Review on the Use of Singular Vectors to Study Error Growth in Numerical Weather Prediction Models

M.Sc. Essay

Li Huang
Atmospheric Science Program
Department of Earth and Ocean Sciences
University of British Columbia

April 20th, 2005

Table of Contents

1. Introduction

Numerical weather prediction (NWP) models possess the property that two or more slightly different initial states, in general, over time develop into states no more similar than two or more randomly observed states of the atmosphere. This inherent error growth is a consequence of the nonlinearity and instability of the atmospheric dynamics (Leith 1978). The finite errors in initial conditions and inevitable model deficiencies imply a limited forecast range to skillful atmospheric predictions. Errors in initialization, errors and simplifications in parameterization schemes, errors due to model formulation, and errors in lateral boundary conditions for limited area models all contribute to the overall errors in NWP forecasts.

Singular vectors (SVs) are one tool for studying numerical predictability. They have been widely used in three different ways (Palmer et al. 1998). First, following the work of Orr (1907) in studying the transition to turbulence in Couette flow, SVs have been applied in geophysical-fluid-dynamics studies to explain particular phenomena, such as extratropical cyclogenesis (e.g., Farrell 1982) and El Nino (Penland and Sardeshmukh 1995).

Second, SVs of the linearized equations of motion have been employed to study questions of the atmosphere-ocean system predictability (Palmer 1996), especially to estimate the evolution of initial errors during the course of a forecast. There have been a number of studies related to the SVs approach in the atmospheric predictability field for both post-mortem cases and operational forecasts. For example, singular vectors were able to explain the dependence of weather forecast errors on the sign of the Pacific-North American mode (Molteni and Palmer 1993); to capture a major transition to blocking (Mureau et al. 1993); and to demonstrate the upscale energy cascade associated with extratropical predictability (e.g., Charney and Shukla 1981). Leutbecher (2003) estimated changes of forecast error produced by a change of the observation network during an assimilation cycle by performing the required calculations in a subspace of a small number of leading SVs. Ehrendorfer and Tribbia (1997) concluded that SVs, constructed using covariance information valid at the initial time in a tangent-linear framework, represented the most efficient approach for predicting the forecast error-covariance matrix valid for the end of the optimization interval.

A third potential use of SVs is considered in defining a strategy for targeting adaptive observations of the atmosphere or the oceans. In such a strategy, observations would be made in particularly flow-dependent "sensitive" regions determined by the location of the singular vectors at initial time (Palmer et al. 1998).

Since Lorenz (1965) first considered singular vectors in a meteorological context in a study of the predictability of a 28-variable atmospheric model and concluded that errors tend to project along non-random directions in phase space, the dynamics of short-term error growth have been best understood in terms of SVs. At larger synoptic scales, at which most of error growth occurs, their behaviors are fairly linear. Rabier et al. (1996) showed that the sensitivity of day-2 forecast error to perturbations in the initial state projects well into the space of dominant SVs. Based on the understanding of SV's direct relevance to the directions of most rapid error growth in NWP models, localized SVs have been used to routinely construct initial perturbations for the ensemble prediction system in European Centre for Medium-Range Weather Forecasts (ECMWF) (Palmer 1993, Molteni and Palmer 1993, Ehrendorfer and Errico 1995).

In this review, I will focus on singular-vector application in ensemble forecasts (belonging to predictability studies of the atmosphere), discuss why SVs can explain directions of error growth in the tangent-linear model frame, how SVs are used as the basis of initial perturbations for operational medium-range ensemble forecasting at ECMWF, and some studies on short-range ensemble prediction system constructed on optimal SVs.

The organization of this essay is as follows. Section 2 gives a review of ensemble forecasting methods in NWP models. Section 3 describes the phase-space evolution of perturbations. Section 4 contains some issues on SV methodology, such as SV formulation, the relationship between SVs and error growth in the phase space, the choice of norms, the nonmodality of SVs, and the comparison of SVs with Lyapunov vectors. SV application on ensemble forecasts is discussed in section 5. A summary is presented in section 6.

2. Review of ensemble forecasting methods

2.1 Early studies

a. Stochastic-dynamic forecasting

In 1969, Epstein introduced the idea of stochastic-dynamic forecasting, which is the first forecasting method to explicitly acknowledge the uncertainty of atmospheric-model predictions. By deriving a continuity equation for the probability density of the model solution of a dynamical model with some assumptions, the first and second moments of the probability distribution (expected means and covariances) could be obtained. Because this approach requires a huge number of forecast equations for modern NWP models, it is completely unfeasible.

b. Monte Carlo forecasting

The Monte Carlo method is the first approach to efficiently construct ensemble forecasts, which are potentially as useful as stochastic forecasts, but are much cheaper. Leith (1974) concluded that an individual deterministic forecast for long lead times has twice the error covariance of a climatological forecast. He also built a regressed forecast, based on the least-squared error, to eventually temper the forecast towards climatology. Instead of regression, which involves considerable work to estimate the matrix of regression coefficients, he pointed out that adequate accuracy would be attained for the best estimate of the forecast with ensemble member sizes as small as 8, for members chosen randomly that all are equally likely. He also thought that the estimation of forecast errors might require a larger number of ensemble members. Finally Leith showed that averaging a Monte Carlo ensemble of forecasts approximates the tempering of the forecasts towards climatology, without the need to perform regression.

In the late 1980's, Errico and Baumhefner (1987), Tribbia and Baumhefner (1988), and Mullen and Baumhefner (1989) followed Leith to produce random initial perturbations. These approximate the actual two-dimensional error of the horizontal fields as estimated by Daley and Mayer (1986), on limited-area domains by using a variation of the spectra method.

c. *Lagged-average forecasting*

In 1983, lagged-average forecasting (LAF) was created as an alternative to the Monte Carlo forecasting by Hoffman and Kalnay. This approach takes advantage of the fact that operational forecast centers produce new forecast runs every τ hours. The way to construct ensemble members is to combine the forecasts initialized at the current initial time, t=0, as well as at previous times, $t = -\tau, -2\tau, \ldots, -(N-1)\tau$ together at the same forecast valid time, so that the initial perturbations are generated automatically from the forecast errors. By comparing the lagged-average forecasting and the Monte Carlo forecasting methods within a simulation system, they found that LAF predicted forecast skill much better, with the correlation between predicted and observed time of crossing the 50% skill level being 0.68 for Monte Carlo forecasting and 0.79 for lagged-average forecasting, even though LAF ensemble average forecast was only slightly better than the Monte Carlo forecast (Kalnay 2003). LAF initial perturbations were not chosen randomly like in Monte Carlo forecasting, but contained dynamical influences. Another advantage is that the initial conditions are computed cost free, because sequential forecasts required are already generated operationally. Toth and Kalnay (1993) pointed out that the disadvantage of LAF is that the forecast error magnitude is not the same for each ensemble member because of the different forecast lead times.

2.2 Operational ensemble forecasting methods

The two leading operational centers in the world, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP), use the singular-vector and breeding methods to build the initial perturbations for medium-range (5-15 day) ensemble forecasting. At the Canadian Meteorological Center (CMC), ensemble forecasts are performed by introducing initial-condition and model errors.

a. Breeding method

In 1993, Toth and Kalnay first introduced the breeding method into ensemble forecasting. A random initial perturbation is introduced into a breeding cycle with a given initial size, then a short-range control forecast and a short-range perturbed forecast are obtained by integrating from the control and from the perturbed initial conditions. The difference between these two forecasts is scaled down so that it has the same amplitude as the initial perturbation, and then is added to the corresponding new analysis state to create a new perturbation (denoted bred vectors). Beyond an initial transient period of 3-4 days after random perturbations were introduced, the perturbations acquired large growth rates, faster than for Monte Carlo forecasting or lagged-average forecasting. Toth and Kalnay argued that the differences represent the fastest growing forecast errors, which are only partially removed in the analysis cycle by the addition of new observations.

b. Singular vectors

The singular-vector approach was based on the linear error-propagator concepts developed by Lorenz (1965). Researchers at ECMWF used the forward tangent version and the adjoint tangent version of the Integrated Forecasting System, a medium range forecast model, to catch the important analysis errors (Buizza, 1997). Firstly, they defined an error-propagator matrix that was determined by the solution of equations in a linearized simplification of the NWP model, where an error term was already added. Secondly, ECMWF used the total energy norm as the initial norm. By calculating the norm of the state vector, they got the eigenvalues of the propagator matrix squared (singular vectors), which identify directions of the greatest error growth. Finally, they applied the adjoint of the linear model to project the errors back onto the initial state to generate perturbations for an ensemble forecast. Similar to the breeding method, the SV approach constitutes its initial perturbations based on the evolving underlying atmospheric flow rather than random errors.

c. Ensembles based on multiple data assimilation

By running an ensemble of data-assimilation systems, Houtekamer et al. (1996) created the initial conditions for an ensemble forecasting system. The errors were randomly added to the observations in different data-assimilation systems, and also different physical-parameterization schemes of the model were included in different ensembles. This ensemble approach is related to the breeding method and it is more general than it. Hamill et al. (2000) have shown for the quasi-geostrophic system that the multiple data assimilation ensemble system performs better than the singular vector or breeding approaches.

d. Multisystem ensemble approach

Most of the ensemble methods above assume a “perfect model”, and focus only on the statistical uncertainty in the initial conditions. Recently, perturbations to the models have been introduced in ensemble forecasting by varying the model parameterizations of subgrid-scale physical processes (Stensrud et al. 1998). It is concluded that an ensemble average of operational global forecasts from different operational centers is much more skilful than the best individual forecast (Fritsch et al., 2000). In mesoscale short-range ensemble forecasting with perturbations in the subgrid-scale “physics”, ensemble skills are improved more dramatically than in global models, since the higher resolution may allow a faster response from growing modes driven by convective instability (Hou et al., 2001). Krishnamurti et al. (1999) have shown that the ensemble skill is significantly improved in a multisystem with correction of the systematic errors by regression.

In the meantime, ensemble forecasts have been used to improve data assimilation. This technique is called the ensemble Kalman Filter (EnKF), and is a special case of a linear filter. An estimate of the forecast error covariance can be obtained from an ensemble of data-assimilation systems.

3. The forecast probability density function

Edward Lorenz (1963) discovered the fundamental theorem of predictability. This theorem says that unstable systems have a finite limit of predictability, and conversely, stable systems are infinitely predictable. Lorenz demonstrated that the atmosphere is a chaotic dynamic system with instability and even if you could create a perfect model, predictability is limited to about two weeks by the sensitivity to the imprecise initial conditions. This explained the primary reason for the limitations of deterministic NWP, which by this time in its development was meeting with some success.

One way to deal with the prediction problem of forecast uncertainty is with a probability density function (PDF). By looking at the time-evolution of the forecast-error PDF in an unstable system, we are able to understand how the initial uncertainty develops and how this system finally loses its predictability (e.g., Palmer 1996, Kalnay 2003).

Fig. 1 shows a schematic illustration of the phase-space evolution of the PDF of analysis error throughout the entire forecast range. We assume that the distribution of the PDF is normal along each phase-space direction at initial time (Fig. 1 (a)), namely the PDF is isotropic. In the early part of the forecast, error growth is governed by linear dynamics. During this period an initially spherical isopleth of the PDF will evolve to bound a m-dimensional ellipsoidal volume (Fig. 1 (b)), where m is the dimension of phase space ($O(10^7)$ for the ECMWF operational forecast model). The major axis of the ellipsoid corresponds to a phase-space direction that defines the dominant finite-time instability of that part of phase space (relative to the analysis error covariance metric). The small arrow shown in Fig. 1 (b) points along the major axis of the ellipsoid. It can be thought of as evolving from the small arrow shown in Fig. 1 (a). The small arrow in Fig. 1 (a) is not parallel to the one in (b). This illustrates the non-modal nature of linear perturbation growth.

The growth of the PDF between Fig. 1 (b) and (c) can be described as a nonlinear evolution of the PDF. In Fig.1 (c) the PDF has deformed from its ellipsoidal shape. The nonlinear deformation will cause the PDF to evolve away from a normal distribution. Fig. 1 (d) schematically describes the situation where the PDF has evolved to cover the entire attractor, so that all predictability has been lost. That means we only know that each original perturbation is within the climatology of possible solutions, but we don't know where, or even in which region of the "attractor" it may be (Kalnay 2003).

In practice, for NWP, there was evidence that errors of about one standard deviation of the analysis error PDF evolve linearly for 2-3 days, and that the 'weakly nonlinear' timescale lasts until about day 7 of the forecast (Hartmann et al. 1995). More recent evidence suggests that the strongly nonlinear growth phase might start sooner, after 2 or 3 days.

4. Singular vector methodology

a. SVs as a representation of directions of error growth

Singular vectors arise when searching for the perturbation that, when added to a given basic state, will achieve maximum growth over a specified time interval. It was noticed that singular vectors have norm-dependent structures (e.g., Palmer 1998).

A nonlinear system can be written as a set of *n* coupled ordinary differential equations:

$$\frac{dX}{dt} = F(X) \qquad X = (x_1, x_2, \dots, x_n) \qquad F = (F_1, F_2, \dots, F_n) \tag{1}$$

After choosing a finite-difference scheme, the above nonlinear equations become a set of difference equations:

$$X(t) = M[X(t_0)] \tag{2}$$

where M is the time integration of the numerical scheme from the initial condition t_0 to time t. It is shown that the numerical solution of a dynamic system is decided only by its initial value once a time-difference scheme is chosen for (1).

Let us consider a small perturbation y of the state vector X. For sufficiently short time intervals, the linear evolution of the perturbation y will be given by

$$\frac{dy}{dt} = Jy \tag{3}$$

where $J = \partial F / \partial X$ is the Jacobian of F.

After adding a small perturbation $y(t_0)$ to (2), and using Taylor series, we have:

$$\begin{aligned} M[X(t_0)+y(t_0)] &= M[X(t_0)] + \frac{\partial M}{\partial X} y(t_0) + O[y(t_0)^2] \\ &= X(t) + y(t) + O[y(t_0)^2] \end{aligned} \tag{4}$$

Let $L(t_0,t) = \frac{\partial M}{\partial X}$, we can get:

$$y(t) = L(t_0,t) y(t_0) \tag{5}$$

In fact, (5) is the integral form of (3). The operator $L(t_0,t)$ is known as the propagator of the forward tangent linear model. It propagates an initial perturbation at initial time t_0 into the final perturbation at time t. $L(t_0,t)$ depends on the basic trajectory $X(t)$ since it is linearized over the flow from t_0 to t, but it does not depend on the perturbation y. If $y(t_0)$ is the typical error in the initial conditions for a weather forecast, then (3) and (5) hold for approximately 2-3 days of integration time.

Consider further a scalar, positive-semidefinite quantity $\tilde{J}$ defined as

$$\tilde{J}(y(t)) = y(t)^T y(t) \tag{6}$$

Substitution of (5) into (6) leads to an expression of $\tilde{J}$ in terms of $y(t_0)$, to be denoted $\hat{J}$, and allows us to state in the following way the maximization problem, leading to the definition of the SVs (e.g., Ehrendorfer and Errico 1995). Maximize

$$\hat{J}(y(t_0)) = (Ly(t_0))^T (Ly(t_0)) \tag{7}$$

subject to the constraint

$$y(t_0)^T y(t_0) = 1 \tag{8}$$

Note that (7) is chosen in such a way that (7) reduces to (8) for the choice $L = I$, which corresponds to the case that no time evolution is considered (I denotes the identity operator). In other words, $\hat{J}$ is restricted to unity initially.

It can be seen, by introducing a Lagrangian and setting its derivatives equal to zero, that the solution to this problem is given by the solution of the eigenproblem

$$(L^*L)v_i = \sigma_i^2 v_i \qquad \text{with} \quad v_i^T v_i = 1 \tag{9}$$

where L^* is the adjoint of L (see more details in Kalnay 2003). If L is represented in matrix form, then L^* is just the matrix transpose of L. The vectors v_i are the eigenvectors of the matrix L^*L, σ_i^2 are the eigenvalues.

Since $L^*L = (L^*L)^T$, the operator L^*L is a symmetric and normal matrix. We can choose its eigenvectors to constitute an orthonormal basis in the tangent space of linear perturbations, with real eigenvalues $\sigma_i^2 \geq 0$.

At the end of the interval of optimization, these eigenvectors evolve to u_i which in turn satisfy the eigenvector equation

$$(LL^*)u_i = \sigma_i^2 u_i \tag{10}$$

Also $\quad u_i = Lv_i \qquad (11)$

Therefore $\quad u_i^T u_i = (Lv_i)^T Lv_i = v_i^T L^T Lv_i = v_i^T \sigma_i^2 v_i = \sigma_i^2 \qquad (12)$

Due to the terminology of linear algebra, the σ_i, ranked in terms of magnitude, are called the singular values of the operator L, and the vectors v_i and u_i are called the right singular vectors of L and the left singular vectors of L, respectively.

Let $y(t_0) = v_i$ in (7) with v_i satisfying (9), we have

$$\hat{J}(y(t_0) = v_i) = \sigma_i^2 v_i^T v_i = \sigma_i^2 \tag{13}$$

Since initially $\hat{J}$ is restricted to unity, the eigenvalues σ_i^2 corresponding to individual SVs indicate by how much the value of $\hat{J}$ increases (or decreases) from t_0 to t. Any vector can be written as a linear combination of singular vectors as follows

$$y(t_0) = \sum_{i=1}^{n} \langle y_0, v_i \rangle v_i \tag{14}$$

$$y(t) = \sum_{i=1}^{n} \langle y_t, u_i \rangle u_i \tag{15}$$

where $\langle x, y \rangle$ is the inner product of two vectors x and y. After using (14) and (11), we get

$$y(t) = L(t_0, t) y(t_0) = \sum_{i=1}^{n} \langle y_0, v_i \rangle u_i \tag{16}$$

Now we take the inner product of (16) and obtain

$$\langle y(t), u_i \rangle = \sigma_i \langle y(t_0), v_i \rangle \tag{17}$$

This indicates that by applying the tangent linear model L each initial vector v_i component will be stretched by an amount equal to the singular value σ_i (or contracted if $\sigma_i < 1$), and the direction will be rotated to that of the evolved vector u_i.

If we use the inner product of a vector with itself to define the norm of this vector, the norm of small error y at initial time can be calculated by

$$\|y(t_0)\|^2 = \langle y(t_0), y(t_0) \rangle = 1 \tag{18}$$

And the norm at optimization time is given by

$$\|y(t)\|^2 = \langle y(t), y(t) \rangle = \langle L^* L y(t_0), y(t_0) \rangle \tag{19}$$

Based on the discussion above, we know that any $\dfrac{y(t)}{\|y(t_0)\|}$ can be written as a linear combination of the eigenvectors u_i and v_i so that we have

$$\max_{y(t_0) \neq 0} \left(\frac{\|y(t)\|}{\|y(t_0)\|} \right) = \sigma_1 \tag{20}$$

It implies that maximum error growth rate over the time interval $t - t_0$ is associated with the dominant singular vectors v_i and u_i at initial time and at optimization time, respectively. Similarly, applying L^* is like running the adjoint model backward, from t_1 to t_0. If we apply the adjoint model to isotropic perturbations with size 1, they also get stretched or contracted by the ratio of σ_i, and rotated into the directions of the v_i (e.g., Palmer 1996, Kalnay 2003).

Following the discussion above, we know that a set of initial perturbations on an isotropic sphere will evolve into an ellipsoid. The u_i define the directions of the axes of the forecast PDF ellipsoid, with u_1 defining the major axis, u_2 the second major axis, and so on. The directions at initial time that evolve into these axes are given by v_1 , v_2 respectively.

For large NWP models (e.g., $O(10^4)$ or more), the eigenvalue problem (equation (9) and equation (10)) cannot be solved directly since it is too time consuming. Instead, iterative techniques provide an alternative possibility if the adjoint propagator has been coded. ECMWF uses either the Lanczos algorithm or the Jacobi-Davidson algorithm to estimate singular vectors (e.g., Strang 1986, Buizza and Palmer 1995, Sleijpen and van der Vorst 1995, Barkmeijer et al. 1998).

b. The choice of norms

Singular vectors are very sensitive to the choice of norms. We can define a norm using a weight matrix W applied to y at the initial time so that the size of initial perturbation is set to 1. In the meantime, we use a different norm to define the size of the perturbation to be maximized at the final time, for example the final norm could be a projection operation operator P. After performing a maximization problem, we have the eigenvalue problem as follows

$$(W^{-1})^T L^T P^T PLW^{-1}\hat{y}(t_0) = \lambda\hat{y}(t_0) \tag{21}$$

subject to the constraint

$$\hat{y}(t_0)\hat{y}(t_0) = 1 \tag{22}$$

where, $\hat{y}(t_0) = Wy(t_0)$

Palmer et al. (1998) tested different matrices generated by energy, enstrophy, and streamfunction squared norms at initial time. They noticed that the use of different initial norms resulted in extremely different initial singular vectors. From two independent sets of calculations based on analyses, and short-range forecast data, it was concluded that of these three choices, energy is the most appropriate metric for the predictability problem. The energy metric is a reasonable first-order estimate of the analysis-error covariance metric. Enstrophy and streamfunction were ruled out as suitable metrics since they could not present some consistency between singular vector and analysis error structure.

Enrendorfer and Tribbia (1997) showed that the initial SVs, constructed with the analysis-error covariance matrix (AECM), evolved into the eigenvectors of the forecast error-covariance matrix. This implies that SVs are optimal in describing the forecast error at the end of the optimization period. However, in realistic contexts, knowledge about AECM (or, more precisely, the positive root of AECM) may be inaccurate, thus limiting the applicability of the SV method in these situations. They thought that, with an operational data-assimilation system, this approach should be available since any initial misspecification in AECM would gradually be removed through accumulating dynamical and observational information.

The second derivative, called Hessian, of the variational analysis cost function (the sum of background and observation cost functions) can be proved to be equal to the inverse of the AECM (Fisher and Courtier 1995). Hence, in terms of the Hessian, the SV computation becomes equivalent to a generalized eigenvector equation (e.g., Palmer et al. 1998, Barkmeijer et al. 1999). Currently the Hessian SVs have been referred to as the second choice in ECMWF ensemble-prediction system. Barkmeijer et al. (1998) found that the use of the analysis-error covariance as the initial norm, instead of the energy norm, resulted in improved results.

For instance, it is possible to set the state vector to zero outside a prescribed area at optimization time, by using a projection operator, and only optimizing energy inside the geographical area. Buizza (1994) found that the ensemble contained perturbations along vectors that would not have been otherwise perturbed because they were ranked too low. The ensemble spread over the European verification area was increased while perturbations outside the interest area were discarded. At present scientists at ECWMF are using different projection operators in their operational medium-range ensemble systems.

In order to study the upscale energy transfer, Palmer (1994) defined a spectral projection operator $P_{[n_1,n_2]}$ as follows

$$\begin{aligned} P_{[n_1,n_2]}x_n &= x_n \quad \text{if } n \in [n_1,n_2] \\ P_{[n_1,n_2]}x_n &= 0 \quad \text{otherwise} \end{aligned} \tag{23}$$

Here $[n_1,n_2]$ denotes the total wavenumber interval $n_1 \le n \le n_2$, and x_n is the wavenumber n component of the spherical harmonic expansion of the state vector. This projection operator can be applied to study SVs whose energy is optimized to a specific wavenumber interval (representing the scale).

c. The nonmodality of SVs

In reality, the linear-evolution operators are never normal ($L^*L \uparrow LL^*$) because of vertical and horizontal shear (e.g., Farrell and Ioannou 1996). That means these eigenvectors of the operator L or adjoint L^* are not normal.

For indefinitely long optimization time, the dominant SVs, at initial time and optimization time, are determined by the first adjoint “eigenmode” and the first “eigenmode” itself, respectively. For finite optimization time, the dominant SVs no longer project onto individual “eigenmodes” or their adjoint “eigenmodes” (see Palmer 1996 for details). They may possess very different properties than exponentially-growing-shape-preserving normal-mode solutions to linear-perturbation equations (e.g., Farrell 1988).

The essential nonmodality of SV evolution was discussed by Buizza and Palmer (1995), with emphasis on the upscale cascade of energy from subcyclone to cyclone scale, and vertical propagation of energy from the baroclinic steering-level to the upper-troposphere jet level. Non-mode growth of perturbation, represented through SVs, is faster than exponential, even if no unstable normal modes are present (e.g., Farrell and Ioannou 1993).

d. Relationship between SVs and Lyapunov vectors

Some studies in low-dimensional dynamical systems have shown that all perturbations, including all singular vectors, evolve towards the leading Lyapunov vectors, which is the attractor in the atmosphere dynamic system (e.g., Trevisan and Legnani 1995, Kalnay 2003).

The global Lyapunov exponents describe the long-term average exponential rate of stretching (or contracting) in the attractor:

$$\lambda_i = \lim_{s \to \bullet} \frac{1}{s} \ln[\sigma_i(t_0 + s)] \tag{24}$$

where $\sigma_i(t_0 + s)$ are the singular values of the linear operator at a finite interval s.

While dealing with atmospheric predictions, we are more interested in the local stability properties of perturbations at a given time and space. The leading local Lyapunov vector (LLV) can be estimated at time t:

$$l_1(t) = \lim_{s \to \bullet} L(t - s, t) y(t - s) \tag{25}$$

And the leading local Lyapunov exponent can be obtained from the rate of change of its norm over a finite period τ:

$$ll_1 \approx \frac{1}{\tau} \ln \left[\frac{\| l_1(t + \tau) \|}{\| l_1(t) \|} \right] \tag{26}$$

In general, there are three differences between SVs and LLVs: 1) SVs depends on the definition of the norm, but LLVs are independent, so LLVs are a fundamental characteristic of dynamic systems; 2) SVs grow much faster than the leading LLVs; 3) SVs are initially off the attractor, but the first few LLVs of low-dimension dynamic systems can span the attractor (e.g., Kalnay 2003).

5. Singular vectors for ensemble forecasting

a. Motivation of ensemble forecasting based on the SV method

The fundamental goal of ensemble forecasting is to produce a forecast PDF of possible future states of the atmosphere from which the true state is consistently a random sample. One way to build an ensemble prediction system (EPS) is by running a NWP model starting from an analysis PDF (namely, initial perturbations). To obtain a good estimate of the forecast PDF, the analysis PDF should represent our uncertainty in the atmosphere's true state and its attractor. However in practice, the initial PDF is only poorly known. Theoretically, for an ideal ensemble with an infinite number of members, it can be proved that truth must show up in the bounds defined by approximate forecast PDF. But, under the limit of computational cost, it is impossible to build an ensemble forecasting system with a huge number of members at present.

The scientists at ECWMF followed the idea of Lorenz (1965), who proposed that "optimal perturbations" that grow the fastest in the short-range are revealed by the largest eigenvalues of the eigenvectors of a symmetric matrix, to create initial perturbations for their EPS by using the SVs approach. The most important benefit resulting from this method is that it allows us to develop a reasonable ensemble based on only limited members since choosing the fastest growing modes should ensure that the true evolution of the atmosphere is consistently portrayed. In addition to these, there are three reasons why the ECMWF medium-range EPS exploits SV perturbations:

i) The sensitivity of day-2 forecast error to perturbations in the initial state projects well into the space of dominant SVs (Rabier et al. 1996).

ii) The evolved SVs are lying in the direction the largest eigenvectors of the forecast-error covariance matrix if the metric is an accurate reflection of the analysis-error covariance matrix (Ehrendorfer and Tribbia 1997, hereafter referred to as ET). It is found that less than 15% of the total number of SVs is needed to recover more than 95% of the total forecast-error variance in ET.

iii) SV perturbations may provide a relatively efficient means of sampling the forecast error PDF in the weakly nonlinear range (e.g., Gelaro et al. 1998).

b. Medium-range ensemble prediction system

So far, there have been three spectral models running operationally at ECMWF (Buizza et al. 1998): 1) T42L31; 2) T_l159L31; and 3) T213L31. By using the tangent linear model for the T42L31 model over a 48-hour optimization interval, the 25 twin pair SV perturbations that form the ECMWF's SV ensemble are obtained as follows [see further details in Buizza et al. (1998) and Molteni et al. (1996)].

First, 25 SVs (only for the Northern Hemisphere), V_j; j=1, 25, are calculated over the Northern Hemisphere by either the Lanczos algorithm or the Jacobi-Davidson algorithm, depending on which initial norm, total energy, or the Hessian of the 3-dimensional variational data assimilation (3DVAR) objective function, is selected. The projection operator P is performed so that only perturbations north of 30^0 (or south of 30^0 for Southern Hemisphere) are included in the EPS. The first four SVs are always selected. Each subsequent SV (from the 5th onwards) is selected only if more than half of its total energy is lying outside the localized regions of the SVs already chosen.

Second, once SVs have been selected, an orthogonal rotation in phase space and re-scaling are applied to generate final perturbations for ensemble forecasts. Rotation processing can ensure that the resulting perturbations P_j have the same globally-averaged energy as the 'original' singular vectors, but smaller local maxima and more uniform spatial distribution. There are two reasons for re-scaling: 1) make initial amplitudes of perturbations similar to analysis-error estimates; and 2) limit the ensemble standard deviation to be comparable to the estimated error of the ensemble mean. In practice, a scaling factor R_n is chosen by experimentation (e.g., R_n=0.6) and then the constants α_{jk} are chosen such that $P_j = \sum_{k=1}^{25} \alpha_{jk} v_k$ and $\|P_j\| \le R_n \|a_e\|$ (where v_k and a_e represent SVs and the approximated analysis error, respectively).

Third, perturbations are added to and subtracted from the control initial conditions, the T42L31 analysis value, to create 50 perturbed initial conditions. Following the steps above, a second set of perturbations are obtained for the Southern Hemisphere.

Finally, after interpolating from T42L31 to T_l159L31 resolution, the ensemble forecast is run up to 10 days with the nonlinear model. In the meantime, a control forecast is performed at either T42L31 (low-resolution) or T213L31 (high-resolution).

With all these advantages of the SV approach in the EPS, there are three notable problems. First, it is computationally expensive to find the optimal perturbations even thought we just pick up a few leading SVs. ECMWF has to run the tangent linear model and its adjoint about three times the number of SVs required. For this reason, the computation of the dominant SVs is done with a lower resolution than the operational model. Second, this method is applied to maximize the linear error growth for a 48-hour optimization interval. Although these SVs represent directions that have grown optimally in the last 48 hours, it is not guaranteed that the optimal perturbations at 48 hours will continue to be the fastest growing modes into the medium range (10 days). Third, SVs are designed to sample the extremes of the analysis PDF instead of providing a purely random sampling. In fact, not all the analysis errors project onto growing modes, so the sampling is limited. This may be one of the reasons why the ensemble spread is lower than desired.

c. Short-range to early-medium range ensemble prediction system

Ensemble prediction systems, which are optimized for the large-scale flow in the medium range, work fairly well at both ECMWF and NCEP (breeding method). In the short-range, the scale and weather parameters of interest are less predictable so their errors may saturate too quickly for an ensemble to be of use. Also the influence of model uncertainty is more significant in the short-range than in the medium-range. If simply using the medium-range EPS for a short-range forecast, such systems might not be optimal for less predictable weather parameters in that range. Currently, short-range ensemble forecast (SREF) is given quite some attention and the value of SREF remains an open question. There are a few ways to design a SREF, such as the multi-regional model (e.g., Du and Tracton 2001), multi-analysis method, the poor man's technique (e.g., Atger 1999), the analysis-centroid mirroring system (Eckel 2003), the targeted ensemble prediction system (TEPS) (Hersbach et al. 2003). Here I introduce only TEPS since it is the only one to be designed with SVs.

The goal of the TEPS, similar to the EPS for the medium-range at ECMWF, is to address the predictability of short-range to early medium-range (up to day 5) weather parameters (Hersbach et al. 2000) by using targeted singular vectors (TSVs) to generate initial perturbations. The total energy is still used to define the initial norm, but this time the projection operator *P* is designed to only obtain perturbations on a specific area (e.g., the European domain), instead of the entire extratropics. The smaller area makes it possible to increase the number of directions that is associated with the weather parameters of interest without increasing the number of ensemble members. Hersbach et al (2003) compared the TSVs to the SVs of the medium-range EPS and found that the first five leading TSVs were very well represented by a set of 25 SVs of the medium-range EPS while the slower growing TSVs were poorly described. However these slower growing TSVs do contain a substantial part of the "explained variance" of the 48 h forecast error above the interest area. For the short-range (3 days), an optimization time of 12 h was used instead of 48 h. The SVs obtained for a 12-hour optimization interval were argued to lead to acceptable perturbations even if they might not present the optimal solution. Also, compared to medium-range EPS, ensemble spread calculated from this short-range ensemble system was increased, having a beneficial effect on statistical properties. Although the SV approach has shown some good results in case studies of the short-range ensemble forecast, more research will be needed to address existing problems.

6. Summary

Based on the discussion above, we believe that SVs are a powerful mathematical tool that has been successfully applied to many fields, including the atmospheric-predictability problem. The dominant singular vectors describe the most rapidly growing structures with respect to a given metric (norm) over a certain optimization interval in a tangent linear sense (Gelaro et al. 1998). The noticeable characteristic is that SVs are dependent on the norm, projection operator and optimization interval. The total energy is the appropriate metric for the predictability since it may be a reasonable first-order estimate of the analysis-error covariance metric. For a medium-range ensemble forecast, the projection operator is designed to cover perturbations on the entire extratropics. However, uncertainties in a small area, which are obtained by the use of a projection operator, show a positive effect on a SREF. Also, based on previous studies, a short optimization interval has been thought to be a better selection than a longer interval when it comes to issues in a SREF (e.g., Hersbach et al. 2003). The nonnormality of SVs allows for the possibility of finite-time growth faster than exponential.

The conclusion that SVs represent the directions of most rapid error growth provides the main rationale for their use in the application of ensemble numerical weather prediction.

References

Atger, F. 1999: The skill of Ensemble Prediction Systems. *Mon. Weather Rev.,* **127,** 1941-1953.

Barkmeijer, J., R. Buizza, and T. N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.,* **125,** 2333-2351.

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

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

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distribution of the ECMWF ensemble prediction system. *Mon. Wea. Rev.,* **125,** 99-119.

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

Buizza R., 1994: Sensitivity of optimal unstable structures. *Quart. J. Roy. Meteor. Soc.,* **120,** 429-451.

Charney, J. G., and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics, J. Lighthill and R. Pearce, Eds.,* Cambridge University Press, 735pp.

Daley, R. and T. Mayer, 1986: Estimates of global analysis error from the global weather experiment observational network. *Mon. Wea. Rev.,* **114**,1642-1653.

Du, J. and Tracton, M. S. 2001: 'Implementation of a real-time short-range ensemble forecasting system at NCEP: An update'. Pp. 335-356 in Preprints of the ninth conference on mesoscale processes, 30 July-2 August 2001, Fort Lauderdale, Florida. *American Meteorological Society,* 45 Beacon Street, Boston MA02108-3693, USA

Eckel, F. A. 2003: Toward an effective short-range ensemble forecasting system. Ph.D. thesis.

Ehrendorfer, M., and R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. *J. Atmos. Sci.,* **52.** 3475-3500.

Ehrendorfer, M., and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.,* **54,** 286-313.

Errico, R. and D. Baumhefner, 1987: Predictability experiments using a high-resolution limited-area model. *Mon. Wea. Rev.,* **113**, 488-504.

Farrell, B. F. and P. J. Ioannou, 1996: Generalised stability theory. Part I: Autonomous operators. *J. Atmos. Sci.,* **53,** 2025-2040.

Farrell, B. F. and P. J. Ioannou, 1993: Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.,* **50,** 4044-4057.

Farrell, B. F., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.,* **45,** 163-172.

Farrell, B. F., 1982: The growth of disturbances in a barolinic flow. *J. Atmos. Sci.,* **39,** 1663-1686.

Fisher, M., and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Tech. Memo. 220, 28 pp. [Available from Research Department, ECMWF Shinfield Park, Reading RG29AX, United Kingdom.]

Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky, 2000: Model consensus. *Wea. Forecasting* **15,** 571-582.

Hamill, T. M., S. L. Mullen, C. Snyder, Z. Toth, D. and P. Baumhefner, 2000: Ensemble Forecasting in the Short to Medium Range: Report from a Workshop. *Bull. Amer. Meteor. Soc.,* **81**, 2653–2664.

Hartmann, D. L., R. Buizza, and T. N. Palmer, 1995: Singular vectors: the effect of spatial scale on linear growth of disturbances. *J. Atmos. Sci.,* **52,** 3885-3894.

Hersbach, H., R. Mureau, J. D. Opsteegh, and J. Barkmeijer, 2000: A short-range to early-medium range Ensemble Prediction System for the European Area. *Mon. Wea. Rev.,* **128,** 3501-3519.

Hersbach, H., R. Mureau, J. D. Opsteegh, and J. Barkmeijer, 2003: Developments of a Targeted Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.,* **129,** 2027-2048.

Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX'98 ensemble forecasts. *Mon. Wea. Rev.* **129**,73-91.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225-1242.

Kalnay, E., 2003: Atmospheric predictability and ensemble forecasting. Chapter 6 in *Atmospheric modeling, Data assimilation and predictability*, Cambridge University press, Cambridge.

Krishnamurti, T. N., and Coauthors, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science,* **285,** 1548-1550.

Leith, C. E., 1978: Objective methods for weather prediction. *Ann. Rev. Fluid Mechanics*, **10,** 107-128.

Leith, C. E., 1974: Theoretical skill of monte carlo forecasts. *Mon. Wea. Rew.,* **102**, 409-418.

Leutbecher, M., 2003: A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing network. *J. Atmos. Sci.,* **60,** 729–742.

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

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130-141.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73-120.

Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.,* **119,** 269-298.

Mullen, S. L. and D. P. Baumhefner, 1989: The impact of initial condition uncertainty on numerical simulations of large-scale explosive cyclogenesis. *Mon. Wea. Rev.,* **117**,2800-2821.

Mureau, F., F. Molteni, and T. M., Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.,* **119,** 299-323.

Orr, W. McF., 1907: Stability or instability of the steady motions of a perfect liquid. *Proc. Roy. Irish Acad.,* **27**, 9-69.

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

Palmer, T. N., 1996: Predictability of the atmosphere and oceans: From days to decades. *Decadal Climate Variability-Dynamics and Predictability*, NATO ASI Series, Subseries I, *Global Environmental Change,* Vol. 44, D. Anderson and J. Willebrand, Eds., Springer, 83-155.

Palmer, T. N., 1994: Chaos and predictability in forecasting the monsoons. *Proc. Indian Natn. Sci. Acad.,* **60A**, 57-66.

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

Sleijpen and van der Vorst, 1995: A Jacobi-Davidson iteration method for linear eigenvalue problems. Universiteit Utrecht, Department of Mathematics, preprint 856.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 1998: Ensemble forecasting of mesoscale convective systems. Preprints, 12th Conf. on Numerical Weather Prediction, Phoenix, AZ, *Amer. Meteor. Soc.,* 265-268.

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

Trevisan, A., and R. Legnani, 1995: Transient error growth and local predictability: a study in the Lorenz system. *Tellus* **47A,** 103-117.

Tribbia, J. and D. Baumhefner, 1988: The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. *Mon. Wea. Rev.,* **116,** 2276-2288.

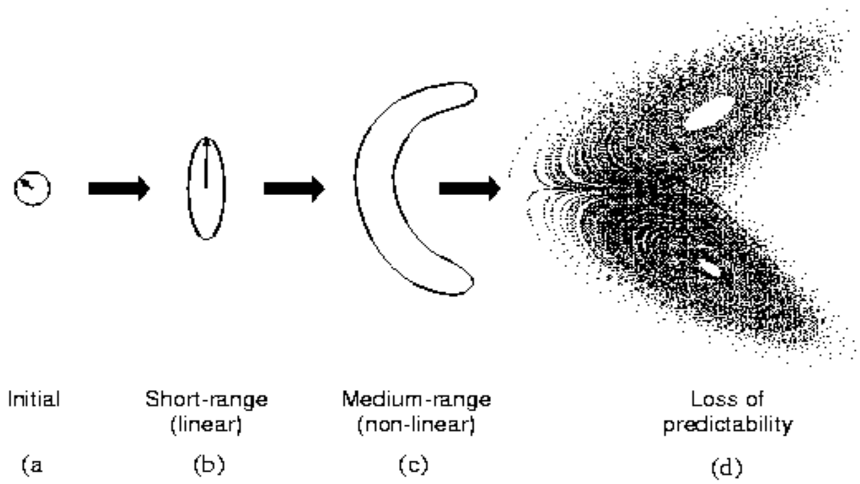

Figure 1: A schematic illustration of the growth of an isopleth of the forecast error probability distribution function, from (a) initial phase,to (b) linear growth phase, to (c) nonlinear growth phase to (d) loss of predictability (from ECMWF website http://www.ecmwf.int/research/ifsdocs_old/ENSEMBLE/index.html)