## 6.5 Operational ensemble forecasting methods

Ensemble forecasting methods differ mostly by the way the initial perturbations are generated, and can be classified into essentially two classes. In the first class, which we can denote "Monte Carlo forecasting" (MCF), the initial perturbations are chosen to be "realistic", i.e., they have horizontal and vertical structures similar to forecast errors, and amplitudes compatible with the estimated analysis uncertainty[1]. However, they are chosen randomly, without regard to the "dynamics of the day".

Errico (1987) and Mullen and Baumhefner (1994) developed a Monte Carlo method that results in very realistic perturbations compatible with the average estimated analysis errors. However, by construction, this type of Monte Carlo forecast does not include finite size "growing errors of the day" which are almost certainly present in the analysis. The experiments of Hollingsworth (1980), Hoffman and Kalnay (1983), and Kalnay and Toth (1996) suggest that random initial perturbations do not grow as fast as the real analysis errors, even if they are in quasi-geostrophic balance.

Recently, a second class of methods was developed, tested and implemented at several operational centers. They are known as the "breeding" method and the "singular vector" (or optimal perturbations) method. In contrast to MCF, they are characterized by the inclusion in the initial perturbations growing errors that depend on the evolving underlying atmospheric flow. Two other methods that are also very promising are based on ensembles of data assimilations (Canada) , and ensembles based on operational systems from different centers, combining different models and data assimilations.

---

[1] The analysis uncertainty depends on the distribution of the observations. It can be estimated from the analysis error covariance (Chapter 5), which depends on the accuracy of the statistical assumptions, or empirically, from the rms differences between two independent analysis cycles (Fig. 6.12). It is available from EnKF!
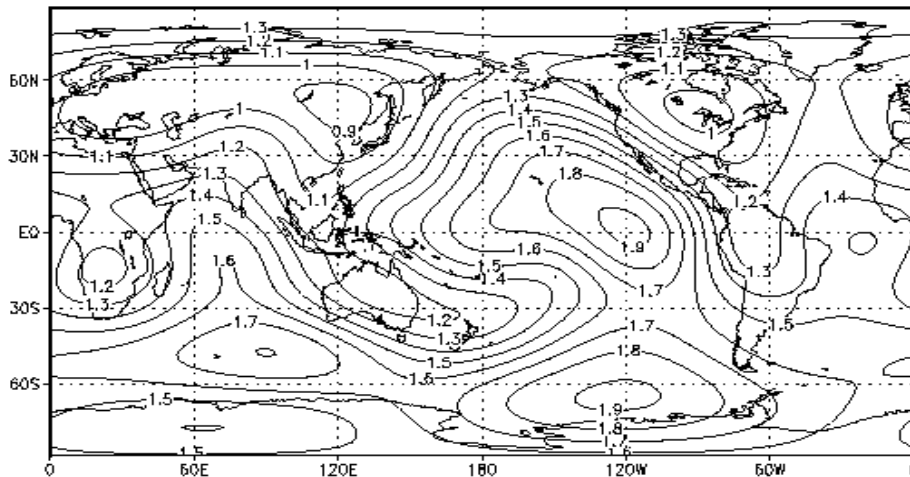
Fig. 6.12: Estimation of the 500hPa geopotential height analysis uncertainty obtained from running two independent analysis cycles, computing their rms difference, and using a filter to retain the planetary scales. The units are arbitrary. Note the minima over and downstream of rawinsonde rich land regions and the maxima over the oceans (Courtesy I. Szunyogh, U of Maryland).

6.5.1 Breeding

Ensemble experiments performed at NCEP during 1991 showed that initial ensemble perturbations based on LAF, SLAF and on the forecast differences (FD) between forecasts verifying at the same initial time, grew much faster than Monte Carlo perturbations with the same overall size and statistical distribution (Kalnay and Toth, 1996). It was apparent that the differences in growth rate were due to the fact that the first group included perturbations that, by construction, "knew" about the evolving underlying dynamics.

Toth and Kalnay (1993, 1996, 1997) created a special operational cycle designed to "breed" fast growing "errors of the day" (Fig. 6.13). Given an evolving atmospheric flow (either a series of atmospheric analyses, or a long model run), a breeding cycle is started by introducing a random initial perturbation ("random seed") with a given initial size (measured with any norm, such as the rms of the geopotential height or the kinetic energy). It should be noted that the random seed is introduced only once.

The same nonlinear model is integrated from the control and from the perturbed initial conditions. From then on, at fixed time intervals (e.g., every 6 hours or every 24 hours), the control forecast is subtracted from the perturbed forecast.

The difference is scaled down so that it has the same amplitude (defined using the same arbitrary norm) as the initial perturbation, and then added to the corresponding new analysis or model state.

It was found that beyond an initial transient period of 3-4 days after random perturbations were introduced, the perturbations generated in the breeding cycle (denoted Bred Vectors or BVs), acquired a large growth rate, faster than the growth rate for MCF or even SLAF or FD.

Toth and Kalnay (1993, 1997) also found that (after the transient period of 3-4 days) the shape or structure of the perturbation BVs did not depend on either the norm used for the rescaling or the length of the scaling period. The BVs did depend on the initial random seed in the sense that regional BV perturbations would have the same shape but different signs, and that in many areas two or more "competing BVs" appeared in cycles originated from different random seeds.

The breeding method is a *nonlinear generalization of the method used to construct Lyapunov vectors* (performing two nonlinear integrations and obtaining the approximately linear perturbation from their difference). Therefore the BVs are related to LVs localized in both space and time, and it is not surprising that they share their lack of dependence on the norm or on the scaling period.

Toth and Kalnay (1993) have argued that the analysis cycle also contains errors that project strongly on the local LVs because they are evolved in time through the forecast used as background, and they are only partially corrected through the use of noisy data (Fig. 6.14). Fig. 6.15 shows a schematic of how breeding cycles are self-propagated from the ensemble forecasts, without any additional cost.

Fig. 6.16 presents a comparison of bred vectors and background errors for a quasi-geostrophic model data assimilation system developed by Morss et al, 2001. The fact that there is a strong resemblance between the errors of the forecast used as a first guess and the bred vectors valid at the same time suggest that the forecast and analysis errors do project strongly on the bred vectors.

Fig. 6.13: Schematic of a breeding cycle run upon an unperturbed (control) model integration. The initial growth after introducing a random initial perturbation is usually very small, but with time, the perturbation is more dominated by growing errors. The initial transient with slow growth lasts about 3 days. The difference of the complete perturbed and control forecasts is scaled back periodically (e.g., every 6 or every 24 hours) to the initial amplitude. The rescaling is done by dividing all differences by the same observed growth (typically about 1.5/day for mid-latitudes). In operational NWP, the unperturbed model integration is substituted by short-range control forecasts started from consecutive analysis fields. The breeding cycle is a nonlinear generalization of the method used to obtain the leading Lyapunov vector. (Adapted from Kalnay and Toth, 1996).

Forecast
values

Initial random
perturbation
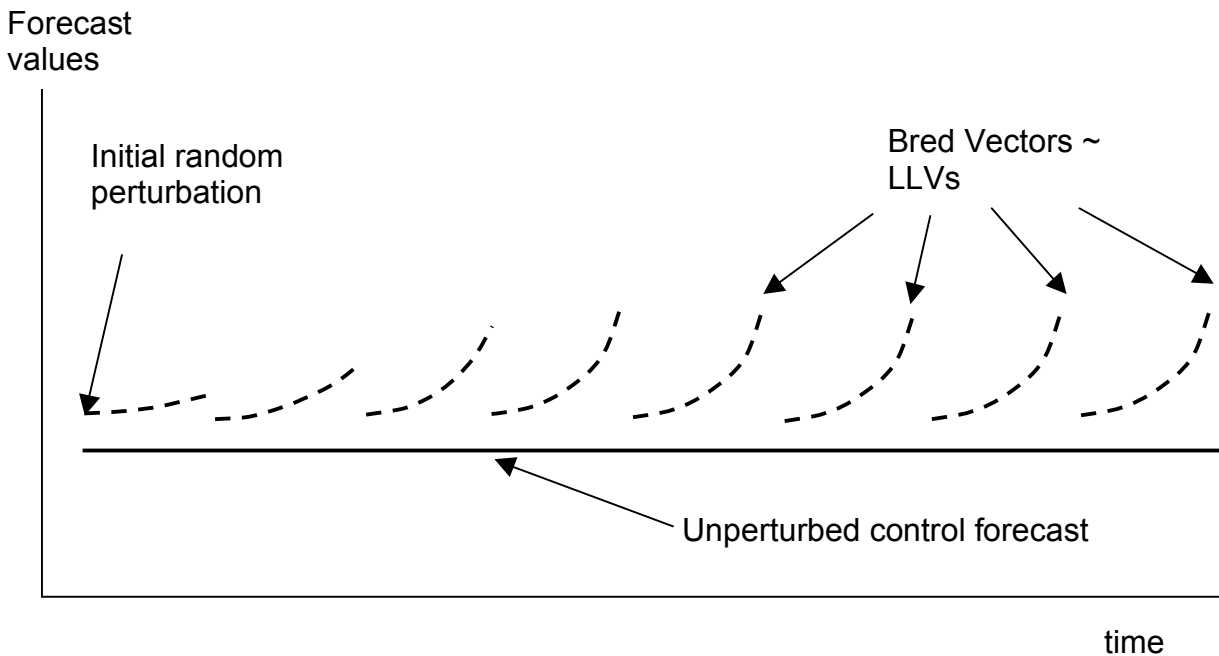
Bred Vectors ~
LLVs

Unperturbed control forecast

time

Fig. 6.14: Schematic of the 6-hour analysis cycle. Indicated on the vertical axis are differences between the true state of the atmosphere (or its observational measurements, burdened with observational errors). The difference between the forecast and the true atmosphere (or to the observations) increases with time in the 6 hour forecast because of the presence of growing errors in the analysis (adapted from Kalnay and Toth, 1996).

Analysis and
forecasts

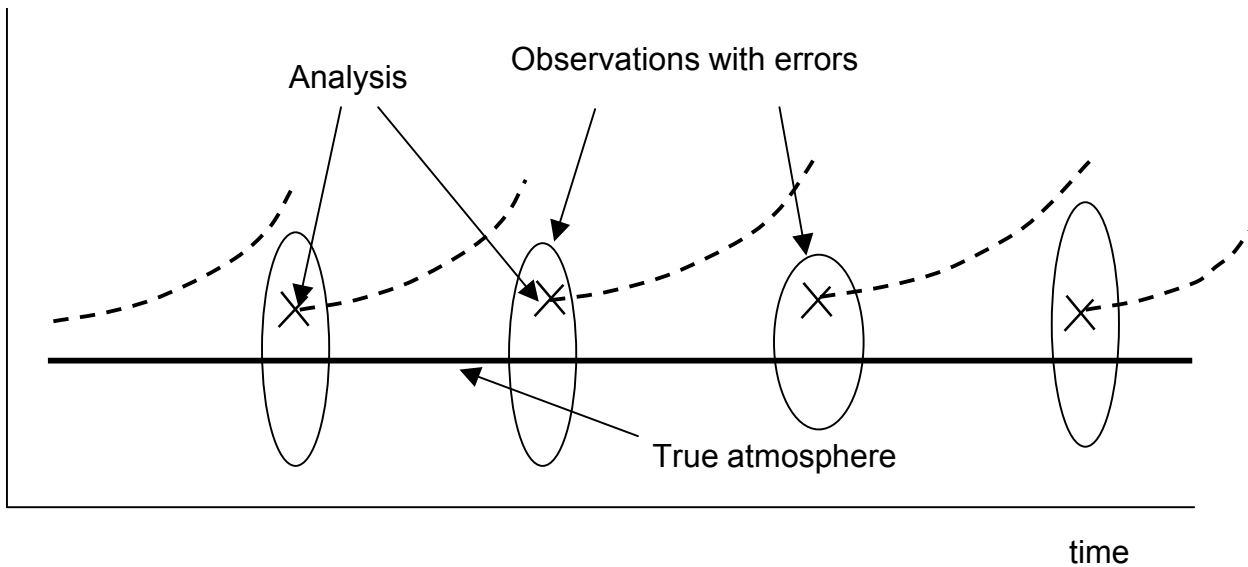Analysis   Observations with errors

True atmosphere

time

Fig. 6.15: Schematic of a self-contained breeding pair of ensemble forecasts. Every day, the one-day forecast from the negative perturbation is subtracted from the one-day forecast from the positive perturbation. This difference is divided by two, and then scaled down (by dividing all variables by the one-day growth, so that difference has the same size as the initial perturbation. The scaled difference is then added and subtracted from the new analysis, generating the initial conditions for the new pair of forecasts. This self-breeding is part of the extended ensemble forecast system, and does not require computer resources to generate initial perturbations beyond running the ensemble forecasts. (Adapted from Toth and Kalnay, 1997).
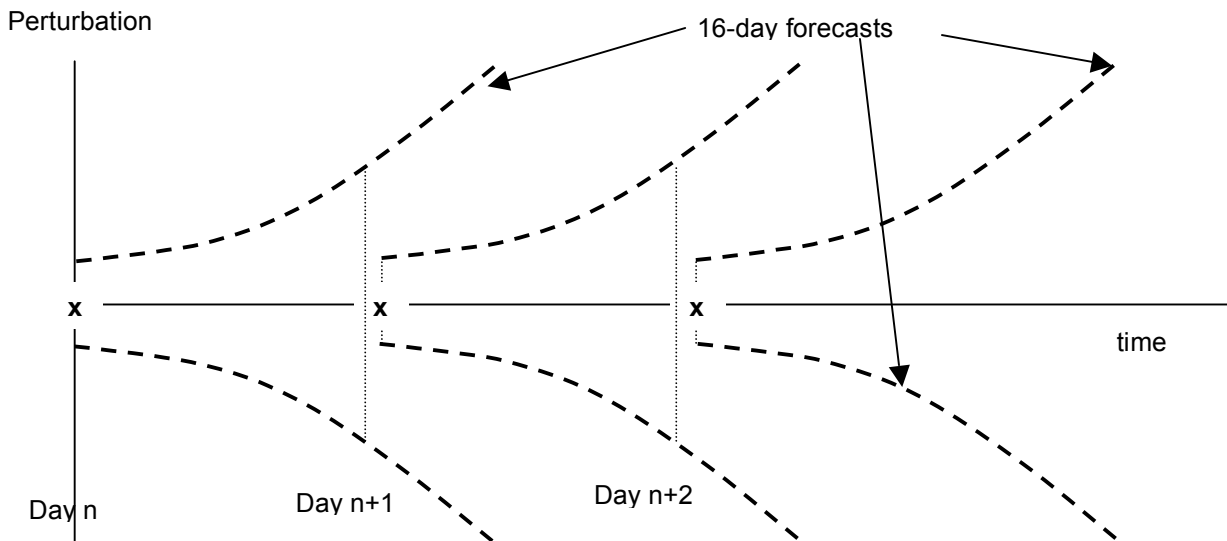
Fig. 6.16: Comparison between the 12 hour forecast error used as background (contours) and a randomly chosen bred vector for a QG data assimilation simulation system. First image at the center level of the model. Second is a vertical cross-section (from Corazza et al, 2001).
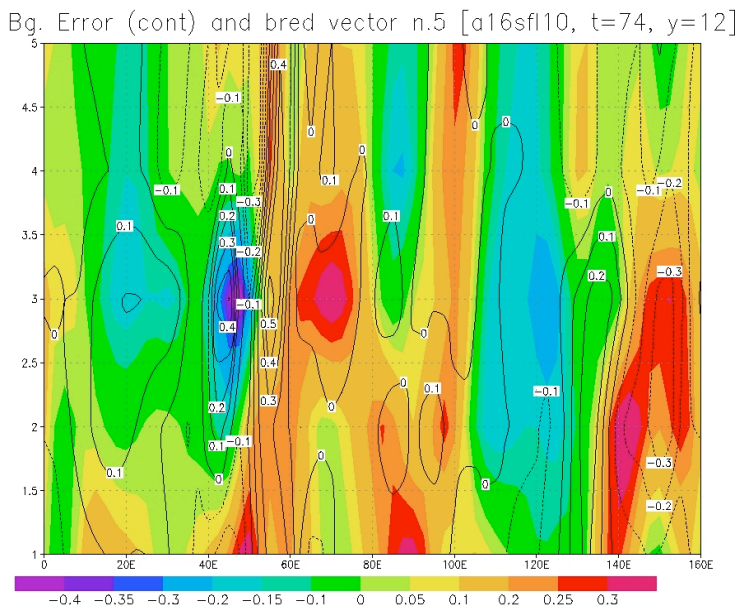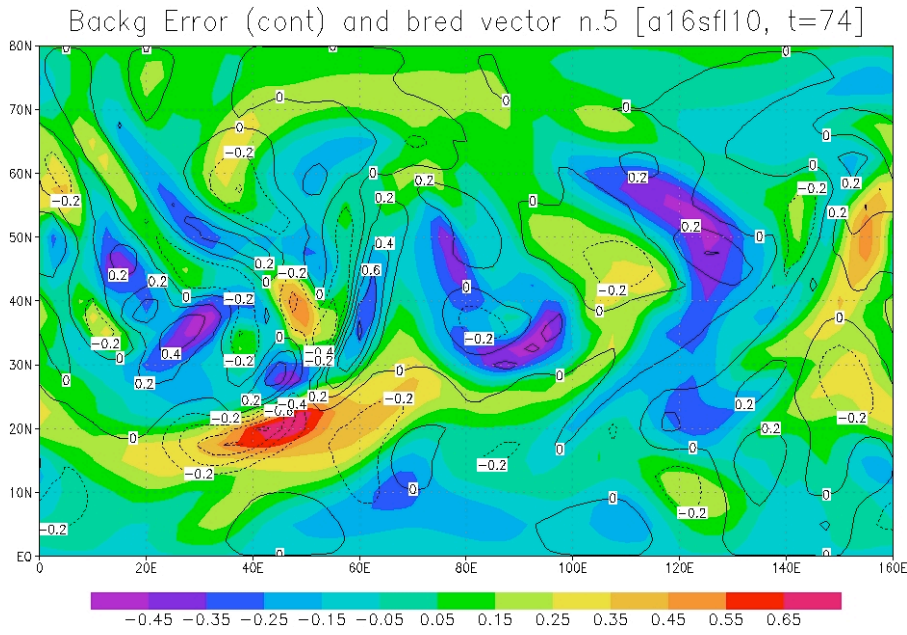
Fig. 6.17: Examples of bred vectors (500 hPa geopotential height field differences, without plotting the zero contour) from the NCEP operational ensemble system valid at 5 March 2000. a) Bred vector 1; b) Bred vector 5. Note that over large parts of the eastern Pacific Ocean and western North America, the two perturbations are very similar but of opposite signs. In other areas the shape of the perturbations is quite different. c) The BV-local dimension of the 5 perturbations subspace (Patil et al, 2001). Only dimensions less or equal than 3 are contoured with a contour interval o 0.25. (Courtesy of DJ Patil).
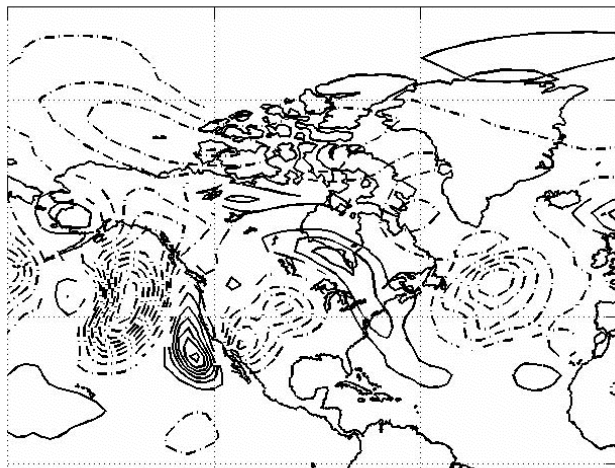
b) Bred
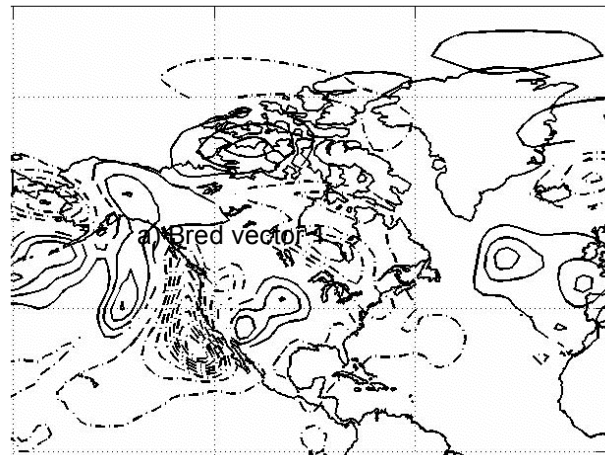
b) Bred
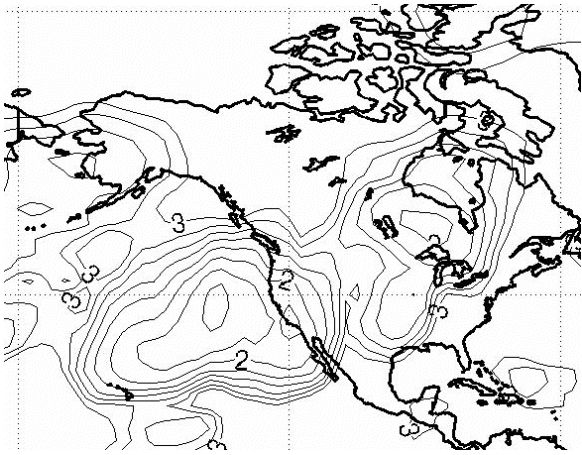
c) Local E-dimension
(contour interval. 0.25)

Figures 6.17 (a, b) show 2 out of 5 bred perturbations corresponding to March 5 2000 at 00UTC. Figure 6.17 (c) presents an estimate of the local dimension of the subspace of the 5 perturbations using (Patil et al, 2001). Only the areas where the local dimension has collapsed from the original 5 independent directions to 3 or less are contoured. Note that there are areas where the independently bred vectors aligned themselves into a smaller subspace. The collapse of the perturbations into fewer dimensions is what one could expect if there are locally growing dominant Lyapunov Vectors. These low-dimensional areas are organized into horizontal and vertical structures and have a lifetime of 4-7 days, similar to that of baroclinic developments.

The breeding ensemble forecasting system was introduced operationally in December 1992 at NCEP, with two pairs of bred vectors (Tracton and Kalnay, 1993). In 1994, seven pairs of self-breeding cycles replaced the original four perturbed forecasts. In addition, a regional rescaling was introduced that allowed larger perturbation amplitudes over ocean than over land proportionally to the estimate of the analysis uncertainty, Fig. 6.12 (Toth and Kalnay, 1997).

TK 93 found that when the initial amplitude was chosen to be within the range of estimated analysis errors (i.e., between 1m and 15m for the 500 hPa geopotential height) the BVs developed most strongly generally in strong baroclinic areas. Their horizontal scale was that of short baroclinic waves, and their hemispherical average growth rate was about 1.5-1.6/day (similar to the estimated growth of analysis errors). However, if the initial amplitude was chosen to be much smaller than the small (10 cm or less), then a different type of BV appeared, associated with convective instabilities, which grew much faster than baroclinic instabilities (at a rate of more than 5/day), but saturated at an amplitude much smaller than the analysis error range (Fig. 6.18).
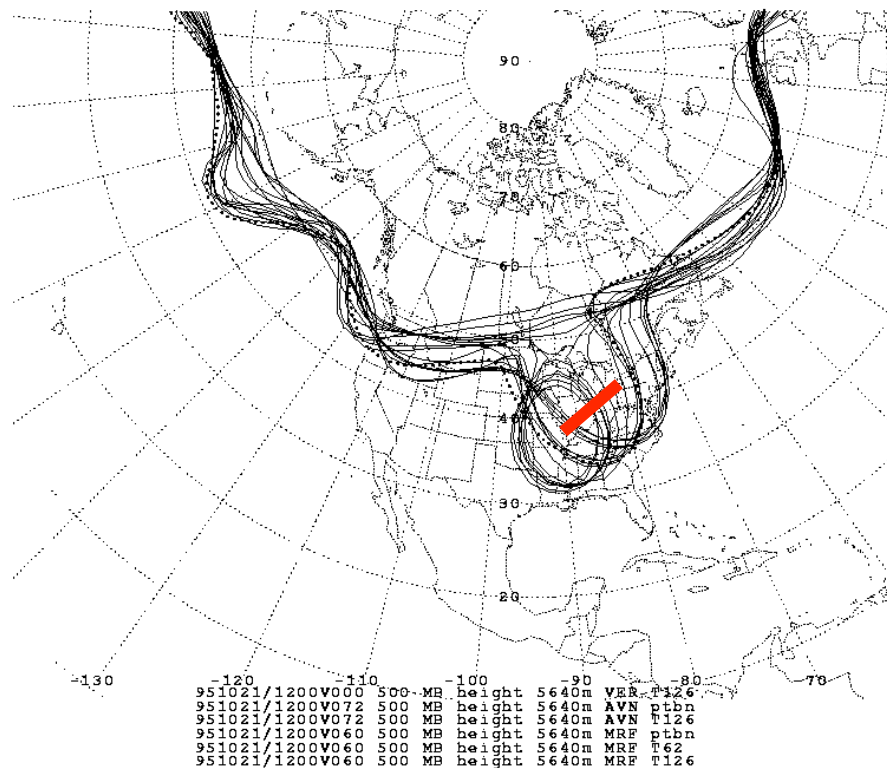
TK93 suggested that the use of nonlinear perturbations in breeding has the advantage of filtering Lyapunov vectors associated with fast

growing but energetically irrelevant instabilities, like convection. This was confirmed by Lorenz (1996), who performed experiments with a low order model containing large amplitude but slowly growing modes coupled with fast growing modes with small amplitude that made them irrelevant. Lorenz found that the use of breeding using finite amplitudes yielded the Lyapunov vectors of the large amplitude, slowly growing vectors, as desired, whereas for very small amplitudes the Lyapunov vectors of the fast system were recovered. The nonlinear saturation of irrelevant fast growing modes is an advantage that suggests the use of breeding for other problems. For example, in the coupled ocean-atmosphere system the slower growing (but very large energy amplitude) coupled ENSO instabilities could perhaps be captured, while eliminating through nonlinear saturation the irrelevant details of weather perturbations.

Fig. 6.18: Schematic of the time evolution of the rms amplitude of high energy baroclinic modes and low-energy convective modes. Note that although initially growing much faster than the baroclinic modes, convective modes saturate at a substantially lower level. These modes are therefore insignificant in the analysis/ensemble perturbation problem, since the errors in the analysis (dashed line) are much larger than the convective saturation level. (Adapted from Toth and Kalnay, 1993).
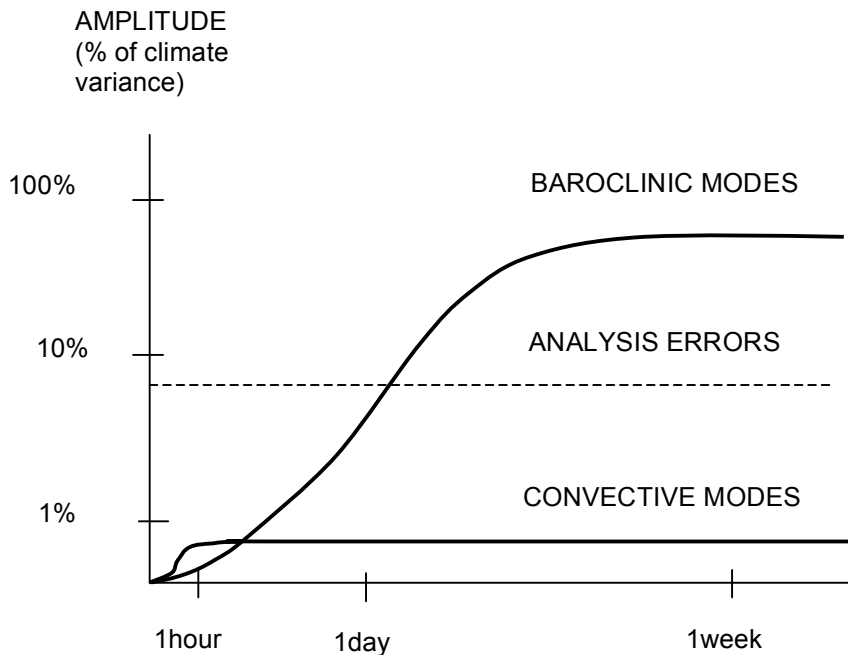
Figs. 1.10a,b (Chapter 1) show two examples of one of the ways information on ensemble forecasts are presented to the users, the "spaghetti plots", or plots showing one contour lines for each forecast. In one case, a 5-day forecast verifying on November 15 1995, the agreement in intensity and location of the contours indicated to the forecasters that this was a very predictable snow storm. In the second case, a 2.5 day forecast verifying on 21 October 1995, the ensemble members show unusually strong divergence in the location of a winter storm, warning the human forecasters that this situation is intrinsically unpredictable. Note that although the ensemble forecasts show a wide divergence in phase, this is also a case in which, in perturbation space, there is a single dimension, since the perturbations align themselves along the same basic shape.

951021/1200V000 500 MB height 5640m VER T126
951021/1200V072 500 MB height 5640m AVN ptbn
951021/1200V072 500 MB height 5640m AVN T126
951021/1200V060 500 MB height 5640m MRF ptbn
951021/1200V060 500 MB height 5640m MRF T62
951021/1200V060 500 MB height 5640m MRF T126

The second case points out the potential value of ensembles in a new area of research: **targeted observations**. In cases like this it should be possible to find the area that originated this region of uncertainty in time to launch new observations for the next analysis cycle, and decrease significantly the forecast error. Finding the area where the observations should be launched can be done through several approaches recently developed. They are the adjoint sensitivity, the use of singular vectors (Rabier et al, 1996, Langland et al, 1995, Pu et al, 1998, and others), the linear inverse of the TLM (Pu et al, 1998), and Ensemble-based Singular Value Decomposition (Bishop and Toth, 1998). These methods were tested during FASTEX (Jan-Feb 1997 in the Atlantic) and NORPEX (winter of 1997-1998 in the North Atlantic, Langland et al, 1999, Pu and Kalnay, 2000). The experience in the North Pacific has been so successful that targeted observations are performed routinely over the Gulf of Alaska every winter (Szunyogh et al, 2000).

Fig. 6.19 shows another example of how the massive information contained in the ensemble forecast can be conveyed to the forecasters. It shows a probabilistic presentation of a 24 hr and a 7-day forecast of precipitation above a threshold of 5mm in 24 hr. The probabilities are simply computed as the percentage of the ensemble members with accumulated precipitation at least as large as the indicated threshold. They both verify on April 6, 2001. Note that the short range forecast has many areas with probabilities equal to zero or above 95%, indicating that all the ensemble members agree that there will be no precipitation or at least 5mm accumulated precipitation. In the 7-day forecast, the areas with maximum probability of precipitation are generally in agreement with the short range forecast, indicating the presence of skill. However, because by this time there are few areas where there is consensus on either rain or no rain among the forecasts, since their solutions have dispersed significantly over a week.

Check: http://www.emc.ncep.noaa.gov/gmb/ens/

Fig. 6.19: Example of a probabilistic forecast of accumulated precipitation greater than 5mm. The probabilities are computed simply as the number of ensemble members with at least the indicated threshold of accumulated precipitation divided by the total number of ensemble forecasts. Both the 24 hr and the 7-day forecast verify on April 6 2001.
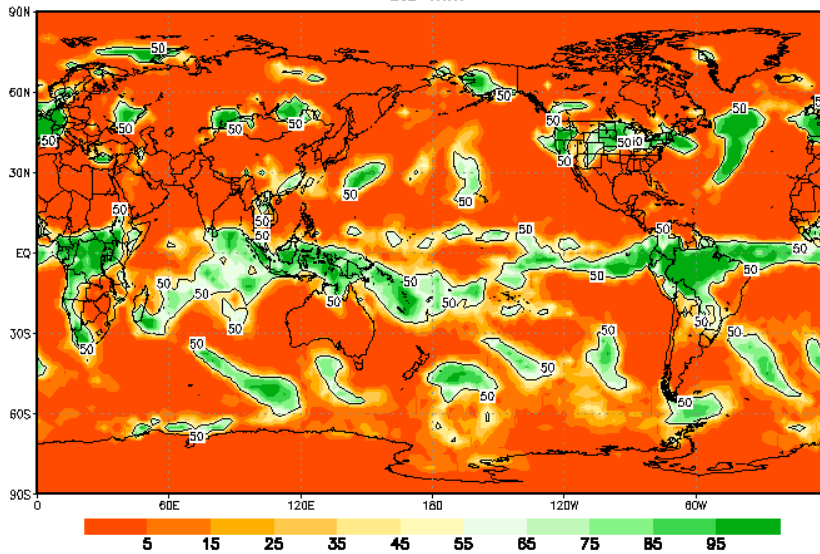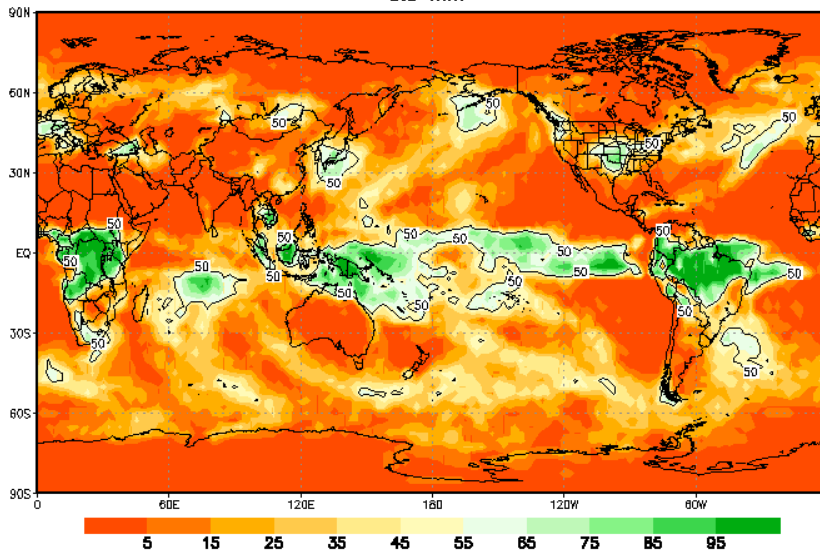
6.5.2 Singular Vectors

(Check http://www.atmos.umd.edu/~ekalnay/PalmerENStm540.pdf for an update on the ECMWF ensemble system)

ECMWF developed and implemented operationally in December 1992 an ensemble forecasting system based on initial perturbations that are linear combinations of the singular vectors of the 48 hr TLM (Molteni et al, 1996, Molteni et al, 1993, Buizza, 1996, Buizza et al, 1997).

As discussed in section 6.3, the singular vectors used to create the initial perturbations are obtained as the leading eigenvectors of

$$(\mathbf{W}^{-1})^T \mathbf{L}^T \mathbf{P}^T \mathbf{P} \mathbf{L} \mathbf{W}^{-1} \hat{\mathbf{y}}(t_0) = \sigma^2 \hat{\mathbf{y}}(t_0) \qquad (1.1)$$

subject to

$$\hat{\mathbf{y}}^T(t_o)\hat{\mathbf{y}}(t_o) = 1 \text{, and } \mathbf{y}(t_o) = \mathbf{W}^{-1}\hat{\mathbf{y}}(t_o). \qquad (1.2)$$

ECMWF originally used as the projection operator $\mathbf{P}$ a symmetric projector operator that includes only forecast perturbations north of 30N, and as the initial norm $\mathbf{W}^{-1}$ the total energy norm. Barkmeijer et al (1998) tested the analysis error covariance as initial norm with good results (the SVs were also closer to bred vectors than with the total energy norm). They also found that the use of evolved vectors (also closer to Lyapunov or bred vectors) resulted in improved results. More recent experiments with a simplified Kalman Filter also resulted in promising results (Fischer et al, 1999).

From (1.1) and (1.2), the initial singular vectors y_i are the perturbations with maximum energy growth north of 30N, for the time interval 0-36 hours (Buizza, 1996), or more recently, 0-48hours. The

method used to obtain the SVs is the Lanczos algorithm (which requires integrating forward with L for a period t, and backward with L^T a number of times about three times the number of singular vectors desired). Fig. 6.20 shows an example corresponding to the initial and final singular vectors #1, 3 and 6. Singular vectors defined with the total energy norm tend to have a maximum initial energy at low levels (about 700 hPa), and their final (evolved) energy at the tropopause level. Fig. 21 shows the corresponding initial and evolved horizontal structure (Buizza, 1996). In 1996 ECMWF used 16 SVs selected among 38 leading SVs, requiring therefore, every day, 3*72*38 hours of integration with either L or L^T. For this reason, the computation was done with a lower resolution (T42/19 level) model. A second set of perturbations was added for the Southern Hemisphere, which originally had no perturbations. The generation of the SH SVs also requires additional computations.

Fig. 6.20: Singular vectors numbers 1 (top panels), 3 (middle panels), and 6 (bottom panels) at initial (left panels) and optimization time (right panels). Each panel shows the singular vector streamfunction at model level 11 (approximately 500 hPa), superimposed to the trajectory 500-hPa geopotential height field. Streamfunction contour interval is 0.5×10^{-8} m^2 s^{-1} for left panels and 20 times larger for the right panels; geopotential height contour interval is 80 m.

951105 PSI - SV= 1   951107 PSI - SV= 1

951105 PSI - SV= 3   951107 PSI - SV= 3

951105 PSI - SV= 6   951107 PSI - SV= 6
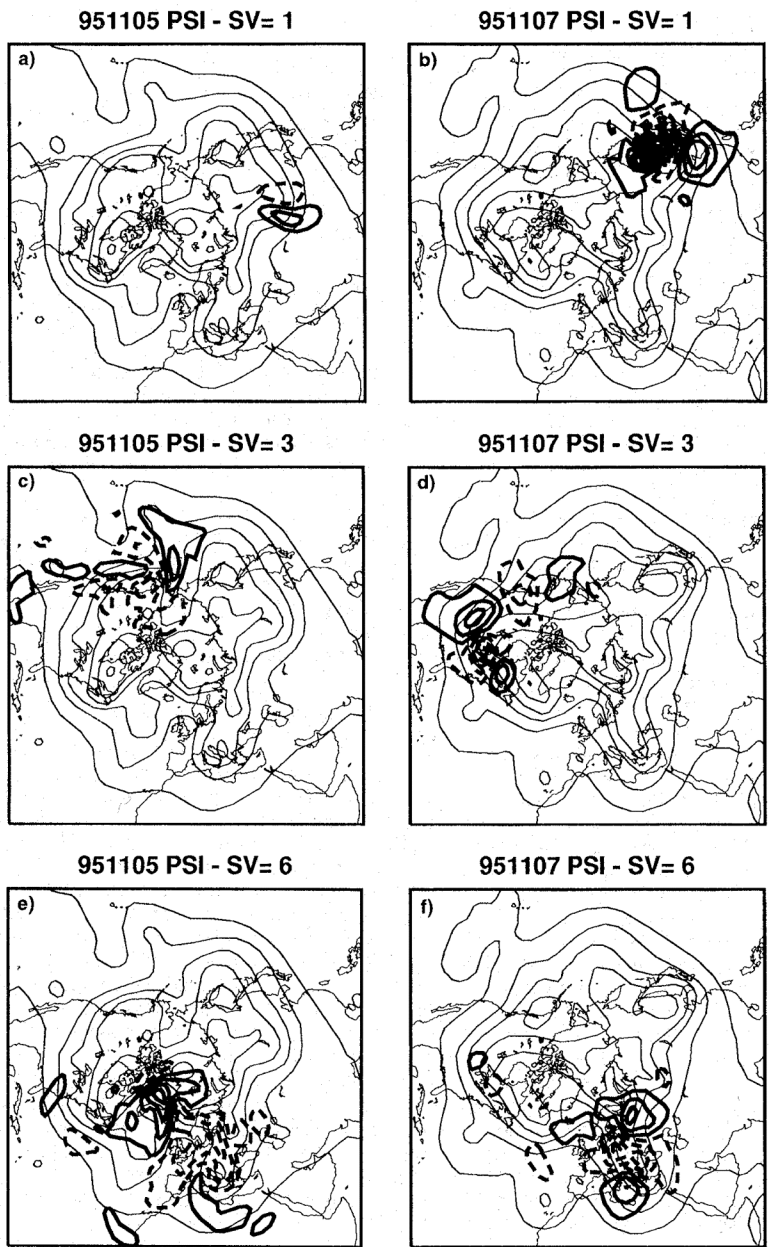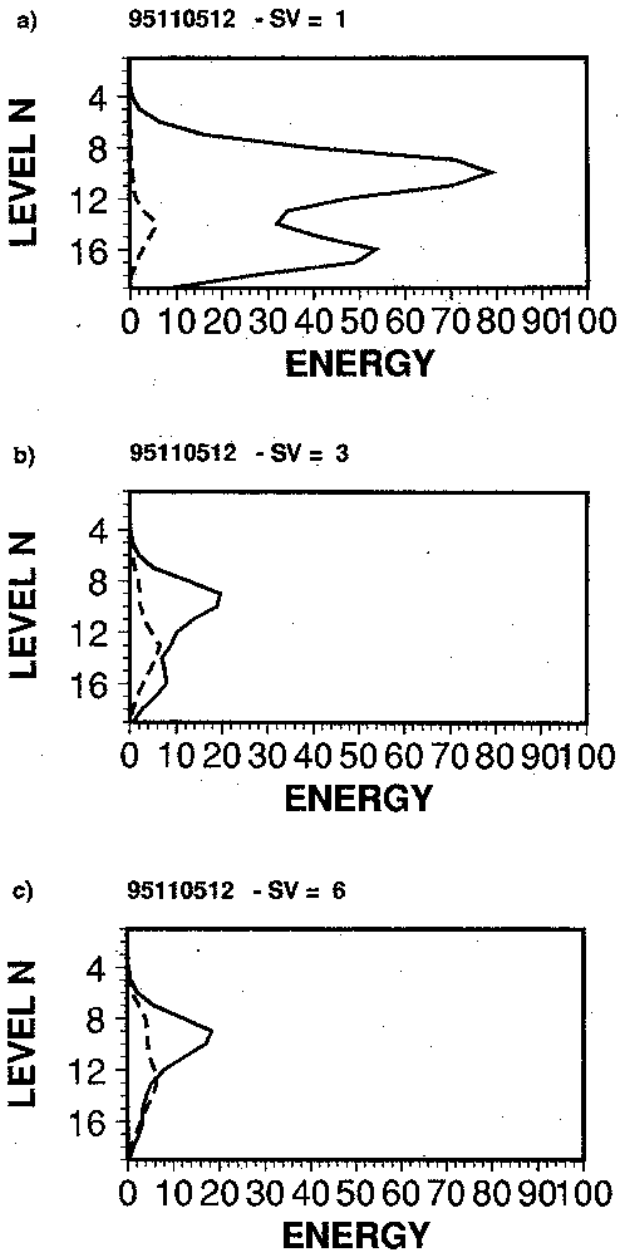
Fig 6.21:. Total energy (m^2 s^{-2}) vertical profile of the (a) first, (b) third, and (c) sixth singular vector of the 5 November 1995, at the initial (dashed line, values multiplied by 100) and optimization (solid line) time. Note that singular vectors are normalized to have unit initial total energy norm. (From Buizza, 1997)

The selection of 16 SVs is such that the first 4 are always selected, and from the 5th on, each subsequent SV is selected if 50% of its energy is located outside the regions where the SVs already selected are localized. Once the 16 SVs are selected, an orthogonal rotation in phase-space and a final re-scaling are performed to construct the ensemble perturbations. The purpose of the phase space rotation is to generate perturbations that have the same globally averaged energy as the SVs but smaller local maxima and more uniform spatial distribution. The rotated SVs are characterized by similar growth rates (at least to 48 hrs). The rotation is defined to minimize the local ratio between the perturbation amplitude and the amplitude of the analysis error estimate of the ECMWF Optimal Interpolation analysis. The rescaling allows local amplitudes up to $\sqrt{1.5}$ larger than the OI error.

The 16 rotated perturbations are 3D fields of temperature, vorticity, divergence and surface pressure (no moisture, since the propagator is "dry", although there is current work on including physical processes in the TLM and adjoint). They are added and subtracted to the control initial conditions to create 33 initial conditions (32+control), from which the ensemble forecast is run with the nonlinear model at T63 resolution.

In 1997 ECMWF changed the system to an ensemble of 50 members (plus control) run at a resolution of T156 (with a linear gaussian grid, since their use of a semilagrangian scheme allow the use of a more efficient linear rather than quadratic grid).   This increase in resolution had a major positive effect on the quality of the ECMWF ensemble forecasting system. In March 1998 ECMWF added to the initial perturbations the evolved (or final) singular vectors from 48 hours before the analysis time. The 2-days evolved SVs are much closer to the Lyapunov vectors (or bred modes) (Barkmeijer et al 1998).

Initially NCEP and ECMWF considered in their ensembles only the errors generated by uncertainties in the initial conditions, and neglecting the additional errors due to the models themselves. This is a reasonable but not perfect assumption for the extratropics (Reynolds et al, 1994). In 1998 ECMWF tested an innovative way to account for the fact that the model has deficiencies (Miller et

Macintosh HD:Users:ekalnay:Documents:AOSC614-DOCS:PPTClasses:ch6_5EnsembleOperational.docCreated on December 7, 2006 1:53 PM

43

al, 1999). The time derivatives of the physical parameterizations are multiplied by gaussian random numbers with a mean of 1.0 and a standard deviation of 0.2, which have a time lag correlation of several hours and horizontal correlation of a few hundred kilometers. This introduction of randomness in the "physics" had a very good impact on the ensemble. It increased the ensemble spread to levels similar to those of the control forecast error, which is a necessary condition if "nature" (the verifying ensemble) is to be a plausible member of the ensemble Toth and Kalnay, 1993).

6.5.3 Ensembles based on multiple data assimilation

Houtekamer (1996) and Houtekamer et al (1998) have developed a very promising ensemble forecasting system based on running an ensemble of data assimilation systems to create the initial conditions. In their different data assimilation systems they add (additional) random errors to the observations and include different parameters in the physical parameterizations of the model in different ensembles. This is a promising approach, related to but more general than breeding. One novel approach introduced by Houtekamer et al (1998) is the use of perturbations in the physical parameterizations in the models used in different analysis cycles. Through a careful combination of changes in major parameterizations, it is possible to use the ensemble forecasts to isolate the impact of particular parameterizations. As indicated by the results of Miller et al (1999), the introduction of uncertainty in the model should improve the efficiency of the ensemble.

Hamill et al (1999) have shown in a quasigeostrophic system that the multiple data assimilation ensemble system performs better than the singular vectors or breeding approaches. The computational cost of creating the initial perturbations is comparable to that of the singular vector approach, whereas in the breeding method the perturbations are obtained as a by-product of the ensemble forecasts themselves.

6.5.4 Multi-system ensemble approach

The ensemble forecasting approach should replicate in the initial perturbations the statistical uncertainty in the initial conditions: ideally, the initial perturbations should be the leading eigenvectors of the analysis error covariance (Ehrendorfer and Tribbia, 1997). Moreover, it should also reflect model imperfections and our uncertainty about model deficiencies. In the standard approaches discussed so far the uncertainty in the initial conditions is introduced through *perturbations added to the control analysis*, which is the best estimate of the initial conditions. As a result, the perturbed ensemble forecasts are, on the average, somewhat less skillful than the control forecast. Similarly, when perturbations are introduced upon the control model parameterizations (Miller et al, 1999, Houtekamer et al, 1998), the model is made slightly worse, since the control model has been tuned to best replicate the evolution of the atmosphere.

A different approach that has become more popular recently is that of a multi-system ensemble. It has long been known that an ensemble average of operational global forecasts from different operational centers is far more skilful than the best individual forecast (e.g., Kalnay and Ham, 1989, Fritsch et al, 2000, and references therein). More recently, it has been shown that this is true also for shorter range ensembles of regional models (Hou et al, 2000), and that the use of multi-systems can therefore extend the utility of ensemble forecasting to the short-range. Krishnamurti et al (1999) have shown that if the multi-system ensemble includes correction of the systematic errors by regression, the quality of the ensemble system is further significantly improved. Krishnamurti et al (1999) call this multiple system approach "super-ensemble".

The advantages of a multi-system ensemble are not surprising. Instead of adding perturbations to the initial analysis, and introducing perturbations into the control model parameterizations, the multi-system approach takes the best (control) initial conditions and the best (control) model estimated at different operational centers that run competitive state-of-the-art operational analyses and model forecasts. Thus the multi-system probably samples the true uncertainty in both the initial conditions and the models better than any perturbation introduced a posteriori into a single operational system.