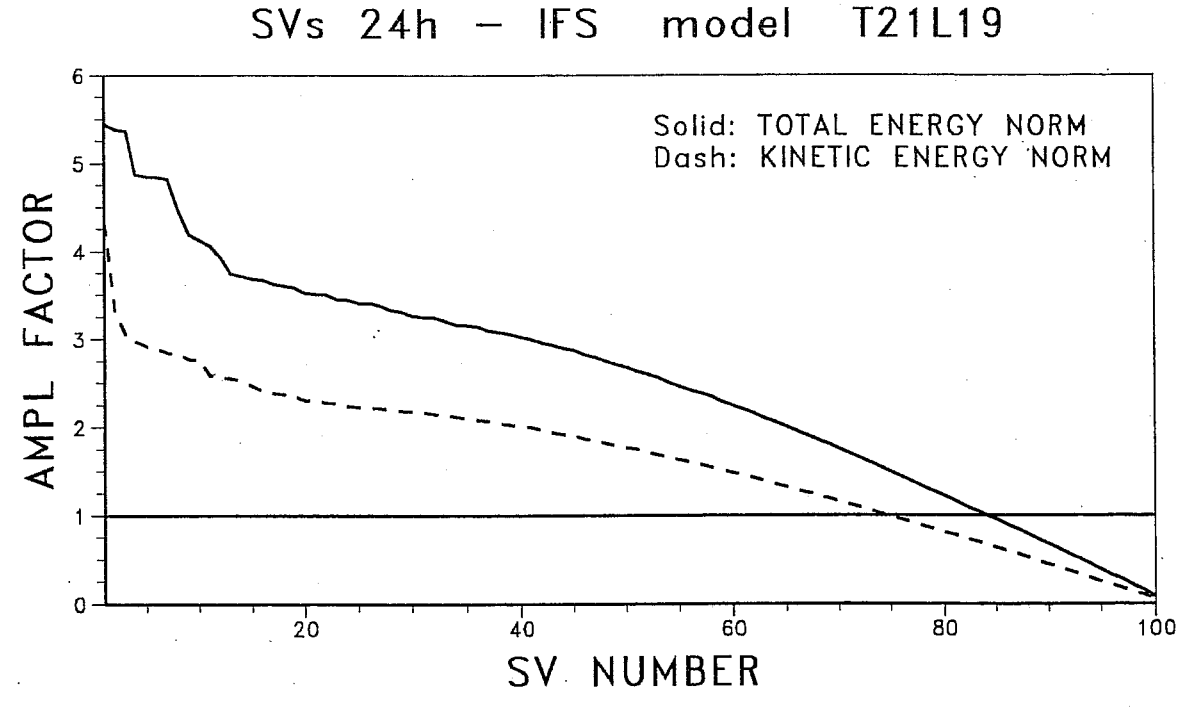


Figure 5. Amplification factors of the SVs optimized over a 24h time interval, computed with the total energy and the kinetic energy norm.

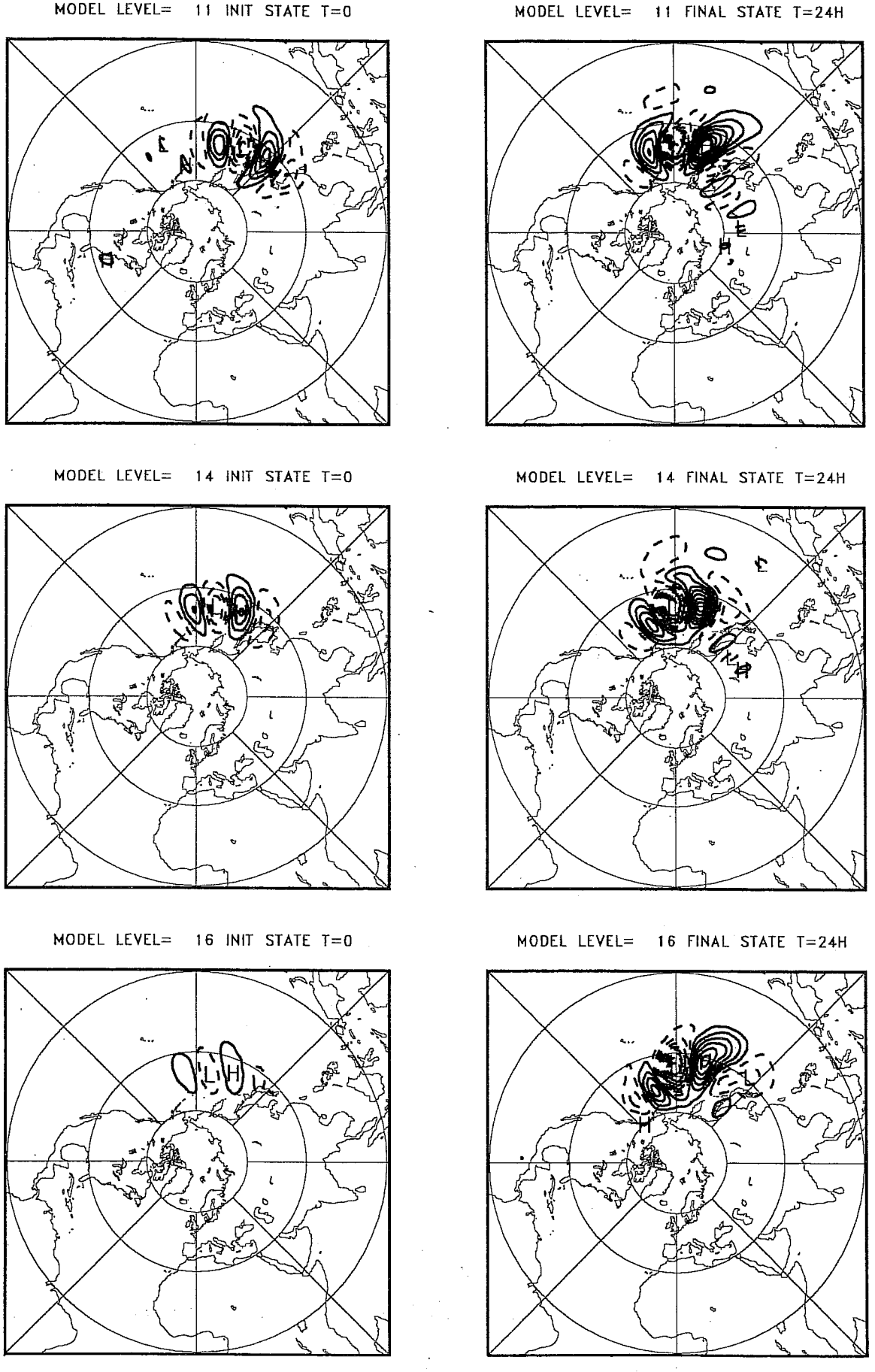


Figure 6. Tangent linear time evolution of the fastest growing SV computed over a 24h time interval, with NMI applied to the first 5 modes. The temperature field has been plotted at initial time t=0, and after 24h tangent linear integration, at three model levels: L=11: approximately 500hPa;
L=14: " 800hPa;
L=16: " 900hPa.

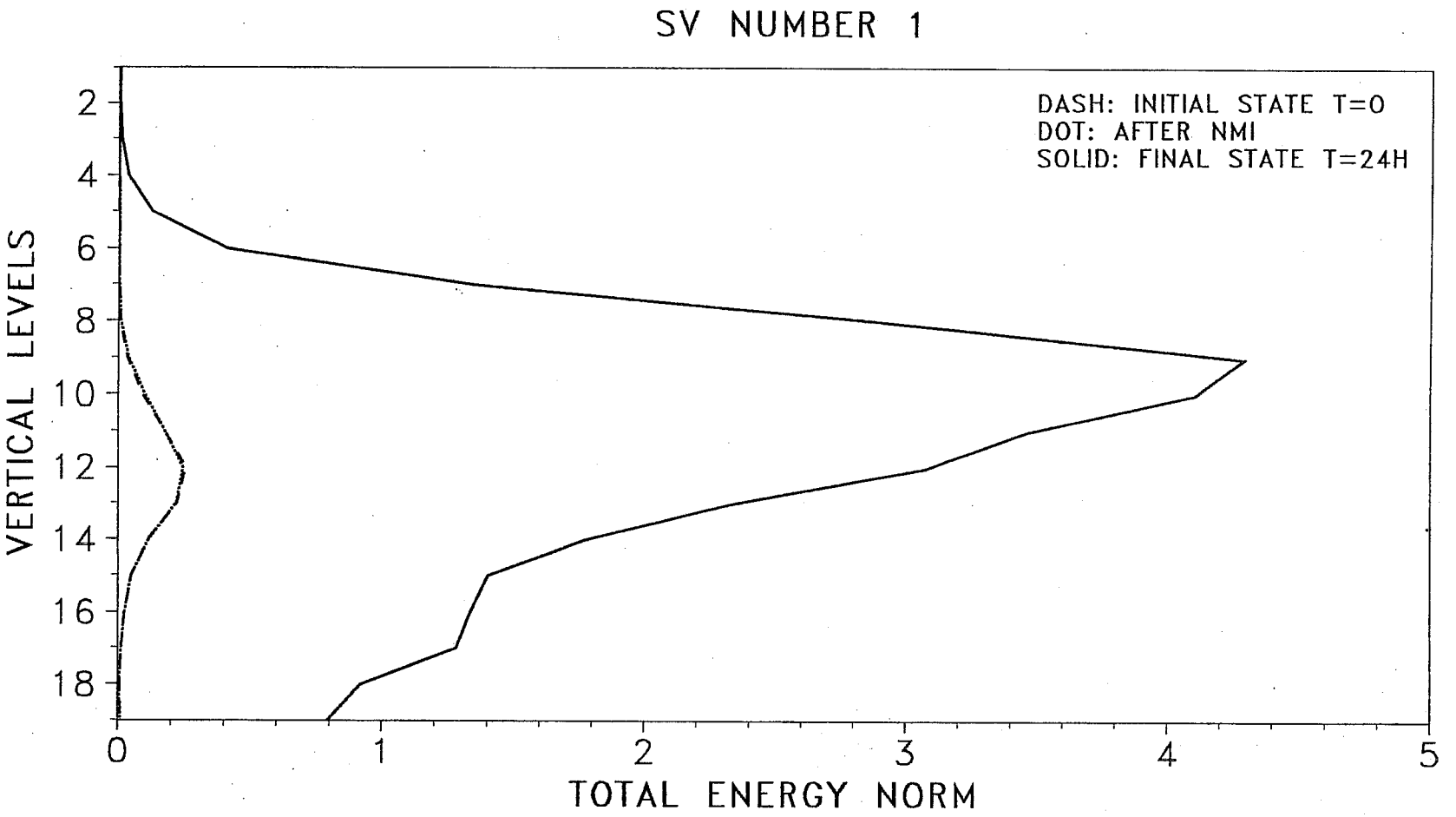


Figure 7. Vertical cross section of the total energy norm of the 1st SV at the initial state, after the application of the NMI procedure, and after 24h tangent linear time integration.

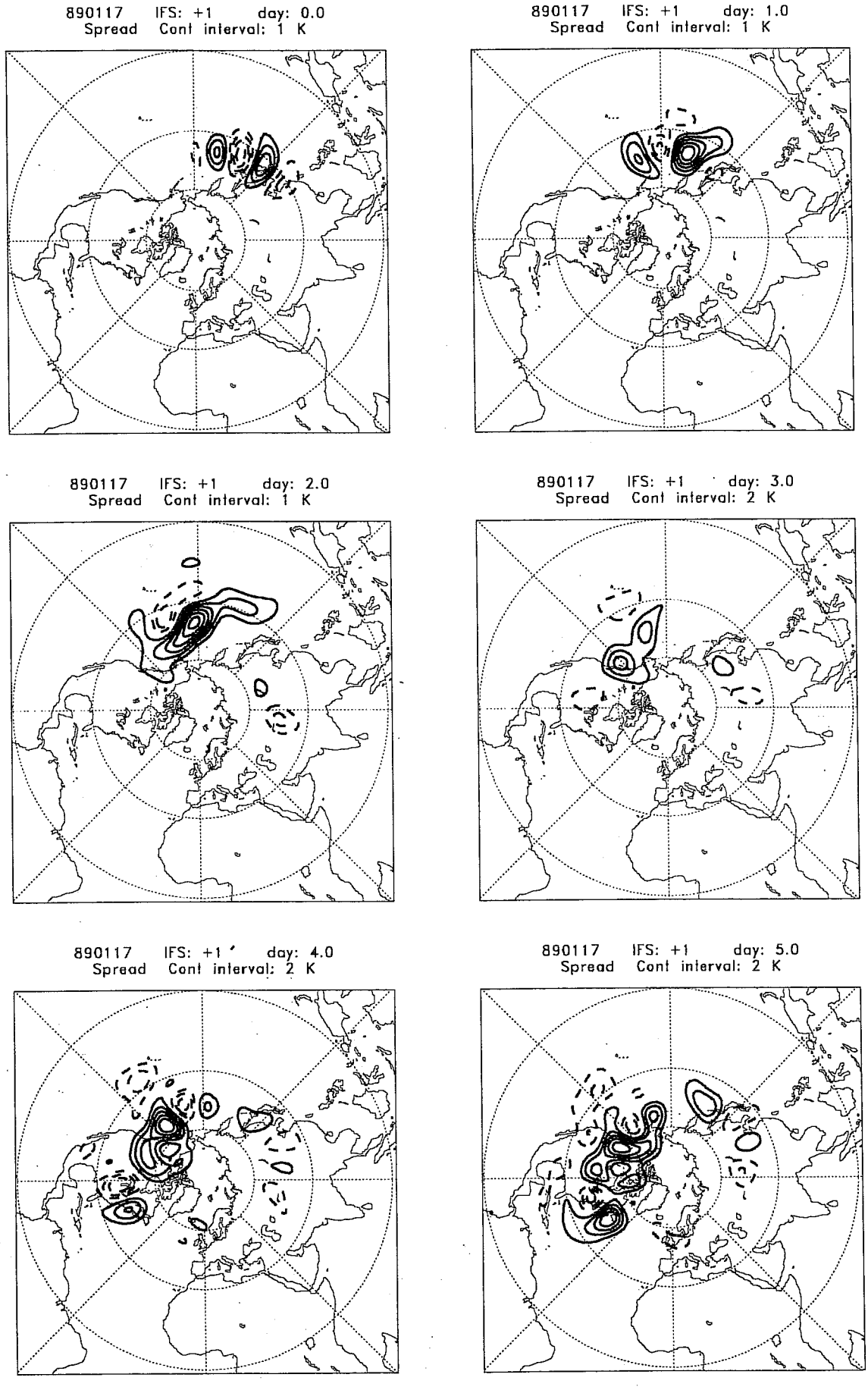


Figure 8. Time evolution of the fastest growing SV generated by the IFS model over a 24h time interval, computed using the T63L19 non-linear version of the ECMWF model. The 500hPa temperature field at different forecast times are shown: starting from the top-left panel at day 0, 1, till day 5.

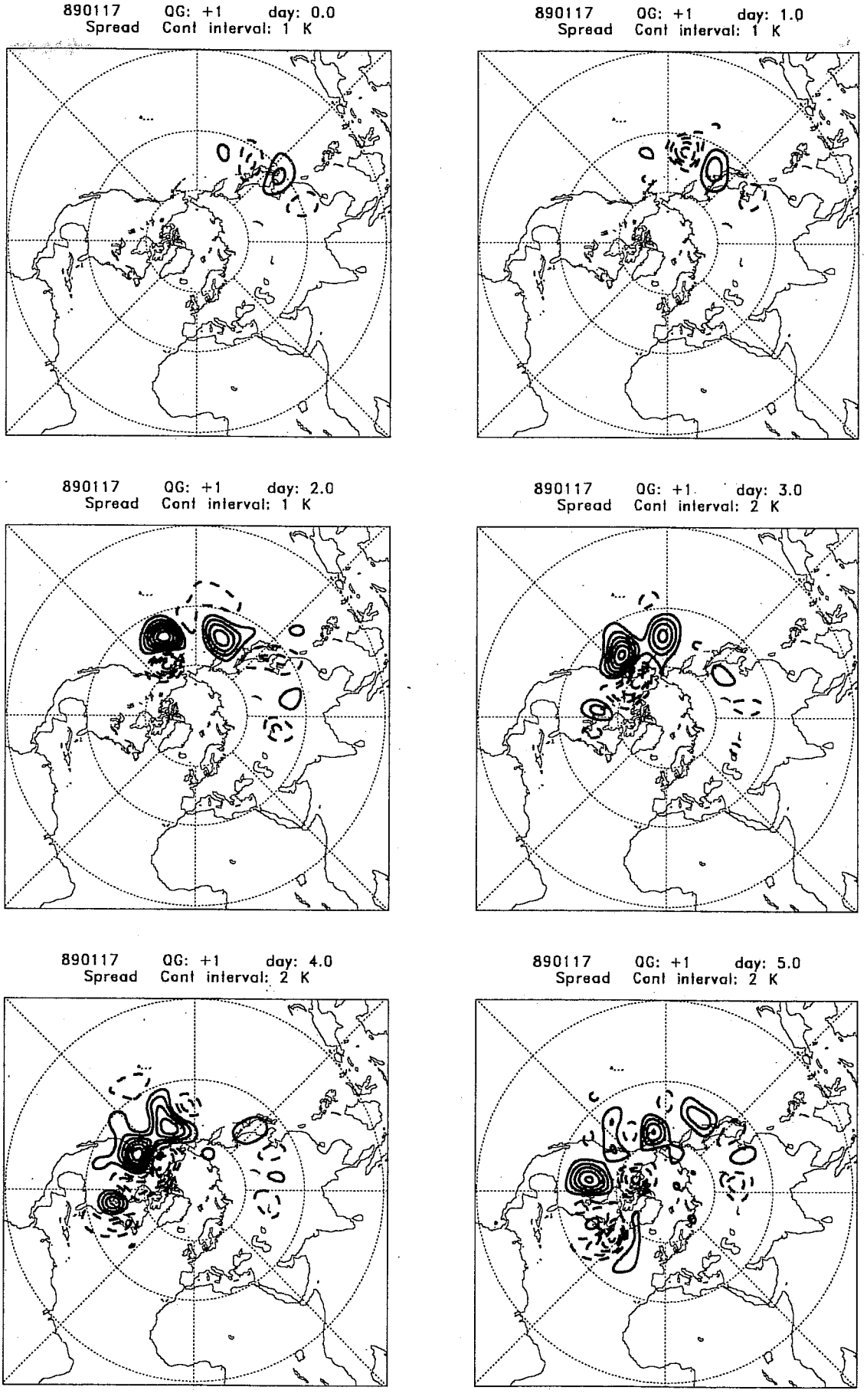


Figure 9. As Fig. 8 but for the T63L19 time evolution of the SV computed by Molteni and Palmer (1992) using a T21L3 QG model (this QG-SV is very similar to the 1st IFS SV).

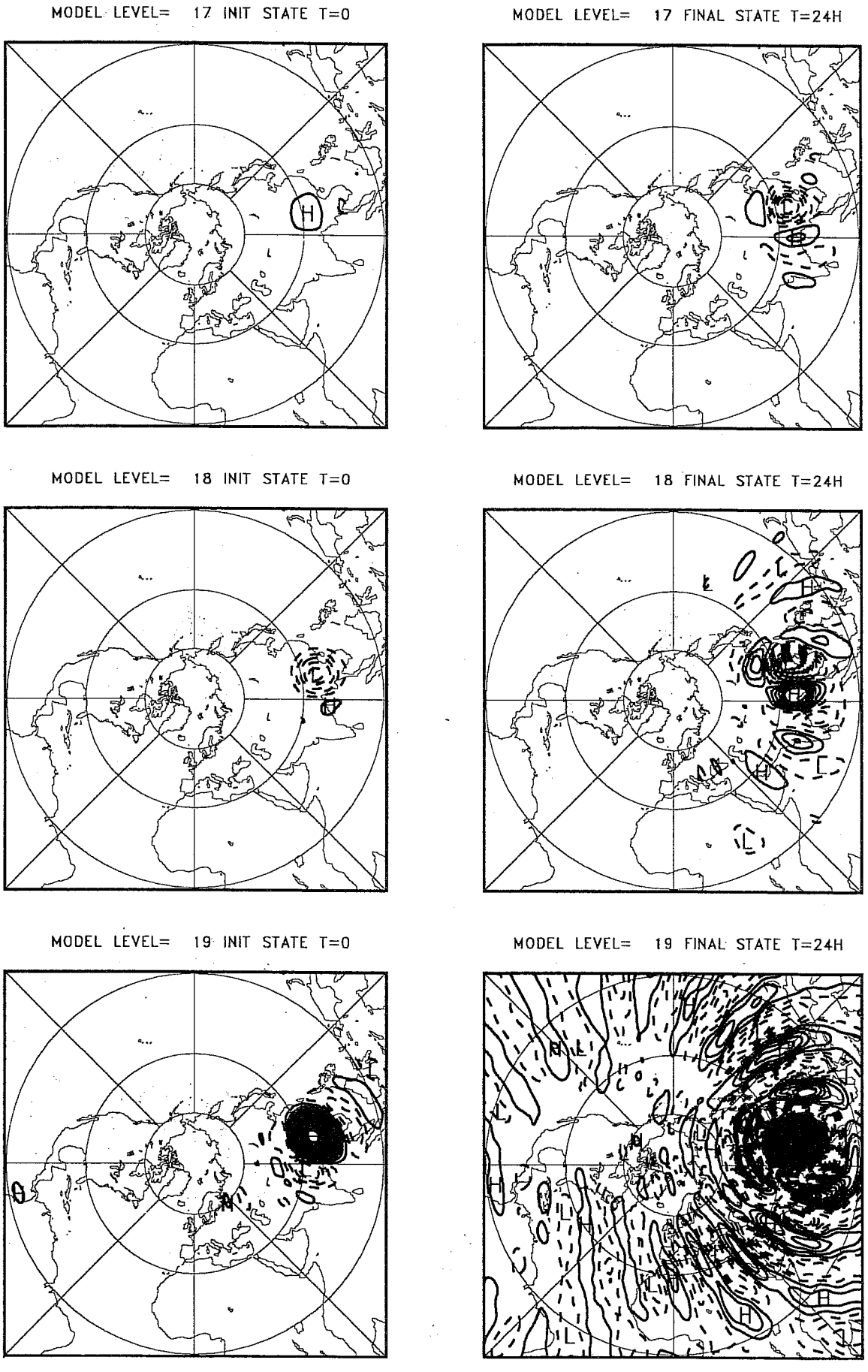


Figure 10. As Fig. 6 but for the time evolution of the 2nd 24h 5-NMI IFS SV: the temperature field has been plotted at the last 3 model levels.

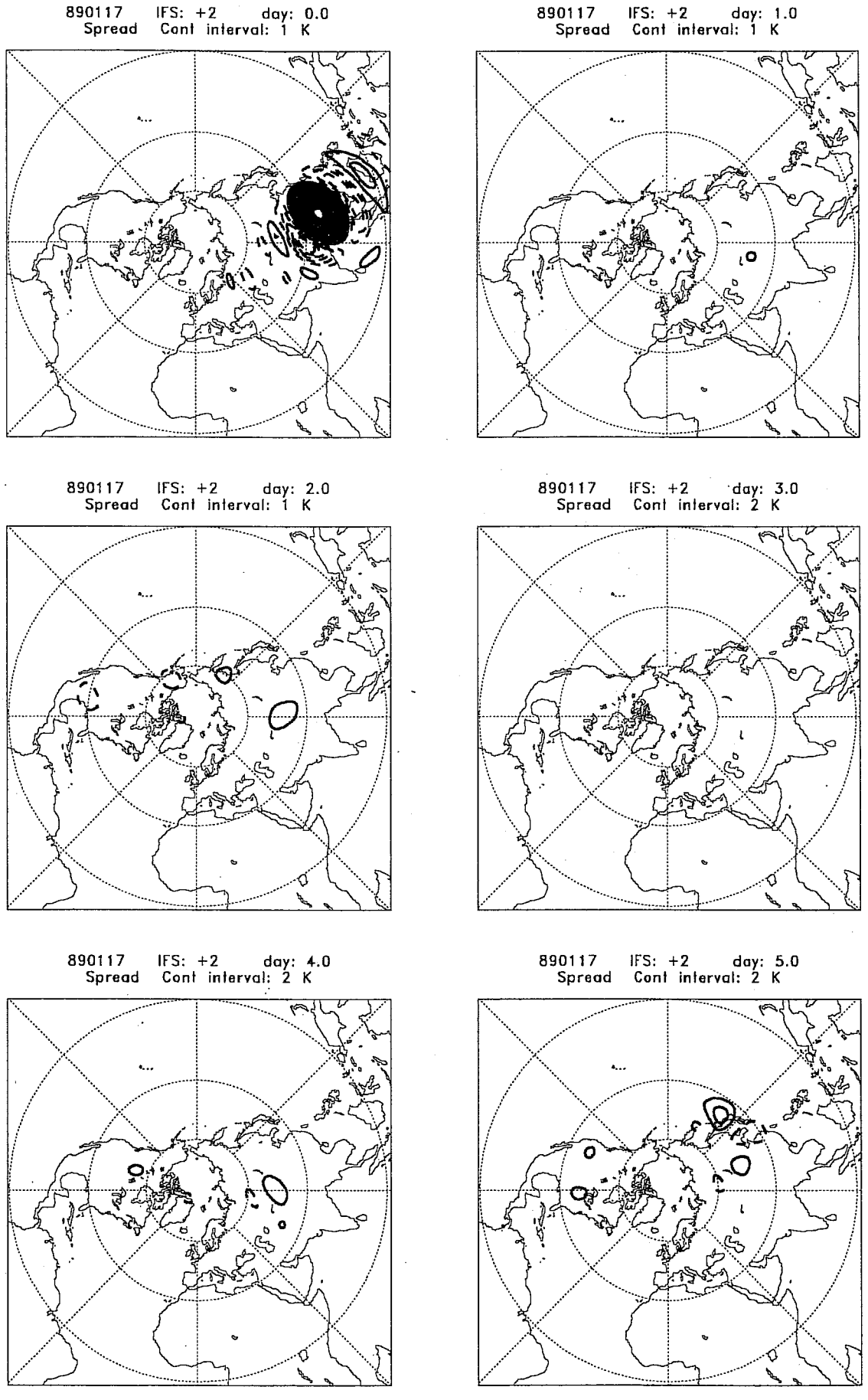


Figure 11. As Fig. 8, but for the time evolution of the 2nd 24h 5-NMI IFS SV. The 1000hPa temperature field at different forecast times are shown: starting from the top-left panel at day 0, 1, till day 5.

REFERENCES

Courtier P., Freydier C., Geleyn J. F., Rabier F., Rochas M., 1991, "The Arpege project at Météo France", ECMWF Seminar Proceedings, Vol. 2, 193-231.

Errico R. M., 1989, "Theory and application of Nonlinear Normal Mode Initialization", NCAR Technical note, NCAR/TN-344+IA.

Golub G. H. and Van Loan C. F., 1983, "Matrix computations", North Oxford Academic Pubbl. Co. Ltd..

Kaniel S., 1966, "Estimates for some computational techniques in linear algebra", Math. Comp. 20, 369-378.

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 XVII, 3, 321-333.

Molteni F. and Palmer T. N., 1992, "Predictability and finite-time instability of the northern winter circulation", submitted to QJRMS.

Mureau R., Molteni F. and Palmer T. N. (1992), "Ensemble prediction using dynamically-conditioned perturbations", submitted to QJRMS.

Nour-Omid B., "The Lanczos algorithm for solution of large generalized eigenproblems", private communication.

Paige C. C.(1971), "The computation of eigenvalues and eigenvectors of very large sparse matrices", PhD thesis, London University.

Rabier F. and Courtier P., 1991, "Four-dimensional assimilation in the presence of baroclinic instability", RD Tech. Mem. N. 183, ECMWF.

Scott D. S., 1978, " Analysis of the symmetric Lanczos process", UCB-ERL Technical Report M78/40, University of California, Berkeley.

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

Wergen W., 1987, "Diabatic Nonlinear Normal Mode Initialization for a spectral model with a hybrid vertical coordinate", ECMWF Technical Report N. 59.

Wiin-Nielsen A. C., 1979, "Normal mode initialization: a comparative study", ECMWF Lecture note N. 1.

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.