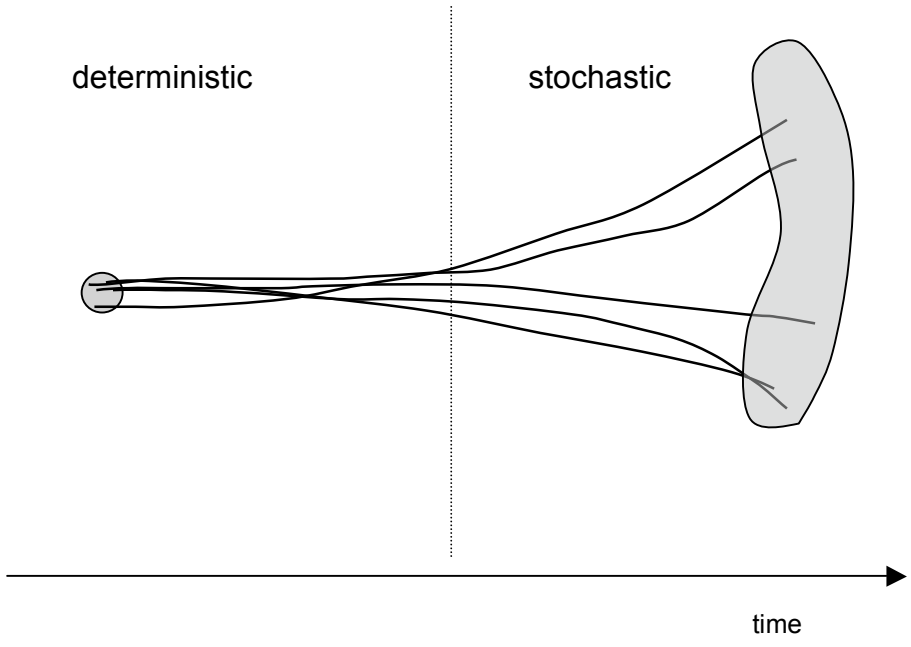
6.4 Ensemble forecasting: early studies

As we saw in previous sections, Lorenz (1963,1965) showed that the forecast skill of atmospheric models depends not only on the accuracy of the initial conditions and on the realism of the model, (as it was generally believed at the time), but also on the instabilities of the flow itself. He demonstrated that any nonlinear dynamical system with instabilities, like the atmosphere, has a finite limit of predictability. The growth of errors due to instabilities implies that the smallest imperfection in the forecast model or the tiniest error in the initial conditions, will *inevitably* lead to a total loss of skill in the weather forecasts after a *finite* forecast length. Lorenz estimated this period limiting weather predictability as about two weeks. With his simple model he also pointed out that predictability is strongly dependent on the evolution of the atmosphere itself: some days the forecasts can remain accurate for a week or longer, and on other days the forecast skill may break down after only 3 days. This discovery made inevitable the realization that NWP needs to account for the stochastic nature of the evolution of the atmosphere (Fig. 6.9). As we saw in the previous section, Lorenz (1965) studied the error growth of a complete "ensemble" of perturbed forecasts, with the ensemble size equal to the dimension of the phase space (one perturbation for each of the 28 model variables). In this paper he introduced into the literature concepts related to singular vectors (SVs) and local Lyapunov vectors (LLVs) discussed in the previous section. This was followed by several early approaches to the problem of accounting for the variable predictability of the atmosphere.

Fig. 6.9: Schematic of ensemble prediction, with individual trajectories drawn for forecasts starting from a representative

set of perturbed initial conditions within a circle representing the uncertainty of the initial conditions (ideally the analysis error covariance) and ending within the range of possible solutions. For the shorter range, the forecasts are close to each other, and they may be considered deterministic, but beyond a certain time, the equally probable forecasts are so different that they must be considered stochastic. The transition time is of the order of 2-3 days for the prediction of large-scale flow, but can be as short as a few hours for mesoscale phenomena like the prediction of individual storms. The transition time is shorter for strongly nonlinear parameters: even for large-scale flow, precipitation forecasts show significant divergence faster than the 500 hPa fields. (Adapted from Tracton and Kalnay, 1993).



6.4.1 Stochastic-dynamical forecasting (SDF)

Historically, the first forecasting method to explicitly acknowledge the uncertainty of atmospheric model predictions was developed by Epstein (1969), who introduced the idea of **stochastic-dynamic forecasting (SDF)**. He derived a continuity equation for the probability density $\varphi(X;t)$ of a model solution X of a dynamical model $\dot{X} = G(X(t))$, where the model has dimension D:

$$\frac{\partial \varphi}{\partial t} + \nabla_D . (\dot{X}\varphi) = 0 \qquad (4.1)$$

This equation indicates that in an ensemble of forecast solutions, "no member of the ensemble may be created or destroyed". An ensemble starting from an infinite number of perturbed integrations spanning the analysis uncertainty gives the "true" probability distribution (with all its moments), but even for a simple low order model, the integration of (4.1) is far too expensive. Therefore Epstein introduced an approximation to predict only the first and second moments of the probability distribution (expected means and covariances) rather than the full probability distribution. Epstein assumed that the model equations are of the form

$$\dot{x}_i = \sum_{j,k} a_{ijk} x_j x_k - \sum_j b_{ij} x_j + c_i \ , \qquad (4.2)$$

,

The forecast equations for the expected first and second moments are given by

$$\dot{\mu}_i = E(\dot{x}_i)$$
$$\dot{\rho}_{ij} = E(x_i \dot{x}_j + \dot{x}_i x_j) \cdot \qquad (4.3)$$

The covariances ρ_{ij} are related to the second order moments by $\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$. Replacing (4.2) into (4.3) gives rise to forecast equations for $\dot{\mu}$ and $\dot{\sigma}$ that contain triple moments $(x_i x_j x_k)$. As done in turbulence models with a second order closure for the triple products (Chapter 4), Epstein introduced a closure assumption for the third order moments around the mean $\tau_{ijk} = E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)]$. He assumed that $\sum_{kl} a_{jkl}\tau_{ikl} + a_{ikl}\tau_{jkl} = 0$, which then gives a closed set of equations for the means and covariances:

$$\dot{\mu}_i = \sum_{jk} a_{ijk}(\sigma_{jk} + \mu_j \mu_k) - \sum_j b_{ij}\mu_j + c_i$$
$$\dot{\sigma}_{ij} = \sum_{kl} a_{jkl}(\mu_k \sigma_{il} + \mu_l \sigma_{ik}) + a_{ikl}(\mu_k \sigma_{jl} + \mu_l \sigma_{jk}) - \sum_k (b_{ik}\sigma_{jk} + b_{jk}\sigma_{ik})$$
(4.4)

Epstein tested these "approximate" stochastic equations for a Lorenz 3-variable model. The "true" probability distribution was computed from a Monte Carlo ensemble of 500 members, and the comparison indicated good agreement, at least for several simulated days. Note that in his case, the number of ensemble members was much larger than the number of degrees of freedom of the model, a situation that would be impossible to replicate with current models with millions of degrees of freedom. In his paper, Epstein also introduced the idea of using stochastic-dynamical

forecasting in the *analysis cycle*, with the background forecast and error covariance provided by stochastic-dynamical forecasts combined with observations that also contain errors (cf. Chapter 5, sections 3-5).

Unfortunately, although the SDF method was introduced as a shortcut to an "infinite" Monte Carlo ensemble, in a model with N degrees of freedom, it requires N(N+1)/2+N forecast equations, equivalent to making about (N+3)/2 model forecasts. Although this was practical with a 3 variable model, it is completely unfeasible for a modern model, with millions of degrees of freedom.

6.4.2 Monte Carlo forecasting (MCF)

In 1978, Leith proposed the idea of performing ensemble forecasting with a limited number m of ensemble members instead of the conventional single (deterministic) control forecast. He also proposed performing an "optimal estimation" of the verification using linear regression on the dynamical forecasts, with optimal weights determined from forecast error covariances (cf. Chapter 5, sections 3-5). Since forecasts lose their skill at longer lead times, and individual forecasts eventually are further away from the verification than the climatology (cf. eqs. (4.5) and (4.6)), optimal estimation of the verification is equivalent to **tempering** (i.e., **hedging** the forecast towards climatology).

He cast his analysis using, instead of model variables, their deviation $\mathbf{u}$ with respect to climatology (also known as forecast **anomalies**). The true state of the atmosphere is denoted $\mathbf{u_o}$, and $\hat{\mathbf{u}}$ denotes an unbiased estimate of $\mathbf{u}_0$, whose expected value (average over many forecasts,

represented by the square brackets) is equal to zero: $<\hat{\mathbf{u}}>=0$.

We can compute the expected error covariance of a climatological forecast (i.e., a forecast of zero anomaly):

$$<(0-\mathbf{u}_o)(0-\mathbf{u}_o)^T>=<\mathbf{u}_o\mathbf{u}_o^T>=\mathbf{U}.$$

(4.5)

A single (deterministic) forecast $\hat{\mathbf{u}}$, on the other hand, has, on the average, an error covariance given by

$$<(\hat{\mathbf{u}}-\mathbf{u}_o)(\hat{\mathbf{u}}-\mathbf{u}_o)^T>=<\hat{\mathbf{u}}\hat{\mathbf{u}}^T+\mathbf{u}_o\mathbf{u}_o^T-\hat{\mathbf{u}}\mathbf{u}_o^T-\mathbf{u}_o\hat{\mathbf{u}}^T>\underset{t\to\infty}{\longrightarrow}2\mathbf{U}$$

(4.6)

This limit occurs because the last two terms in the brackets go to zero as the forecasts become decorrelated with the true atmosphere at long lead times, and we assume that the model covariance is also unbiased. This indicates that for long lead times an individual deterministic forecast has twice the error covariance than a climatological forecast. Therefore, a "regressed" forecast, tempered towards climatology, must be better than a single deterministic forecast (in a least square error sense), with an error covariance that asymptotes to $\mathbf{U}$, and not $2\mathbf{U}$.

A regressed forecast $\hat{\mathbf{u}}_o=\hat{\mathbf{u}}\mathbf{A}$ is obtained by linear regression, minimizing the square of the regressed error $\varepsilon^T\varepsilon=<(\mathbf{u}_o-\hat{\mathbf{u}}\mathbf{A})^T(\mathbf{u}_o-\hat{\mathbf{u}}\mathbf{A})>$ with respect to the elements of the matrix of constant regression coefficients **A**.

As we did in the derivation of the optimal weight matrix for the observational increments in section 5.4, we make use of the linear regression formulas: if the linear prediction equation is $\hat{\mathbf{y}} = \mathbf{xA}$, then the error is given by $\varepsilon = \mathbf{y} - \mathbf{xA}$. The matrix of the derivatives of the (scalar) squared error $\varepsilon^T \varepsilon$ with respect to each element of $\mathbf{A}$ is given by $\frac{\partial \varepsilon^T \varepsilon}{\partial \mathbf{A}} = -2\mathbf{x}^T(\mathbf{y} - \mathbf{xA}) = 0$, which gives the normal equation $\mathbf{x}^T\mathbf{y} = \mathbf{x}^T\mathbf{xA}$, or $\mathbf{A} = (\mathbf{x}^T\mathbf{x})^{-1}(\mathbf{x}^T\mathbf{y})$. Applying this to the regressed forecast we obtain $< \hat{\mathbf{u}}^T(\mathbf{u}_o - \hat{\mathbf{u}}\mathbf{A}) >= 0$, or

$$\mathbf{A} =< \hat{\mathbf{u}}^T\hat{\mathbf{u}} >^{-1} < \hat{\mathbf{u}}^T\mathbf{u}_o > \qquad (4.7)$$

Estimating the required forecast statistics in (4.7) involves considerable work. The size of the regression matrix is usually large compared to the size of the sample available to estimate it, and in order to reduce the number of parameters to be estimated additional approximations are needed (e.g., by parameterizing error growth, Hoffman and Kalnay, 1983).

Now, let's consider instead of regression an ensemble of *m* forecasts computed from perturbations $\mathbf{r}_i$ to the initial best estimate (analysis) $\hat{\mathbf{u}}$. Ideally, the perturbations should be chosen so that their outer product is a good estimate of the initial error covariance (i.e., the analysis error covariance $< \mathbf{rr}^T >= \mathbf{P}_a$, as suggested in the schematic Fig. 6.9). In practice, however, the analysis error covariance can only be approximately estimated (e.g., Barkmeijer, 1998).

If $\overline{\mathbf{u}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{u}_i$ is the average of an ensemble of *m* forecasts, then its error covariance evolves like

$$<(\overline{\mathbf{u}}-\mathbf{u}_o)(\overline{\mathbf{u}}-\mathbf{u}_o)^T>=<(\overline{\mathbf{u}}\,\overline{\mathbf{u}}^T+\mathbf{u}_o\mathbf{u}_o{}^T-\overline{\mathbf{u}}\mathbf{u}_o{}^T-\mathbf{u}_o\overline{\mathbf{u}}^T> \underset{t\to\infty}{\longrightarrow}(1+\frac{1}{m})\mathbf{U} \qquad (4.8)$$

since the last two terms in the brackets go to zero at long time leads, and the first one evolves like

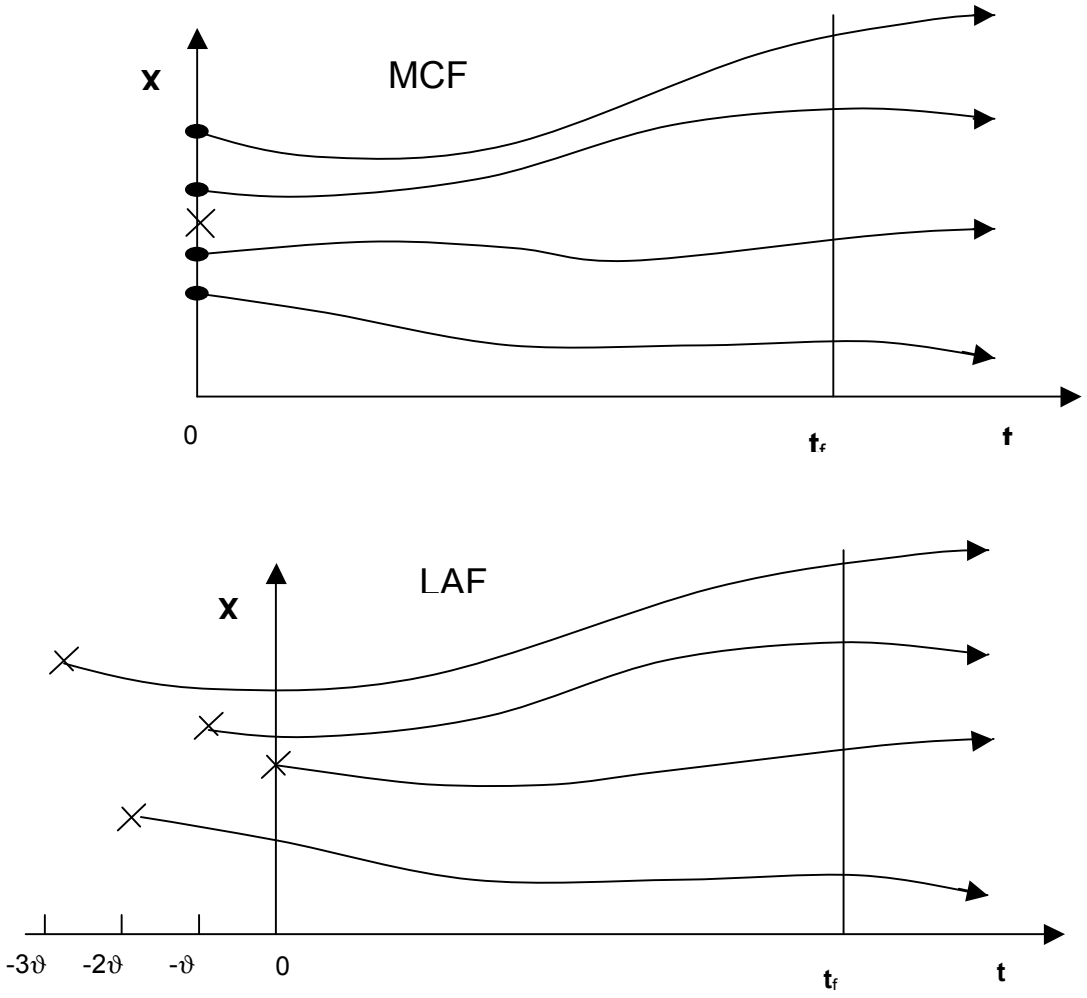
$$<\overline{\mathbf{u}}\,\overline{\mathbf{u}}^T>=\frac{1}{m}\sum_{i=1}^{m}\mathbf{u}_i\frac{1}{m}\sum_{j=1}^{m}\mathbf{u}_j{}^T\underset{t\to\infty}{\longrightarrow}\frac{m}{m^2}\mathbf{U} \qquad (4.9)$$

Equation (4.8) shows that averaging a Monte Carlo ensemble of forecasts approximates the tempering of the forecasts towards climatology, *without the need to perform regression*. It suggests that such tempering may be substantially achieved with a relatively small number of ensemble members (compare (4.8) with (4.5) and (4.6)). Leith used an analytical turbulence model to test this hypothesis, and concluded that a Monte Carlo forecasting procedure represents a practical, computable approximation to the stochastic-dynamic forecasts proposed by Epstein (1969). He suggested that adequate accuracy would be obtained for the best estimate of the forecast (i.e., *the ensemble mean*) with sample sizes as small as 8, but that the estimation of *forecast errors* may require a larger number of ensemble members. MCF is thus a feasible approach for ensemble forecasting, requiring only a definition of the initial perturbations and *m* forecasts.

6.4.3 Lagged Average Forecasting (LAF)

In 1983, Hoffman and Kalnay proposed Lagged Average Forecasting (LAF) as an alternative to Monte Carlo Forecasting (MCF), in which the forecasts initialized at the current initial time, *t=0*, as well as at previous times, $t = -\tau, -2\tau, \ldots -(N-1)\tau$ are combined to form an ensemble (see schematic Fig. 6.10). In an operational set up, τ is typically 6, 12 or 24 hours, so that the forecasts are already available, and the perturbations are generated automatically from the forecast errors. Since the ensemble is comprised of forecasts of different "age", Hoffman and Kalnay (1983) weighted them according to their expected error, which they estimated by parameterizing the observed error covariance growth. They compared the LAF and MCF methods within a simulation system, using a primitive equations model as "nature", and a quasi-geostrophic model to perform the “forecasts”. In this way they allowed for model errors, unlike the previous "identical twin” experiments that assumed a perfect model. They "observed" the required variables and introduced random "observation errors" every 6 hours and performed many ensemble forecast experiments separated by 50 days of integration. They compared the results of single forecasts (ordinary dynamical forecasts, ODF), MCF, LAF and tempered ODF (tODF), as well as persistence-climatology forecast (the most skillful baseline forecast).

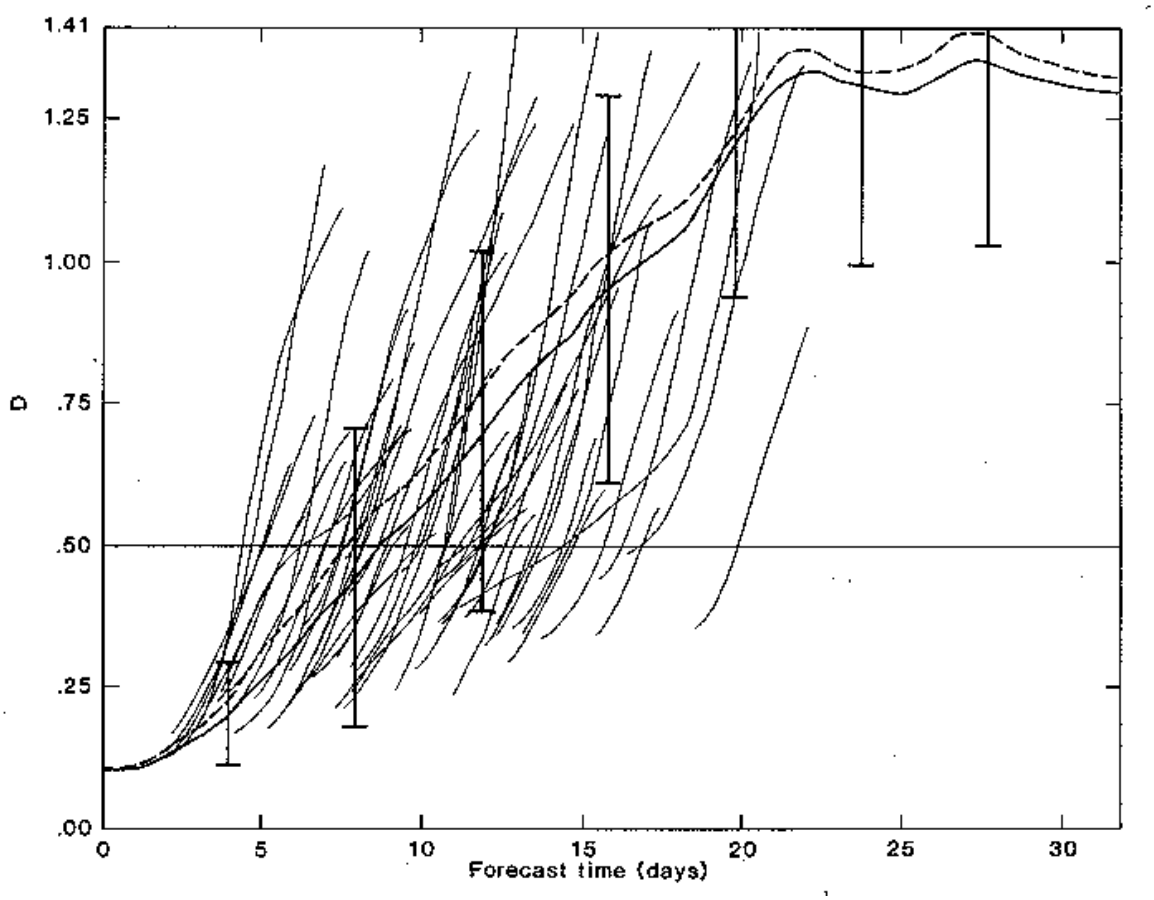
Fig. 6.10: Schematic time evolutions of Monte Carlo Forecasts (top) and Lagged Averaged Forecasts (bottom). The abscissa is forecast time *t*, and the ordinate is the value of a forecast variable *X*. The crosses represent analyses obtained at time intervals τ , and the dots, randomly perturbed initial conditions; t_f is a particular forecast time. The initial "perturbation" for the LAF is the previous forecasts' error at the initial time. (Adapted from Hoffman and Kalnay, 1983).



Hoffman and Kalnay looked at the error growth of individual forecasts (Fig. 6.11). Note in this figure that the individual forecast errors grow slowly and then at a certain time there is a rapid error growth until nonlinear saturation takes place (only the period of rapid growth is plotted). Note also that the forecast errors saturate around $\sqrt{2}$ of the climatological variability, as indicated by equation (4.6).

In these simulated forecasts, like in real weather forecasts, the forecast skill exhibits a lot of day-to-day variability. The rapid growth takes place at a time that varies from a minimum of 5 days to a maximum of 20 days. Hoffman and Kalnay tested the ability of the ensemble to predict the time at which the forecast error crossed 50% of the climatological standard deviation. They used as predictor the spread of the ensembles (standard deviation with respect to their mean). They found that the LAF ensemble average forecast was only slightly better than the MCF, but the advantage of LAF in predicting forecast skill was much more apparent, with the correlation between predicted and observed time of crossing the 50% level being 0.68 for MCF and 0.79 for LAF.

Fig. 6.11: Time evolution of D, the individual forecast errors scaled by the climatological forecast error, plotted during the period the forecast error crossed D=0.5. Also forecasted are two measures of average forecast error. Adapted from Hoffman and Kalnay (1983).



The advantages of LAF with respect to MCF are probably due to the fact that LAF perturbations in the initial conditions were not *randomly chosen* errors like in MCF but included dynamical influences and therefore contained

"*errors of the day*". This is because the perturbations are generated from actual forecast errors and therefore they are influenced by the evolution of the underlying background large-scale flow.

LAF has been frequently used for experimental ensemble forecasting, both for medium range and climate prediction. However, the statistics required to estimate the weights of the members of the LAF ensemble according to their “age” are very hard to obtain, so that except for the study by Dalcher et al (1986), all the LAF members have been generally given equal weight. The advantages of LAF are a) some of the forecasts are already available in operational centers; b) it is very simple to perform and does not require special generation of perturbations; and c) the perturbations contain "errors of the day" (Lyapunov vectors). LAF has also major disadvantages: a) a large LAF ensemble would have to include excessively “old” forecasts; b) without the use of optimal weights, the LAF ensemble average may be tainted by the older forecasts.

Ebisuzaki and Kalnay (1991) introduced a variant of LAF denoted Scaled Lagged Average Forecasting (SLAF) which reduces these two disadvantages. The perturbations are obtained computing the forecast error of forecasts started at $t = -j\tau, j = 1,...,N-1$, and multiplying these errors by $\pm 1/j$. This assumes that the errors grow approximately linearly with time during the first 2-3 days, and that the perturbations can be subtracted from and not just added to the analysis. The advantages of SLAF are a) the initial perturbations of the ensemble members are all of approximately the same size (this can be enforced using a more sophisticated rescaling than linear growth), and b) their number is doubled with respect to LAF, so that only shorter-

range forecasts are needed to create SLAF. In practice, it has been observed that pairs of initial perturbations with opposite sign, as done in SLAF, yield better ensemble forecasts, presumably because the Lyapunov vectors within the analysis errors can have either sign, whereas LAF tends to maintain a single sign in the error. Experiments with the NCEP global model showed that SLAF ensembles were better than LAF ensembles (Ebisuzaki and Kalnay, 1991). This method is also easier to implement in regional ensemble forecasts, since it generates boundary condition perturbations consistent with the interior perturbations (Hou et al, 2001).