PREDICTION OF THE UNCERTAINTY OF NUMERICAL WEATHER FORECASTS: PROBLEMS AND APPROACHES

Martin Ehrendorfer*

European Centre for Medium–Range Weather Forecasts
Reading, England, United Kingdom

ABSTRACT

Several aspects and approaches related to the prediction of the uncertainty of numerical weather forecasts are considered. These considerations are centered around the problem of how to propagate in time an initial probability density function given a numerical weather prediction model. In studying the time evolution of the probability density function, the main emphasis is placed here on the time evolution of the forecast error covariance structure. Properties of singular vectors and of eigenvectors of error covariance matrices are discussed and illustrated in the framework of a quasigeostrophic model. In this context, a singular–vector–based Monte Carlo method is described and illustrated. Issues related to nonlinearities are discussed briefly.

---

1. Introduction and motivation

One of the major limiting factors on the accuracy of short– to medium–range weather forecasts that is achievable when forecasts are made with dynamical models of the atmosphere are errors in the specification of the initial state of numerical weather prediction (NWP) models. In addition to errors in the specification of the initial state of the model, errors in the model formulation itself degrade the quality of forecasts. The importance of these error sources in degrading the quality of forecasts is related to the fact that errors introduced in atmospheric models are, in general, growing, as demonstrated by Lorenz (1963, 1982). Both of these errors are – in conjunction with different atmospheric flow patterns – also responsible for the day–to–day variability of atmospheric predictability and of the skill of

* permanent affiliation: Institute for Meteorology and Geophysics, University of Vienna, Vienna, Austria.

individual forecasts. The uncertainty introduced into NWP forecasts through these two error sources, and specifically its quantification or prediction has been the subject of many studies conducted in parallel to the continuous development of NWP models (e.g., Epstein 1969, Leith 1974). However, even before the advent of NWP models the potential presence of the above two error sources in a mathematically–physically based forecasting process was recognized (Bjerknes 1904), as was the unavoidable uncertainty of weather forecasts and the necessity for its subsequent quantification (e.g., Cooke 1906, von Myrbach 1913; for an overview on the early history of probability forecasts, see Murphy 1998).

The early recognition of the implications of such misspecification of the initial model state on ensuing forecasts is illustrated here by a quotation from an interview with P.D. Thompson (1987), recounting difficulties in providing good initial model states in the late 1950s/early 1960s: "At that time, about the only information we had from over the Pacific Ocean, at least near the continental United States, was from weather ship *Papa* in the Gulf of Alaska. Weather ship *Papa* was not terribly reliable. There were times when the reports did not come in at all. There was a big hole there, and no data at all, and one simply had to guess what was there. One had to make a forecast anyway, so you would make a forecast, but on some occasions, the data came in, but late. We were able to go back and put in the data, then re-run the forecast with something like correct initial conditions. Then we compared the forecast with and without the data from the weather ship *Papa*. What we discovered that we might do was characteristically get errors of perhaps 150 meters at 500 millibar height as a result of not having the data, but a couple of days later these would show up as errors of perhaps 240 meters over southern Canada. Now that's a pretty big error. Then, I began to realize that this was a particularly flagrant example of what happens when the initial state is not correct, but that everyday, the initial state is in some degree incorrect. That is to say that the initial analysis is based on a rather small sampling of data which is rather widely scattered in space that's transmitted digitally, so there's some roundoff error. But even worse, there are large errors of interpolation, depending on how far apart the stations are. So even if the analysis we made were the most probable, or the best analysis you could possibly make under the circumstances, there were many neighboring initial states that were almost equally probable. So then the question was, well, suppose that you consider not one deterministic prediction, but a whole ensemble of deterministic predictions, each starting with a slightly incorrect initial state, which is more or less randomly distributed around the most probable initial state." (see also, Thompson 1957).

For a general and consistent treatment of the above–mentioned error sources it becomes necessary to view the atmospheric prediction problem in a probabilistic framework,

such as outlined in the above quotation, leading to the concept of the specification of the initial model state in terms of a probability density function (pdf) that evolves in time under the atmospheric dynamics. *Approaches for studying this time evolution of an initial pdf, as well as difficulties encountered in this process are the subject of this paper.* In the discussion of such approaches particular attention will be given to the problem of deriving efficiently second–order moment (variance/covariance) information from the time–evolving pdf. The even more challenging problem of deriving probabilities for the occurrence of certain events (i.e., for the likelihood of the variables involved to lie within certain ranges) from the time–evolving pdf is not discussed in detail here.

The organization of the paper is as follows. The general problem of uncertainty prediction is briefly restated in mathematical terms in section 2. Possible approaches to this general problem are discussed in section 3 in general terms. Due to central role occupied by the Liouville equation in the process of uncertainty prediction, two results connected to this equation are discussed in section 4. In section 5, the specific problem of variance/covariance prediction is considered. The discussion is centered around the so–called singular vectors and their relation to the eigenstructure of the forecast error covariance matrix. Results relevant to this problem are presented in this section for a quasigeostrophic model; also, a singular–vector–based Monte Carlo approach to covariance prediction is described and illustrated. The paper is concluded with a discussion of open questions. In particular, ideas for relaxing linearity assumptions in covariance prediction, and the computation of singular vectors are discussed.

2. Uncertainty prediction in NWP

The general problem of predicting the uncertainty of a forecast produced with a numerical weather prediction NWP model may be stated as follows. To introduce the notation, an NWP model is represented here in an abstract manner as an operator $\mathcal{M}_t$, in general nonlinear, that integrates an initial condition $\mathbf{x}_0$ forward up to time t with the result being the model state $\mathbf{x}_t$:

$$\mathbf{x}_t = \mathcal{M}_t(\mathbf{x}_0). \tag{2.1}$$

As mentioned in the introduction, the operator $\mathcal{M}_t$ seems to possess the property of being able to produce quite different results at time t given only slight changes in the state at the initial time. In other words, atmospheric models exhibit sensitive dependence on initial conditions which is a property of a chaotic system (see, also, Lorenz 1996). However, one also relies on the notion that the mapping described in (2.1) is in some sense unique; that is, a given initial condition results in a specific state at time t. Or, stated differently, it is not possible on the basis of (2.1) that two *different* initial states result in the (exact) same state at time t.

Given (2.1), the problem of predicting uncertainty may be stated as: predict the statistics or the probability density function of $\mathbf{x}_t$ given that the initial state, $\mathbf{x}_0$, is uncertain. This latter uncertainty may, for example, be described in one of the following ways:

$$\mathbf{x}_0 \sim \mathcal{N}_n(\mathbf{x}_0^c, \mathbf{V}), \tag{2.2}$$

$$E(\mathbf{x}_0) = \mathbf{x}_0^c \qquad D(\mathbf{x}_0) = \mathbf{V}, \tag{2.3}$$

$$\rho(\mathbf{x}, t = 0) = \rho_0(\mathbf{x}). \tag{2.4}$$

Here, (2.2) specifies the full initial pdf in terms of an n–dimensional *normal* pdf with parameters $\mathbf{x}_0^c$ and $\mathbf{V}$ (n will denote the dimension of the phase space). In (2.3), the information about the uncertainty in the initial state is restricted to knowledge of its first and second moments (E denotes the expectation operator, and D the dispersion operator). In (2.4), the function ρ_0 describes the initial pdf in a generic form. An obvious choice for $\mathbf{x}_0^c$ is the analysed state of the atmosphere (the best estimate available), and for $\mathbf{V}$ its associated analysis error covariance matrix.

The problem as posed here can be approached in various ways, as described in the next section. It should be mentioned, however, that even a solution to this problem will, in general, be incomplete in the sense that it will not contain all the desired information about the uncertainty of the forecast $\mathbf{x}_t$, because of model errors. Model errors have been neglected in writing (2.1), and their effect is, as discussed in section 1, to introduce further uncertainty during the model integration, beyond the initial state uncertainty. It is, in principle, possible to include a model error term in (2.1), or in the analogous equations below (see also next section). Such a term has not been included, however, since the emphasis here is on prediction of uncertainty arising from initial state uncertainty.

3. Possible approaches to uncertainty prediction in NWP

The problem posed in the previous section (without the complication of model errors) is known in the statistical literature as the problem of finding the pdf of a transformed variable (here, $\mathbf{x}_t$) given the pdf of the untransformed variable (here, $\mathbf{x}_0$); see, for example, Berliner (1998), DeGroot (1987). Clearly, however, the complicated form of the transformation operator $\mathcal{M}_t$, as well as its dimensionality, are leading to additional complications. This view on the problem posed will be taken up briefly again in section 4.1.

An alternative description for the time evolution of the pdf $\rho_t(\mathbf{x})$ is provided by the so–called *Liouville equation (LE)* (e.g., Thompson 1972, Epstein 1969, Fortak 1973), which is a **linear** partial differential equation in the single unknown ρ_t (i.e., the pdf at time t). The LE describes the time evolution of the pdf ρ_t in phase space, when the governing phase space dynamics is given by the (possibly nonlinear) dynamical system:

$$\frac{d}{dt}\mathbf{x} = \mathbf{\Phi}(\mathbf{x}), \tag{3.1}$$

where the vector–valued function $\mathbf{\Phi}$ describes the (autonomous), generally nonlinear, dynamics governing the time evolution of the state vector. Note that (3.1) may be interpreted as prescribing "velocities" in phase space. In this situation, the LE may be written in the following format:

$$\frac{\partial \rho_t(\mathbf{x})}{\partial t} + \sum_{k=1}^{n} \mathbf{\Phi}_k(\mathbf{x}) \frac{\partial \rho_t(\mathbf{x})}{\partial \mathbf{x}_k} = -\psi(\mathbf{x})\rho_t(\mathbf{x}), \tag{3.2}$$

where:

$$\psi(\mathbf{x}) \equiv \sum_{k=1}^{n} \frac{\partial \mathbf{\Phi}_k(\mathbf{x})}{\partial \mathbf{x}_k} \tag{3.3}$$

denotes the divergence of the flow in phase space, n (as before) indicates the dimension of the model phase space, and subscript k identifies the kth component of the relevant vector. The LE (3.2) states that the local rate of change of the pdf at any point in phase space must be exactly balanced by the net probability flux across the surface of a small, but finite, volume element surrounding that point in phase space. As such, the LE is entirely analogous to the well–known continuity equation in fluid mechanics.

In analogy to the continuity equation in fluid mechanics, the LE, written in the form (3.2), may also be interpreted as expressing the conservation of the phase space integral of the pdf. Such conservation prohibits the spontaneous creation or disappearance of trajectories in phase space (analogous to the creation or destruction of mass). In other words, every point in phase space at a given time t must uniquely correspond to a point in phase space at another time (e.g., the initial time; see also, Lin and Segel 1988). Such uniqueness will be observed for the governing dynamics (3.1) (due to the properties of first–order ordinary differential equations; see, e.g., Nicolis 1995). This relationship between the points on a given trajectory is expressed here in the following form:

$$\mathbf{x} = \mathbf{x}(\mathbf{\Xi}, t) \quad \Leftrightarrow \quad \mathbf{\Xi} = \mathbf{\Xi}(\mathbf{x}, t), \tag{3.4}$$

where $\mathbf{\Xi}$ is the (unique) point in phase space that, under the dynamics (3.1), is mapped into the point $\mathbf{x}$ as time evolves from $t = 0$ to t. Due to its linearity, the solution to the

LE for the initial condition:

$$\rho_{t=0}(\mathbf{x}) = \rho_0(\mathbf{x}), \tag{3.5}$$

may be explicitly expressed in the following form:

$$\rho_t(\mathbf{x}) = \rho_0(\Xi) \underbrace{\exp\Big[-\int_0^t \psi[\mathbf{x}(\Xi, t')]dt'\Big]}_{\equiv h(\mathbf{x},t)}. \tag{3.6}$$

Thus, in view of (3.6), it suffices, in principle, to evaluate (3.6) and to compute from this solution any desired statistics of ρ_t. Needless to say, there are various complications prohibiting this direct approach. The most prominent of these complications is the high dimensionality of the phase space in which the partial differential equation (3.2) must be considered for realistic NWP models. In the context of the solution (3.6), it is important to note that, for given $\mathbf{x}$ and t, as arguments for the pdf ρ, Ξ must be regarded as a function of $\mathbf{x}$ and t (see also, Ehrendorfer 1994a, b). Thus, the factor h introduced here (see also, section 4.1), also only depends on $\mathbf{x}$ and t.

Before proceeding, it is worth mentioning that the LE generalizes to a form of the so–called Fokker–Planck equation for the situation that stochastic forcing terms are included in (3.1) as representations of model error (Thompson 1983). Also, the formulation of the LE for discrete–time dynamical systems is known as Frobenius–Perron equation (Nicolis 1995).

As an illustration of the above, the solution of the LE for the one–dimensional system:

$$\frac{d}{dt}x = x - x^3 \tag{3.7}$$

is considered for two different choices of the initial pdf. In this case the LE takes on the form:

$$\frac{\partial \rho_t(x)}{\partial t} + (x - x^3)\frac{\partial \rho_t(x)}{\partial x} = -(1 - 3x^2)\rho_t(x). \tag{3.8}$$

The solution to (3.8), evaluated by using (3.6), is shown in Fig. 1, for two initial pdfs that are both (univariately) normal with the same variance, but with different means; specifically, in (a) $x_0 \sim \mathcal{N}(0.05, 0.01^2)$, and in (b) $x_0 \sim \mathcal{N}(0.1, 0.01^2)$. The solution to (3.8) is plotted in Fig. 1 at time increments of 0.2. It can be seen that the initially symmetric pdf becomes bimodal quite rapidly, with the pdf approaching the two stable equilibrium points at ± 1.

In view of the difficulties involved with the direct computation of the solution of the pdf over the entire phase space, Epstein (1969) proposed to solve instead the equations

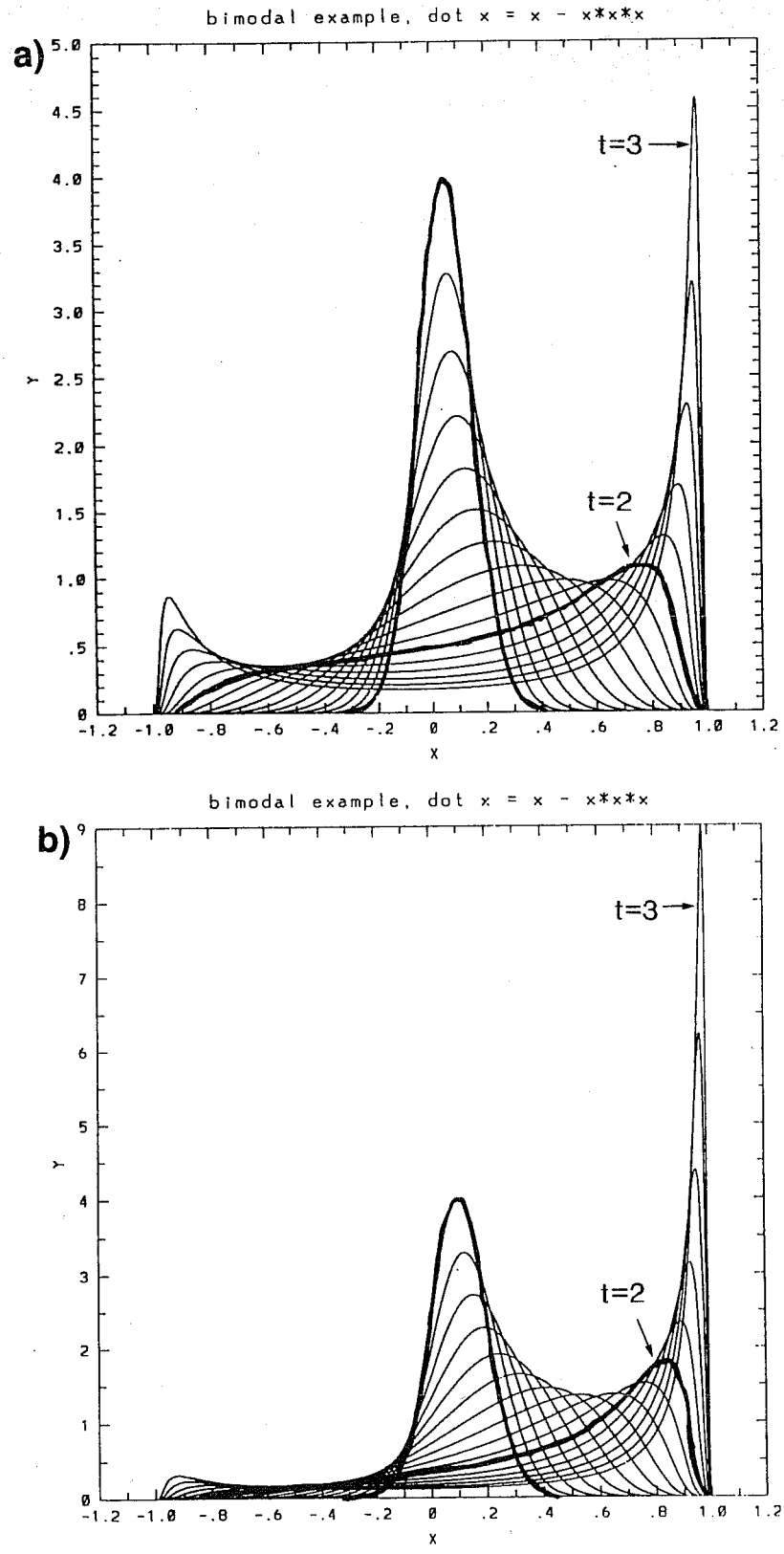

Fig. 1. Solution of the Liouville equation for the dynamical system (3.7) for two different initial probability density functions ρ_0: (a) $\mathcal{N}(0.05, 0.01^2)$ (b) $\mathcal{N}(0.1, 0.01^2)$. The time evolving pdfs are plotted at time increments of 0.2.

governing the time–evolution of first, second, and possibly higher moments of the pdf of the model state vector. These equations are referred to as the stochastic–dynamic equations, and the procedure of solving these equations as *stochastic dynamic prediction* (see also, Fortak 1973, Fleming 1971, Pitcher 1977).

This approach requires the derivation of the stochastic–dynamic equations and their subsequent integration. In the presence of nonlinear governing dynamics (2.1), or (3.1), the prognostic equations for moments of different order are coupled and one has to solve the closure problem. However, the similarity of the stochastic–dynamic equations and the equations used in the extended Kalman filter (e.g., Bouttier 1996) seems to make this approach highly attractive. In addition, all the terms appearing in the mean and the variance/covariance equation seem computable (in principle) given the forecast model and its tangent–linear model (neglecting the coupling with the third–moment equation, however). The ability of directly integrating the equations for mean and second–order moments may therefore make the method of stochastic dynamic prediction quite attractive. Some of these connections are discussed in more detail in the last section, where also some of the differences between the Kalman filter equations and the stochastic–dynamic equations will be pointed out (see, also, Cohn 1993).

As an alternative to stochastic dynamic prediction, Leith (1974) proposed the integration of the NWP model (2.1) from a set of randomly chosen initial conditions all consistent with the uncertainty in the analysed state. This *Monte Carlo (MC) approach* is presently at the basis of all currently operational efforts that are aimed at gaining information about the time–evolved pdf. Various possibilities are under investigation for devising sampling strategies from the initial pdf that are more efficient than a random sampling strategy, mainly in view of limitations on the sample size. Presently, such multiple model integrations are carried out operationally at, for example, the European Centre for Medium–Range Weather Forecasts (ECMWF; see, Molteni et al. 1996, Palmer et al. 1997), and at the National Centers for Environmental Prediction (NCEP; see, Toth and Kalnay 1997). At this point reference is also made to the recent overview on uncertainty prediction by Ehrendorfer (1997).

It should be mentioned that the theoretical basis for the MC approach to predict the moments of the time–evolved pdf lies in the fact that the moments (or, more generally, functions) of a transformed random variable can either be computed by using the pdf associated with the transformed variable, or by using the pdf of the untransformed variable (see, DeGroot 1987). As an illustration of this fact, consider the computation of the mean

of $\mathbf{x}_t$ through either of the following expressions:

$$E[\mathbf{x}_t] = \int \mathbf{x}\rho_t(\mathbf{x})d\mathbf{x}, \tag{3.9}$$

$$E[\mathbf{x}_t] = E[\mathcal{M}_t(\mathbf{x}_0)] = \int \mathcal{M}_t(\mathbf{x})\rho_0(\mathbf{x})d\mathbf{x}, \tag{3.10}$$

where the integration extends over phase space. These equations may immediately be generalized to functions of $\mathbf{x}_t$, leading to the time–evolved variance/covariance structure and other statistics. The MC approach exploits expression (3.10).

4. Comments on the Liouville equation

In view of the central role played by the LE in the context of uncertainty prediction, two observations related to this equation will be discussed in this section.

4.1. A likelihood ratio

The general expression that relates the pdf of a (nonlinearly) transformed variable to the pdf of the original variable involves the Jacobian determinant J of this transformation (see, e.g., Anderson 1958, DeGroot 1987). If this general expression is applied in the context of the notations introduced in the previous sections, the ratio of the pdfs introduced in section 3 can be expressed as:

$$\frac{\rho_t(\mathbf{x})}{\rho_0(\Xi)} = \frac{1}{|J|} = \left[\left|\det \mathbf{M}_{(\Xi,t)}\right|\right]^{-1} \tag{4.1}$$

where the Jacobian determinant J:

$$J \equiv \det\left(\frac{\partial \mathbf{x}(\Xi,t)}{\partial \Xi}|_{\Xi}\right) \tag{4.2}$$

is the determinant of the Jacobian matrix (i.e., the matrix of derivatives) of the transformation (2.1) (see also (3.4)), evaluated at Ξ (double vertical bars, as in (4.1), denote the absolute value). The uniqueness of the transformation, as expressed in (3.4), is a necessary condition for J to be nonzero. It is noted in passing that if the transformation between the two variables changes the physical units of the variables involved, this will be reflected by a unit–carrying determinant J; more generally, a pdf reflects the physical units of the variables described, which becomes also very clear from the expressions (3.9) and (3.10).

The operator $\mathbf{M}_{(\Xi,t)}$ appearing in the second equality in (4.1) is the resolvent of the tangent–linear model, usually considered as not carrying physical dimensions. The dependence of the resolvent on the point of linearization, Ξ, and time t is indicated through the subscripts. The connection between J and $\mathbf{M}_{(\Xi,t)}$ (i.e., J is the determinant of the tangent–linear model) becomes clear by noting that the Jacobian matrix of derivatives appearing in the Jacobian determinant (4.2) also appears in the Taylor explansion of $\mathcal{M}_t(\mathbf{x}_0 + \mathbf{z}_0)$ where it is referred to as the resolvent of the tangent–linear model:

$$\mathcal{M}_t(\mathbf{x}_0 + \mathbf{z}_0) = \mathcal{M}_t(\mathbf{x}_0) + \frac{\partial \mathcal{M}_t(\mathbf{x})}{\partial \mathbf{x}}\Big|_{\mathbf{x}_0} \mathbf{z}_0 + \text{h.o.t.} \tag{4.3a}$$

$$= \mathcal{M}_t(\mathbf{x}_0) + \mathbf{M}_{(\mathbf{x}_0,t)}\mathbf{z}_0 + \text{h.o.t.} \tag{4.3b}$$

where h.o.t. indicates higher–order terms, and $\mathbf{z}_0$ represents an initial perturbation. Further, the combination of (4.1) and (3.6) leads to the following expression:

$$\frac{\rho_t(\mathbf{x})}{\rho_0(\Xi)} = \Big[\big|\det \mathbf{M}_{(\Xi,t)}\big|\Big]^{-1} = h(\mathbf{x},t). \tag{4.4}$$

The pdf–ratio appearing on the left–hand–side in (4.4), called here *likelihood ratio*, indicates by how much the value of the initial pdf taken at Ξ is different from the value of the pdf ρ_t taken at $\mathbf{x}$. This ratio may also be interpreted as the change of the pdf along the trajectory from Ξ to $\mathbf{x}$, since Ξ evolves into $\mathbf{x}$ over time (see (3.4)).

As expressed by (4.4), this likelihood ratio h may be computed in two different, but entirely equivalent, ways. First, according to the definition of h in (3.6), it may be evaluated as the time integral of the divergence in phase space. Second, according to (4.1), it may be evaluated by computing the determinant of the relevant tangent–linear resolvent. By eq. (4.4), these two approaches are equivalent, which may intuitively be understood by noting that both the divergence and the determinant expressions are related to the change over time of a phase space volume (see, Arnold 1989). In addition, it may be seen by referring back to (3.6) that $h = 1$ for $\psi = 0$ (no divergence of the flow; see also below).

The time dependence of h is illustrated in Fig. 2 for the model proposed by Lorenz (1984) as a prototype of the general circulation:

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -y^2 - z^2 - ax + aF \\ xy - bxz - y + G \\ bxy + xz - z \end{pmatrix}, \tag{4.5}$$

with the following parameters:

$$a = 0.25 \qquad b = 4.0 \qquad F = 8.0 \qquad G = 1.25 \tag{4.6}$$

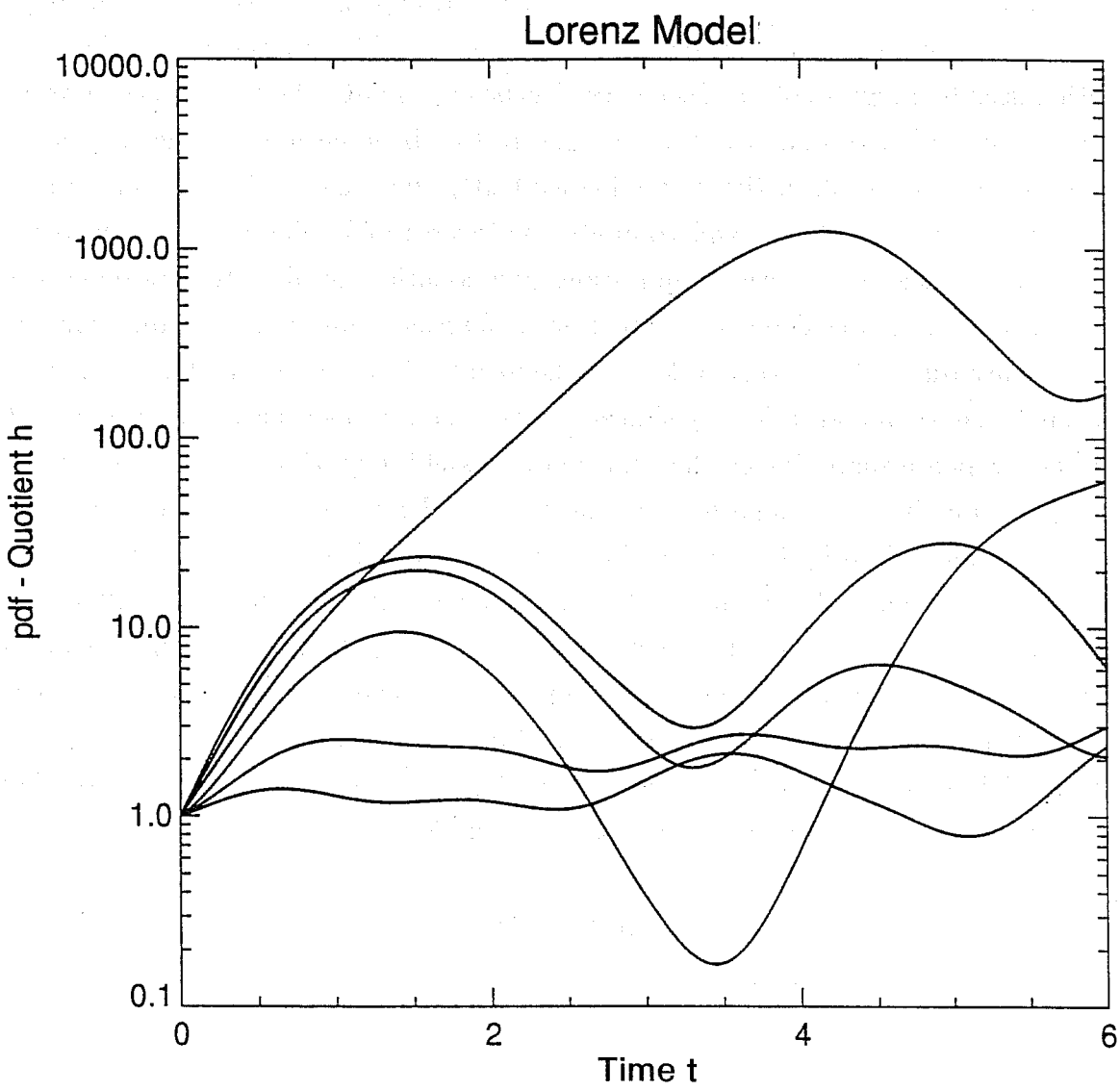

Fig. 2. The likelihood ratio h, defined in (3.6), for the dynamical system (4.5), as a function of time, for six different initial conditions Ξ.

and the following expression for the divergence in phase space:

$$\psi(x,y,z) = 2x - 2 - a \quad \left(\equiv \frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z}\right). \tag{4.7}$$

Note that the phase space divergence of this system depends on the location in phase space through the variable x. The curves shown in Fig. 2 differ by the choice of the initial condition used to integrate (4.5). Each curve has been computed by both of the approaches outlined above: (i) determine the time integral of the divergence along the trajectory up to time t according to (3.6), (ii) determine explicitly the tangent–linear model resolvent (for a given initial condition), and compute the inverse of its determinant according to (4.4). It can be seen that both computations give identical results. It may be seen that the likelihood ratio, considered as a function of forecast time, is also a function of the starting point; due to the nature of the likelihood ratio it is always one for $t = 0$. On the other hand, the likelihood ratio may simply be considered as measuring the (accumulated) divergence in phase space during the course of a model integration. Finally, note that for long integration times, the logarithm of the likelihood ratio h (refer to its definition in (3.6)), when normalized by the integration time t, is related to the sum of the Lyapunov exponents of a dynamical system (as this sum is also a measure of average phase space divergence). For the system (4.5), with parameters as in (4.6), the sum of Lyapunov exponents is negative (i.e., the system is dissipative), with the leading Lyapunov exponent being positive (i.e., the system is chaotic); see also, Talagrand (1996).

It is of interest to consider the above developments in the following linear (autonomous) context, where (3.1) is assumed to be of the form:

$$\frac{d}{dt}\mathbf{x} = \mathbf{A}\mathbf{x}, \tag{4.8}$$

with $\mathbf{A}$ time–independent, which implies that (2.1) takes on the form:

$$\mathbf{x}_t = \mathbf{L}_t \mathbf{x}_0, \tag{4.9}$$

where the resolvent $\mathbf{L}_t$ can now be expressed as the following matrix exponential:

$$\mathbf{L}_t = \exp(\mathbf{A}t). \tag{4.10}$$

In this case the (position–independent) function ψ simplifies to the trace of $\mathbf{A}$ (i.e., the sum of the diagonal elements of $\mathbf{A}$), which may also be written as the sum of the eigenvalues of $\mathbf{A}$ denoted a_k:

$$\psi = \text{trace}\ (\mathbf{A}) = \sum_{k=1}^{n} a_k\,. \tag{4.11}$$

For convenience it is assumed that $\mathbf{A}$ is diagonalizable. Then, the likelihood ratio, defined in (3.6), is only a function of time, and takes on the form:

$$h(t) = \exp\Big(-\int_0^t \sum_{k=1}^n a_k dt'\Big) = \exp\Big(-t\sum_{k=1}^n a_k\Big) = \prod_{k=1}^n e^{-a_k t} =$$
$$= \prod_{k=1}^n \big(e^{a_k t}\big)^{-1} = \prod_{k=1}^n l_k^{-1} = \Big(\prod_{k=1}^n l_k\Big)^{-1} = (\det \mathbf{L}_t)^{-1}, \tag{4.12}$$

where l_k denotes the eigenvalues of $\mathbf{L}_t$ (note that the operators $\mathbf{L}_t$ and $\mathbf{A}t$ are dimensionless; see also section 5.2.). In writing (4.12), the eigenvalues of $\mathbf{L}_t$ have been inferred from (4.10) from the eigenvalues of $\mathbf{A}$ (with diagonalizable $\mathbf{A}$); further, the fact is used that the determinant of a matrix is the product of its eigenvalues. Note that (4.12) is a special case of (4.4), for the linear governing dynamics (4.8). Clearly, in this case the transformation of pdfs is particularly simple, since the likelihood ratio is independent from the position in phase space. Note also that $\det \mathbf{L}_t$ is positive in the present situation.

4.2. The product of growth rates

The second observation related to the LE concerns the product of growth rates of a complete set of perturbations, for the class of dynamical systems with vanishing phase space velocity divergence. With the previous definition, it is, by eq. (4.4), in general true that:

$$\det \mathbf{M}_{(\mathbf{x}_0,t)} = \pm 1/h(\mathbf{x},t)\,. \tag{4.13}$$

Denote by $\mathbf{z}_t$ the difference of two nonlinear (perturbed minus unperturbed) model integrations at time t, which, by (4.3b), is given as:

$$\mathbf{z}_t \equiv \mathcal{M}_t(\mathbf{x}_0+\mathbf{z}_0) - \mathcal{M}_t(\mathbf{x}_0) \equiv \mathcal{H}_{(\mathbf{x}_0,t)}(\mathbf{z}_0) = \mathbf{M}_{(\mathbf{x}_0,t)}\mathbf{z}_0 + \text{h.o.t.} \tag{4.14}$$

where $\mathcal{H}_{(\mathbf{x}_0,t)}$ (depending on initial condition $\mathbf{x}_0$ and time t) is the *fully nonlinear error evolution operator*. If the higher–order terms in (4.14) are neglected, the *tangent–linear approximation* is made, and the corresponding result is denoted as $\mathbf{z}_t^{\mathrm{TL}}$:

$$\mathbf{z}_t^{\mathrm{TL}} \equiv \mathbf{M}_{(\mathbf{x}_0,t)}\mathbf{z}_0\,. \tag{4.15}$$

Note that the degree of agreement (i.e., the size of h.o.t.) between acting with $\mathcal{H}_{(\mathbf{x}_0,t)}$ or $\mathbf{M}_{(\mathbf{x}_0,t)}$ on an initial perturbation $\mathbf{z}_0$ will be, in general, a function of t, size and structure of $\mathbf{z}_0$, as well as of the linearization point $\mathbf{x}_0$ (see, e.g., Buizza 1995). The subscripts on

operators $\mathcal{H}$ and $\mathbf{M}$ will be dropped from now on, unless they are explicitly required for clarity. Consider at this point the optimization problem:

$$\max_{\mathbf{z}_0} \; J^{\mathrm{TL}} = \langle \mathbf{z}_t^{\mathrm{TL}}; \mathbf{z}_t^{\mathrm{TL}} \rangle = \langle \mathbf{M}\mathbf{z}_0; \mathbf{M}\mathbf{z}_0 \rangle \qquad \text{subject to :} \qquad \langle \mathbf{z}_0; \mathbf{z}_0 \rangle = 1, \tag{4.16}$$

where the pointed brackets denote an inner product used to measure perturbation sizes. Problem (4.16) leads to the following eigenproblem:

$$(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{M}^{\mathrm{T}}(\mathbf{C}^{\mathrm{T}}\mathbf{C})\mathbf{M}\mathbf{z}_0 = \lambda \mathbf{z}_0 \qquad \text{with :} \qquad \mathbf{z}_0^{\mathrm{T}}(\mathbf{C}^{\mathrm{T}}\mathbf{C})\mathbf{z}_0 = 1, \tag{4.17}$$

where the positive–definite matrix $\mathbf{C}^{\mathrm{T}}\mathbf{C}$ is used to denote the inner product explicitly. The eigenvector of (4.17) associated with the largest eigenvalue solves (4.16) (see also section 5). Given the eigendecomposition indicated in (4.17), the corresponding eigenvalues λ_k may immediately be used to compute the determinant of $\mathbf{M}_{(\mathbf{x}_0,t)}$ by forming their product:

$$\prod_{k=1}^{n} \lambda_k = \det\Big[(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{M}^{\mathrm{T}}(\mathbf{C}^{\mathrm{T}}\mathbf{C})\mathbf{M}\Big] = \det\Big(\mathbf{M}^{\mathrm{T}}\mathbf{M}\Big) = \big(\det \mathbf{M}\big)^2, \tag{4.18}$$

where basic properties of determinants have been used (Strang 1993). Combining the square of (4.13) with the result (4.18) one obtains:

$$h^2(\mathbf{x},t) = \Big(\prod_{k=1}^{n} \lambda_k\Big)^{-1}. \tag{4.19}$$

The result (4.19) states that the squared likelihood ratio is equal to the inverse of the product of the eigenvalues obtained by solving (4.17). This result is not surprising since the computation of the λ_k amounts to determining the determinant of the tangent–linear resolvent (see also (4.4)). Nevertheless, the result (4.19), which is perfectly general, points to the possibility of computing the likelihood ratio, as used in (4.4), for any nonlinear model for which the tangent–linear resolvent is available. Recognizing the limitation that this computation proceeds by specifying Ξ and t, rather than $\mathbf{x}$ and t (which is what one would really like to specify; see (4.4)), such a computation of the likelihood ratio amounts – under this limitation – to solving the LE for given t for the value of the pdf at one particular point in phase space. Obviously, the cost of this approach is very high, as it requires the computation of the entire λ_k–spectrum. In addition, it is recognized from expression (4.19) that this equation for the likelihood ratio is not very useful for deriving any bounds on h^2 if only the leading portion of the λ_k–spectrum is known (this very leading portion is available operationally at ECMWF), since the trailing part of the spectrum will

become increasingly important (if the relevant λ_k are small enough). Note finally that no assumptions about the initial pdf are entering in such a computation of the likelihood ratio.

Nevertheless, even in view of the limited usefulness of (4.19), this expression still allows making a general statement for dynamical systems with $\psi = 0$ (vanishing phase space velocity divergence). In that case, by (3.6), $h = 1$, and consequently, by (4.19):

$$\psi = 0 \qquad \rightarrow \qquad \prod_{k=1}^{n} \lambda_k = 1. \tag{4.20}$$

This result states that for a dynamical system with $\psi = 0$, the product of eigenvalues resulting from the optimization problem (4.16) must be one. Note, however, the following subtlety: since h, defined in (3.6), is the time integral over the divergence at different points in phase space (along a trajectory), it will be strictly one only if the time integration necessary to reach these different points in phase space preserves the $\psi = 0$ property of the continuous system. In other words, evaluating h, on the basis of (4.19), by computing the λ_k through solving (4.17) with a discrete tangent–linear resolvent may potentially lead to $h^2 \neq 1$, even for a system with $\psi = 0$ (such as the unforced barotropic vorticity equation), if the time integration destroys the property of volume conservation (see, Egger 1996).

Now, two final remarks are made at this point, where it assumed that $\psi = 0$ and it is indeed found (numerically) that $h = 1$. In this case, the first remark concerns the symmetry (or asymmetry) of the λ_k–spectrum. In fact, it is conjectured (without proof) that even under these very special circumstances the λ_k spectrum is, in general, asymmetric so that:

$$\lambda_k \neq \frac{1}{\lambda_{n-k+1}} \qquad k = 1, 2, ..., \frac{n}{2}. \tag{4.21}$$

For example, the largest eigenvalue is not, in general, the inverse of the smallest eigenvalue. It might be of interest to study the restrictions one has to impose (e.g., on the basic state) to achieve (or to be able to prove analytically) symmetry. The second remark concerns the sum of the λ_k. Here, it is immediately evident that these special circumstances ($\psi = 0$ and $h = 1$) imply that the normalized sum of eigenvalues is greater or equal than one:

$$\frac{1}{n} \sum_{k=1}^{n} \lambda_k \geq 1\,, \tag{4.22}$$

where equality holds only if all λ_k are equal to one. The result (4.22) follows from the fact that the arithmetic mean is always greater (or equal) than the geometric mean. Consequently, average growth must occur.

5. Prediction of forecast error covariances: Singular vectors

5.1. General remarks

In view of the difficulties encountered when attempting to predict the entire pdf, as outlined through the approaches in section 3, attention may be focused on the matrix of second moments of the time–evolved pdf, also called the forecast error covariance matrix, defined as:

$$\mathbf{S}_t = E\Big(\mathbf{P}\big[\mathbf{x}_t - E(\mathbf{x}_t)\big]\Big)\Big(\mathbf{P}\big[\mathbf{x}_t - E(\mathbf{x}_t)\big]\Big)^{\mathrm{T}}. \tag{5.1}$$

As before, $\mathbf{x}_t$ denotes the model state at time t, and the expectation operator E indicates an integration over phase space with respect to the time–evolved pdf (see eq. (3.9)). The operator $\mathbf{P}$ indicates a weighting and/or projection of the state vector at time t, to be specified later. The expression obtained by (5.1) for $t = t_0$ and $\mathbf{P} = \mathbf{I}$ defines the analysis error covariance matrix $\mathbf{V}$ in agreement with (2.2) or (2.3). The problem is then to determine the time–evolution of $\mathbf{V}$ under the dynamics (2.1).

Clearly, since $\mathbf{S}_t$ consists of the product of deviations, namely, $\mathbf{x}_t - E(\mathbf{x}_t)$, there will be a time period over which the evolution of $\mathbf{S}_t$ is governed by linear (tangent–linear) dynamics. The length of this time period depends, among other things (see, e.g., Buizza 1995), on the (average) magnitude of $\mathbf{x}_t - E(\mathbf{x}_t)$ at t_0. A geometrical interpretation of this special situation may be given by assuming that $\mathbf{V}$ is used to define an ellipsoid in n–dimensional space through the equation:

$$\mathbf{u}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{u} = c^2\,, \tag{5.2}$$

where c^2 is a positive constant. Then, in this linear regime (and only in the linear regime), the time–evolved counterpart of (5.2), obtained by replacing $\mathbf{V}$ by its time–evolved counterpart, will also be an ellipsoid. However, the principal axes of the time–evolved ellipsoid will, in general, change in length and direction. When the tangent–linear regime is left, nonlinear effects will add curvature and other distorting effects on the time–evolved ellipsoid.

An illustration of these ideas is given in Fig. 3, taken from Ehrendorfer (1997), which shows the complete solution of the LE, namely the pdf, for the dynamical system (4.5) with the parameters (4.6), at four consecutive times (each separated by one nondimensional time step). In producing the results shown in Fig. 3, the initial pdf $\rho_0(\mathbf{x})$ was taken as multivariate normal with diagnonal covariance structure. The pdfs are presented in this figure through the surfaces that separate values of ρ_t larger than one (inside the surface)

from those that are smaller than one (outside the surface); clearly, in this graphical display, ρ_0 appears at a sphere. Note that the definition of these surfaces implies that they are not material surfaces. It may be seen from Fig. 3 that the time–evolving pdf is stretching rapidly during the first nondimensional time step, with little or no curvature becoming apparent. At later times, nonlinear effects become important. Note that these nonlinear effects are fully included at all stages during the solution of the LE, but they are simply negligible at initial times.

5.2. Theoretical aspects

(a) An optimization problem

The geometrical ideas discussed in the previous subsection may be used as guidance in thinking about the design of schemes that are more efficient than simple random sampling, in order to learn about the time evolution of the pdf.

Under the assumption that perturbations are governed by tangent–linear dynamics (see, eq. (4.15)), the tangent–linear equivalent to (5.1) becomes:

$$\mathbf{S}_t^{\mathrm{TL}} = \mathbf{P}\mathbf{M}_t\mathbf{V}\mathbf{M}_t^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}, \tag{5.3}$$

which will be called the tangent–linearly approximated forecast error covariance matrix (see also, Ehrendorfer and Tribbia 1997; note that $\mathbf{S}_t^{\mathrm{TL}}$ is equal to $\mathbf{V}$ for $t = 0$ and $\mathbf{P} = \mathbf{I}$). It is assumed that $\mathbf{V}$ is symmetric and positive definite (and consequently invertible); both properties are natural for a covariance matrix. From eq. (5.3) the geometrical argument made in the previous subsection concerning the deformation of ellipsoids becomes also clear.

One possible approach to predicting the forecast error covariance matrix efficiently is to study the set of vectors that evolve into the eigenvectors of $\mathbf{S}_t^{\mathrm{TL}}$. Guided by this idea, the following maximization problem is considered:

$$\max_{\mathbf{z}_0} \quad \hat{J}(\mathbf{z}_0; t) = \left(\mathbf{P}\mathbf{M}_t\mathbf{z}_0\right)^{\mathrm{T}}\left(\mathbf{P}\mathbf{M}_t\mathbf{z}_0\right) \quad \text{subject to:} \quad \mathbf{z}_0^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{z}_0 = 1, \tag{5.4}$$

where $\mathbf{z}_0$ is a perturbation of suitable physical dimensions to be propagated by the tangent–linear model $\mathbf{M}_t$ (see (4.1), (4.3), (4.14)). The solution to (5.4) is the eigenvector associated with the largest eigenvalue of the eigenproblem (see, e.g., Mardia et al. 1982):

$$\mathbf{V}\mathbf{M}_t^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{P}\mathbf{M}_t\mathbf{z}_0 = \lambda\mathbf{z}_0 \qquad \text{subject to:} \qquad \mathbf{z}_0^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{z}_0 = 1. \tag{5.5}$$

The set of vectors $\mathbf{Z}_0$ solving (5.5) will also be referred to as *initial singular vectors* (SVs). Given the solution to (5.5), it is evident that the set of time–evolved vectors defined by:

$$\mathbf{Z}_t \equiv \mathbf{P}\mathbf{M}_t\mathbf{Z}_0 \tag{5.6}$$

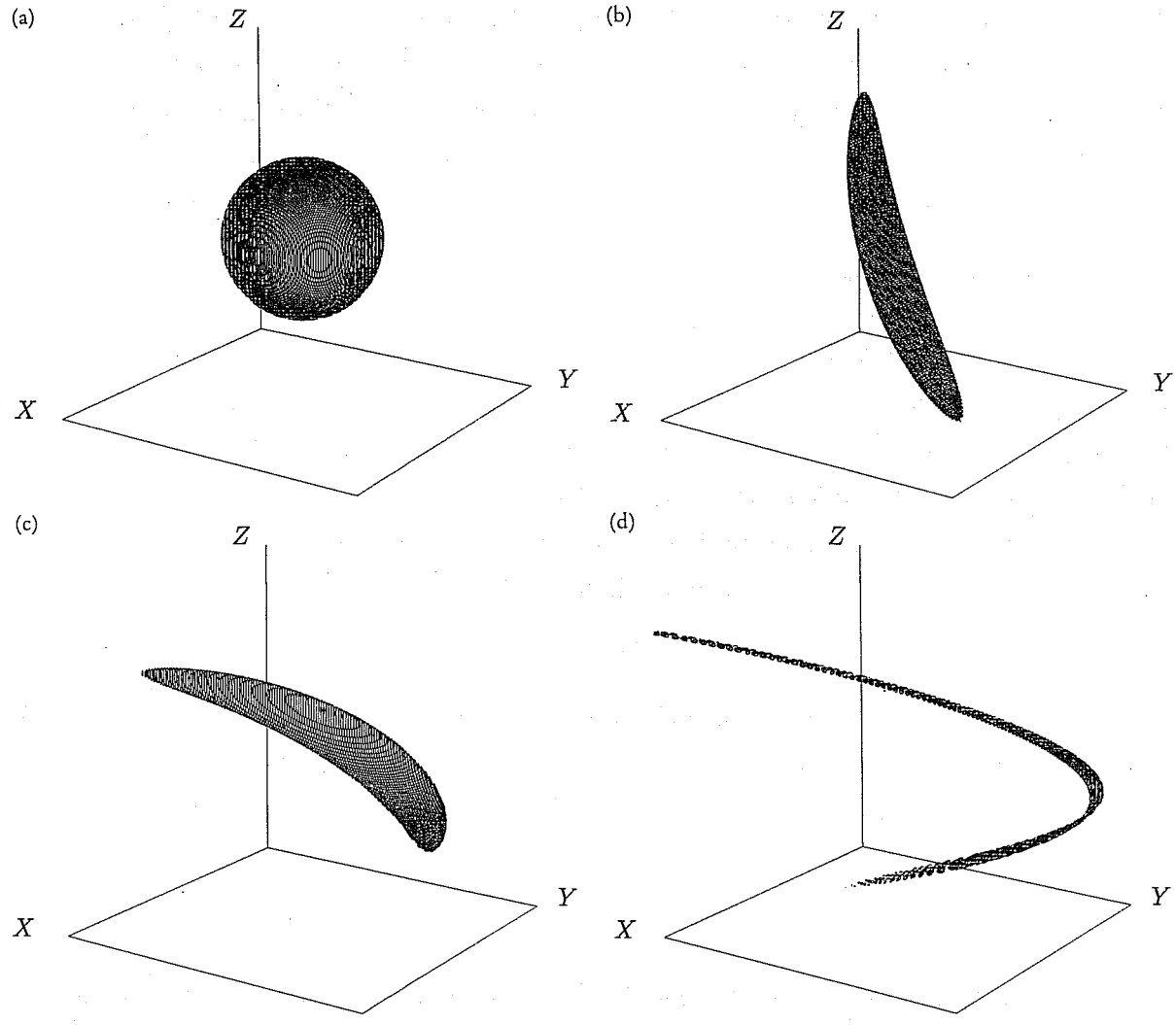

Fig. 3. The solution of the Liouville equation for the dynamical system (4.5), plotted in the schematic way described in section 5.1, at time increments of one nondimensional time step (from Ehrendorfer 1997). The initial pdf is Gaussian, with diagonal covariance structure.

is the set of eigenvectors of $\mathbf{S}_t^{\mathrm{TL}}$. This may be seen immediately, since applying $\mathbf{PM}_t$ to (5.5) results in:

$$\mathbf{PM}_t\Big(\mathbf{VM}_t^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{PM}_t\mathbf{Z}_0\Big) = \mathbf{PM}_t\Big(\mathbf{Z}_0\Lambda\Big)\,, \tag{5.7}$$

which is the same as (using (5.3)):

$$\mathbf{S}_t^{\mathrm{TL}}\mathbf{Z}_t = \mathbf{Z}_t\Lambda. \tag{5.8}$$

The orthogonality relation of the *time–evolved SVs* $\mathbf{Z}_t$ follows from the orthogonality relationship of $\mathbf{Z}_0$:

$$\mathbf{Z}_t^{\mathrm{T}}\mathbf{Z}_t = \Big(\mathbf{PM}_t\mathbf{Z}_0\Big)^{\mathrm{T}}\Big(\mathbf{PM}_t\mathbf{Z}_0\Big) = \mathbf{Z}_0^{\mathrm{T}}\underbrace{\mathbf{M}_t^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{PM}_t\mathbf{Z}_0} = \mathbf{Z}_0^{\mathrm{T}}\underbrace{\mathbf{V}^{-1}\mathbf{Z}_0\Lambda} = \Lambda, \tag{5.9}$$

where the equality of the underbraced terms follows from (5.5). Consequently, by (5.8) and (5.9), the reconstruction of $\mathbf{S}_t^{\mathrm{TL}}$ through its set of eigenvectors is given by:

$$\mathbf{S}_t^{\mathrm{TL}} = \mathbf{Z}_t\Lambda\mathbf{Z}_t^{-1} = \mathbf{Z}_t\mathbf{Z}_t^{\mathrm{T}}\,, \tag{5.10}$$

since $\mathbf{Z}_t^{\mathrm{T}} = \Lambda\mathbf{Z}_t^{-1}$. Particular attention must be given to the way in which the evolved SVs enter eq. (5.10), as a result of their normalization (5.9). It is of interest to note that (5.9) written in the form:

$$\mathbf{Z}_t^{\mathrm{T}}\mathbf{Z}_t\Lambda^{-1} = \mathbf{I} \tag{5.11}$$

implies that the time–evolved SVs satisfy:

$$\mathbf{Z}_t^{\mathrm{T}}\big(\mathbf{S}_t^{\mathrm{TL}}\big)^{-1}\mathbf{Z}_t = \mathbf{I} \tag{5.12}$$

which is immediately seen (from (5.10)) by noting that $\mathbf{Z}_t$ is also the set of eigenvectors to the inverse of $\mathbf{S}_t^{\mathrm{TL}}$:

$$\big(\mathbf{S}_t^{\mathrm{TL}}\big)^{-1}\mathbf{Z}_t = \mathbf{Z}_t\Lambda^{-1}. \tag{5.13}$$

Eq. (5.8), or (5.10), shows that the time–evolved SVs are the eigenvectors of the forecast error covariance matrix defined in (5.3). Assuming that the set $\mathbf{Z}_t$ is ordered corresponding to the associated eigenvalues λ_i (from largest to smallest), it is evident that the fraction of total variance $v_{k,t}^{\mathrm{TL}}$ in $\mathbf{S}_t^{\mathrm{TL}}$ accounted for by the first k vectors $\mathbf{Z}_t$, at time t, is given by:

$$v_{k,t}^{\mathrm{TL}} = \big[\mathrm{trace}(\mathbf{S}_t^{\mathrm{TL}})\big]^{-1}\sum_{i=1}^{k}\lambda_i = \Big(\sum_{i=1}^{n}\lambda_i\Big)^{-1}\sum_{i=1}^{k}\lambda_i\,. \tag{5.14}$$

(b) A remark on the eigenstructure of covariance matrices

It is at this point important to consider briefly the following dimensional argument relevant when one tries to determine the eigenstructure of a covariance matrix. Consider a state vector consisting of (for simplicity) only vorticity and temperature components (assumed to be measured in s^{-1} and K). In that case the set $\mathbf{Z}_0$, as well as the set:

$$\mathbf{Z}_t^- \equiv \mathbf{M}_t \mathbf{Z}_0 \tag{5.15}$$

consists of vectors carrying these dimensions (it is assumed that the tangent–linear model $\mathbf{M}_t$ does not change dimensions). The set $\mathbf{Z}_t^-$ is the set of evolved SVs $\mathbf{Z}_t$, as defined in (5.6), without the final norm applied. Now, it becomes obvious that in order to pose a physically meaningful problem in (5.4), $\mathbf{P}$ must be chosen such that the different physical units (s^{-1} and K) of $\mathbf{Z}_t^-$ are appropriately converted into a physically meaningful quantity (e.g., energy per unit mass). In the present example, this can be achieved, for example, through a diagonal matrix $\mathbf{P}$ with entries of dimension m (length) and $ms^{-1}K^{-1}$, respectively. If this is done, then the vectors in $\mathbf{Z}_t$ (see (5.6)) possess the same physical unit (velocity in the present example), and $\hat{J}$, as well as all entries of $\mathbf{S}_t^{\mathrm{TL}}$ possess the same physical dimensions (energy per unit mass in the present example). Note clearly that the entries of $\mathbf{V}$ do not possess the same physical units, as it is defined as the covariance matrix of the state vector. Note also that this property of $\mathbf{V}$ makes the constraint in (5.4) purely nondimensional. This necessity to account for different physical units in the state vector in the formulation of $\hat{J}$ can also be understood more clearly by considering directly the eigenproblem (5.8), asking for the eigendecomposition of $\mathbf{S}_t^{\mathrm{TL}}$. There it is seen immediately that this eigendecomposition can only be physically meaningful if the entries of $\mathbf{S}_t^{\mathrm{TL}}$ all have the same physical units, which is, in view of the definition (5.3), only achieveable if $\mathbf{P}$ carries the correct dimensions. This argument becomes even clearer by noting that the trace of $\mathbf{S}_t^{\mathrm{TL}}$ is equal to the sum of its eigenvalues (see (5.14)), which can only be achieved (in a physically meaningful way) if the elements (on the diagonal) of $\mathbf{S}_t^{\mathrm{TL}}$ have the same unit. In a more abstract sense, the problem can be summarized by saying that one cannot ask in a meaningful way for the eigendecomposition of a covariance matrix of different variables, since such an eigendecomposition depends always on the units used for measuring the variables. For example, it is not meaningful to ask for the eigenstructure of $\mathbf{V}$ (note that (5.8) simplifies to this question for $\mathbf{M}_t$ and $\mathbf{P}$ both taken identity), because this amounts to asking, in the present example, for the eigenstructure of a matrix with entries of s^{-2} (vorticity variances), K^2 (temperature variances) and $s^{-1}K$ (vorticity–temperature covariances).

Obviously, this problem has been recognized before. Trefethen and Bau (1997) state that *eigenvalue problems make sense only when the range and the domain spaces are the same* (which amounts to ensuring dimensional correctness as outlined above). The problem is also referred to as the lack of scale–invariance when considering covariance matrices in multivariate statistical problems (Mardia et al. 1982), where it may be interpreted as the dependence of the eigenstructure of a covariance matrix on the units in which the variables are measured. Similarly, Wilks (1995) points out that it is important to preconsider the physical units used to measure the variables entering a principal component analysis due to the effect of these units on the covariance matrix.

Summarizing, it is unavoidable to take care of the physical dimensions of the variables properly, in order for the eigenproblem (5.8) to be meaningful (i.e., for the eigenvalues to have meaningful physical units). Or, stated differently, as soon as the eigenstructure of a covariance matrix is considered, a norm choice must be made, and it is made – either explicitly or implicitly – as soon as the eigenstructure is computed. In the above example, when $\mathbf{P}$ is correctly considered, the eigenvalues appearing in Λ in (5.8) also have dimension of energy per unit mass. In a univariate context (i.e., only one variable, for example, vorticity) no dimensional problems occur (as measuring vorticity in different units will not qualitatively change the solution to the eigenproblem (5.8)). However, one still has to consider the possibility of giving different relative weights to the components of the (vorticity) state vector. Clearly, in the above multivariate example of vorticity and temperature, $\mathbf{P}$ may still be set to identity, but the computed solution in this case does then in effect amount to assuming that $\mathbf{P}$ has been taken diagonal with entries of 1s and $1\mathrm{K}^{-1}$ respectively (or any other dimensional or nondimensional multiple thereof). Consequently, if the outlined dimensional problem is disregarded in this way, the eigenproblem (5.8) can still be solved numerically, but then one is implicitly assuming that units of vorticity and temperature variances can be added (in the above example). Clearly then, the solution will change, if temperature is measured instead in degrees Fahrenheit or in centigrades.

Note, however, that it is always possible to compute a dimensionally correct square–root of a covariance matrix (e.g., $\mathbf{V}$), or its dimensionally correct inverse even in the case when its entries do have different physical units. If the square–root is computed through an eigendecomposition in this case, the computed square–root depends on the units used to measure the variables. After this discussion, it is assumed from now on that the entries of $\mathbf{S}_t^{\mathrm{TL}}$ have the same physical units (as mentioned above that will always be the case in a univariate situation; in the multivariate situation it may be achieved by proper specification of $\mathbf{P}$). To keep generality, the operator $\mathbf{P}$ has been introduced.

(c) The optimization problem with $\mathbf{M}_t = \mathbf{I}$

In order to understand the nature of the eigendecomposition (5.8) more clearly, it is instructive at this point to consider the special situation that $\mathbf{M}_t = \mathbf{I}$. In this case, by (5.3), $\mathbf{S}_t^{\mathrm{TL}}$ is just $\mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}}$, and problem (5.5) reduces to the computation of the eigenvectors of $\mathbf{V}\mathbf{P}^{\mathrm{T}}\mathbf{P}$ denoted as $\tilde{\tilde{\mathbf{Z}}}$ and associated eigenvalues $\tilde{\Lambda}$:

$$\mathbf{V}\mathbf{P}^{\mathrm{T}}\mathbf{P}\tilde{\tilde{\mathbf{Z}}} = \tilde{\tilde{\mathbf{Z}}}\tilde{\Lambda} \qquad \text{subject to:} \qquad \tilde{\tilde{\mathbf{Z}}}^{\mathrm{T}}\mathbf{V}^{-1}\tilde{\tilde{\mathbf{Z}}} = \mathbf{I}. \tag{5.16a, b}$$

Eq. (5.16) is the special case of (5.5) for no dynamics present, and it leads directly to the determination of the eigenstructure of the unevolved covariance matrix:

$$\mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}}\tilde{\mathbf{Z}} = \tilde{\mathbf{Z}}\tilde{\Lambda}, \tag{5.17}$$

where:

$$\tilde{\mathbf{Z}} \equiv \mathbf{P}\tilde{\tilde{\mathbf{Z}}} \tag{5.18}$$

corresponds directly to $\mathbf{Z}_t$ in (5.6), specialized to this situation, but has been denoted by a different symbol here (note similarly that $\tilde{\tilde{\mathbf{Z}}}$ corresponds to $\mathbf{Z}_0$ in the general situation). Clearly, in analogy to (5.9) and (5.10) it is true here that:

$$\tilde{\mathbf{Z}}^{\mathrm{T}}\tilde{\mathbf{Z}} = \tilde{\Lambda} \qquad \text{and} \qquad \mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}} = \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{\mathrm{T}}. \tag{5.19a, b}$$

(d) Some additional remarks

Before proceeding, some additional remarks are made with the intention to illustrate some of the foregoing developments.

(i) First, note that – in the case of active dynamics, $\mathbf{M}_t \neq \mathbf{I}$ – the eigenvectors $\tilde{\mathbf{Z}}$ of the unevolved covariance matrix (in the dimensionally correct, see above, form (5.17)) are *not* the initial SVs $\mathbf{Z}_0$, since the initial SVs must satisfy (5.5), and the set $\tilde{\mathbf{Z}}$ must satisfy (5.17). Consequently, it is clearly seen that the eigenvectors $\mathbf{Z}_t$ of $\mathbf{S}_t^{\mathrm{TL}}$ are not the time–evolved eigenvectors of $\mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}}$.

(ii) However, as the optimization time t becomes smaller, the set $\mathbf{Z}_t$ is approaching the set $\tilde{\mathbf{Z}}$, as $\mathbf{S}_t^{\mathrm{TL}}$ is approaching $\mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}}$, and, at the same time, $\mathbf{Z}_0$ is approaching $\tilde{\tilde{\mathbf{Z}}}$. Thus, the analogy of (5.19b) and (5.10) suggests to consider the comparison of the relevant eigenstructures, namely, $\tilde{\Lambda}$ and Λ and associated eigenvectors $\tilde{\mathbf{Z}}$ and $\mathbf{Z}_t$, as an indicatation of how far the dynamics $\mathbf{M}_t$ have removed $\mathbf{S}_t^{\mathrm{TL}}$ from its equivalent, $\mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}}$, at initial time. Note that this consideration shows that it is, in particular, probably not adequate to compare the initial SVs $\mathbf{Z}_0$ with the time–evolved SVs $\mathbf{Z}_t$ because the former are *not* the

eigenvectors of a covariance matrix, whereas the latter are. In particular, in analogy to (5.14), the initial fraction of variance, accounted for by the first k vectors $\tilde{\mathbf{Z}}$, is defined as:

$$v_{k,0} = \left[\mathrm{trace}(\mathbf{PVP}^{\mathrm{T}})\right]^{-1} \sum_{i=1}^{k} \tilde{\lambda}_i \,. \tag{5.20}$$

(iii) At this point it becomes most clear that the two normalization constraints in (5.16b) and (5.19a) are equivalent, in the sense that either one of these two when taken together with the eigenvector equation (5.17), and the definition (5.18), implies the other one. In other words, given that $\tilde{\mathbf{Z}}$ satisfies eq. (5.17), plus the normalization (5.19a), then it follows immediately that $\tilde{\mathbf{Z}}$ satisfies (5.16b), rewritten conveniently as:

$$\tilde{\mathbf{Z}}^{\mathrm{T}} \left(\mathbf{PVP}^{\mathrm{T}}\right)^{-1} \tilde{\mathbf{Z}} = \mathbf{I}. \tag{5.21}$$

On the other hand, given that $\tilde{\mathbf{Z}}$ satisfies (5.17), together with (5.21) (which is (5.16b)), then (5.19a) follows. Eqs. (5.16b) (or, equivalently (5.21)) and (5.19a) are just two different ways to write the normalization constraint for the eigenvectors $\tilde{\mathbf{Z}}$ of $\mathbf{PVP}^{\mathrm{T}}$, defined through the eigenvector equation (5.17). Note that (5.19b) follows immediately, given, for example, (5.17) and (5.19a). The very same remarks apply to the normalization of the eigenvectors $\mathbf{Z}_t$ of $\mathbf{S}_t^{\mathrm{TL}}$ as may be seen from equations (5.8), (5.9), and (5.12). It is emphasized that this remark implies that one may solve (5.17) directly (subject to, e.g., eq. (5.19a)) which leads directly to $\tilde{\mathbf{Z}}$, and the representation (5.19b); then, (5.18) may be used to determine $\tilde{\tilde{\mathbf{Z}}}$.

Note that this approach, as just outlined, is computationally different from the approach that is commonly used to arrive at the eigenstructure of $\mathbf{S}_t^{\mathrm{TL}}$ in (5.8). Here, first the initial SVs, namely, the set $\mathbf{Z}_0$ is determined, by solving (5.5) (see also below; this amounts to solving (5.16) in the above context), from which $\mathbf{Z}_t$, defined in (5.6), is derived; this set $\mathbf{Z}_t$ satisfies automatically (5.9), (5.10), and (5.12). Clearly, in view of the foregoing comments, this latter procedure is entirely equivalent to directly solving for the eigenstructure of $\mathbf{S}_t^{\mathrm{TL}}$ defined in (5.8), subject to (5.9). One of the reasons for not proceeding in this way in the context of computing singular vectors in atmospheric models is that one is then faced with the difficulty of deriving $\mathbf{Z}_0$ from $\mathbf{Z}_t$ by (5.6), as this requires inverting $\mathbf{M}_t$. And, in the context of computing singular vectors in atmospheric models, it is of primary interest to have the initial SVs $\mathbf{Z}_0$ available (see also below).

(iv) Eq. (5.17) shows the specialization of (5.8) to the situation that $\mathbf{M}_t = \mathbf{I}$. Clearly, one does not necessarily have to consider the eigenstructure of $\mathbf{V}$ in the same norm as the

eigenstructure of $\mathbf{M}_t\mathbf{V}\mathbf{M}_t^{\mathrm{T}}$, which is the case in (5.17). To introduce this generalization, (5.4) is generalized to:

$$\max_{\mathbf{z}_0} \quad \bar{J}(\mathbf{z}_0;t) = \Big(\mathbf{P}\mathbf{M}_t\mathbf{P}^{-1}\mathbf{Q}\mathbf{z}_0\Big)^{\mathrm{T}}\Big(\mathbf{P}\mathbf{M}_t\mathbf{P}^{-1}\mathbf{Q}\mathbf{z}_0\Big) \quad \text{subject to:} \quad \mathbf{z}_0^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{z}_0 = 1, \tag{5.22}$$

where $\mathbf{Q}$ is a norm–defining operator with properties analogous to the properties of $\mathbf{P}$, leading directly to the eigenstructure of the time–evolved covariance matrix in the following form:

$$\Big(\mathbf{P}\mathbf{M}_t\mathbf{P}^{-1}\mathbf{Q}\Big)\mathbf{V}\Big(\mathbf{P}\mathbf{M}_t\mathbf{P}^{-1}\mathbf{Q}\Big)^{\mathrm{T}}\bar{\mathbf{Z}}_t = \bar{\mathbf{Z}}_t\bar{\Lambda}\,, \tag{5.23}$$

where the relevant eigenvectors $\bar{\mathbf{Z}}_t$ are obtained from the solution to (5.22), called $\bar{\mathbf{Z}}_0$, as:

$$\bar{\mathbf{Z}}_t = \mathbf{P}\mathbf{M}_t\mathbf{P}^{-1}\mathbf{Q}\bar{\mathbf{Z}}_0. \tag{5.24}$$

Equations analogous to (5.9) and (5.10) also hold. For the choice $\mathbf{Q} = \mathbf{P}$, eqs. (5.22) – (5.24) simplify to the description following eq. (5.3). However, for $\mathbf{Q} \neq \mathbf{P}$, it may be seen that when $\mathbf{M}_t = \mathbf{I}$ is taken in (5.23), the initial covariance matrix $\mathbf{V}$ is considered in the norm $\mathbf{Q}$ (see eq. (5.17)). However, this generalization will not be considered here further, but the future investigation of this generalization might be of interest.

(v) Referring back to remark (i), it is noted that even though the initial SVs $\mathbf{Z}_0$ are not the eigenvectors of $\mathbf{V}$, they may still be used to reconstruct $\mathbf{V}$, because, using the constraint in (5.5), it may be seen immediately that it is true that:

$$\mathbf{Z}_0\mathbf{Z}_0^{\mathrm{T}} = \mathbf{Z}_0\big(\mathbf{V}^{-1}\mathbf{Z}_0\big)^{-1} = \mathbf{Z}_0\mathbf{Z}_0^{-1}\big(\mathbf{V}^{-1}\big)^{-1} = \mathbf{V}. \tag{5.25}$$

Note, however, that this representation of $\mathbf{V}$ is, at least potentially (and also depending on the length of the optimization time interval t), quite different from its reconstruction through its eigenvectors as expressed in (5.19). Eq. (5.25) may also be quite different from an analogous form derived from (5.16b):

$$\mathbf{V} = \tilde{\tilde{\mathbf{Z}}}\tilde{\tilde{\mathbf{Z}}}^{\mathrm{T}}. \tag{5.26}$$

Nevertheless, the representation (5.25) has the highly appealing property that its (tangent–linearly) time–evolved counterpart is the *eigendecomposition* of $\mathbf{S}_t^{\mathrm{TL}}$ (see (5.3), (5.6), and (5.10)). Consequently, if interest if focused on the forecast error covariance matrix $\mathbf{S}_t^{\mathrm{TL}}$ the decomposition (5.25) of $\mathbf{V}$ should be of primary interest rather than an analogous decomposition based on the eigenstructure of $\mathbf{V}$ (such as, e.g., (5.17), (5.26)). Note in this context some similarities and differences to the approach suggested by Bouttier (1996). In

a sense then, the decomposition (5.25) is somewhat similar to a Cholesky decomposition of $\mathbf{V}$ (e.g., Trefethen and Bau 1997), the difference being, of course, that $\mathbf{Z}_0$ is, in general, *not* lower–triangular (or, upper–triangular). The decomposition in (5.25) will be considered in more detail in the next subsection.

(vi) It is noted for completeness that the solution to (5.5) may be obtained by solving the following standard eigenproblem:

$$\left(\mathbf{P}\mathbf{M}_t\mathbf{V}^{1/2}\right)^{\mathrm{T}} \underbrace{\left(\mathbf{P}\mathbf{M}_t\mathbf{V}^{1/2}\right)}_{\equiv\mathbf{G}_t} \hat{\mathbf{z}}_0 = \lambda\hat{\mathbf{z}}_0 \quad \text{with:} \quad \hat{\mathbf{z}}_0^{\mathrm{T}}\hat{\mathbf{z}}_0 = 1\,, \tag{5.27}$$

followed by the backtransformation:

$$\mathbf{z}_0 = \mathbf{V}^{1/2}\hat{\mathbf{z}}_0. \tag{5.28}$$

The computation (5.27) and (5.28) yields the initial SVs $\mathbf{Z}_0$, as computed from $\hat{\mathbf{Z}}_0$, explicitly. It requires the availability of the square–root of $\mathbf{V}$. Alternatively, one may solve (5.8) directly for $\mathbf{Z}_t$ (see also the comments made in (iii)); in this case, the set $\mathbf{Z}_0$ does not become explicitly available. However, this latter approach is leading directly to the eigenstructure of $\mathbf{S}_t^{\mathrm{TL}}$. It requires, however, the availability of $\mathbf{V}$ either explicitly or in operator form.

It is at this point that the term singular vector becomes most obvious, since now the singular value decomposition (e.g., Golub and Van Loan 1989, Trefethen and Bau 1997; see, also, Chu et al. 1997) of $\mathbf{G}_t$ may be written as:

$$\mathbf{G}_t = \left(\mathbf{Z}_t\Lambda^{-1/2}\right)\Lambda^{1/2}\hat{\mathbf{Z}}_0^{\mathrm{T}}\,, \tag{5.29}$$

showing that $\hat{\mathbf{Z}}_0$ are the right singular vectors of the norm–augmented tangent–linear operator $\mathbf{G}_t$, with the relevant time evolution given by (5.6), and the "standard" normalizations (see (5.9) and (5.27)):

$$\left(\mathbf{Z}_t\Lambda^{-1/2}\right)^{\mathrm{T}}\left(\mathbf{Z}_t\Lambda^{-1/2}\right) = \mathbf{I} \qquad \text{and} \qquad \hat{\mathbf{Z}}_0^{\mathrm{T}}\hat{\mathbf{Z}}_0 = \mathbf{I}. \tag{5.30}$$

Now, clearly, the right singular vectors $\hat{\mathbf{Z}}_0$ and the squares of the singular values, $\Lambda^{1/2}$, determine the eigendecomposition (5.27):

$$\mathbf{G}_t^{\mathrm{T}}\mathbf{G}_t = \hat{\mathbf{Z}}_0\Lambda\hat{\mathbf{Z}}_0^{\mathrm{T}}. \tag{5.31}$$

5.3. An SV–based Monte Carlo technique

Given information about initial uncertainty in the form (2.3), a large sample consistent with this initial uncertainty may be evolved in time to estimate the time–evolved statistics according to (3.10). Generating the initial realizations requires knowledge of the initial pdf. If only the first two moments are known, a possible (and probably most reasonable) choice is the normal pdf, as written in eq. (2.2). Such a choice is probably quite reasonable, if only the first two moments are available (no other error characteristics known), in view of the Central Limit Theorem. A Monte Carlo sampling technique based on the normal–distribution assumption, as well as on the initial SVs defined in section 5.2 is considered here.

Under the assumption that the vector $\mathbf{p}$ is multivariate standard normal:

$$\mathbf{p} \sim \mathcal{N}_n(0, \mathbb{I}), \tag{5.32}$$

which implies that the components of $\mathbf{p}$ are mutually independent and univariately normally distributed (and vice versa; see, e.g., Rao 1965), the random vector $\mathbf{q}$ defined through:

$$\mathbf{q} = \mathbf{x}_0^c + \mathbf{V}^{1/2}\mathbf{p} \tag{5.33}$$

is normally distributed with parameters as indicated in (2.2). Here $\mathbf{V}^{1/2}$ represents a square–root of $\mathbf{V}$. Such a square–root is not uniquely defined, but will always exist since $\mathbf{V}$ is positive definite.

The SV–based Monte Carlo technique suggested here is based on the square–root representation of $\mathbf{V}$ described in eq. (5.25). Given this representation of $\mathbf{V}$, it is clearly possible to choose as the square–root of $\mathbf{V}$ its decomposition by the initial SVs $\mathbf{Z}_0$. More specifically, if only the first k SVs, collected in the matrix $\mathbf{Z}_0^{(k)}$, as defined by the problem (5.4), are available, the random vector $\mathbf{q}^{(k)}$ is constructed in analogy to (5.33) as:

$$\mathbf{q}^{(k)} = \mathbf{x}_0^c + \mathbf{Z}_0^{(k)}\mathbf{p}\,, \tag{5.34}$$

which amounts to an approximation of the full initial covariance structure in the following form:

$$\mathbf{V}^{(k)} = (\mathbf{Z}_0^{(k)})(\mathbf{Z}_0^{(k)})^{\mathrm{T}}. \tag{5.35}$$

The matrix $\mathbf{Z}_0^{(k)}$ is $n \times n$ and contains the k leading SVs in its first k columns, with the remaining $n - k$ columns being zero:

$$\mathbf{Z}_0^{(k)} = \begin{pmatrix} . & \cdots & . & . & \cdots & . \\ \mathbf{z}_0^{(1)} & \cdots & \mathbf{z}_0^{(k)} & \mathbf{0} & \cdots & \mathbf{0} \\ . & \cdots & . & . & \cdots & . \end{pmatrix}. \tag{5.36}$$

Clearly, (5.34) becomes eq. (5.33), just as (5.35) becomes (5.25), in the case that all n SVs are available to define a square–root of $\mathbf{V}$. Clearly, for the specification (5.34), the covariance of $\mathbf{q}^{(k)}$ is:

$$\mathrm{cov}(\mathbf{q}^{(k)}) = (\mathbf{Z}_0^{(k)})(\mathbf{Z}_0^{(k)})^{\mathrm{T}} = \mathbf{V}^{(k)}\,, \tag{5.37}$$

which when evolved *tangent–linearly* in time, see eqs. (5.3), (5.6), and (5.10), leads to the approximation of $\mathbf{S}^{\mathrm{TL}}$ by its first k eigenvectors. However, here $\mathbf{q}^{(k)}$ represents a realization from the initial pdf that can be evolved nonlinearly, which will clearly be beneficial as the tangent–linear approximation becomes less valid. However, this choice of sampling from the initial pdf hides in a sense the trailing SVs from the covariance structure, so that the vectors $\mathbf{q}^{(k)}$ obtained in the sampling process only "see" the part of the initial covariance structure that, when time evolved, contributes most to $\mathbf{S}^{\mathrm{TL}}$. In a sense then, sampling based on (5.34) represents an attempt to combine the appealing properties of the SVs with a possible way to overcome limitations posed by tangent–linear restrictions, and it exploits explicitly the fact that the initial SVs evolve into the eigenvectors of $\mathbf{S}^{\mathrm{TL}}$, by writing the initial covariance in terms of the initial SVs (rather than in terms of the eigenvectors of the initial covariance $\mathbf{V}$; recall that these eigenvectors and the initial–time SVs are different – see 5.2.d(i)). Note that $\mathbf{V}^{(k)}$, the covariance structure used for the sampling process, is rank–deficient (or, singular) for $k < n$, but that it is still non–negative definite; this rank deficiency is not considered to be an important limitation at this point at all (see, e.g., Rao (1965) for the definition of the multivariate normal pdf in this case). Note also that in the tangent–linear regime the technique clearly simplifies to the tangent–linear prediction of $\mathbf{S}^{\mathrm{TL}}$.

The rationale for the SV–based MC technique is similar to the generation of initial perturbed states as carried out at ECMWF (see, e.g., Molteni et al. 1996). The matrix–vector product $\mathbf{Z}_0^{(k)}\mathbf{p}$ in eq. (5.34) may be interpreted as performing the linear combination of initial SVs and the subsequent rotation. Note that in the expression (5.34) both steps are carried out simultaneously.

So far the description of this proposed technique has been perfectly general. However, there is an apparent problem with scaling the initial variances, since the expression (5.35) will be clearly "variance–deficient" when compared to the full initial variance structure, in the sense that the variance represented by $\mathbf{V}^{(k)}$ is smaller than the variance represented by $\mathbf{V}^{(n)}$. In order to resolve this issue in such a way that the procedure coincides with the full initial covariance procedure for $k = n$, a scaling is suggested here that is based on the comparison between the initial total variance explained by the first k initial SVs, denoted $v_{k,0}^t$, and the initial total variance carried by all initial SVs, denoted $v_{n,0}^t$ (relevant spectra

are shown in section 5.4). Clearly, as discussed in section 5.2, the assessment of these total variances again requires consideration of a metric for the simultaneous treatment of different variables. In view of the discussion in section 5.2 (see, e.g., eqs. (5.14) and (5.20)), this metric is taken to be $\mathbf{P}$.

Given k initial SVs, the total variance contained in the matrix $\mathbf{V}^{(k)}$, denoted $v^t_{k,0}$ may be directly computed from the set $\mathbf{Z}^{(k)}_0$ by noting:

$$v^t_{k,0} \equiv \mathrm{trace}(\mathbf{P}\mathbf{V}^{(k)}\mathbf{P}^{\mathrm{T}}) = \mathrm{trace}\Big(\mathbf{P}(\mathbf{Z}^{(k)}_0)(\mathbf{Z}^{(k)}_0)^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\Big) =$$

$$= \mathrm{trace}\Big((\mathbf{P}\mathbf{Z}^{(k)}_0)(\mathbf{P}\mathbf{Z}^{(k)}_0)^{\mathrm{T}}\Big) = \sum_{i=1}^{n}\sum_{l=1}^{k}\Big((\mathbf{P}\mathbf{Z}^{(k)}_0)_{il}\Big)^2 \equiv \| \mathbf{P}\mathbf{Z}^{(k)}_0 \|^2_F , \tag{5.38}$$

where the Frobenius matrix norm of $\mathbf{P}\mathbf{Z}^{(k)}_0$ has been introduced. Note that it is true for the Frobenius norm of any matrix $\mathbf{A}$ that $\| \mathbf{A} \|^2_F = \mathrm{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = \mathrm{trace}(\mathbf{A}\mathbf{A}^{\mathrm{T}})$ (e.g., Parlett 1998, Trefethen and Bau 1997). This property is also evident in eq. (5.38). Taking equation (5.38) for the case $k = n$ (considering the situation that all n initial SVs are available) gives the total initial variance as follows:

$$v^t_{n,0} \equiv \mathrm{trace}(\mathbf{P}\mathbf{V}^{(n)}\mathbf{P}^{\mathrm{T}}) = \mathrm{trace}(\mathbf{P}\mathbf{V}\mathbf{P}^{\mathrm{T}}) = \mathrm{trace}(\mathbf{P}\mathbf{Z}_0\mathbf{Z}_0^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}) =$$

$$= \mathrm{trace}\Big(\mathbf{P}\mathbf{Z}_0(\mathbf{P}\mathbf{Z}_0)^{\mathrm{T}}\Big) = \sum_{i=1}^{n}\sum_{l=1}^{n}\Big((\mathbf{P}\mathbf{Z}_0)_{il}\Big)^2 \equiv \| \mathbf{P}\mathbf{Z}_0 \|^2_F \, . \tag{5.39}$$

Now, these equations show clearly that the factor α defined by:

$$\alpha \equiv \frac{v^t_{n,0}}{v^t_{k,0}} \geq 1 \tag{5.40}$$

is the factor by which the total variance in $\mathbf{V}$ is larger than the variance in $\mathbf{V}^{(k)}$ (both in the metric $\mathbf{P}$). Note that this factor goes to one, as k becomes larger. To make the truncated variance $v^t_{k,0}$ equal to the full initial variance it seems therefore possible to define a rescaled set of initial SVs as:

$$\hat{\mathbf{Z}}^{(k)}_0 \equiv (\sqrt{\alpha^s})\mathbf{Z}^{(k)}_0 , \tag{5.41}$$

where the exponent s is intended to take on only values of zero and one. Using (5.40) in (5.41), together with the properties of the Frobenius norm, and (5.38), it may be shown immediately that, for s=1:

$$\| \mathbf{P}\hat{\mathbf{Z}}^{(k)}_0 \|^2_F = \alpha \, \| \mathbf{P}\mathbf{Z}^{(k)}_0 \|^2_F = \frac{v^t_{n,0}}{v^t_{k,0}} v^t_{k,0} = v^t_{n,0} , \tag{5.42}$$

indicating that building the truncated initial covariance matrix $\mathbf{V}^{(k)}$ through the scaled set $\hat{\mathbf{Z}}_0^{(k)}$ (see eq. (5.35)) leads to an initial covariance structure containing the full total initial variance. Note that this scaling process does not require the availability of all SVs, but it does require an estimate of $v_{n,0}^t$. In view of (5.41), eq. (5.34) is rewritten in the following form:

$$\hat{\mathbf{q}}^{(k)} = \mathbf{x}_0^c + \hat{\mathbf{Z}}_0^{(k)}\mathbf{p} = \mathbf{x}_0^c + (\sqrt{\alpha^s})\mathbf{Z}_0^{(k)}\mathbf{p}\,, \tag{5.43}$$

which, in analogy to (5.37), shows that the covariance structure is defined by $\mathbf{V}^{(k)}$:

$$\mathrm{cov}(\hat{\mathbf{q}}^{(k)}) = \alpha^s(\mathbf{Z}_0^{(k)})(\mathbf{Z}_0^{(k)})^{\mathrm{T}} = \alpha^s\mathbf{V}^{(k)}\,, \tag{5.44}$$

but that, in addition, for $s = 1$, the total variance level carried by $\hat{\mathbf{q}}^{(k)}$ is the same as in the full initial covariance $\mathbf{V}$ (see (5.38)):

$$\mathrm{trace}\,(\mathbf{P}\,\mathrm{cov}(\hat{\mathbf{q}}^{(k)})\,\mathbf{P}^{\mathrm{T}}) = \alpha\,\mathrm{trace}\,(\mathbf{P}\mathbf{V}^{(k)}\mathbf{P}^{\mathrm{T}}) = \alpha v_{k,0}^t = v_{n,0}^t\,. \tag{5.45}$$

The use of (5.43) to generate initial perturbations will produce perturbations consistent with the initial covariance structure as assessed by the first k SVs (see (5.35)). Note that the rescaling process, as outlined above, is conservative in the sense that for both choices, $s = 0$ or $s = 1$, it approaches the correct variance level when k approaches n: for $s = 1$, the variance level is always kept precisely at $v_{n,0}^t$, whereas for $s = 0$, $v_{n,0}^t$ is (slowly) approached (from below) as more and more SVs are included. However, for $s = 1$, the rescaling might be a bit dangerous insofar, as it gives some over–proportional attention to the fastest–growing initial SVs, which might be expected to lead to an overestimation of the final variance structure when the realizations $\hat{\mathbf{q}}^{(k)}$ are time–evolved. Alternatively, one may consider taking $s = 0$ which will not do any scaling at the initial time. Note that both approaches, namely $s = 0$ and $s = 1$ become identical as k approaches n, since in this case α tends to one. In section 5.4, results are presented for no rescaling, $s = 0$, since the overemphasis of the first SVs might not be appropriate.

In conclusion it is noted that this rescaling procedure outlined above reflects again the basic question of how to properly scale initial perturbations in the nonlinear regime.

5.4. Results

In this section some of the ideas presented in sections 5.2 and 5.3 are illustrated within the framework of the quasigeostrophic (QG) atmospheric prediction model developed by Marshall and Molteni (1993). The case studied here in detail is the 48–hour period beginning on 17/02/1997/1200GMT. Results for a second case, 03/01/1994/0000GMT, will be

mentioned briefly. In the formalism introduced in the previous chapters the state variable here is vorticity at the three model levels (more precisely, the spectral expansion coefficients of vorticity). Thus, all covariance matrices are also defined in spectral space and are valid for the spectral components of vorticity. It might be mentioned for completeness that the state variable is taken to be vorticity and not the deviation of quasigeostrophic potential vorticity from planetary vorticity (vorticity and this deviation are assumed to be related through a linear operator that can be inverted). Since the state vector consists of only one physical variable, the dimensional problems associated with the computation of the eigenstructure of covariance matrices (as described in section 5.2.) do not appear with their full complications. The choice $\mathbf{P} = \mathbf{I}$ is therefore a valid choice in this context (see eq. (5.3)).

First, the influence of different norm specifications on the properties of SVs is investigated. SVs were computed for five different combinations of initial and final norm for the two cases (see Table 1). In order to refer to these five combinations the third letter of the internal identifier will be used. In experiment (a) (kinetic energy at both initial and final times) the problem described by eq. (4.17) is solved, where $\mathbf{C}^{\mathrm{T}}\mathbf{C}$ now denotes the kinetic energy metric. In experiments (c), (d), (g), and (h), problem (5.4) was solved; here in (c) and (g), the operator $\mathbf{P}^{\mathrm{T}}\mathbf{P}$ again denotes the kinetic energy metric, whereas in (d) and (h) $\mathbf{P}^{\mathrm{T}}\mathbf{P}$ is taken as identity. In solving problem (5.4), two different specifications for the initial covariance matrix $\mathbf{V}$ were chosen: in (c) and (d) $\mathbf{V}_0$ denotes a "flat" initial covariance structure corresponding to the background formulation for vorticity in the integrated forecast system (IFS) at ECMWF. In (g) and (h) $\mathbf{V}_{48}$ was used which is a tangent–linear prediction of $\mathbf{V}_0$ started 48 hours prior to the beginning of the optimization time interval (in carrying out this tangent–linear prediction the QG model was linearized around a basic state trajectory produced with the QG model for the relevant 48 hour time interval).

As mentioned above, in all experiments described here the dimensional variable $\mathbf{z}_0$ (see (4.17) and (5.4)) is vorticity (in spectral representation) at the three levels in the QG model. As a consequence, the initial covariance matrix $\mathbf{V}$ (also valid in spectral space) has dimensions of s^{-2} and the dimensional problems mentioned in section 5.2b do not occur with their full complications. More precisely, the choice $\mathbf{P} = \mathbf{I}$ is valid in this context; taking $\mathbf{P} \neq \mathbf{I}$ will amount to applying different weights (all of the same physical dimension) to the state vector. In any of the pictures further below, the display of perturbations to this state variable (like eigenvectors and singular vectors) is as follows. The state perturbation $\mathbf{z}_0$ is first scaled such that its kinetic energy is one (i.e., $\mathbf{z}_0^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{z}_0 = 1$). Then, this scaled perturbation is converted (from vorticity) to (dimensionless) streamfunction and displayed in physical space (the reason for this latter conversion from vorticity to streamfunction is

Table 1. Amplifications for the five leading SVs in terms of kinetic energy for various combinations of initial and final norms. In the head of each column the initial and the final norms are indicated, as well as the internal experiment identifier. The optimization time interval is 48 hours.

case 1: 17/02/1997/1200GMT.					
	KE/KE	$\mathbf{V}_0$/KE	$\mathbf{V}_0$/I	$\mathbf{V}_{48}$/KE	$\mathbf{V}_{48}$/I
	C7aa	C7cb	C7db	C7gb	C7hb
SV 1	151.31	79.50	78.37	18.28	18.81
SV 2	149.12	71.47	71.11	20.24	21.38
SV 3	138.75	50.64	45.11	19.64	17.51
SV 4	117.32	40.63	35.79	17.10	14.41
SV 5	104.32	34.37	31.66	10.07	11.37
case 2: 03/01/1994/0000GMT.					
	KE/KE	$\mathbf{V}_0$/KE	$\mathbf{V}_0$/I	$\mathbf{V}_{48}$/KE	$\mathbf{V}_{48}$/I
	H7ab	H7cb	H7db	H7gb	H7hb
SV 1	98.04	31.78	30.43	6.94	6.78
SV 2	80.75	31.23	24.20	7.42	5.77
SV 3	67.10	27.19	24.17	7.72	4.89
SV 4	55.82	25.45	24.30	4.82	7.41
SV 5	52.00	23.80	19.08	6.43	5.80

that it produces smoother fields). The fractions contributed by a given perturbation field at each model level to the normalized value of unit kinetic energy are printed in the first line of each subpanel (ordered from top to bottom). Dimensionless variables are obtained by using the radius of the earth as length scale, and $(2\Omega)^{-1}$ as time scale, where Ω is the angular speed of rotation of the earth.

In addition to the results of Table 1, the three leading SVs are shown in Figs. 4 (KE/KE), 5 ($\mathbf{V}_0$/KE), and 6 ($\mathbf{V}_{48}$/KE), in the format described above, for both initial and final times. Note that the final SVs plotted here are the fields $\mathbf{Z}_t^-$, defined in (5.15) (i.e., if the final norm is different from the identity, it was not applied after time–evolving the initial SV). For the amplifications of these structures reference is made to Table 1. It is of interest to note that the amplifications *decrease* as the initial norm is changed from kinetic energy to $\mathbf{V}_0$ and then $\mathbf{V}_{48}$ (experiments a, c, g). The reason for this result was not analysed in detail, but it could be that with a more structured initial constraint (i.e., $\mathbf{V}_{48}$) there is less potential for growth.

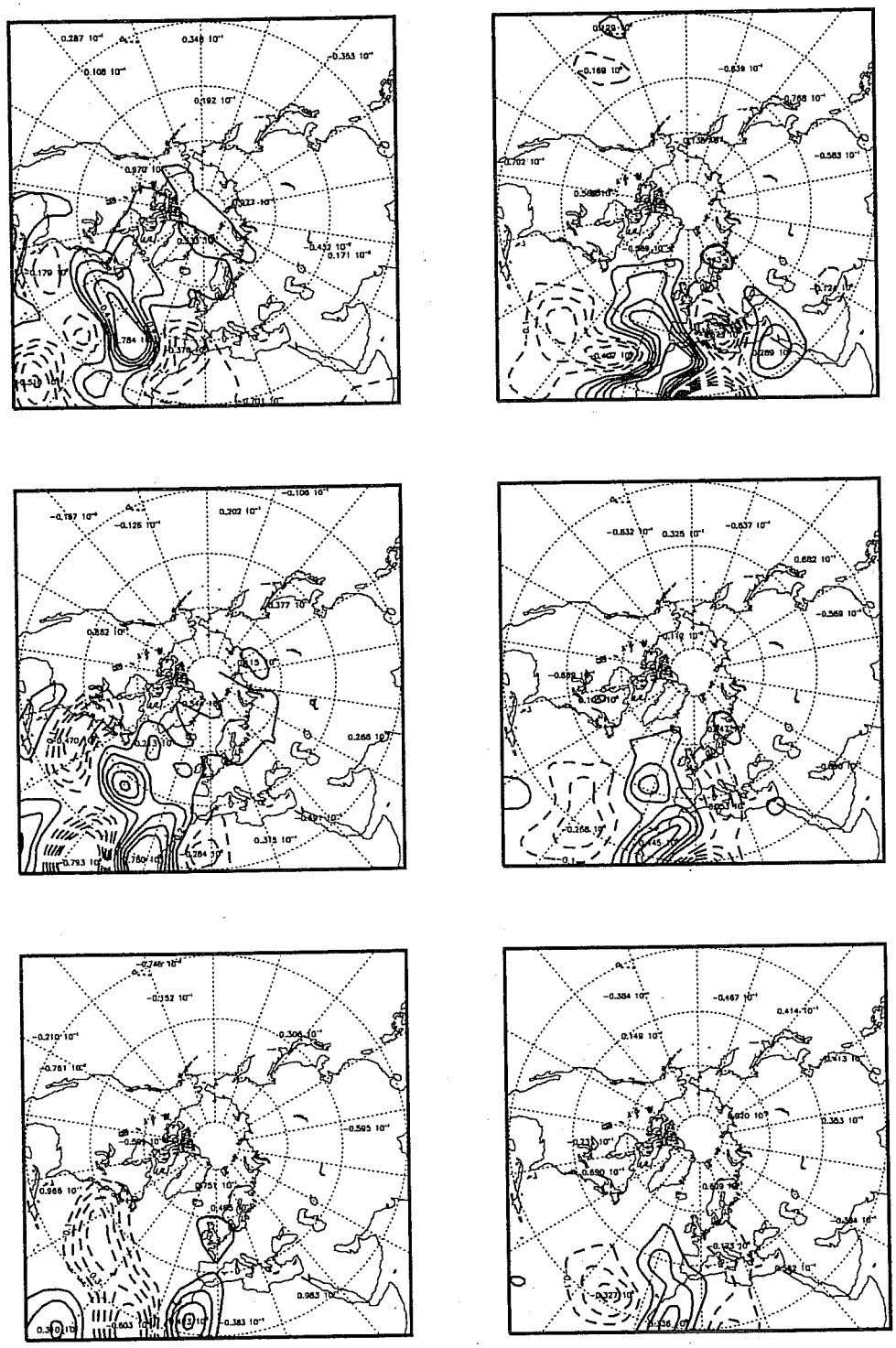

Fig. 4. The leading three singular vectors for case 1 (i.e., 17/02/1997/1200 GMT) for an optimization time of 48 hours computed with quasigeostrophic dynamics, at initial (left column) and final (right column) times, at the three model levels. Panel (a) SV 1, (b) SV 2, (c) SV 3. Initial and final norm is kinetic energy. For the scaling convention of the SVs refer to section 5.4. Level contributions to unit kinetic energy are printed in the first line of each subpanel.

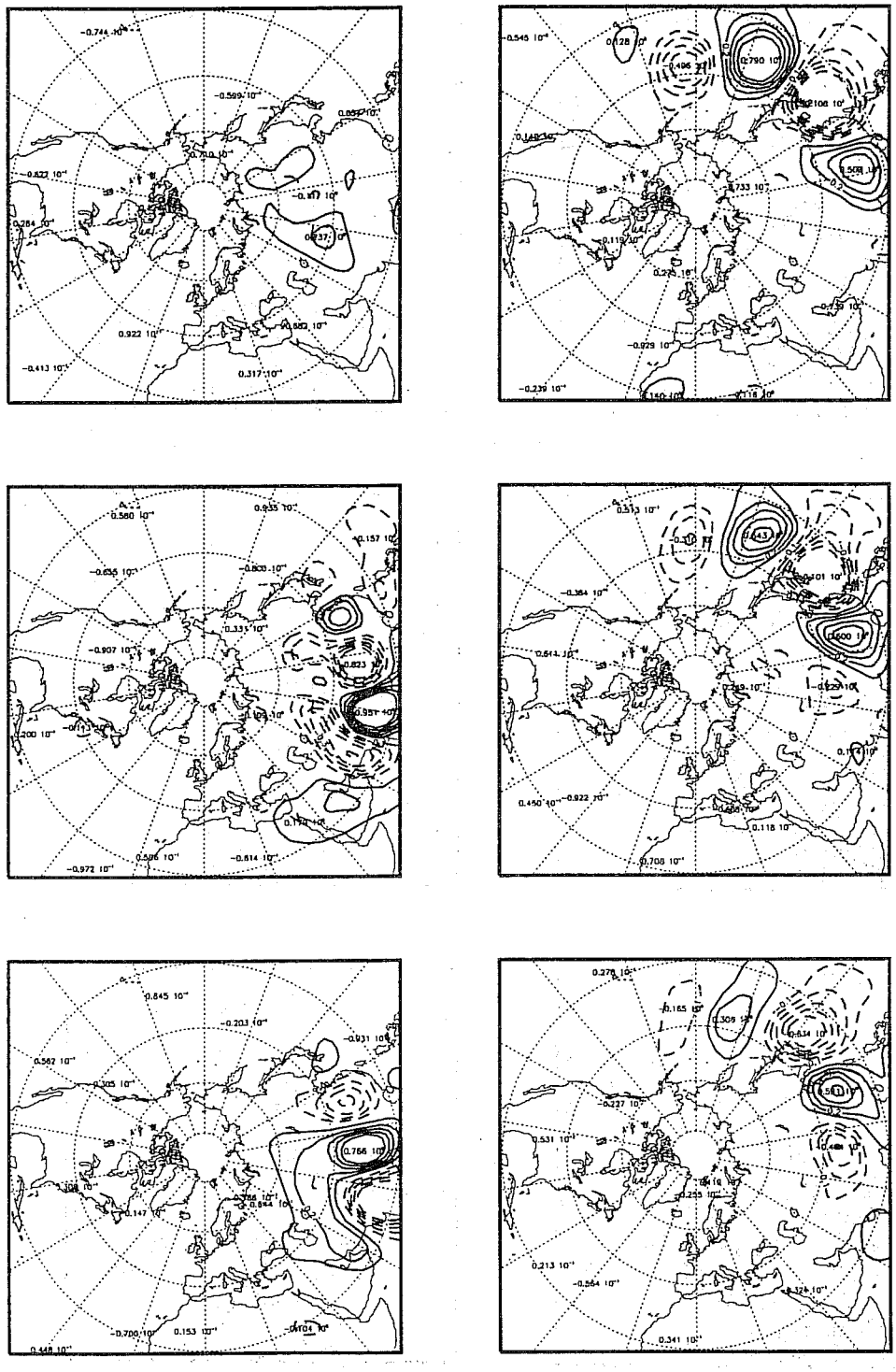

Fig. 4. (b).

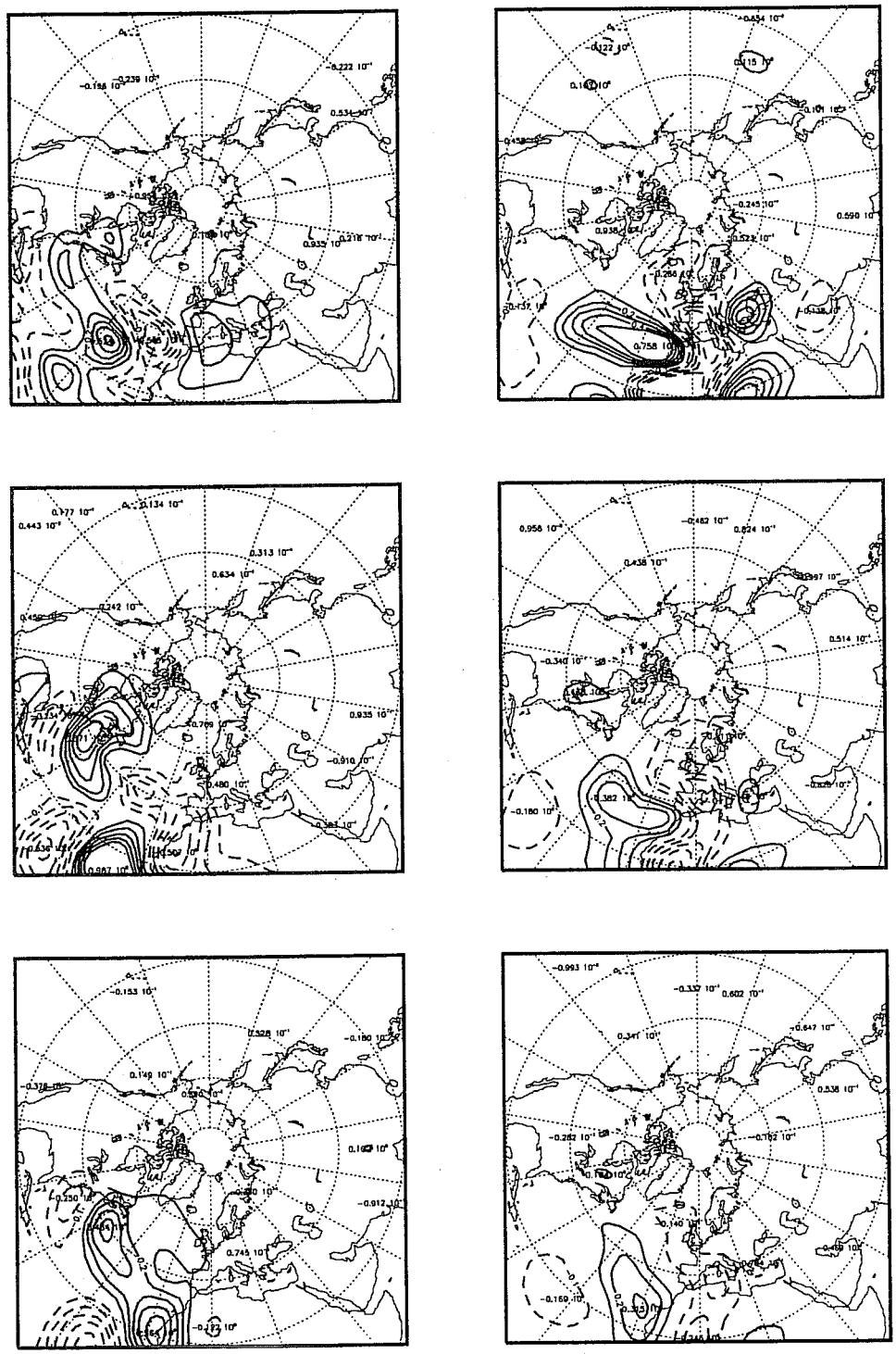

Fig. 4. (c).

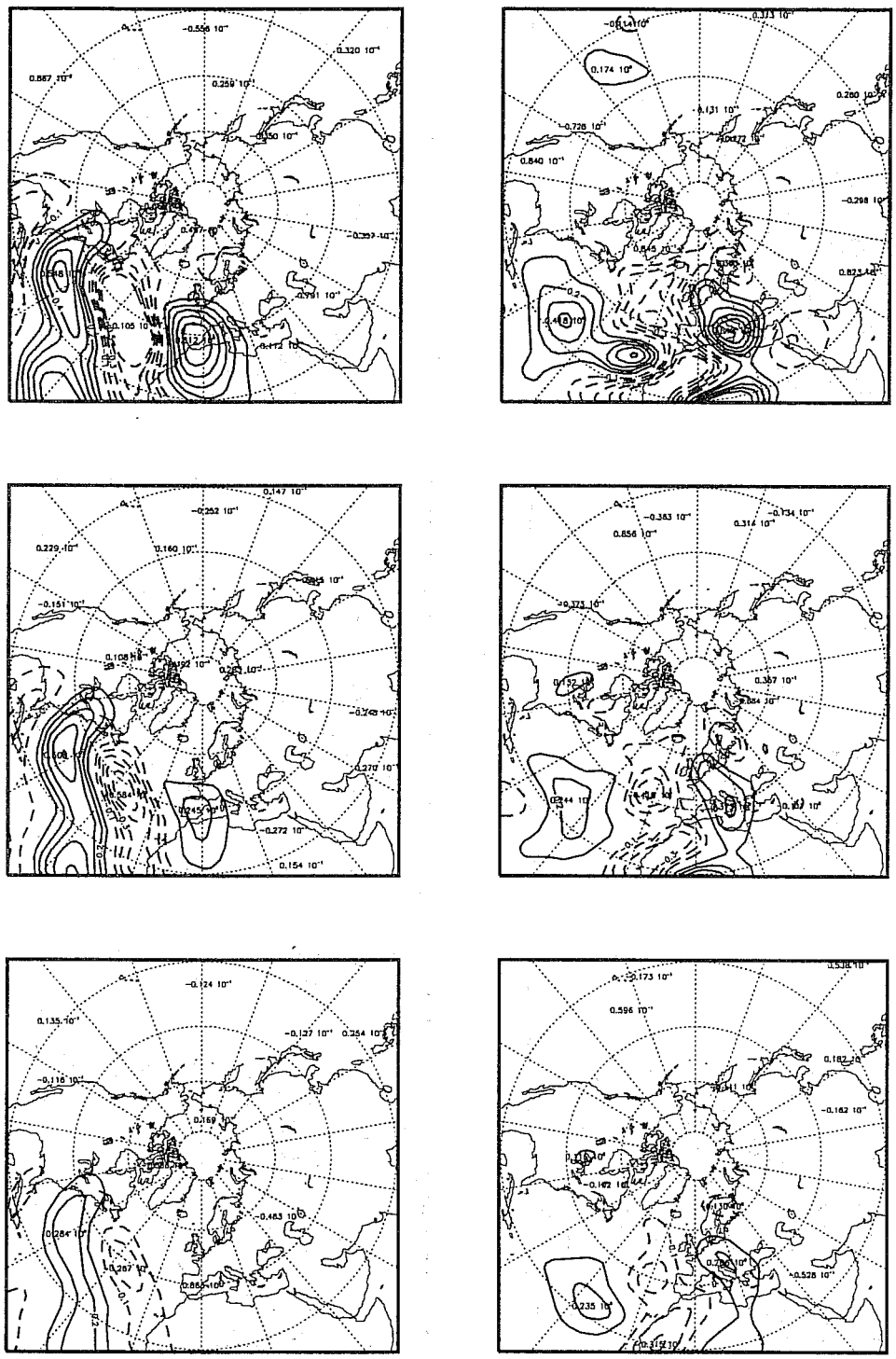

Fig. 5. As Fig. 4, except for using $\mathbf{V}_0$ to define the initial norm.

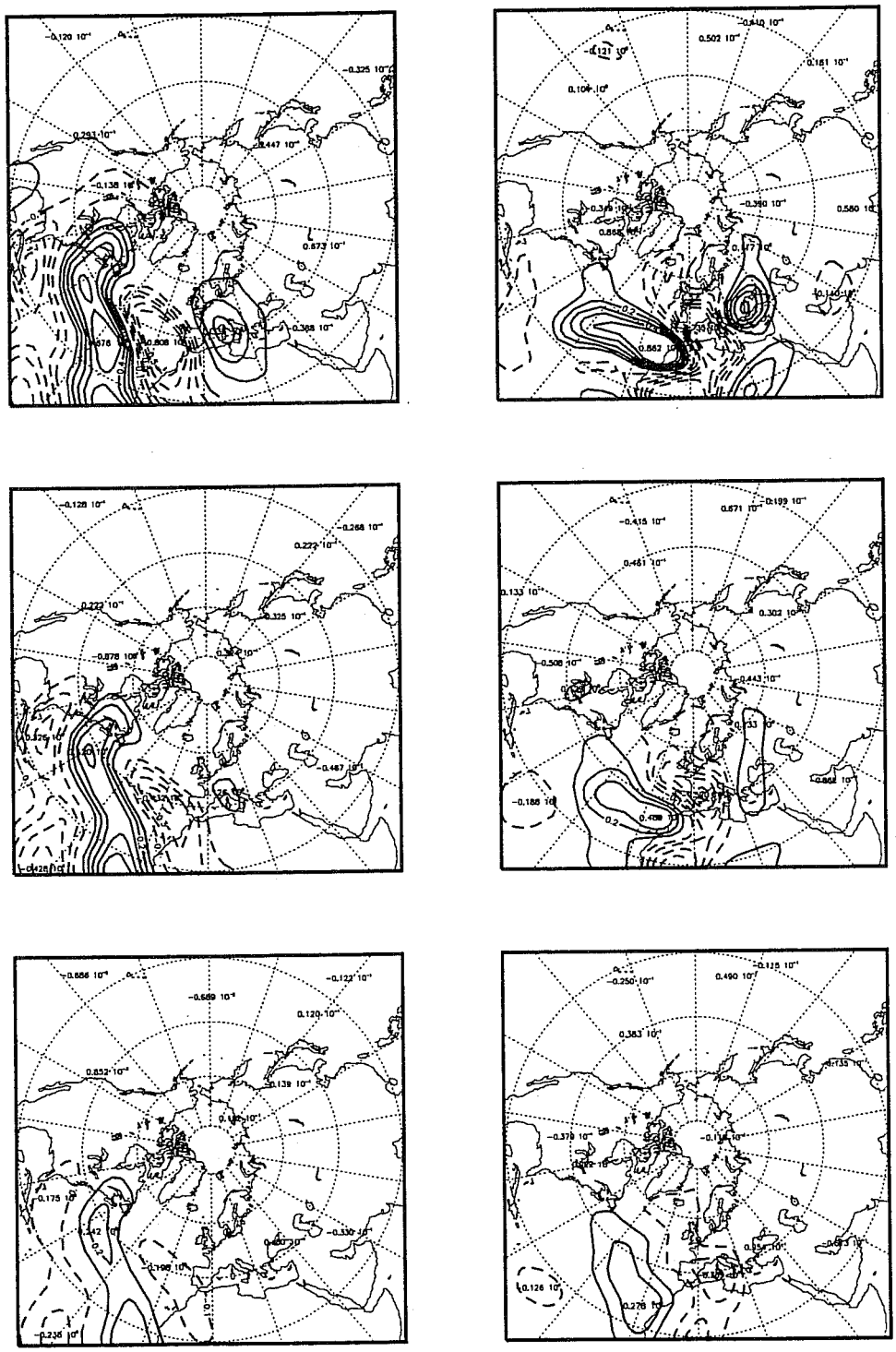

Fig. 5. (b).

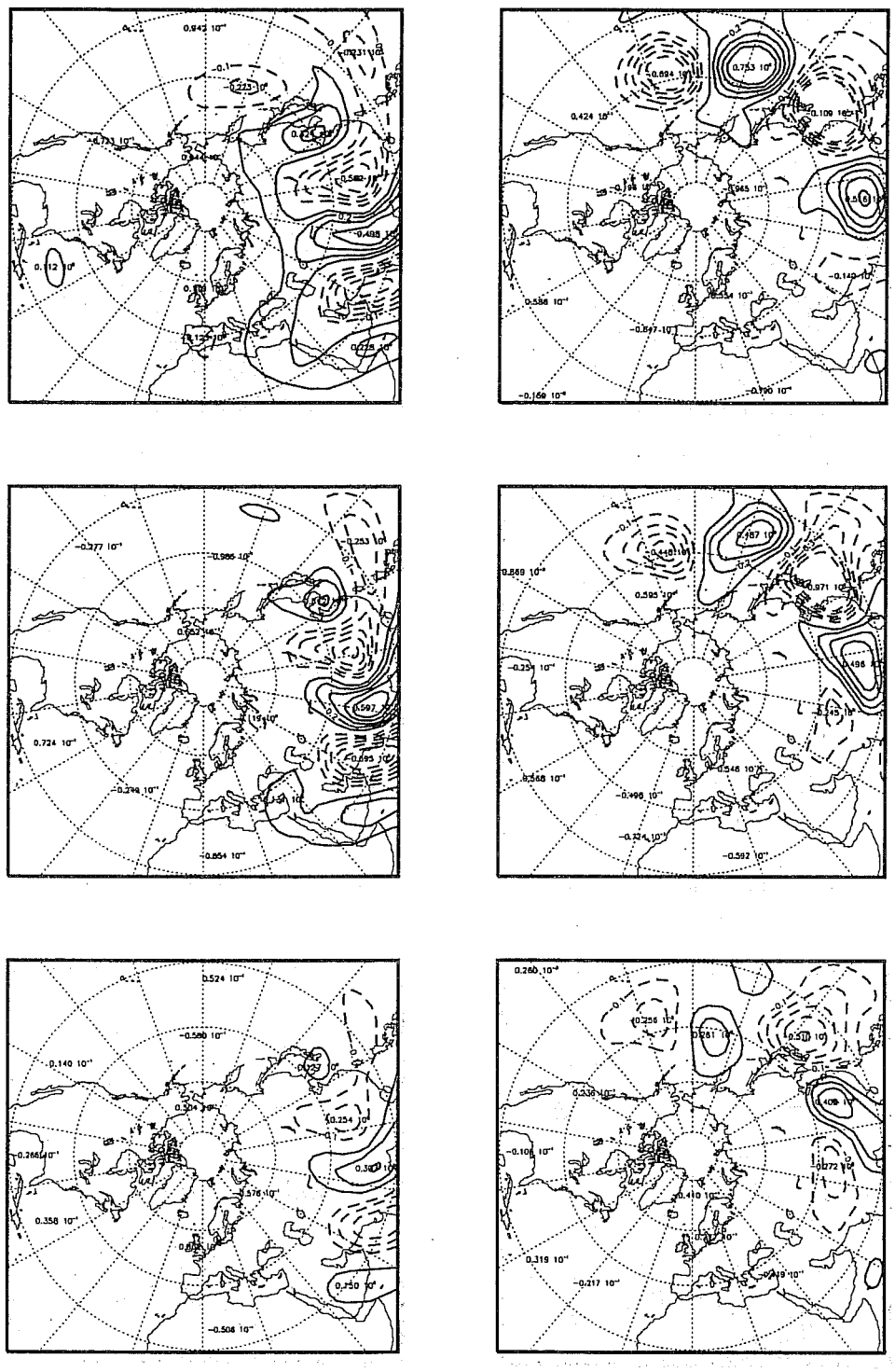

Fig. 5. (c).

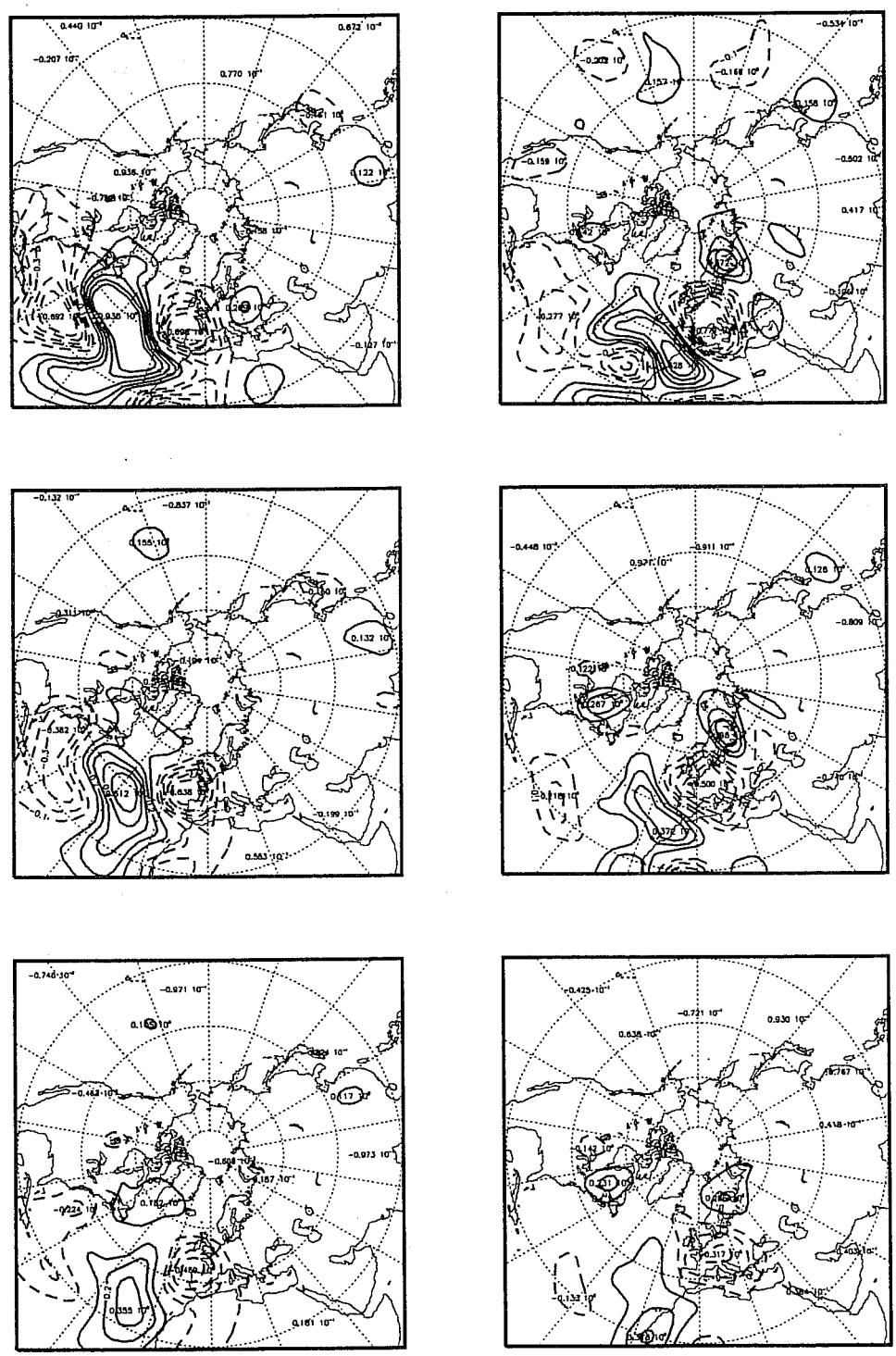

Fig. 6. As Fig. 4, except for using $\mathbf{V}_{48}$ to define the initial norm.

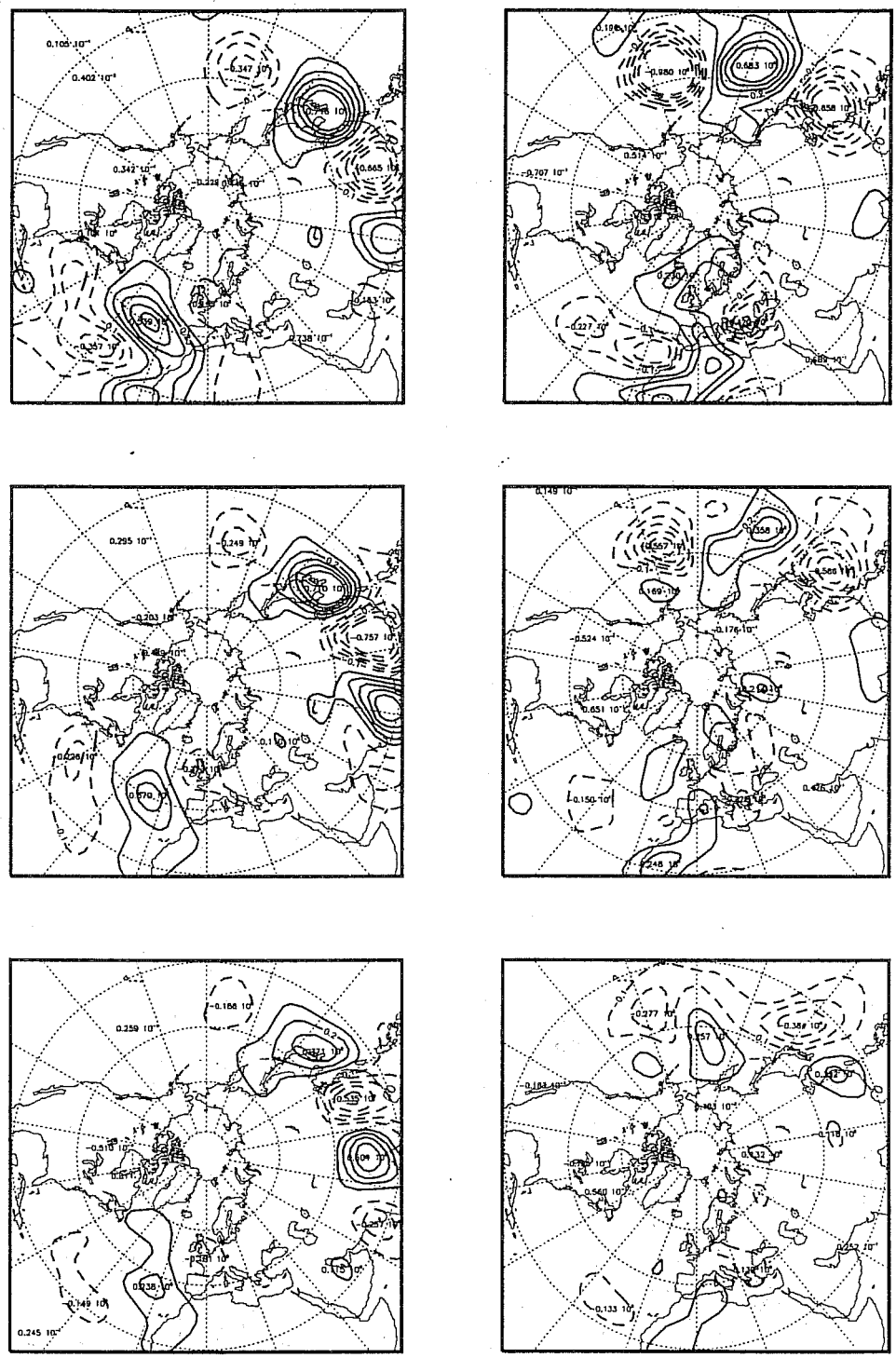

Fig. 6. (b).

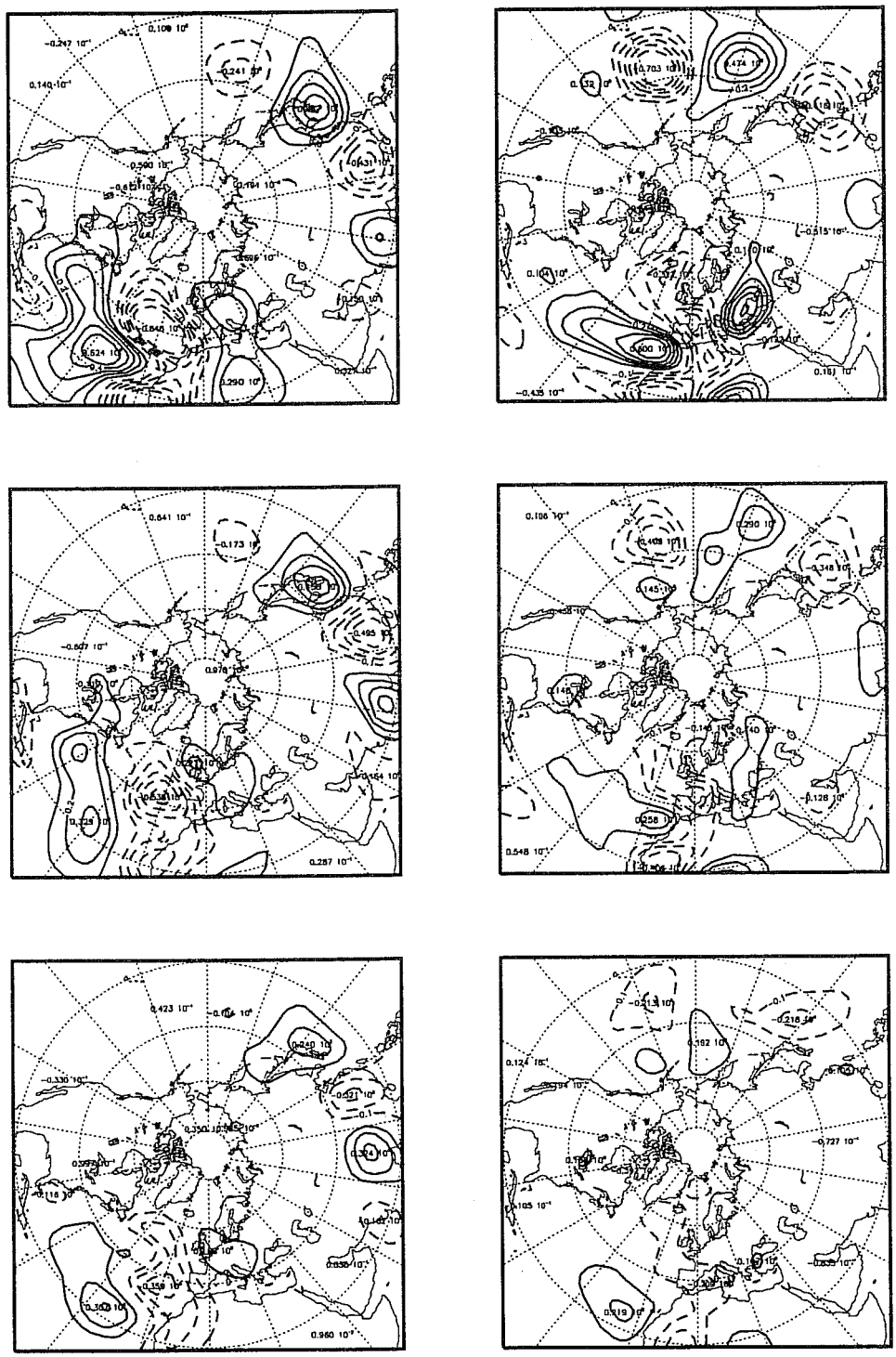

Fig. 6. (c).

In Fig. 7, the variances in physical space implied by $\mathbf{V}_{48}$ valid for case 1 are displayed at the three model levels. The structure that is apparent is a result of the characteristics of the flow. Note that the fields at the three model levels are drawn with the same coutours, and that the contours are logarithmically equally spaced (the factor in going from one contour to the next is $10^{0.15}$). All other subsequent variance maps are plotted with these same contours. The total variance contained in $\mathbf{V}_{48}$ is 8.301×10^{-10} s^{-2}, whereas the total variance contained in $\mathbf{V}_0$ is 3.301×10^{-10} s^{-2}. Note that it can be seen from Fig. 7 that the variance decreases from the top to the bottom levels (the contributions by each level to the total variance are printed in the first line of these and all subsequent variance maps).

Experiment C7hb will now be considered in some more detail. In this case the operator $\mathbf{P}$ is the identity, so that the solution to (5.5) leads directly to the eigenstructure of $\mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$. Due to the fact that the only variable considered here is vorticity, no dimensional problems are occuring, and it is possible to look directly at the eigenstructure of initial and final covariance matrix. In Fig. 8, the spectra $\tilde{\Lambda}$ (eq. (5.17); the initial spectrum, dashed) and Λ (eq. (5.8); the final spectrum, solid) are plotted both in an absolute manner (panel a; plotted are the logarithms of the eigenvalues), and in a cumulative manner (panel b). It is noted that the spectrum becomes steeper during the time evolution. Note that in panel (b) the two curves are relative to two different total variances (namely, 8.301×10^{-10} s^{-2} for the dashed curve, see above; and 43.841×10^{-10} s^{-2} for the solid curve). Fig. 8b shows that about 200 eigenvectors of $\mathbf{V}_{48}$ are needed to account for approximately 80% of the variance contained in $\mathbf{V}_{48}$ (dashed curve), whereas 200 eigenvectors account for more than 90% of $\mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$ (solid curve). It must be noted that this large fraction of variance is partly related to the loss of accuracy of the tangent–linear approximation (see below); this loss of accuracy will become apparent when comparing the value 43.841×10^{-10} s^{-2} with the respective value obtained by nonlinear integrations.

As pointed out in section 5.2, it is, in addition to the spectrum of $\mathbf{V}_{48}$, also of interest to look at the eigenvectors of $\mathbf{V}_{48}$ (taken as the initial covariance in the present experiment C7hb). The first three eigenvectors (i.e., the structures $\tilde{\mathbf{Z}}$ from eq. (5.19)) are shown in Fig. 9: they account for 4.17%, 3.44%, and 3.32% of the total initial variance of 8.301×10^{-10} s^{-2} (note that this information is also contained in the dashed line in Fig. 8b). These eigenvectors have been obtained by solving (5.17) with $\mathbf{P} = \mathbf{I}$. Note that eigenvectors 2 and 3 have similar structures, with *different* signs over the east coast of Asia, but the *same* sign over the Atlantic.

The eigenvectors of $\mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$ (i.e., the evolved SVs of experiment C7hb; i.e., the structures $\mathbf{Z}_t$ defined in (5.6), (5.8)) are shown in Fig. 10, together with the initial SVs (left columns). Note that in this case a comparison of the structures $\mathbf{Z}_t$ (Fig. 10, right

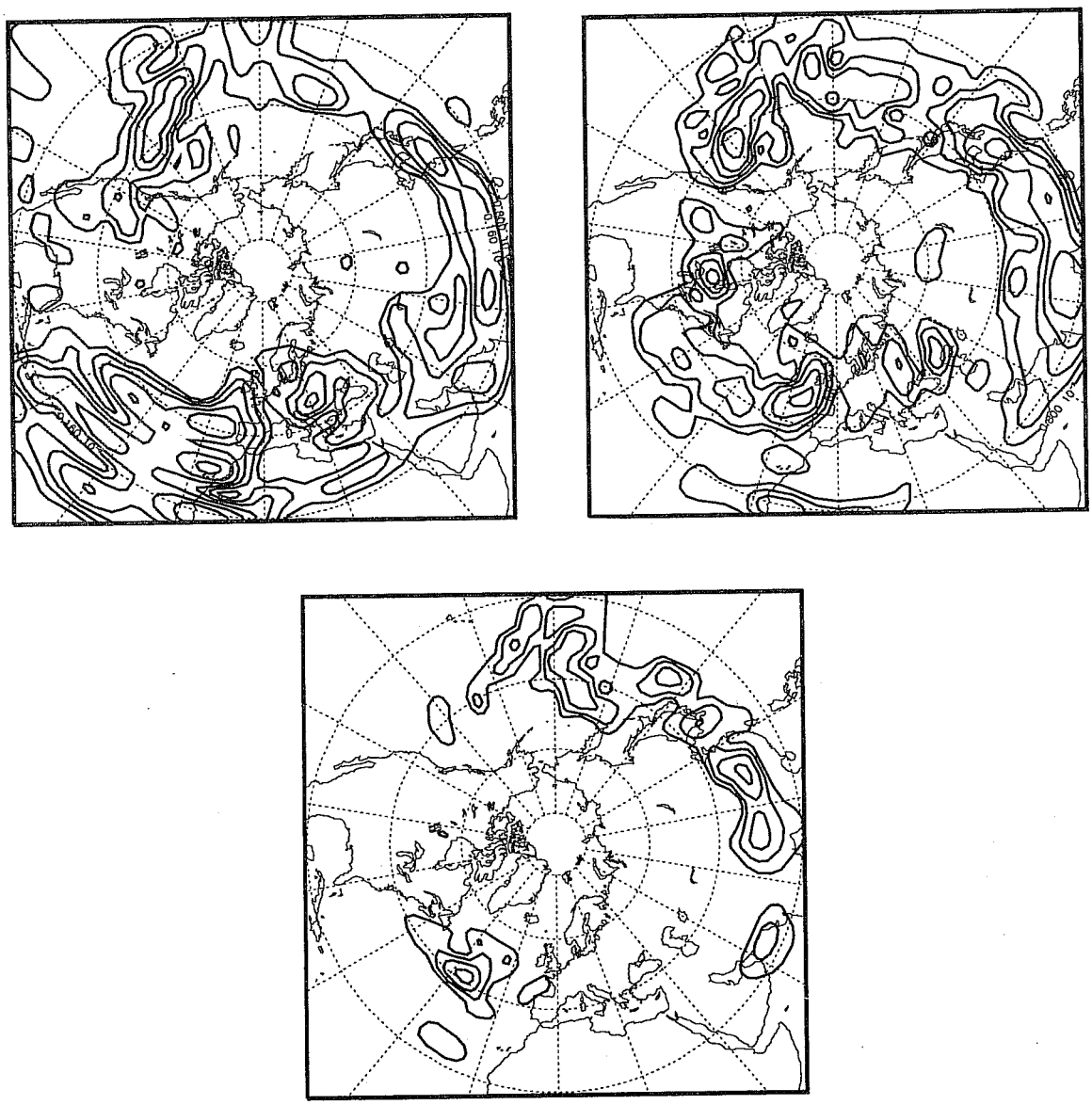

Fig. 7. The variance structure described by $\mathbf{V}_{48}$ at the three model levels. $\mathbf{V}_{48}$ is the result of a 48–hour tangent–linear quasi–geostrophic covariance prediction from an isotropic and homogeneous covariance structure at 15/02/1997/1200 GMT. Note that the same contours are used at the three levels, and that contours are logarithmically equally spaced (the factor between two contours is $10^{0.15}$). Level contributions to total variance are printed in the first line of each subpanel.

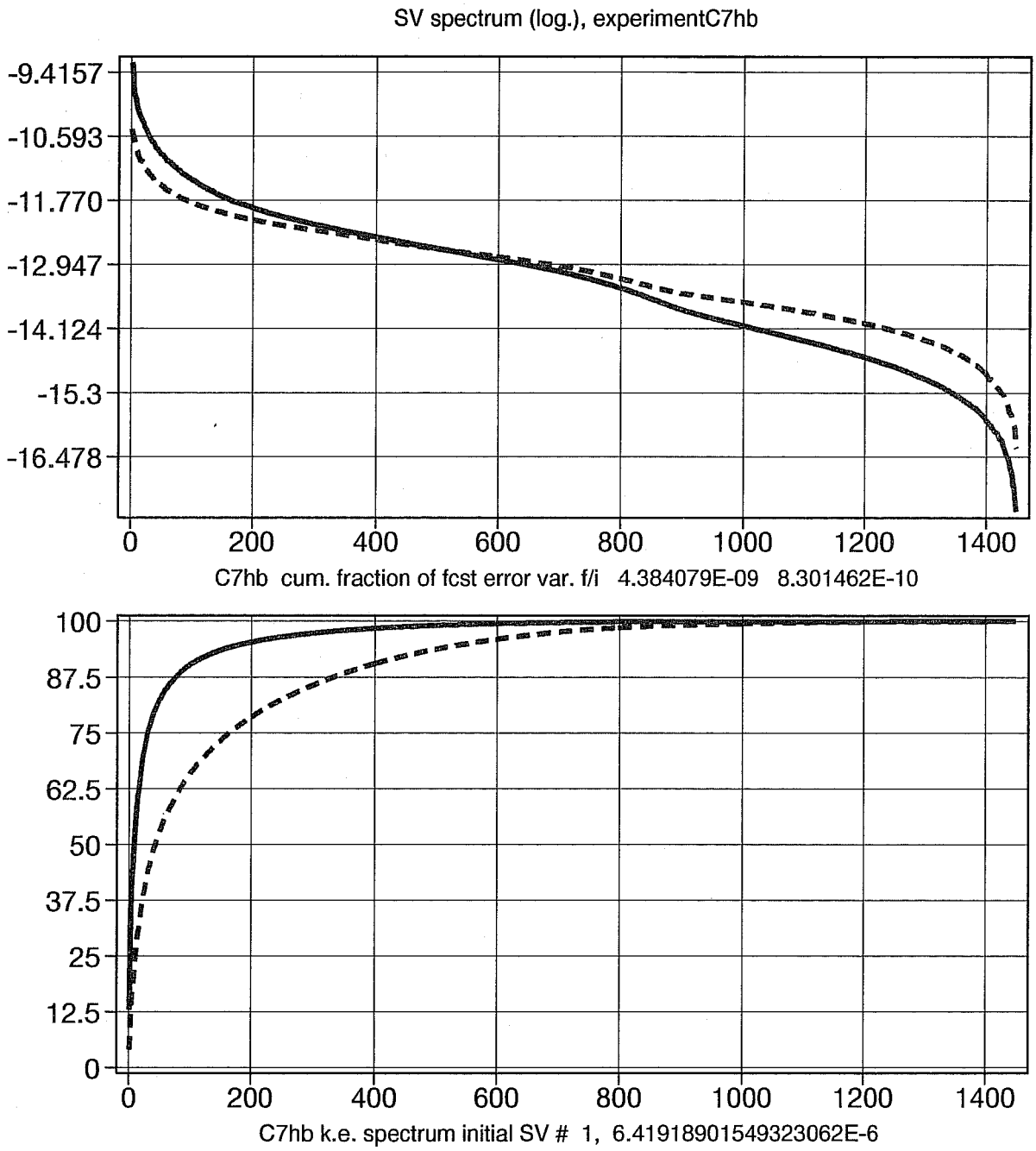

Fig. 8. The eigenspectra of $\mathbf{V}_{48}$ (dashed) and $\mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$ (solid) plotted (a) in terms of their logarithms, and (b) in a cumulative manner. $\mathbf{M}$ denotes the tangent–linear quasi–geostrophic propagator (see eq. (4.14)) valid for the 48–hour interval starting at 17/02/1997/1200 GMT.

columns) with the structures plotted in Fig. 9 yields insight into how much the initial covariance has changed (by assessing how much its eigenstructure has changed). It is noted that the eigenstructures are quite different: especially the second and the third eigenvector of $\mathbf{V}_{48}$ (Fig. 9b, c) are quite different from the second and third time–evolved SV (Fig. 10b, c). The fraction of variance of $\mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$ accounted for by the first three time–evolved SVs is 13.00%, 12.00%, and 9.5%, respectively (this can also be seen from the solid line Fig. 8b).

Whereas the final SVs, shown in Fig. 10 (right columns), may be compared with the eigenvectors of $\mathbf{V}_{48}$ (shown in Fig. 9), the SVs of Fig. 10 (experiment C7hb) may also be compared with the SVs shown in Fig. 6 (experiment C7gb; these two experiments differ only by the choice of the final norm; see Tab. 1). Note that by the plotting convention the final norm is not applied, since the fields $\mathbf{Z}_t^-$ are plotted. Quite interestingly, the first SVs are quite similar both at initial and final times (Fig. 6a and 10a), indicating that the final norm does not seem to matter too much for the fastest–growing structure (similar behavior was also found in a primitive–equation model; J. Barkmeijer, pers. comm.). The subsequent structures (i.e., SVs 2 and 3) are more different, as is necessary since two different optimization problems are solved (i.e., two different final norms).

Given the eigenvectors (i.e., time–evolved SVs) of $\mathbf{S}^{\mathrm{TL}}$ in experiment C7hb (see Fig. 10), it is now clearly possible to reconstruct $\mathbf{S}^{\mathrm{TL}}$, see eq. (5.3), by a given number of SVs, through eq. (5.10). The result is shown in Fig. 11, for (a) 10 SVs, (b) 100 SVs, (c) 1449 SVs, with 54.9%, 90.4%, and 100% of explained variance, respectively (see also solid curve in Fig. 8b). Again, the variance fields are plotted at the three levels, with the same contour intervals as used in Fig. 7. It is apparent that the fields are quite noisy, and that they are bigger by a factor of about five ($\approx 43.841 \times 10^{-10}$ s^{-2}/8.301×10^{-10} s^{-2}) than the initial variances (see Fig. 7). Also, it can be seen that the reconstruction of these variance fields is quite rapid (i.e., the general pattern is already visible with 10 SVs) and that adding more SVs leads to somewhat smoother variance fields.

In order to assess the possibility of predicting $\mathbf{S}^{\mathrm{TL}}$ through the leading time–evolved SVs, the forecast error covariance matrix $\mathbf{S}_t$ as defined in (5.1) was computed through a Monte Carlo experiment (for $\mathbf{P}=\mathbf{I}$, in analogy to experiment C7hb). This computation was carried out by taking the initial pdf as normal (see also section 5.3), with covariance structure $\mathbf{V}_{48}$ and a given mean $\mathbf{x}_0^c$ (this mean is the same as the starting point for the basic state integration used to define the tangent–linear model). The covariance structure was then estimated from a time–evolved (over 48 h = the optimization time interval in the present experiments) random sample of size M from this initial pdf according to eq. (3.10). The results obtained for $M = 5000$ and $M = 10$ are shown in Fig. 12a, b,

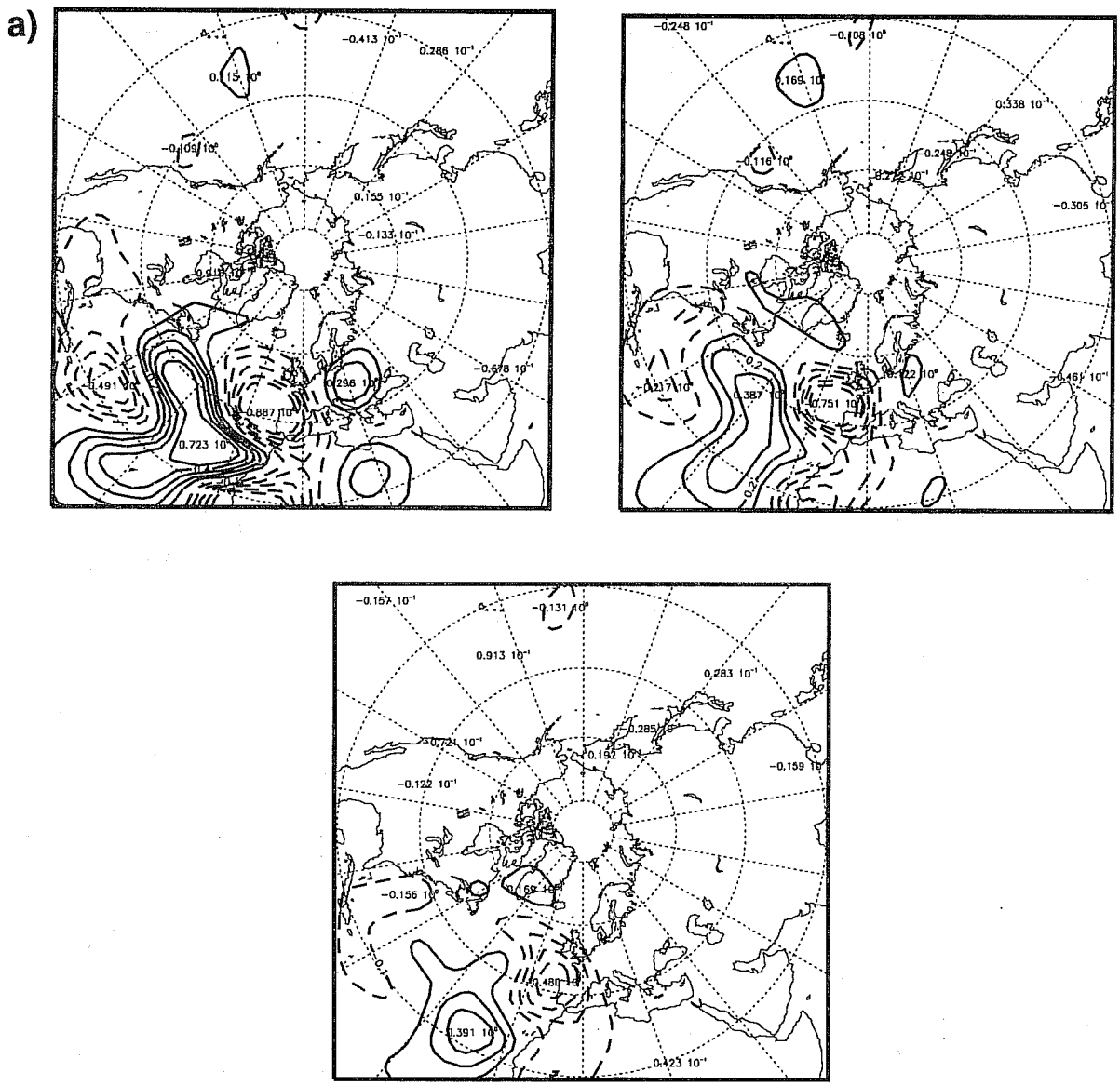

Fig. 9. The three leading eigenvectors of $\mathbf{V}_{48}$ (i.e., the structures $\tilde{\mathbf{Z}}$ in eq. (5.19)). The explained variances are (a) 4.17%, (b) 3.44%, (c) 3.32%. The plotting convention is as in Fig. 4.

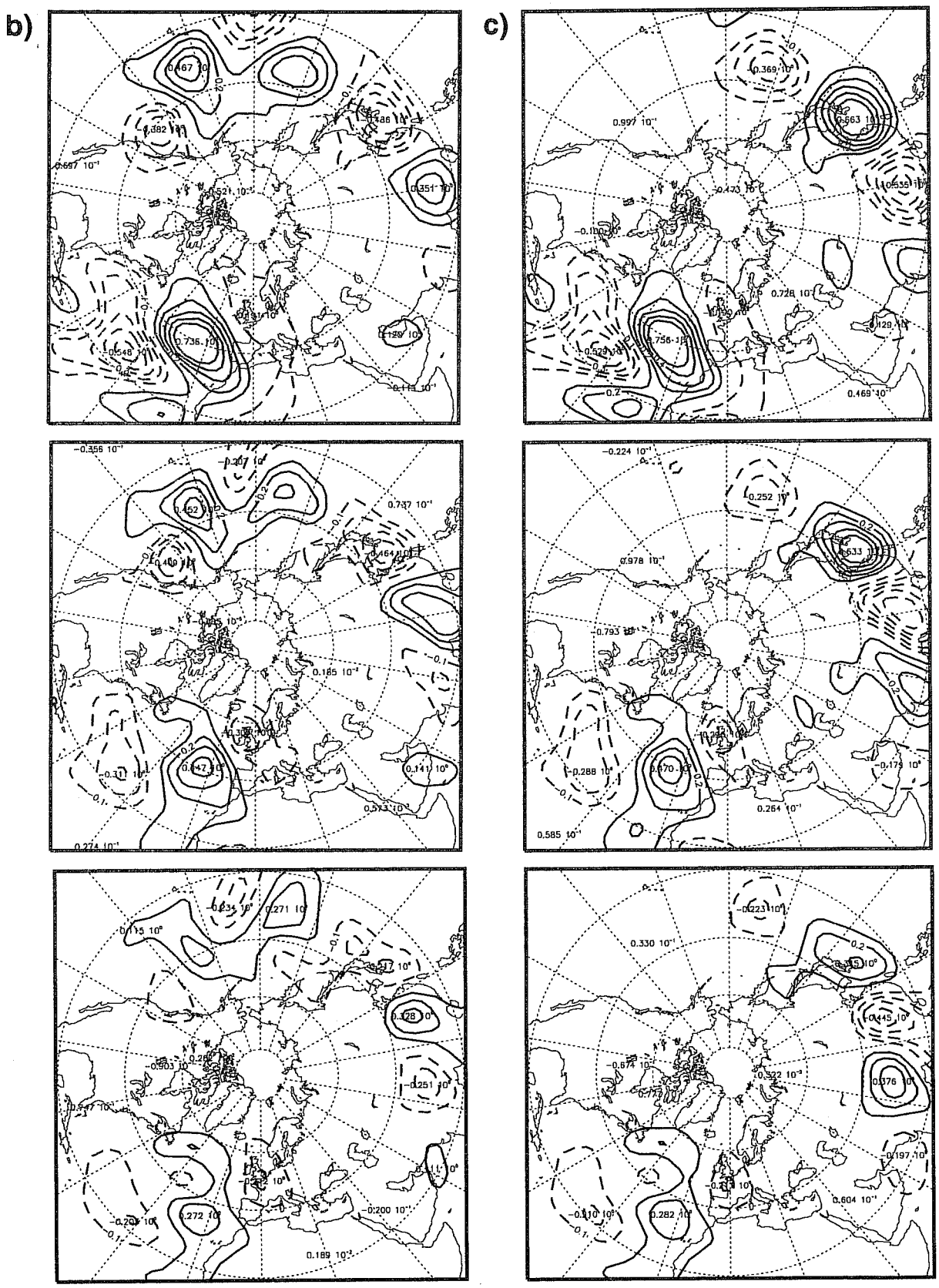

Fig. 9. (b), (c).

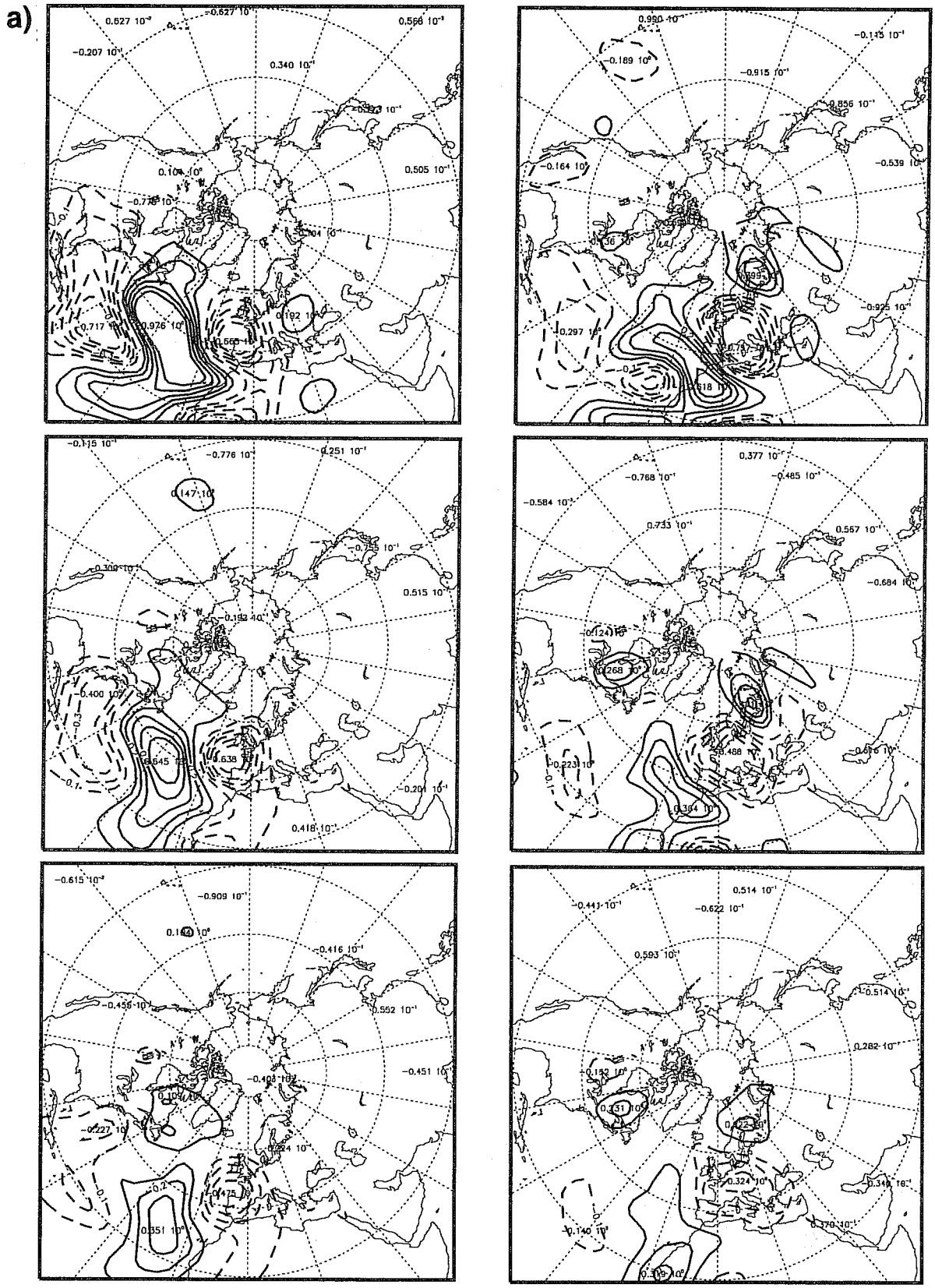

Fig. 10. Initial and final SVs of experiment C7hb. Left columns show $\mathbf{Z}_0$ (eq. (5.4)), right columns show $\mathbf{Z}_t$ (eq. 5.6)). The vectors shown in the right columns are the eigenvectors of $\mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$. Plotting convention as in Fig. 4. Panel (a) SV 1, (b) SV 2, (c) SV 3.

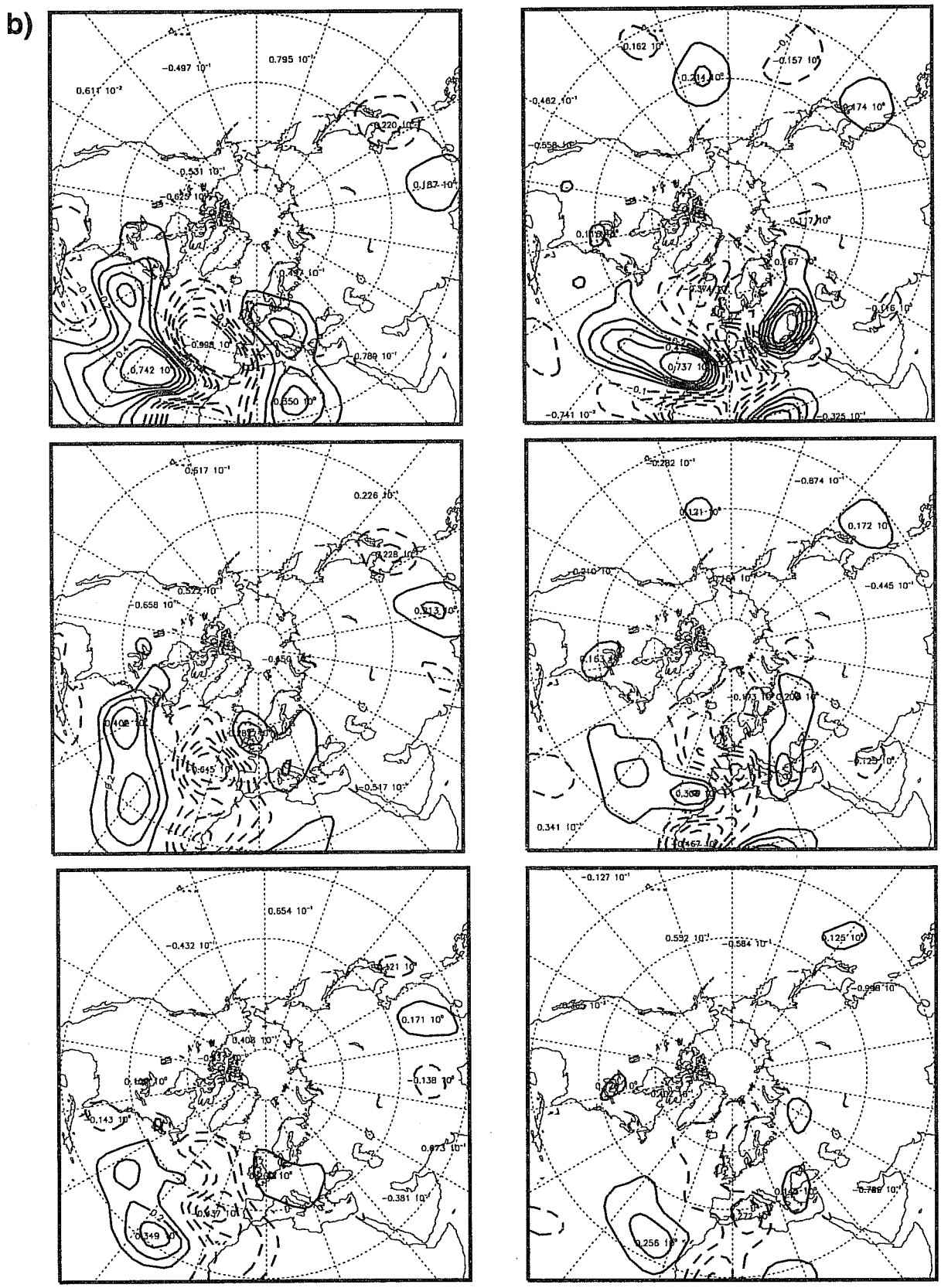

Fig. 10. (b).

c)

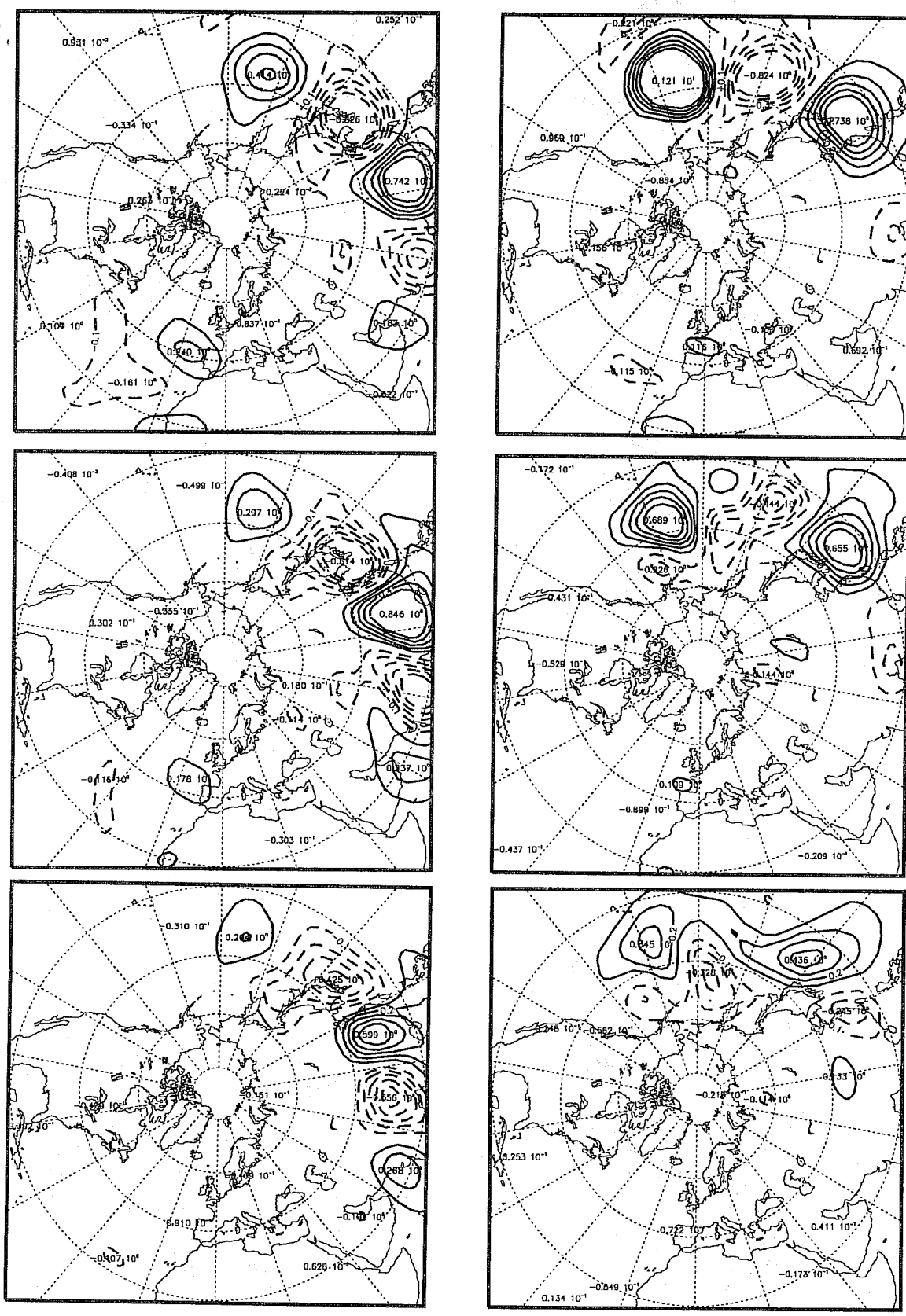

Fig. 10. (c).

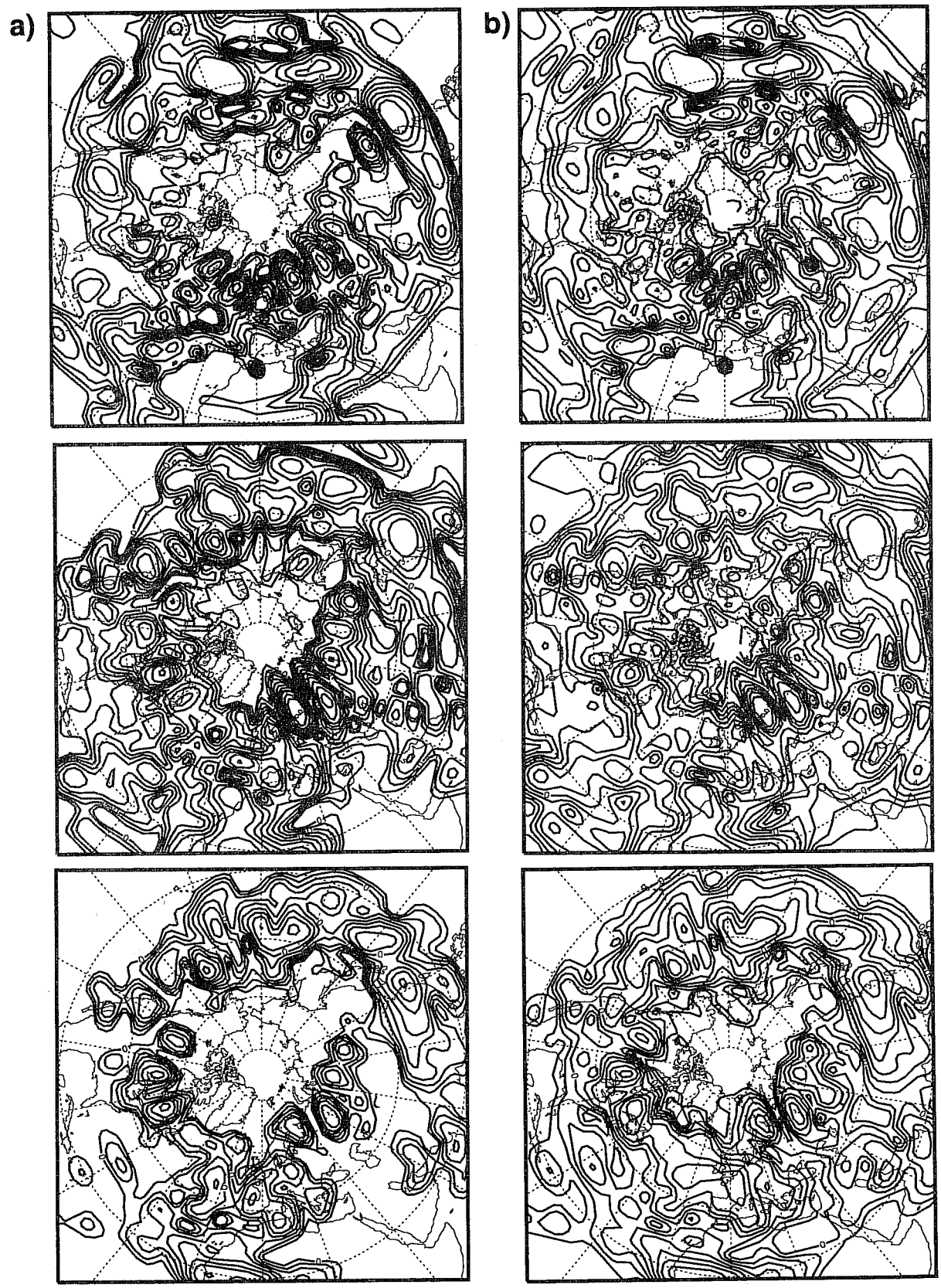

Fig. 11. Reconstruction (i.e., prediction) of $\mathbf{S}^{\mathrm{TL}} = \mathbf{M}\mathbf{V}_{48}\mathbf{M}^{\mathrm{T}}$ through (a) 10, (b) 100, (c) 1449 time evolved SVs, $\mathbf{Z}_t$ (see, eq. (5.10)), shown in terms of variance fields at the three model levels. Plotting convention as in Fig. 7.

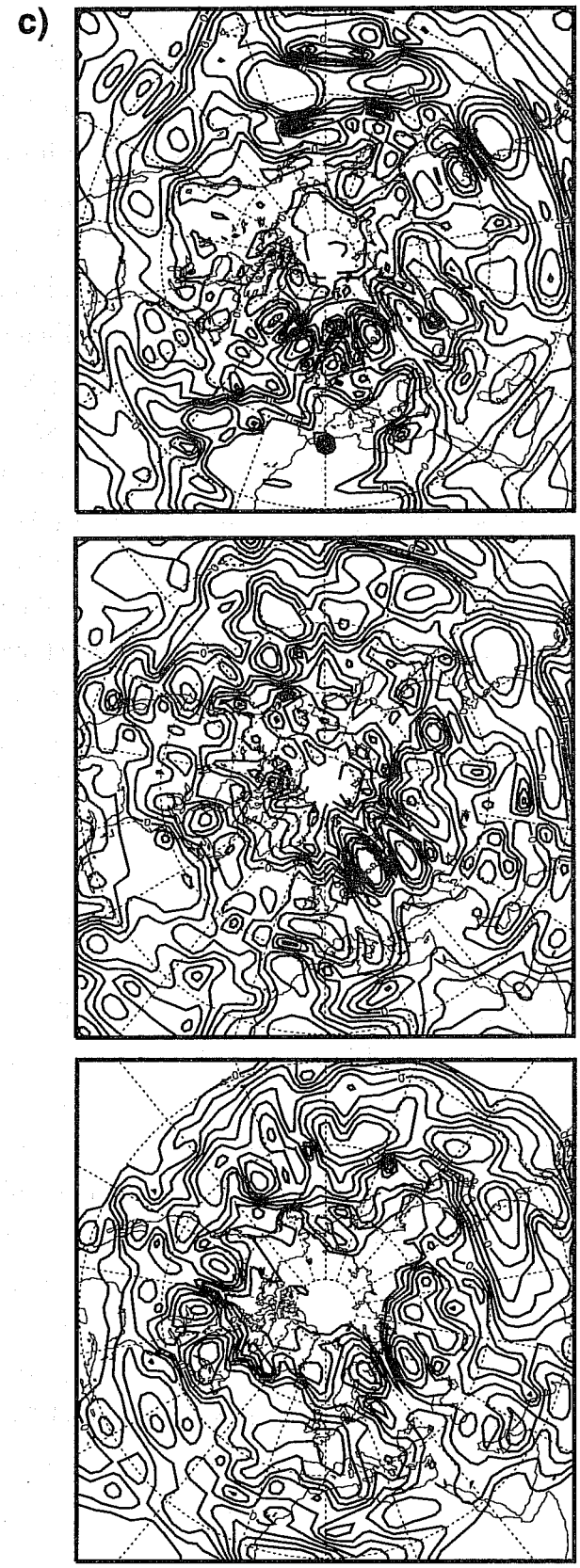

Fig. 11. (c).

respectively (again plotted with the contour interval chosen for the variance plots in Fig. 7 and 11). Note that the $M = 10$–field is quite noisy, but is already picking up some of the structure present in the "true" forecast error covariance field shown in Fig. 12a. Note that the same remark applies also for the reconstruction of $\mathbf{S}^{\mathrm{TL}}$ shown in Fig. 11a (with 10 SVs). It is noted in particular that the nonlinearly estimated total final variance (with $M = 5000$) is 17.956×10^{-10} s^{-2}, indicating a total *nonlinear* variance growth by about a factor of 2 (the initial total variance is 8.301×10^{-10} s^{-2}; see above). Thus, the tangent–linear total variance amplification by a factor of about five is unrealistically large, and must be traced back to a rather inaccurate tangent–linear approximation (given the optimization time considered here as well as typical initial perturbation sizes induced through the consideration of the initial covariance structure $\mathbf{V}_{48}$).

In the context of Fig. 12a it is noted that if in an MC experiment the time integration is carried out over a rather long time (e.g., 8 days), the final covariance structure becomes quite independent from the parameter choices to define the initial pdf (2.2). Covariance structures were computed for $M = 5000$ for an integration time of 8 days, and initial states $\mathbf{x}_0^c$ for both cases 1 and 2. For both cases, the initial covariance structure was prescribed as either $\mathbf{V}_0$ or $\mathbf{V}_{48}$. The final covariance structures (not shown) are all quite similar, which is thought to be a reflection of the forcing and dissipation terms applied in the QG model. In these experiments, the total final variance levels were found to be between 18.141×10^{-10} s^{-2} and $24.785 \times \times 10^{-10}$ s^{-2} (these may be compared with the values reported above for 48 hours).

Before proceeding, the reconstruction of $\mathbf{S}^{\mathrm{TL}}$ by a small number of SVs is shown in Fig. 13 (at 500 hPa), for the situation that the tangent–linear approximation is more accurate (experiment C7db) than in experiment C7hb. Fig. 13a shows the final variance reconstructed with 10 SVs (see eq. (5.10)), Fig. 13b shows a Monte Carlo estimate (obtained as described above) based on $M = 10$, and Fig. 13c shows a Monte Carlo estimate based on $M = 5000$ (the optimization time is again 48 h). In this case, the tangent–linearly estimated final variance is 8.347×10^{-10} s^{-2}, compared to the nonlinearly estimated final variance of 6.888×10^{-10} s^{-2} (for $M = 5000$). Comparison of these numbers shows that the tangent–linear approximation is now more accurate: consequently, prediction through 10 SVs (accounting for 23.79 % of the above tangent–linear final variance; Fig. 13a) is approaching the result shown in Fig. 13c more closely. Note again, that the Monte Carlo estimate for $M = 10$ shows some of the features present in panel Fig. 13c. Figs. 13d, 13e, and 13f show the vorticity covariances of one grid point (located at the 500 hPa model level at approximately zero degrees longitude, and 40 degrees northern latitude) with all other grid points (in the Northern hemisphere at that model level), derived from the predicted

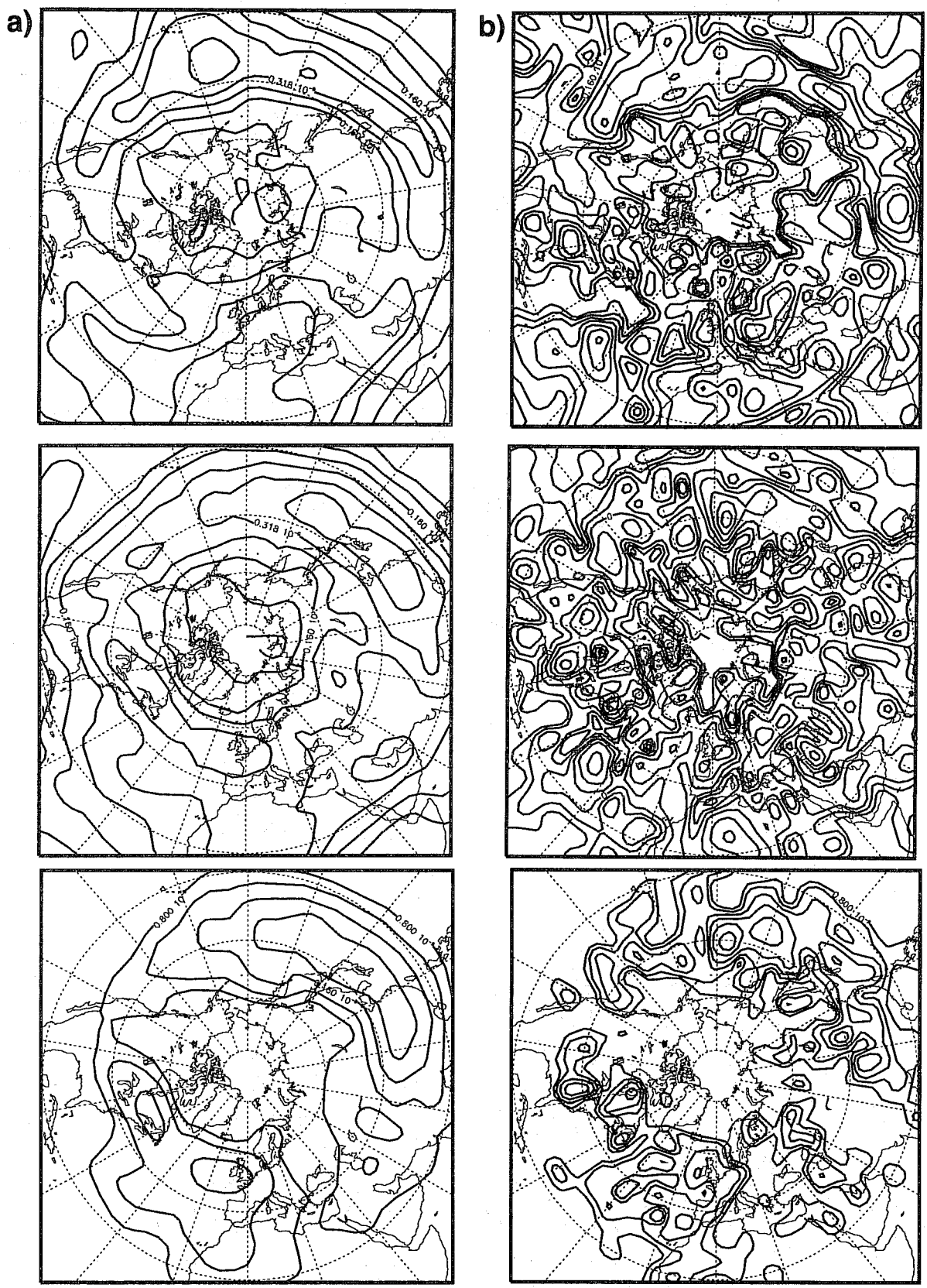

Fig. 12. Nonlinearly predicted forecast error covariance matrix $\mathbf{S}_t$ for experiment C7hb, shown in terms of variance fields at the three model levels. Sample size of the Monte Carlo experiment is (a) $M = 5000$ (left column), (b) $M = 10$ (right column). Plotting convention as in Fig. 7.

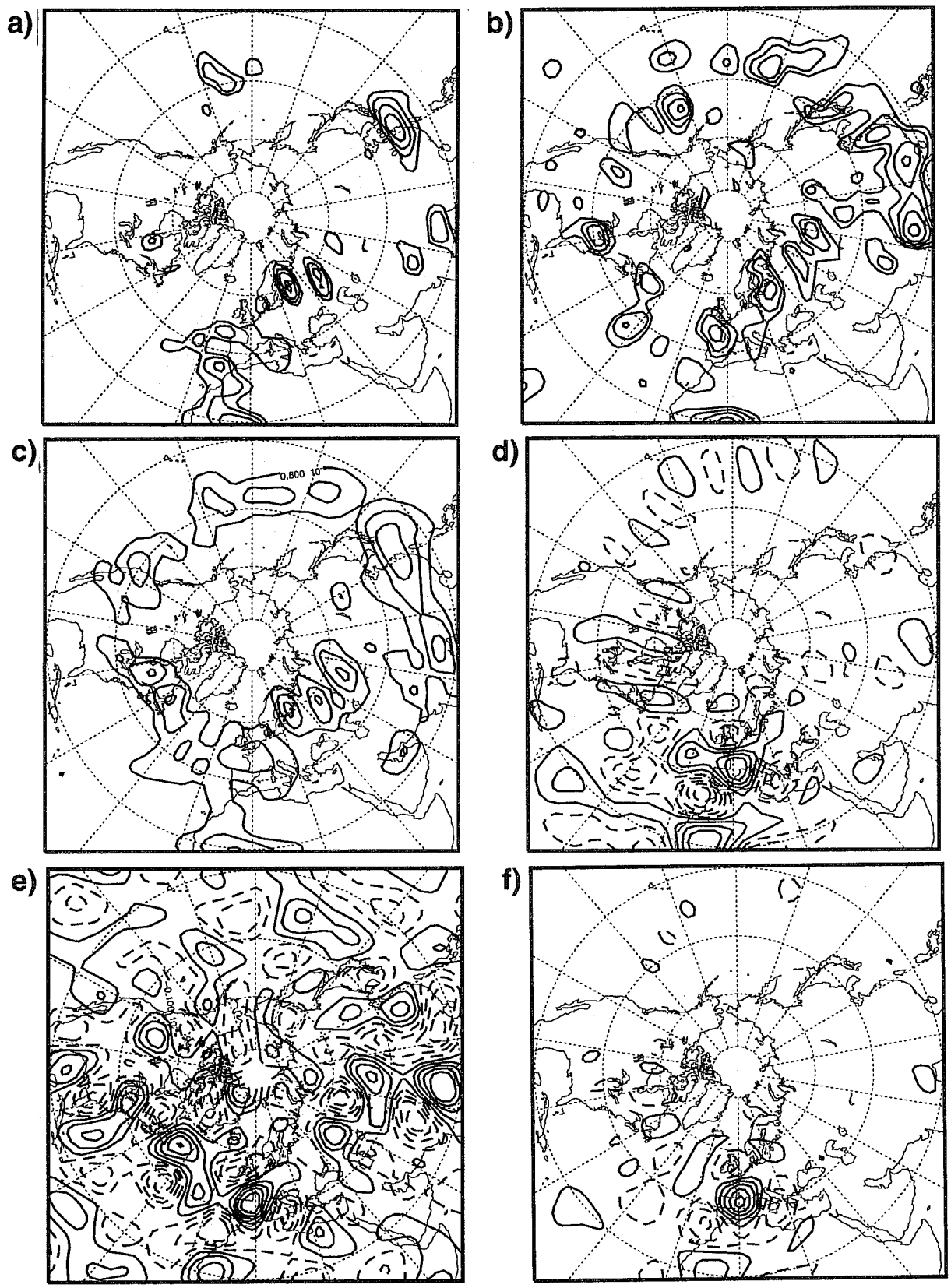

Fig. 13. Variance fields at the 500 hPa model level predicted through (a) 10 SVs, (b) an $M = 10$ MC experiment, and (c) an $M = 5000$ MC experiment, for experiment C7db. Panels (d), (e), and (f) correspond to (a), (b), and (c), respectively, but show the vorticity covariances of one model grid point (at approximately zero degrees longitude and 40 degrees northern latitude at the 500 hPa level) with all other grid points at the 500 hPa model level.

– through either experiment C7db (Fig. 13d), or a Monte Carlo experiment with $M = 10$ (Fig. 13e), $M = 5000$ (Fig. 13f) – covariance structure. Here, it can be seen that the covariance reconstruction through 10 SVs approaches the $M = 5000$ result quite closely (as opposed to the $M = 10$ Monte Carlo result).

Returning back to experiment C7hb, Figs. 14 and 15 give information about the eigenstructure of the (nonlinearly obtained, $M = 5000$) forecast error covariance matrix $\mathbf{S}_t$. Clearly, this covariance matrix $\mathbf{S}_t$ may studied in a totally analogous manner to, for example, $\mathbf{V}$, for example, along eq. (5.17) (recall that $\mathbf{P} = \mathbf{I}$ here). Consequently, Fig. 14 shows the spectrum of $\mathbf{S}_t$, and Fig. 15 shows its first three eigenvectors (note that logarithms of the eigenvalues are plotted in Fig. 14a). This eigenstructure will be referred to as the nonlinear eigenstructure. As a first step, the nonlinear (Fig. 14) and the tangent–linear spectra (solid curves in Fig. 8) may be compared. Note again, that the curve in Fig. 14b is drawn relative to the nonlinear total variance of 17.956×10^{-10} s^{-2} (see above). Comparison of the solid curves shows the degree of nonlinearity. Clearly, it takes somewhere around 600 nonlinear eigenvectors to recover 95% of variance in the nonlinear case, whereas the tangent–linear result indicates that only 200 time–evolved SVs are needed. For example, the nonlinear variance accounted for by the leading three eigenvectors of $\mathbf{S}_t$ is 1.63%, 1.37%, and 1.18%, respectively (compare these with the numbers given above for $\mathbf{S}_t^{\mathrm{TL}}$ in the context of discussing Fig. 10). This comparison illustrates to what degree the tangent–linear results may be misleading when compared to the fully nonlinear situation. One may also directly compare the eigenvalues in the present situation: the dashed curve in Fig. 8a (also referring to logarithms), and the solid curve in Fig. 14: the steepness of the nonlinear spectrum is quite similar to the steepness of the initial spectrum. The same comparison may be done for the eigenvectors: comparing the nonlinear eigenvectors (Fig. 15) with Fig. 9 (the eigenvectors of the initial covariance structure) shows how the covariance structure changes nonlinearly over time, in terms of its eigenstructure. Similarly, comparing Fig. 15 with the time–evolved SVs in Fig. 10 shows the degree of nonlinearity (in terms of the eigenstructure of the two covariance matrices) which is quite substantial in the present situation.

Clearly, Fig. 14 and Fig. 15 show the best that can be done in reconstructing $\mathbf{S}_t$ through its eigenstructure. A question that is then occuring naturally is: *is there a way to obtain the structures at initial time that are evolving into these eigenvectors of* $\mathbf{S}_t$*?* In the tangent–linear context, this question is answered by the structures $\mathbf{Z}_0$ obtained by solving (5.5). In the nonlinear context obtaining the answer to this question may be extremely difficult; this issue is discussed some more in section 6.

This section is closed by mentioning some preliminary results obtained with the SV–

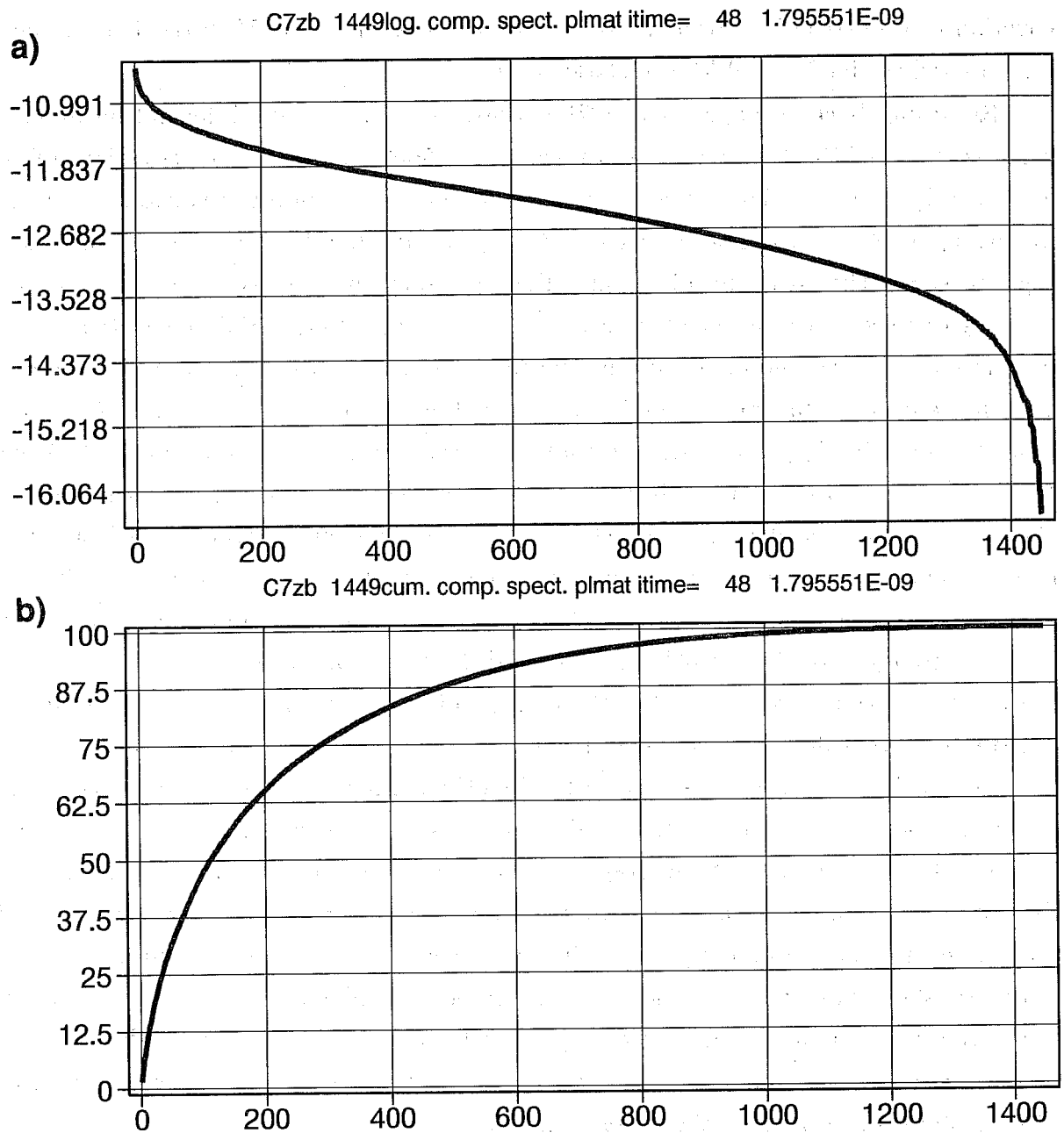

Fig. 14. The eigenspectrum of the nonlinearly predicted forecast error covariance matrix of Fig. 12a ($M = 5000$), (a) in terms of logarithms of eigenvalues, and (b) in a cumulative manner.

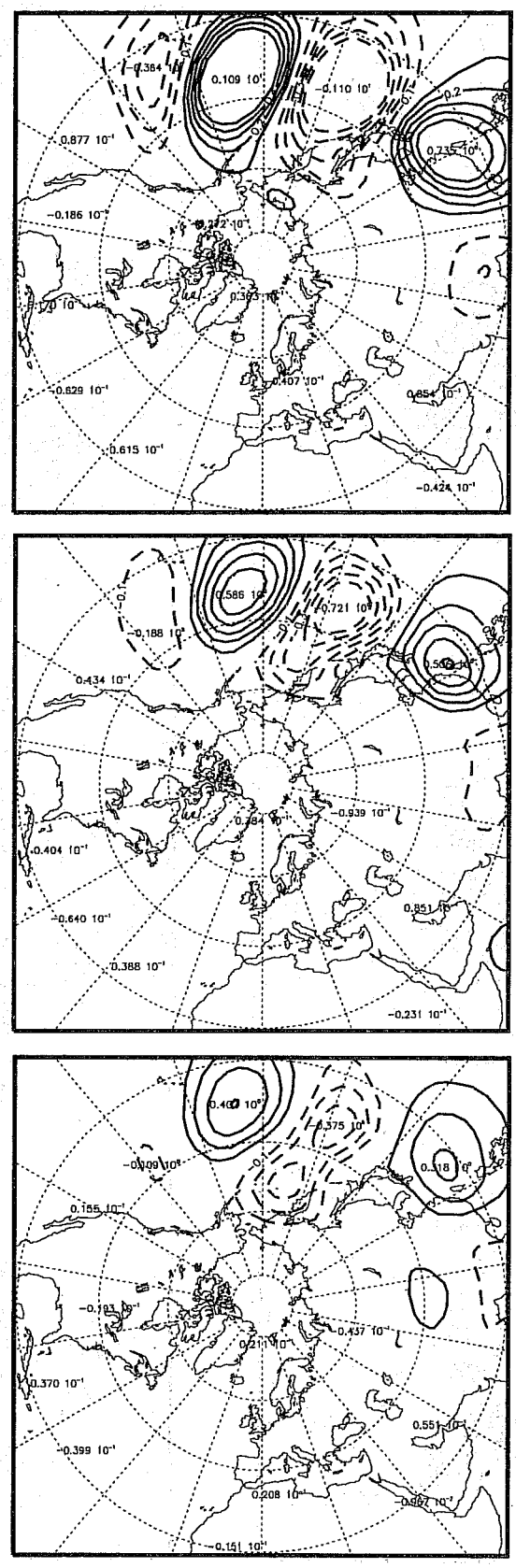

Fig. 15. The three leading eigenvectors of the nonlinearly predicted forecast error covariance matrix of Fig. 12a ($M = 5000$). The explained variances are (a) 1.63%, (b) 1.37%, (c) 1.18%. Plotting convention as in Fig. 4.

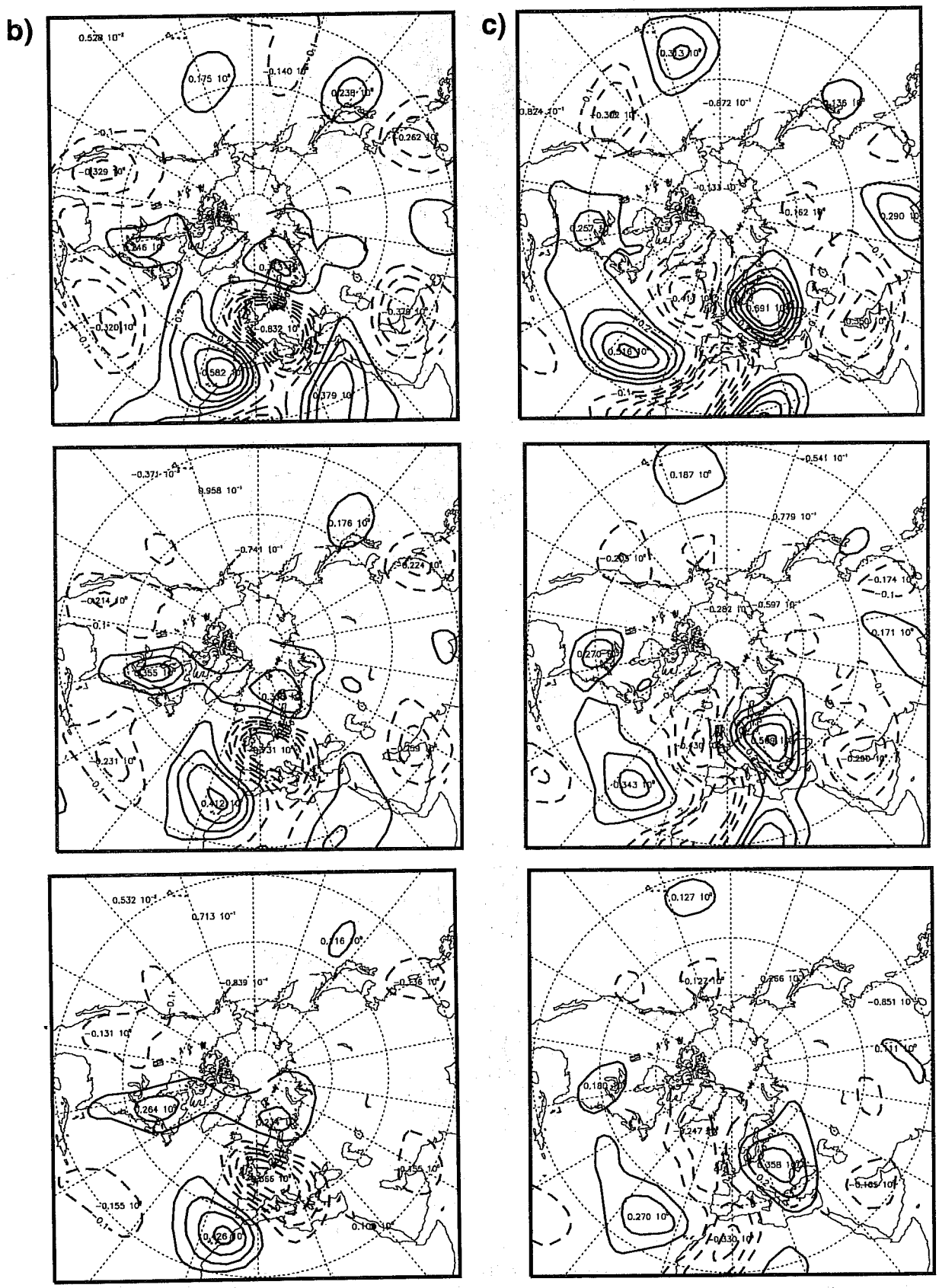

Fig. 15. (b), (c).

based Monte Carlo technique described in section 5.3, for experiment C7hb. In eq. (5.35), the choices $k = 10$, $k = 100$ and $k = 500$ are made. As explained in section 5.3, it is immediately possible to assess the fraction of variance accounted for by any leading subset of SVs $\mathbf{Z}_0^{(k)}$ as a function of k. The relevant cumulative picture is shown in Fig. 16. In the notation of section 5.3, this figure shows α^{-1} as a function of k in a cumulative sense. Note that the (tangent–linearly) time–evolved counterpart of this figure is the solid curve in Fig. 8b. Note in particular that – as pointed out in connection with eq. (5.25) – this cumulative variance picture is quite different from the dashed curve in Fig. 8b (representing the variance reconstruction of the initial covariance through its *eigenstructure*); these curves are directly comparable, and are also normalized to the same value, namely the total initial variance (i.e., 8.301×10^{-10} s^{-2}; see also eqs. (5.20), as well as (5.38), and (5.39)). In the present case, the fractions for the above k values in Fig. 16 are 19.53 %, 57.76 %, and 87.94 % for the $\mathbf{Z}_0^{(k)}$ reconstruction (5.35) of $\mathbf{V}_{48}$. The respective numbers for the *eigendecomposition* of $\mathbf{V}_{48}$ (dashed curve in Fig. 8b) are 25.16 %, 65.92 %, and 93.77 %, showing – by necessity – the superiority of the *eigendecomposition*. Recall, however, that the decomposition (5.25), or (5.35) *evolves* into an eigenstructure (in contrast to the decomposition (5.19b); see also, 5.2.d(i)).

Before proceeding to the discussion of the results obtained with the SV–based Monte Carlo technique, the picture corresponding to the solid curve in Fig. 8b is shown for experiment C7gb in Fig. 17. Fig. 17 shows the cumulative variance accounted for by the structures $\mathbf{Z}_t^-$ which are in experiment C7gb different from $\mathbf{Z}_t$ (see eqs. (5.6) and (5.15)), because the optimization is done with kinetic energy as the final norm in C7gb. Note again, that both curves are normalized with the same value of 43.841×10^{-10} s^{-2}, as the full sets $\mathbf{Z}_t$ of C7hb and $\mathbf{Z}_t^-$ of C7gb are both possible and complete representations of $\mathbf{M}_t\mathbf{V}_{48}\mathbf{M}_t^{\mathrm{T}}$. Note, however, that the curve in Fig. 8b (C7hb) is (slightly) steeper than the curve in Fig. 17, because here the set $\mathbf{Z}_t$ is directly the set of eigenvectors of $\mathbf{M}_t\mathbf{V}_{48}\mathbf{M}_t^{\mathrm{T}}$ (and therefore possessing the optimality property). The set $\mathbf{Z}_t^-$ of C7gb is derived from the eigenvectors of $\mathbf{P}\mathbf{M}_t\mathbf{V}_{48}\mathbf{M}_t^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}$ (with $\mathbf{P}$ standing for kinetic energy) and therefore cannot possess the optimality property for $\mathbf{M}_t\mathbf{V}_{48}\mathbf{M}_t^{\mathrm{T}}$. This is reflected by the fact that the curve in Fig. 8b lies always (slightly) above the curve in Fig. 17. For example, at $k = 10$, the numbers are 54.60741 (C7gb), 54.93016 (C7hb), for $k = 100$ they are 89.95155 (C7gb), 90.35872 (C7hb), and for $k = 500$ they are 98.89042 (C7gb), 99.08510 (C7hb). Note that this example illustrates a general result, in that an optimization with the final norm taken identity (under the caveat of the dimensional problems mentioned in section 5.2) must *always* lead to a "faster" reconstruction of $\mathbf{M}_t\mathbf{V}\mathbf{M}_t^{\mathrm{T}}$ than any other final norm choice.

In Fig. 18, the initial variance fields $\mathbf{V}_{48}^{(k)}$, relevant for experiment C7hb, see eq.

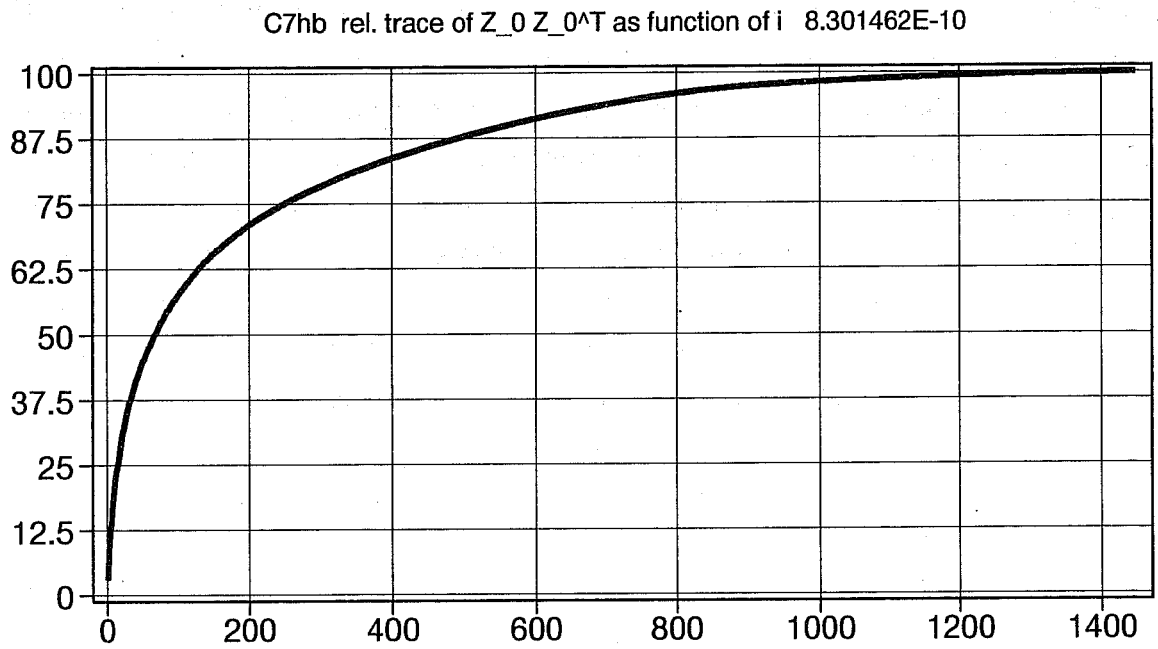

Fig. 16. Fraction of total initial variance explained through k initial SVs (see eq. (5.35)), in a cumulative manner, for experiment C7hb. The figure shows α^{-1}, defined in eq. (5.40), as a function of k.

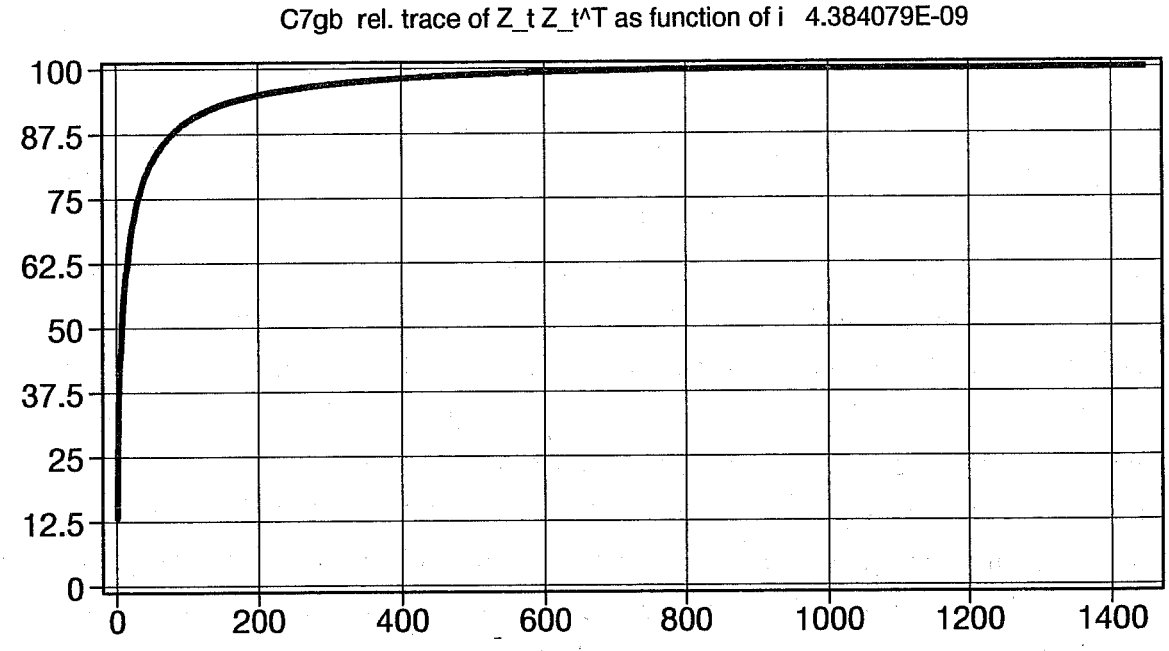

Fig. 17. Fraction of total final variance explained through k time evolved SVs $\mathbf{Z}_t^-$ (see eq. (5.15)), in a cumulative manner, for experiment C7gb.

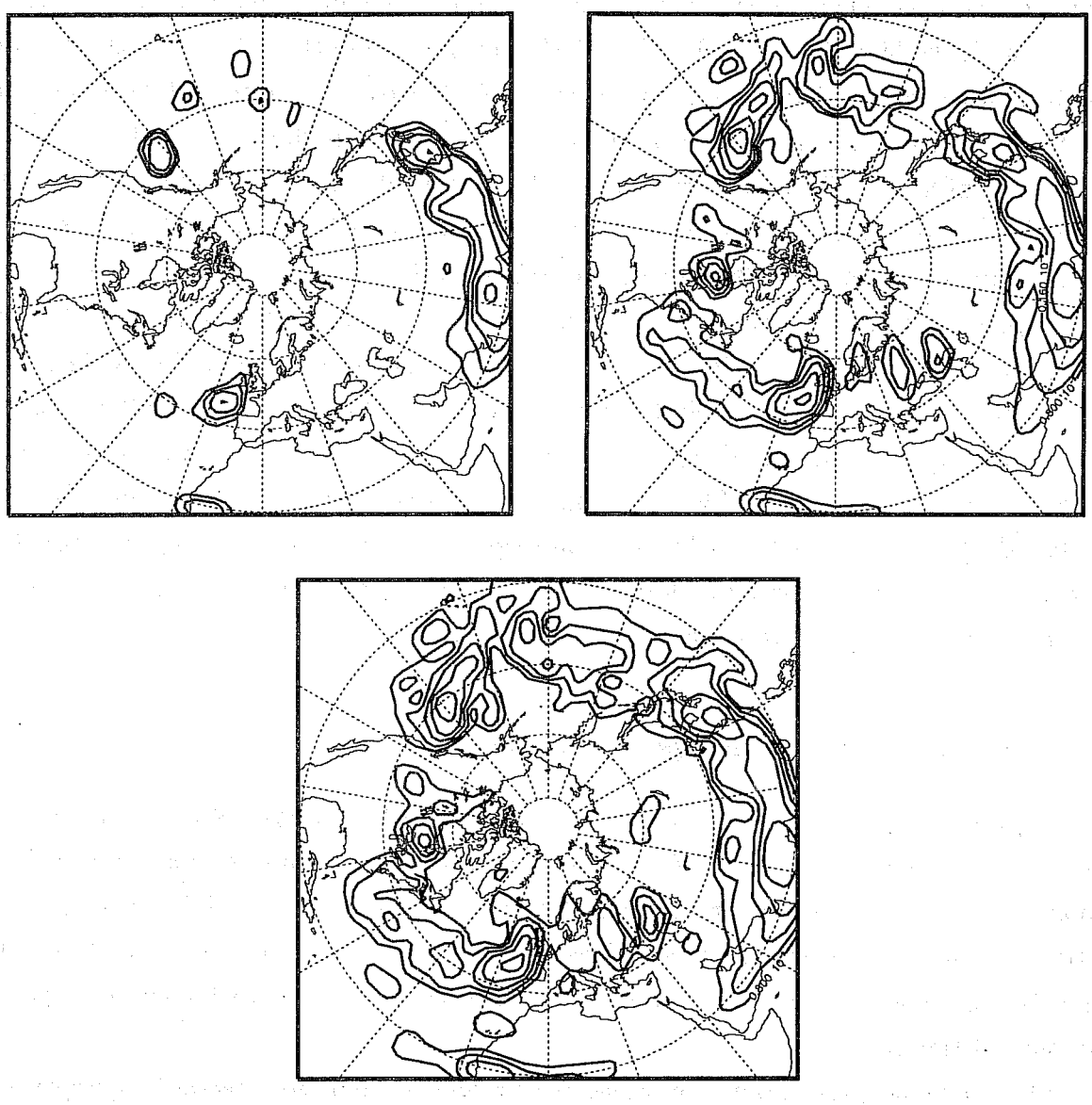

Fig. 18. Initial variance fields, $\mathbf{V}_{48}^{(k)}$ (see eq. (5.35)), for experiment C7hb, at the 500 hPa level, as a function of the number of initial SVs $\mathbf{Z}_0$ retained to represent the initial variance fields: (a) $k = 10$ (19.53%), (b) $k = 100$ (57.76%), (c) $k = 500$ (87.94%). Numbers in parentheses give cumulative fraction of initial variance (see also Fig. 16). Plotting convention as in Fig. 7.

(5.35), are shown for the above three k–values at the middle level of the model (500 hPa). These initial variances are for the scaling parameter $s = 0$ (see eq. (5.44)). These are the variance fields (with their associated covariances; not shown) that the SV–based MC technique "sees" when sampling from the initial pdf according to eq. (5.43), (5.44). The degree of "completeness" of these truncated variance fields may be assessed by comparing Fig. 18 with the 500 hPa – panel in Fig. 7. For each of these three k–values the SV–based MC technique was carried out for sample sizes of $M = 10$, $M = 100$, and $M = 5000$, with $s = 0$. The results are shown in Fig. 19 in terms of the (nonlinearly) predicted variance at the 500 hPa level (in each row of Fig. 19, k is fixed; in each column of Fig. 19, M is fixed). All the panels in Fig. 19 should be compared with the 500 hPa picture in Fig. 12a (obtained with $M = 5000$ and the complete initial covariance structure $\mathbf{V}_{48}^{(n)}$). One result apparent from Fig. 19 is that it is beneficial to increase the sample size M even if k is quite small (first row of Fig. 19). Note that the right–most panel in the last row in Fig. 19 approaches asymptotically the true result. Note that taking a small sample from the low–rank approximation of $\mathbf{V}_{48}$ (panel in upper–left corner in Fig. 19) gives already a rough idea of the true result. Clearly, it may be of interest to study the properties of this SV–based Monte Carlo technique (also with respect to scaling with $s = 1$), in more detail; for example, in terms of its more detailed dependence on the values k and M chosen, and also in terms of the predicted *covariance* structures.

6. Concluding remarks and open questions

In this paper some aspects related to the prediction of the uncertainty of numerical forecast were considered. These consideration were centered around the problem of how to propagate in time an initial probability density function given an NWP model. In the restricted context of this paper, model errors were neglected, and in studying the time evolution of the pdf the main emphasis was placed on the time evolution of the covariance structure of the pdf.

After reviewing the general problem of covariance prediction, possible approaches to this problem based on the Liouville equation, the stochastic dynamic equations, or the Monte Carlo approach were briefly discussed in section 3. The more detailed consideration of the Liouville equation in section 4 indicated some interesting properties of the solution of the Liouville equation related to tangent–linear versions of NWP models, and the product of growth rates of singular vectors. It may be of interest to study both of the results discussed in section 4 somewhat further.

The discussion in section 5 concentrated around the prediction of the forecast error

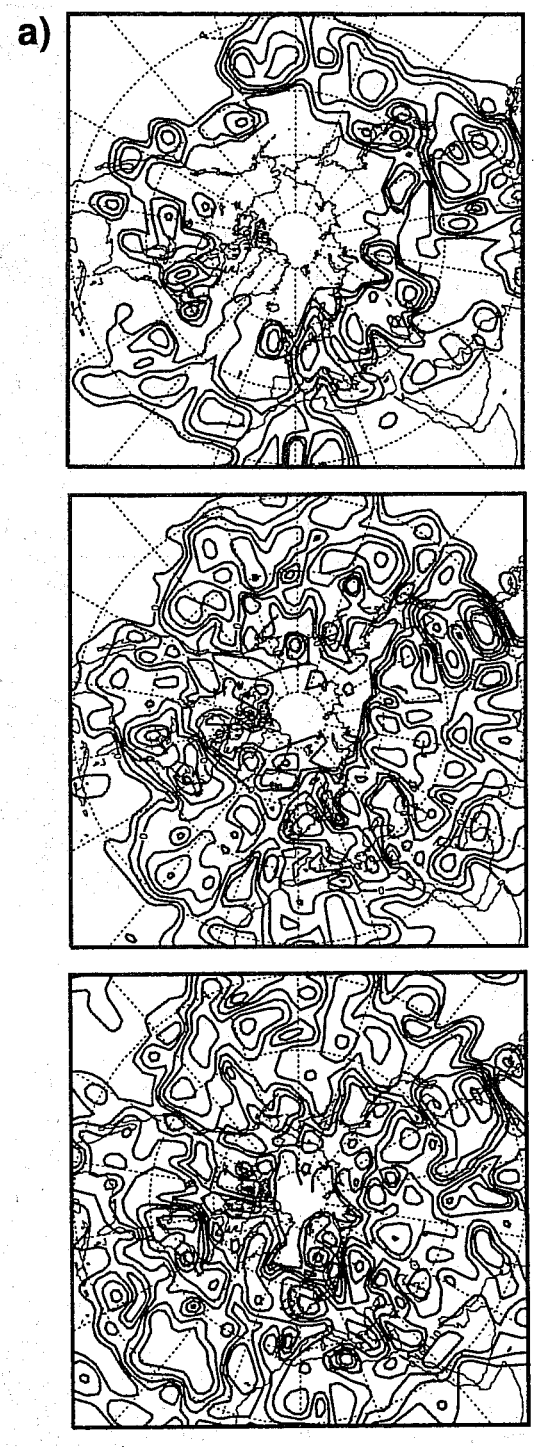

Fig. 19. Assessment of the SV–based Monte Carlo technique described in section 5.3. Shown are, in the context of experiment C7hb, variance fields at the 500 hPa model level evaluated from nonlinearly predicted forecast error covariance matrices, as a function of the number k of initial SVs $\mathbf{Z}_0$ retained to describe the initial covariance structure ($k = 10$ (top), 100 (middle), 500 (bottom) in each column), and as a function of the number of Monte Carlo integrations: (a) $M = 10$, (b) M=100, (c) $M = 5000$. All panels are approximations to the 500 hPa model level picture in Fig. 12a. Plotting convention as in Fig. 7.

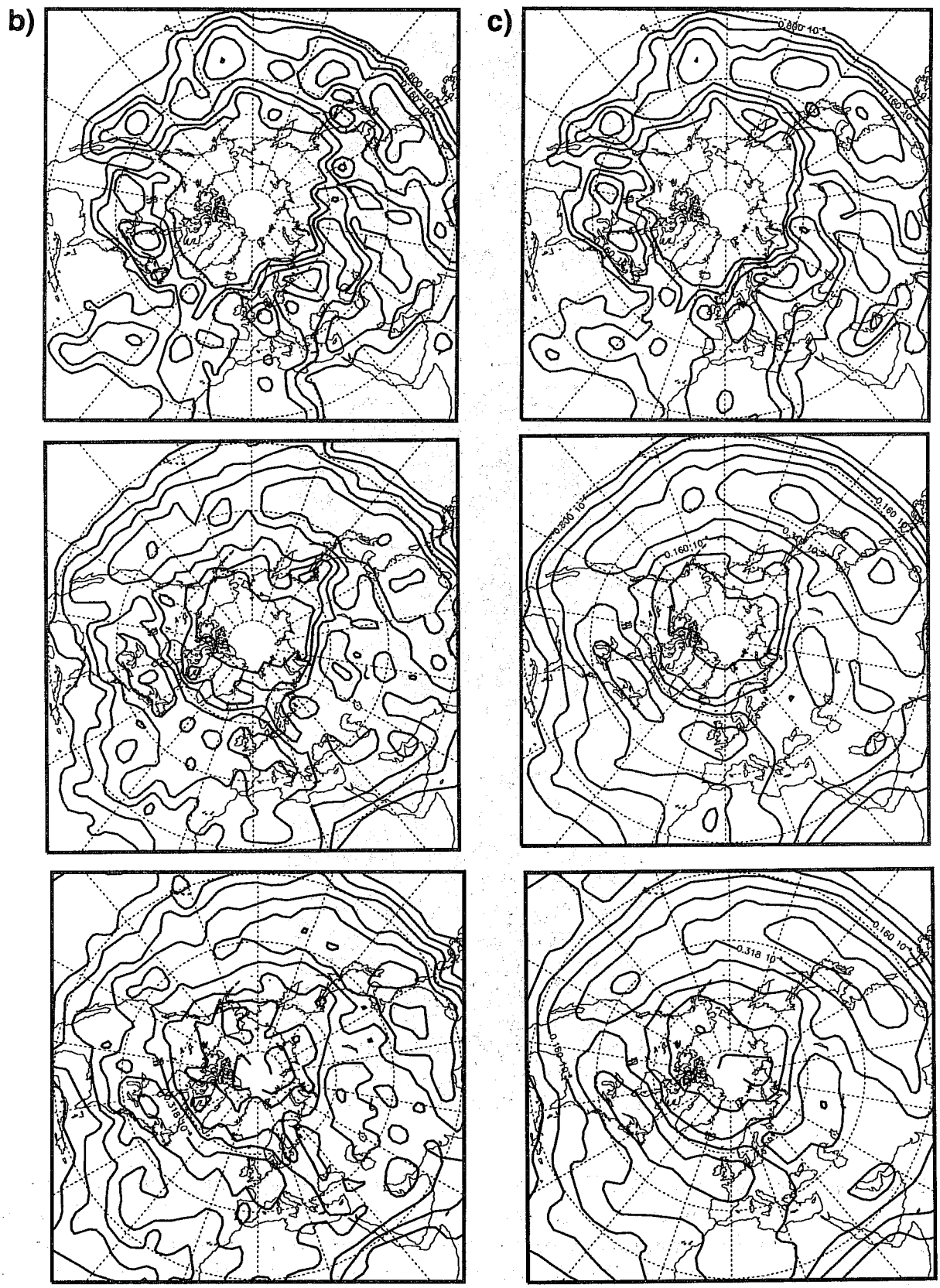

Fig. 19. (b) (c).

covariance structure as defined in eq. (5.1). This matrix carries the complete second–order information of the time–evolved pdf, and therefore its properties (be it in some approximated form) are of primary interest in making statements about the uncertainty of an NWP forecast. In section 5, various relationships between eigenvectors of covariance matrices, and the so–called singular vectors were discussed and illustrated in the context of a three–level QG model. It was emphasized that it is unavoidable to specify a norm (or metric), either explicitly, or implicitly, as soon as the concept of eigenvectors (or singular vectors) is applied in the context of covariance matrices, or (tangent–linear) model resolvents. Tangent–linear predictions of covariance structures were compared with nonlinear predictions derived from large–ensemble MC integrations. An SV–based MC procedure was discussed and briefly illustrated: this procedure relies on the representation of the initial covariance structure in terms of the initial SVs, rather than in terms of the eigenvectors of the initial covariance structure (or another square–root representation).

The paper is concluded here with some additional remarks, some of which are referring to the issue of nonlinearity. First, in the context of the SV–based MC procedure, described in section 5.3, it is clearly possible to choose for the computation of the SVs used in (5.35), or (5.43), an optimization time that is somewhat shorter than the subsequent integration time used for the nonlinear evolution of the model states defined through (5.43). In that case the tangent–linear approximation is more accurate, but the SVs cannot fully develop into the eigenvectors of $\mathbf{S}_t^{\mathrm{TL}}$. Studying that issue systematically is complicated by the appearance of nonlinearities.

Second, one of the central equations used as a starting point to study covariance evolution is eq. (5.3), expressing the tangent–linearly approximated form of eq. (5.1). This equation is not only the starting point for SV computations (e.g., Barkmeijer et al. 1997), but it is also central in the covariance prediction step in the Kalman filter (e.g., Bouttier 1996). In reconsidering (5.3), it is obvious that the use of *linear* (tangent–linear) operators to describe the evolution of perturbations limits the validity of such error evolution depending on the time intervals chosen and the magnitude of the initial perturbation. Making this approximation amounts to neglecting the higher–order terms indicated in (4.14). It appears, however, that these higher–order terms can be accounted for – to some degree, at least – at several places. For example, consider rewriting equation (5.3) in the form:

$$\mathbf{S}_t^{\mathrm{TL}} = \left(\mathbf{P}\mathbf{M}_t\mathbf{V}^{1/2}\right)\left((\mathbf{V}^{1/2})^{\mathrm{T}}\mathbf{M}_t^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\right) = \left(\mathbf{P}\mathbf{M}_t\mathbf{V}^{1/2}\right)\left(\mathbf{P}\mathbf{M}_t\mathbf{V}^{1/2}\right)^{\mathrm{T}} =$$

$$= \left(\mathbf{P}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathbf{M}_t\mathbf{v}_1^s & \mathbf{M}_t\mathbf{v}_2^s & \ldots & \mathbf{M}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}\right)\left(\mathbf{P}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathbf{M}_t\mathbf{v}_1^s & \mathbf{M}_t\mathbf{v}_2^s & \ldots & \mathbf{M}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}\right)^{\mathrm{T}}, \quad (6.1)$$

where $\mathbf{v}_i^s$ denotes the ith vector appearing in a square–root representation of $\mathbf{V}$. The form of (6.1) clearly suggests to consider replacing the tangent–linear operator $\mathbf{M}_t$ (as it operates on the set of perturbations $\mathbf{v}_i^s$) by a nonlinear operator. Referring back to (4.14), a possibility for such a formulation would be:

$$\mathbf{S}_t^{\mathrm{NL}} = \left(\mathbf{P}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathcal{H}_t\mathbf{v}_1^s & \mathcal{H}_t\mathbf{v}_2^s & \ldots & \mathcal{H}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}\right)\left(\mathbf{P}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathcal{H}_t\mathbf{v}_1^s & \mathcal{H}_t\mathbf{v}_2^s & \ldots & \mathcal{H}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}\right)^{\mathrm{T}} =$$

$$= \mathbf{P}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathcal{H}_t\mathbf{v}_1^s & \mathcal{H}_t\mathbf{v}_2^s & \ldots & \mathcal{H}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathcal{H}_t\mathbf{v}_1^s & \mathcal{H}_t\mathbf{v}_2^s & \ldots & \mathcal{H}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}, \quad (6.2)$$

where the fully nonlinear error evolution operator $\mathcal{H}_{\mathbf{x}_0,t}$ has been defined in (4.14). Note that the computation of the action of $\mathcal{H}_{\mathbf{x}_0,t}$ on a given vector $\mathbf{v}_i^s$ is only approximately half as costly as the computation of the tangent–linear operator $\mathbf{M}_{\mathbf{x}_0,t}$. The forecast error covariance matrix $\mathbf{S}_t^{\mathrm{NL}}$ will be positive definite and symmetric, by its definition (6.2).

It seems unclear, however, in which way the eigenstructure of $\mathbf{S}_t^{\mathrm{NL}}$ will relate to the concept of singular vectors. It is also important to note that expression (6.2) is not the same as the definition (5.1), because (6.2) does not allow for the nonlinear evolution of the mean, $E(\mathbf{x}_0)$ (see below). In addition, it must also be noted that $\mathbf{S}_t^{\mathrm{NL}}$, evaluated according to (6.2), will depend on the particular form of the square–root chosen for $\mathbf{V}$ – in contrast to eq. (6.1): here the result $\mathbf{S}_t^{\mathrm{TL}}$ will be always the same, no matter how $\mathbf{V}^{1/2}$ was chosen (assuming, of course, that all n vectors in $\mathbf{V}^{1/2}$ are used).

In a sense then, eq. (6.2) defines a nonlinear operator $\mathbf{H}_t$ (that is clearly not representable in a matrix form) to propagate the analysis error covariance matrix:

$$\mathbf{S}_t^{\mathrm{NL}} = \mathbf{P}\mathbf{V}_t^{\mathrm{NL}}\mathbf{P}^{\mathrm{T}}, \quad (6.3)$$

where the nonlinearly time–evolved analysis error covariance matrix is:

$$\mathbf{V}_t^{\mathrm{NL}} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathcal{H}_t\mathbf{v}_1^s & \mathcal{H}_t\mathbf{v}_2^s & \ldots & \mathcal{H}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \mathcal{H}_t\mathbf{v}_1^s & \mathcal{H}_t\mathbf{v}_2^s & \ldots & \mathcal{H}_t\mathbf{v}_n^s \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}^{\mathrm{T}} \equiv \mathbf{H}_t(\mathbf{V}). \quad (6.4)$$

As a first assessment of the structure of $\mathbf{S}_t^{\mathrm{NL}}$, expression (6.2) was evaluated for the situation of experiment C7hb (see Table 1 in section 5.4), taking as $\mathbf{V}^{1/2}$ the (unique) lower–triangular Cholesky decomposition of $\mathbf{V}$ (i.e., $\mathbf{V}_{48}$ in this experiment). The result is shown

in Fig. 20 (using $\mathcal{H}_t$), and it is directly comparable with Fig. 11c (using $\mathbf{M}_t$). It is immediately evident that both figures are highly similar, with, however, the total final variance in $\mathbf{S}_t^{\mathrm{NL}}$ of $43.796 \times \times 10^{-10}$ s^{-2} being just insignificantly smaller than the relevant value in $\mathbf{S}_t^{\mathrm{TL}}$ ($43.841 \times \times 10^{-10}$ s^{-2}; see section 5.4). This result is an indication that the tangent–linear approximation is highly accurate (over 48 hours) when tested on the columns of the lower–triangular Cholesky decomposition of $\mathbf{V}$.

However, as discussed in section 5.4, the inaccuracy of the tangent–linear approximation implied large differences between $\mathbf{S}_t^{\mathrm{TL}}$ (Fig. 11c) and the result $\mathbf{S}_t$ (Fig. 12a) obtained through the MC approach (with $M = 5000$). These differences must then traced to the fact that – in the presence of nonlinearities in the model (2.1) – the mean over many (nonlinearly evolved) realizations is different from the nonlinear time–evolution of the initial mean $\mathbf{x}_0^c$:

$$E(\mathbf{x}_t) \approx \overline{\mathcal{M}_t(\mathbf{x}_0^i)} \neq \mathcal{M}_t(\mathbf{x}_0^c)\,, \tag{6.5}$$

where $\mathbf{x}_0^i$ represents a realization from the initial pdf (2.2), the overbar indicates the averaging over all realizations i (intended to estimate $E(\mathbf{x}_t)$ according to eq. (3.10)), and $\mathcal{M}_t$ is the nonlinear model operator introduced in (2.1). By referring back to (4.14), it is noted that it is $\mathcal{M}_t(\mathbf{x}_0^c)$ that is used to define $\mathcal{H}_t$. It is speculated that using $\overline{\mathcal{M}_t(\mathbf{x}_0^i)}$ instead to define $\mathcal{H}_t$ rather than using $\mathcal{M}_t(\mathbf{x}_0^c)$ might make $\mathbf{S}_t^{\mathrm{NL}}$ more similar to $\mathbf{S}_t$; note, however, that $\overline{\mathcal{M}_t(\mathbf{x}_0^i)}$ only becomes available through performing MC integrations.

The difference between the mean state at time t and the time–evolved mean, as expressed in (6.5), is indeed substantial in the present experiment (at a forecasting time of 48 hours). This result may also be given the interpretation that the coupling term in the stochastic–dynamic equation for the mean with the second–order stochastic–dynamic equation is not negligible (note that the MC approach can be interpreted as integrating the stochastic dynamic equations with perfect closure). The same experiment was carried out for a forecasting time of just 6 hours: in this case, $\mathbf{S}_t$, $\mathbf{S}_t^{\mathrm{NL}}$, and $\mathbf{S}_t^{\mathrm{TL}}$ are all very similar, as are $\overline{\mathcal{M}_t(\mathbf{x}_0^i)}$, and $\mathcal{M}_t(\mathbf{x}_0^c)$, indicating only weak coupling between the first–order and second–order stochastic dynamic equations.

In view of this discussion it might be of interest to look closer at nonlinear predictions of the forecast error covariance matrix, and, in particular, at the coupling between the first– and second–order stochastic dynamic equations (see, also, Cohn 1993).

More generally, in dealing with nonlinear effects in approaches to uncertainty prediction, it may be important and useful to describe nonlinearities through a specific set of indicators, such as, for example: (i) the difference between mean forecast and control forecast (see also, Hersbach et al. 1997), (ii) the difference between $\mathbf{S}_t^{\mathrm{TL}}$ and $\mathbf{S}_t^{\mathrm{NL}}$ (see

Fig. 20. The forecast error covariance matrix, for experiment C7hb, in the (partly nonlinear) formulation $\mathbf{S}_t^{\mathrm{NL}}$ (see, eq. 6.3)), shown in terms of variance fields at the three model levels. Plotting convention as in Fig. 7.

eqs. (6.1), (6.2)) for different factorizations of $\mathbf{V}$, (iii) the difference between $\mathbf{S}_t^{\mathrm{TL}}$ and $\mathbf{S}_t$ (see eqs. (5.3) and (5.1)), (iv) the difference between nonlinearly and tangent–linearly evolved perturbations (see eq. (4.14)) for perturbations consistent in structure and size with the analysis error covariance matrix $\mathbf{V}$. Assessing different aspects of nonlinearity (see also, Buizza 1995) may provide new insight for finding approximate ways for their inclusion in uncertainty prediction. Clearly, further studies will be required to study in more detail the difficulties encountered in covariance prediction when model nonlinearities become important.

A point related to the effect of nonlinearities concerns the extension of the "linear" optimization problem (4.16) to the "nonlinear" optimization problem:

$$\max_{\mathbf{z}_0} J = \langle \mathbf{z}_t; \mathbf{z}_t \rangle = \langle \mathcal{H}_{(\mathbf{x}_0,t)}(\mathbf{z}_0); \mathcal{H}_{(\mathbf{x}_0,t)}(\mathbf{z}_0) \rangle \qquad \text{subject to :} \qquad \langle \mathbf{z}_0; \mathbf{z}_0 \rangle = 1, \qquad (6.6)$$

where $\mathcal{H}$ has been introduced in (4.14). The solution to (6.6) (see also, Barkmeijer 1996) will clearly include a better representation of nonlinear effects than the solution to (4.16). However, it seems less clear how a solution (or, solutions) to (6.6) relate to the fully nonlinearly obtained forecast error covariance matrix $\mathbf{S}_t$. Obviously, solving (6.6) is considerably more difficult than solving (4.16), since the solution to (6.6) can no longer be obtained by solving an eigenproblem. However, it is noted here that this nonlinear problem (6.6) bears some relationship to the variational problems considered in variational data assimilation (e.g., Courtier 1997). In such variational data assimilation contexts a cost function depending *nonlinearly* on a control variable is *minimized*. The cost function is, however, quadratic in certain variables that are nonlinearly related to the control variable.

A strong connection also exists between atmospheric data assimilation and the uncertainty prediction problem through the appearance of the analysis error covariance matrix in the definition of the singular vectors (see, e.g., eq. (5.4)). The analysis error covariance matrix $\mathbf{V}$ is, in principle, a product of a data assimilation algorithm, and through its inclusion in the computation of SVs important properties of the initial pdf are reflected by the growth of SV perturbations (see, also, Barkmeijer et al. 1997, Palmer et al. 1998). Perturbation growth obtained on the basis of such inclusion of $\mathbf{V}$ will also represent in a more realistic way the stability of the associated atmospheric flow, due to the constraints on the initial perturbations imposed by analysis error covariance information. In the study of atmospheric predictability, such stability investigations in terms of singular vectors are an important generalization of stability studies in terms of the so–called normal modes (see also, Farrell and Ioannou 1996a, b; Hall and Sardeshmukh 1998). Even though the analysis error covariance matrix $\mathbf{V}$ is, in principle, a product of a data assimilation algorithm, the direct assessment of $\mathbf{V}$ in an operational environment presents a computationally highly

expensive component within the data assimilation algorithm. Consequently, its representation in operational contexts is only possible in approximated forms (e.g., no or only very limited flow dependence; see also, Riishøjgaard 1998). The sequential assimilation algorithm described by the Kalman filter (e.g., Gauthier et al. 1993) provides a way to relax some of these assumptions, resulting, for example, in a flow–dependent $\mathbf{V}$. Consequently, in this framework of the Kalman filter, the uncertainty prediction problem can be studied more realistically in the sense that *both* the first and second moments of the initial pdf (2.3) become flow–dependent, with the possibility of obtaining more realistic flow–dependent predicted first– and second–order statistics, representing the time–evolved probability density function.

Acknowledgments. This manuscript was prepared during a one–year visit of the author as a consultant at the European Centre for Medium–Range Weather Forecasts (ECMWF). The financial support by ECMWF that made this visit possible is gratefully acknowledged. I thank Jan Barkmeijer, ECMWF, for many discussions on topics presented in this manuscript.

References

Anderson, T.W., 1958: An introduction to multivariate statistical analysis. Wiley, 374 pp.

Arnold, V.I., 1989: *Mathematical methods of classical mechanics.* Springer, 508 pp.

Barkmeijer, J., 1996: Constructing fast–growing perturbations for the nonlinear regime. *Journal of the Atmospheric Sciences,* **53**, 2838–2851.

Barkmeijer, J., R. Buizza, and T.N. Palmer, 1997: 3D–VAR Hessian singular vectors and their use in the ECMWF ensemble prediction system. Proceedings ECMWF Workshop on Predictability, 20 – 23 October 1997, this volume.

Berliner, L.M., 1998: Monte Carlo based ensemble forecasting. *Statistics and Computation,* submitted.

Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorologische Zeitschrift,* **21**, 1–7.

Bouttier, F., 1996: Application of Kalman filtering to numerical weather prediction. Proceedings 1996 ECMWF Seminar on data assimilation and Workshop on non–linear aspects of data assimilation, 61–90.

Buizza, R., 1995: Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. *Quarterly Journal of the Royal Meteorological Society,* **121**, 1705–1738.

Chu, M.T., R.E. Funderlic, and G.H. Golub, 1997: On a variational formulation of the generalized singular value decomposition. *SIAM Journal on Matrix Analysis and Applications,* **18**, 1082–1092.

Cohn, S.E., 1993: Dynamics of short–term univariate forecast error covariances. *Monthly Weather Review,* **121**, 3123–3149.

Cooke, W.E., 1906: Forecasts and verifications in western Australia. *Monthly Weather Review,* **34**, 23–24.

Courtier, P., 1997: Dual formulation of four–dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society,* **123**, 2449–2461.

DeGroot, M.H., 1987: *Probability and Statistics.* Addison–Wesley, 723 pp.

Egger, J., 1996: Volume conservation in phase space: a fresh look at numerical integration schemes. *Monthly Weather Review,* **124**, 1955–1964.

Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Monthly Weather Review,* **122**, 703–713.

Ehrendorfer, M., 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *Monthly Weather Review,* **122**, 714–728.

Ehrendorfer, M., 1997: Predicting the uncertainty of numerical weather forecasts: a review. *Meteorologische Zeitschrift, N.F.,* **6**, 147–183.

Ehrendorfer, M., and J.J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *Journal of the Atmospheric Sciences,* **54**, 286–313.

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

Farrell, B.F., and P.J. Ioannou, 1996a: Generalized stability theory. Part I: autonomous operators. *Journal of the Atmospheric Sciences,* **53**, 2025–2040.

Farrell, B.F., and P.J. Ioannou, 1996b: Generalized stability theory. Part II: nonautonomous operators. *Journal of the Atmospheric Sciences,* **53**, 2041–2053.

Fleming, R.J., 1971: On stochastic dynamic prediction. I. The energetics of uncertainty and the question of closure. *Monthly Weather Review,* **99**, 851–872.

Fortak, H., 1973: Prinzipielle Grenzen der deterministischen Vorhersagbarkeit atmosphärischer Prozesse. *Annalen der Meteorologie (Neue Folge),* **6**, 111–120.

Gauthier, P., P. Courtier, and P. Moll, 1993: Assimilation of simulated wind Lidar data with a Kalman filter. *Monthly Weather Review,* **121**, 1803–1820.

Golub, G.H., and C.F. Van Loan, 1989: *Matrix computations.* The Johns Hopkins University Press, 642 pp.

Hall, N.M.J., and P.D. Sardeshmukh, 1998: Is the time–mean northern hemisphere flow baroclinically unstable? *Journal of the Atmospheric Sciences,* **55**, 41–56.

Hersbach, H., R. Mureau, J.D. Opsteegh, and J. Barkmeijer, 1997: An EPS for the short and early medium range. Proceedings ECMWF Workshop on Predictability, 20 – 23 October 1997, this volume.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review,* **102**, 409–418.

Lin, C.C., and L.A. Segel, 1988: *Mathematics applied to deterministic problems in the natural sciences.* Society for Industrial and Applied Mathematics, 609 pp.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences,* **20**, 130–141.

Lorenz, E.N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus,* **34**, 505–513.

Lorenz, E.N., 1984: Irregularity: a fundamental property of the atmosphere. *Tellus,* **36A**, 98–110.

Lorenz, E.N., 1996: The evolution of dynamic meteorology. In *Historical Essays on Meteorology 1919–1995*, J.R. Fleming, Ed., American Meteorological Society, 3–19.

Mardia, K.V., J.T. Kent, and J.M. Bibby, 1982: *Multivariate analysis.* 521 pp.

Marshall, J., and F. Molteni, 1993: Towards a dynamical understanding of planetary–scale flow regimes. *Journal of the Atmospheric Sciences,* **50**, 1792–1818.

Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society,* **122**, 73–119.

Murphy, A.H., 1998: The early history of probability forecasts: some extensions and clarifications. *Weather and Forecasting,* **13**, 5–15.

Nicolis, G., 1995: *Introduction to nonlinear science.* Cambridge University Press, 254 pp.

Palmer, T.N., R. Buizza, and F. Lalaurette, 1997: Performance of the ECMWF ensemble prediction system. Proceedings ECMWF Workshop on Predictability, 20 – 23 October 1997, this volume.

Palmer, T.N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences,* **55**, 633–653.

Pitcher, E.J., 1977: Application of stochastic dynamic prediction to real data. *Journal of the Atmospheric Sciences,* **34**, 3–21.

Rao, C.R., 1965: *Linear statistical inference and its applications.* Wiley, 625 pp.

Riishøjgaard, L.P., 1998: A direct way of specifying flow–dependent background error correlations for meteorological analysis systems. *Tellus,* **50A**, 42–57.

Strang, G., 1993: *Introduction to linear algebra.* Wellesley–Cambridge Press, 472 pp.

Talagrand, O., 1996: Basic facts about chaotic motions. Proceedings 1996 ECMWF Seminar on data assimilation and Workshop on non–linear aspects of data assimilation, 593–594.

Thompson, P.D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus,* **9**, 275–295.

Thompson, P.D., 1972: Some exact statistics of two–dimensional viscous flow with random forcing. *Journal of Fluid Mechanics,* **55**, 711–717.

Thompson, P.D., 1983: Equilibrium statistics of two–dimensional viscous flows with arbitrary random forcing. *Physics of Fluids,* **26**, 3461–3470.

Thompson, P.D., 1987: Oral history. Interview by J. Tribbia, A. Kasahara, 52 pp. [Available from the National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, U.S.A.].

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review,* **125**, 3297–3319.

Trefethen, L.N., and B. Bau, III, 1997: *Numerical linear algebra.* Society for Industrial and Applied Mathematics. 361 pp.

Von Myrbach, O., 1913: Verläßlichkeit von Wetterprognosen (Ein Vorschlag). *Meteorologische Zeitschrift,* **30**, 496–497.

Wilks, D.S., 1995: *Statistical methods in the atmospheric sciences.* Academic Press, 467 pp.