Q. J. R. Meteorol. Soc. (1994), **120**, pp. 387–428 551.509.313.5

Predictability experiments with persistence forecasts in a red-noise atmosphere

By KLAUS FRAEDRICH* and CHRISTINE ZIEHMANN-SCHLUMBOHM
Universität Hamburg, Germany

(Received 16 November 1992; revised 22 June 1993)

SUMMARY

Individual and lagged ensemble forecasts of persistence in a red-noise atmosphere are analysed to obtain information on predictability experiments performed in an imperfect model/ensemble environment. By examining the lead-time-dependent error budgets of individual and ensemble forecasts, various measures of predictability are analytically determined: the initial and saturation error, the error growth rates, the limit of predictability, error and squared error distributions depending on initial conditions, and the systematic and non-systematic error, etc. Furthermore, weather-regime dependent predictability can be studied directly by using different autocorrelation time-scales of the red-noise atmosphere. Finally, ensembles of lagged forecasts are constructed to analyse the relation between the forecast errors of the ensemble mean and the dispersion within the ensemble. Despite the simplicity of this external predictability experiment, the error budgets show features that may be qualitatively compared with those of numerical weather-prediction and climate-model systems.

1. INTRODUCTION

Predictability characterizes the weather or climate system's sensitive dependence on initial and boundary conditions. Predictions of the first kind (Lorenz 1975), which show sensitive dependence on initial conditions, are dominated by internally occurring instabilities at fixed boundary conditions. These forecasts are related to the practical aspects of weather prediction, revealing either chaotic or random properties of the weather system. Predictions of the second kind describe the response of the system to changing boundary conditions. Such predictions are associated with the structural stability and, from a practical point of view, related to long-range or climate-anomaly forecasting and, therefore, to the static properties of the weather or climate attractor.

Predictability is analysed by the error budget that describes the time evolution of forecast errors in terms of the (squared) distance between the forecast and its verification. The error-budget analyses are analogous to the diffusion process where, in a first step, only the distances between the diffusing particles are of relevance: single-particle diffusion corresponds to an analysis of the verification trajectory only, where the distance from the origin represents the error growth of a persistence forecast; two-particle diffusion provides the frame for an individual forecast evolving in relation to its verification; and, finally, the ensemble forecast gives the picture of a cloud of forecast trajectories, which disperses near an individual verification trajectory. One may even go one step further in this analogy and apply the kinematics of the diffusion or mixing process by not only analysing changing distances but also deformation, rotation and other properties, which may eventually lead to a mathematical theory of predictability.

Predictability experiments provide the data for diagnosing the error budget. Experiments of external (or practical) predictability are directly linked with the practical task of weather or climate forecasting; the bias due to differences between the model and the real climate is one of the problems met in analysing the predictability of numerical models. They are predictability experiments performed in an imperfect model environment. Intrinsic (internal or theoretical) predictability, in contrast, is related to error

* Corresponding author: Meteorologisches Institut, Universität Hamburg, Bundesstr. 55, D-20146 Hamburg, Germany.

budgets due to small perturbations in the initial and/or boundary conditions generated in identical model atmospheres ('identical twins'); this leads to predictabiity experiments in the perfect model environment.

An approach to incorporate predictability aspects into practical forecasting is the method of stochastic–dynamic prediction, by which not only the expected forecast fields but also the related statistical moments or distributions are determined (Epstein 1969; Gleeson 1970). Because of the complexity of this approach a more practical technique has been suggested that leads to similar results: Monte Carlo forecasts (Leith 1974; Seidman 1981; Hayashi 1986). A convenient variant utilizes lagged forecast ensembles that are commonly available at the national weather services (Hoffmann and Kalnay 1983); they may be considered as a parametrization of the stochastic–dynamic predictions or of the Monte Carlo forecasts. Many Monte Carlo and lagged ensemble predictability experiments have been performed with numerical weather-prediction (NWP) models in both perfect and imperfect forecast environments (Lorenz 1982; Hoffman and Kalnay 1983; Roads 1987, 1988; Deque 1988; Murphy 1988; Dalcher *et al.* 1988; Chen 1989; Brankovic *et al.* 1990, etc). In the real atmosphere similar studies have been made that diagnose ensembles of past-weather analogues, or nearest neighbours in phase space (for example, Lorenz 1969).

In the following we are concerned with predictability analyses in the imperfect model–imperfect forecast ensemble environment. However, instead of diagnosing lagged ensembles of complex NWPs for real weather systems, we analyse lagged persistence forecasts in a substitute red-noise atmosphere. In this way we demonstrate the predictability analysis methods in a toy predictability experiment consisting of the persistence-forecast model and the red-noise substitute atmosphere. The red-noise process has been used in many studies. Because of its simplicity it serves as a null-hypothesis for statistical tests, a stochastic model to parametrize various aspects of fluctuations associated with general circulation and climate time-scales, and it provides a description of the mutual dependence of stochastic forcing and deterministic response (for example, Dole and Gordon 1983; Gutzler and Mo 1983; Trenberth 1984). Persistence forecasts are a fundamental reference prediction scheme, particularly in long- and short-range forecasting, and a characteristic of atmospheric dynamics (Munk 1960; Legras and Ghil 1985; Horel 1985; Trenberth 1985; Saha and van den Dool 1988; Schubert *et al.* 1992, etc.).

From a different perspective we like to draw an analogy to the analysis of the similarly complex climate problem: a simple, say zero-dimensional, climate model can serve as a prototype model for climate analysis, which involves internal and structural stability, sensitivity studies and the response to stochastic forcing. We hope to introduce persistence forecasts in a red-noise atmosphere as a similar prototype system to study the external predictability problem posed by the imperfect model–imperfect ensemble forecast experiments. (For a study of predictability in a perfect model environment see, for example, Palmer (1993).) In section 2 the substitute atmosphere is described as a first-order autoregressive process; the error budget and error distribution associated with individual persistence forecasts are discussed. Section 3 describes the predictability analysis of an ensemble of lagged persistence forecasts. Section 4 continues with the forecast-error budget, depending on the initial anomaly conditions. Section 5 describes an ensemble of two-lagged forecasts, as a special case, to provide a maximum of simplicity. In section 6 the ensemble dispersion is discussed. Furthermore, the results of the imperfect model forecast-verification system of persistence in red-noise is linked to NWP predictability experiments and observations to emphasize results that have potential practical relevance. Finally the conclusions are given in section 7.

2. Persistence forecasts in a red-noise atmosphere: the error budget

Before analysing the predictability or error budget of persistence forecasts, the deterministic and stochastic properties of the substitute red-noise atmosphere, for which forecasts are being made, are discussed. This substitute is a first-order autoregressive process, which is one of the simplest non-trivial processes that simulates many observed aspects of the variability in the atmosphere and climate system (see, for example, Dole and Gordon 1983; van den Dool and Chervin 1986). It has also been used in predicting some dynamical features of the weather and climate.

(*a*) *Autoregressive process (first order)*

The atmospheric dynamics, $X(t)$, are represented by fluctuations, $X(t) = \langle X \rangle + X'(t)$, about zero climate mean $\langle X \rangle = 0$. Time (sample) averaging is denoted by $\langle\rangle$, and the prime, describing anomalies, will be deleted in the following. Red-noise fluctuations are introduced by a first-order autoregressive process, AR(1), which serves as a substitutre atmosphere. This stochastic process consists of a deterministic part, $X(t)$, and an additive random part, z_i:

$$\left.\begin{aligned} X(t) &= aX(t-1) + z_1 \\ X(t) &= a^r X(t-r) + \sum_{i=0}^{r-1} a^i z_{r-i} \end{aligned}\right\} \tag{2.1}$$

with the lag-one autocorrelation $a = \langle \{X(t) - X(t-1)\}^2 \rangle$ and the time lag r. Given the Gaussian white-noise random forcing, z_i, with zero mean $\langle z_i \rangle = 0$, variance $s_z^2 = \langle z_i^2 \rangle$ and vanishing lag-correlation, $\langle z_i z_j \rangle = 0$ if $i \neq j$, then the variance of the red-noise atmosphere, $s_x^2 = \langle (X\text{-}\langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$, can be related to the intensity (or variance) of the white-noise random forcing: $s_z^2 = s_x^2(1 - a^2)$, describing the fraction $1 - a^2$ of the total variability, s_x^2, which will be set to unity in the following. The deterministic part of the fluctuations contributes the remaining variance, $s_x^2 a^2$, so that a signal-to-noise ratio is $a^2/(1 - a^2)$.

(*b*) *Integral time-scale—weather regimes*

A time-scale, τ, which is based on the time integral over the autocorrelation can be assigned to the anomaly fluctuations, $C_{xx}(r) = \langle X(t)\, X(t+r) \rangle / s_x^2 = a^r$. This leads to a geometric series, $\Sigma_{r=0}^{r=N-1} a^r = (1 - a^N)/(1 - a)$ which, for $N \to \infty$, gives the integral time-scale:

$$\tau = \frac{1}{1-a}. \tag{2.2}$$

The integral time-scale is a measure of the lifetime of the red-noise anomaly fluctuations in the sense that it reveals how quickly a single realization loses memory of its initial state on average. For $a = 0.8$ (or 0.3) the integral time-scale is 5 (or 1.4) unit time steps. The time-scale, τ, grows with increasing lag-one autocorrelation (or red-noise parameter), a. Thus red-noise regimes in this model atmosphere are simultaneously characterized by the integral time-scale, $1 < \tau < \infty$, the intensity of the stochastic forcing, $0 < s_z^2 < 1$, and the autocorrelation, $0 < a < 1$, with $a = (1 - s_z^2/s_x^2)^{1/2} = 1 - 1/\tau$. That is, large (small) red-noise parameters, a, describe regimes with large (small) integral time-scales τ, which are associated with relatively small (large) intensities of the stochastic forcing s_z^2.

In this sense a 'weather regime' may be parametrized in the substitute red-noise atmosphere by an autocorrelation, a, which determines the magnitude of the integral time-scale, τ, etc. Time series observed in the real atmosphere share some of these red-noise properties; in particular, if one assumes that weather regimes are distinguished by different lag-one autocorrelations. In this sense the quasi-stationary behaviour in the large scales (deterministic part, $a^2 s_x^2$) can be integrally associated with an organized behaviour of the synoptic scales (random part, s_z^2) (see, for example, Reinhold and Pierrehumbert 1982). Therefore, it is not surprising that red noise has been used in many studies as a substitute (or surrogate) atmosphere or, vice versa, to fit atmospheric data for interpretation or test purposes to this prototype stochastic–dynamic process (see, for example, Munk 1960; Dole and Gordon 1983; Trenberth 1984; van den Dool and Chervin 1986).

(c) *Individual persistence forecasts*

A fundamentally important reference forecast is persistence, 'because only forecasts better than persistence have skill in the forecast of the time derivative—the essence of forecasting' (van den Dool 1989). Persistence predicts the future weather states (verifications), $X(t)$, by the initially observed state $X(t_0)$; that is, a persistence forecast, F, commences at the time $t_0 = t - r$ with the observation $X(t_0) = X(t - r)$ and is evaluated after the lead time or forecast range, r, by the observation (verification) $X(t)$:

$$F(r) = X(t - r). \tag{2.3}$$

Forecasts and their corresponding verifications are commonly analysed in terms of pairs of trajectories evolving in state space, whose squared Euclidean distance corresponds to the squared error. Error analyses of persistence forecasts can, therefore, be interpreted as the structure function statistics associated with a degenerate trajectory pair (that is a single trajectory), because only the individual verification trajectory, $X(t)$, and its squared distance from the initial position is considered: $|X(t) - X(t_0)|^2$. In the red-noise atmosphere the error budget of individual persistence forecasts (subscript 1) is described by the following evolution of the error variance, where the average is taken over all verification pairs: $E_1(r) = \langle \{X(t) - F(r)\}^2 \rangle = \langle \{X(t) - X(t - r)\}^2 \rangle = \langle X^2(t) \rangle + \langle X^2(t - r) \rangle - 2\langle X(t)\, X(t - r) \rangle$:

$$E_1(r) = 2s_x^2 (1 - a^r). \tag{2.4}$$

The error budget is shown in Fig. 1 for $a = 0.8$. The following features are noted:

(i) The vanishing initial error, $E_1(r = 0) = 0$, is one of the advantages of persistence forecasting. Therefore, it has been introduced as a forecast guidance which, if combined with another independent prediction scheme, can improve the forecast accuracy. In long-range forecasting this is achieved by persistence plus an analogue trend prediction (see, for example, Livezey *et al.* 1990); in short-term forecasting of tropical rainfall the daily persistence is combined with a shorter-term Markov chain to yield a high accuracy (Fraedrich and Leslie 1990).

(ii) As the autocorrelation, a^r, vanishes at lead times $r \to \infty$, the limiting persistence-error variance approaches the saturation level $E_1(r \to \infty) = 2s_x^2$. That is, in this limit, the persistence error corresponds to the mean difference between two randomly chosen 'weather' states.

(iii) The probability distributions of both the error, $e = X(t) - F(r)$, and the squared error, e^2, provide important additional information lost by the time (sample) averaging,

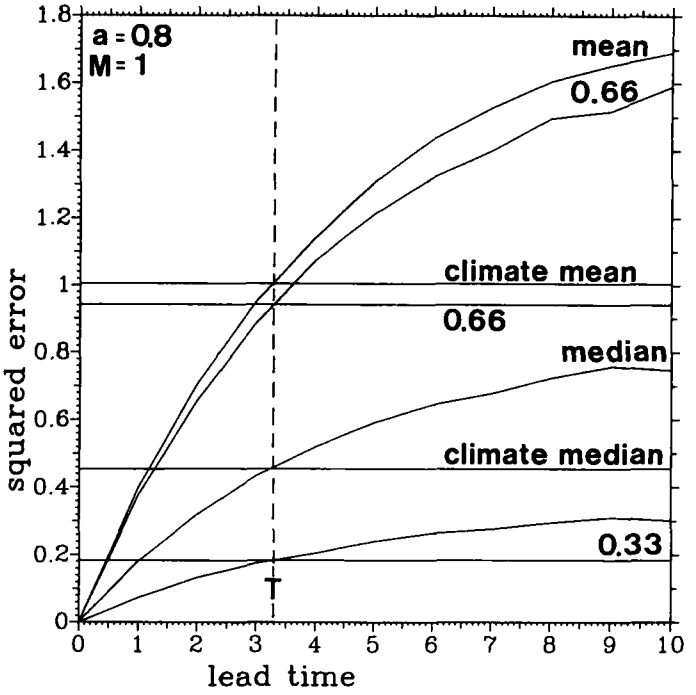


Figure 1. Error budget of individual persistence forecasts ($M = 1$) in a red-noise atmosphere (autocorrelation $a = 0.8$). The time evolution of the error variance, the median and the upper and lower terciles (0.66; 0.33) of the squared-error distribution. The same is shown for climate-mean forecasts (horizontal lines). The predictability limit, T, is also indicated.

$\langle\rangle$, over all forecast-verification samples, $X - F$. The normalized error of persistence in a red-noise atmosphere, $y = \{X(t) - F(r)\}/\sqrt{E_1(r)} = e/\sqrt{E_1(r)}$ with $E_1(r) = \langle e^2 \rangle$, is lead-time dependent and Gaussian distributed with zero mean and unit variance, $E_1(r)$; the squared error e^2 is chi-squared distributed with one degree of freedom. Thus, with the normalized error random variable, $y = (X - F)/\sqrt{E}$, the densities of the error, $g(y)$, and the squared error, $f(y^2)$, yield a normal and a chi-square distribution, respectively:

$$\left.\begin{aligned} g(y) &= \frac{1}{\sqrt{(2\pi E)}} \cdot \exp\left(-\frac{y^2}{2}\right) \\ f(y^2) &= \frac{1}{2^{n/2}\,\Gamma(n/2)} \cdot (y^2)^{(n/2)-1} \cdot \exp\left(-\frac{y^2}{2}\right) \end{aligned}\right\} \tag{2.5}$$

with the mean $\langle y^2 \rangle = n$, variance $\langle (y^2 - \langle y^2 \rangle)^2 \rangle = 2n$, and $n = 1$ degrees of freedom (d.o.f.); hence the Gamma-function, $\Gamma\{(n = 1)/2\} = \sqrt{\pi}$. The squared-error distribution, $f(y^2)$, can be displayed in terms of the quantiles associated with the $f(y^2)$ density. The mean, median and the upper and lower terciles (thirds) are shown in Fig. 1 (for persistence and also for climate forecasts). For a sufficiently small number of d.o.f.s the median of the squared-error distribution may be a more appropriate error measure than the mean (which is the error variance). For increasing d.o.f.s this discussion may become less relevant, because the distribution tends towards a Gaussian one as it is observed in multivariate weather maps composed of normally distributed variables.

(iv) At a fixed lead time, r^*, the forecast error, $E_1(r^*)$, is associated with the autocorrelation regime, $a = (1 - E_1(r^*)/s_x^2)^{1/r^*}$. Because of the underlying stochastic fluctuations, a small autocorrelation (or integral time-scale, τ) leads to a large error variance (and vice versa), which is also affecting the error growth rate.

(v) The rate of change of the error, $dE_1/dr = E_{1r}(r)$, yields (with $da^r/dr = a^r \ln a$):

$$\left.\begin{aligned} E_{1r}(r) &= 2 s_x^2 a^r \ln(1/a) \\ E_{1r}(r=0) &= 2 s_x^2 \ln(1/a) \end{aligned}\right\} \tag{2.6}$$

where $\ln(1/a) = -\ln a = -\ln(1 - 1/\tau) \sim 1/\tau$ for large integral time-scales. That is, the initial error growth is inversely proportional to the life span of the weather regimes. For $a = 0.8$ (or 0.3) the initial growth rate of the normalized error, $E_{1r}(r=0)/2s_x^2$ is $\ln(1/a) = 0.45$ (or 2.41) per unit time step. That is, the smaller the autocorrelation a, or integral time-scale τ, the larger the stochastic forcing s_z^2 and the larger the initial rate of error growth, $E_{1r}(r=0)$. Furthermore, substituting $a = (1 - E_1(r)/s_x^2)^{1/r}$ shows that the rate, E_{1r}, decreases with increasing magnitude of the error, $E_1(r)$. This leads to the often misinterpreted generalization that large errors grow slower, because this holds only for predictability experiments in the same forecast environment (a = constant) where the error growth decreases when approaching saturation. This needs to be modified when regimes ($0 < a < 1$) change.

(*d*) *Predictability limit*

The climate forecast error (or variance), s_x^2, serves as a predictability threshold. That is, predictions at lead time $r > T$ have passed the predictability limit at $E(r = T) = s_x^2$ if their error variance, $E(r)$, exceeds that of the climate forecast. For individual persistence forecasts, $E_1(r)$, the limit of predictability, $r = T$, is derived from (2.4) with $a^T = 1/2$:

$$T = \frac{\ln 2}{\ln(1/a)} \sim \tau \ln 2. \tag{2.7}$$

This predictability limit has a close connection to the integral (or life) time-scales τ of weather regimes, if τ is large. With $\ln(1/a) = -\ln a = -\ln(1 - 1/\tau) \sim 1/\tau$, the predictability limit of individual persistence forecasts is proportional to the integral time-scale of the red-noise regimes: $T \sim \tau \ln 2$. It is smaller than the lifetime-scale since the persistence forecast is not a perfect model but contains a systematic error. Only if the threshold is defined by $E_1(r = T) = 2(1 - 1/e)s_x^2 \sim 1.26\, s_x^2$, does $T \sim \tau$. For $a = 0.8$ (or 0.3) the predictability limit is reached after $T = 3.1$ (or 0.6) unit time steps.

That is, the persistence–red-noise (or forecast-verification) system also reveals the well known effective forecast range for complex physical systems, which is finite and limited by the life span of its most energetic phenomenon (see, for example, Tennekes 1991). Here it should be noted that the condition for the deterministic signal being larger (or more energetic) than the noise, $a^2/(1 - a^2) > 1$ or $a > 2^{-1/2}$, coincides with the condition of a sufficiently large life span of the weather regime, if $\tau > 3.4$. . . . (2.2), satisfying the approximation for $1/\tau$ being small (in 2.7).

(*e*) *Systematic and non-systematic errors*

In an imperfect model environment the mean square error can be separated into systematic, *SE*, and non-systematic or random components, *RE*:

$$\left.\begin{aligned} E &= SE + RE \\ SE &= \langle\langle e\rangle^2\rangle = \langle(\langle F\rangle - \langle X\rangle)^2\rangle \\ RE &= \langle(e - \langle e\rangle)^2\rangle = \langle((F - X) - (\langle F\rangle - \langle X\rangle))^2\rangle \end{aligned}\right\} \tag{2.8}$$

where the error at lead time r is $e = X(t) - F(r)$. Now the systematic error and the random error of persistence forecasts in the red-noise atmosphere can be determined.

The initial anomaly X_0 yields the persistence forecast $F(r) = X_0$ for lead time r. The associated verification time series commencing from this anomaly at the rth step backward is $X(t) = a^r X(t - r) + \Sigma_{i=0}^{r-1} a^i z_{r-i}$. Now averaging over a sample of the forecast–verification pairs, $\langle\rangle$, conditional on a fixed initial anomaly, X_0, yields the conditional forecast error and its associated growth rate:

$$\left.\begin{aligned} E_1(r|X_0) &= \langle X_0^2\rangle (1 - a^r)^2 + s_x^2 (1 - a^{2r}) \\ E_{1r}(r|X_0) &= 2\langle X_0^2\rangle (1 - a^r)\, a^r \ln(1/a) + 2s_x^2 a^{2r} \ln(1/a). \end{aligned}\right\} \tag{2.9}$$

The sample averaged persistence forecast conditional at the initial anomaly $\langle F(r)\rangle = \langle X(t)\rangle|_{X_0} = X_0$ is associated with the following average verification, which commences from the same anomaly X_0, $\langle X(t)\rangle = a^r \langle X(t - r)\rangle + \langle \Sigma_{i=0}^{r-1} a^i z_{r-i}\rangle = a^r \langle X(t - r)\rangle = a^r X_0$, where the average of the last term vanishes. Thus the systematic error $\langle e\rangle|_{X_0} = X_0(1 - a^r)$. Now averaging over all initial or conditional anomalies X_0 leads to the unconditional error budget (2.4) and also to the distinction between the forecast error's systematic and non-systematic or random components. That is $\langle X_0^2\rangle = s_x^2$ in (2.9) is interpreted as the variance of all possible initial anomalies (and not as the square of an anomaly), leading to the random and systematic errors, $RE = s_x^2(1 - a^{2r})$ and $SE = s_x^2(1 - a^r)^2$:

(i) The systematic error is smaller than the random error, and both approach unity for infinitely large lead times. At the limit of predictability (2.7), the systematic (non-systematic) error attains 1/4 (3/4) of the climate variance, $SE(r = T) = s_x^2/4$ ($RE = 1 - SE = 3s_x^2/4$). The initial error growth rate vanishes for systematic errors, $\mathrm{d}SE/\mathrm{d}r = 0$ for $r = 0$, but is finite for the random part, which, therefore, determines the total initial error growth of persistence forecasts.

(ii) NWP experiments show qualitatively similar results that, however, reveal considerably smaller systematic errors (Dalcher and Kalnay (1987) show systematic errors of about 20% of the total). Here it should be mentioned that the analogy to NWP predictability experiments can be improved by ensemble averaging (sections 3 and 4).

(iii) Finally, the mean squared error (error variance) of the individual persistence forecasts conditional on the initial anomaly is shown in Fig. 2 for a varying red-noise parameter, $a = 0.1, \ldots 0.9$ and $r = 1$. At zero initial anomaly (and forecasts initialized near the climate mean) there is no systematic error and the total forecast error is minimal.

(*f*) *Summary*

Individual persistence forecasts in the red-noise atmosphere are analytically analysed as a toy predictability experiment for an imperfect model. There is, in a qualitative sense, similarity with the results of external (or practical) predictability experiments based on NWP models:

(1) Persistence forecasts reach a saturation level which, as in NWP experiments, corresponds to the mean difference between two randomly chosen weather states; the error growth depends on the weather regime; and the conditional forecast errors are smaller

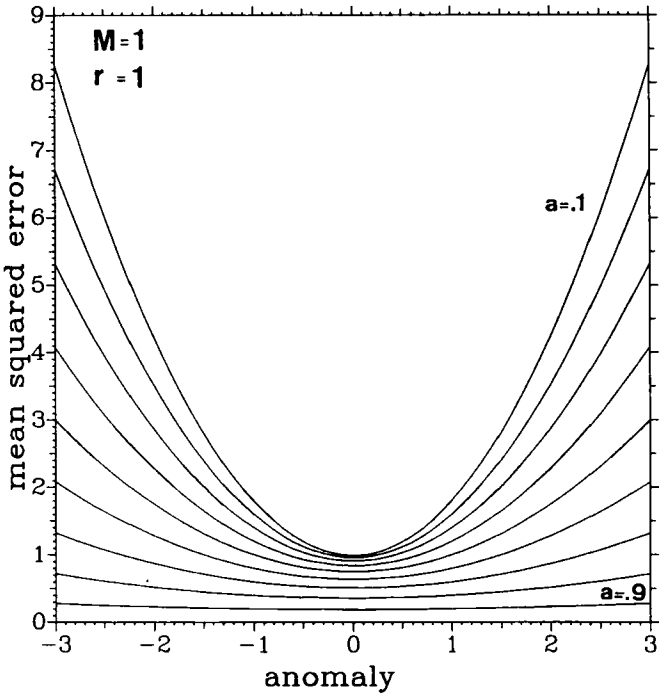


Figure 2. Mean squared error at lead time $r = 1$ of the individual persistence forecasts ($M = 1$) depending on the initial anomaly for different red-noise regimes, autocorrelation a.

when they are initialized close to the climate mean (or near zero anomaly). The error growth depends on the weather regime, with the predictability limit being proportional to the integral time-scale. Furthermore, the distinction between systematic and non-systematic errors allows direct insight into the complex relations within the forecast-verification system as, for example, occurring in the NWP-atmosphere predictability experiment. Finally, besides being analytical, this toy experiment has the additional advantage that weather-regime dependence (a = constant but variable) can be studied explicitly.

(2) Other simple forecast schemes do not show these features. Damped persistence, $X(t) = aX(t-1)$, for example, reaches only half of the saturation limit (natural variance), does not exhibit systematic errors and, therefore, conditional forecast errors are independent of the magnitude of the anomaly.

The predictability features of the persistence red-noise system make it an ideal toy for further analysis. Therefore, the predictability experiments are extended to ensemble forecasts to investigate the two main goals of ensemble forecasting: improvement *and* prediction of the forecast skill. Although older data in a first-order substitute atmosphere do not add (in the unconditional mean) to the prediction itself, they do conditionally and, in particular, to a spread–skill relation.

3. Ensemble-mean persistence forecasts

Ensemble-mean forecasts of one model with different initial conditions are performed to achieve three goals: to improve the forecast skill, to predict the forecast skill

and, ultimately, to provide a realistic probability distribution for expected atmospheric states. The motivation to achieve the first two aims is based on the theoretical perfect model–perfect ensemble scenario (Leith 1974). The basic mathematical background is easily deduced (see Brankovic *et al.* 1990).

Let F_i be an individual forecast by one member of the ensemble forecasts ($i = 1, \ldots, M$). For a given field variable X, the mean squared difference of X from the forecast members, F_i, gives

$$[(F_i - X)^2] = ([F_i] - X)^2 + [(F_i - [F_i])^2] \tag{3.1}$$

with $[(F_i - [F_i] + [F_i] - X)^2] = [([F_i] - X)^2] + [(F_i - [F_i])^2] + 2[(F_i - [F_i])([F_i] - X)]$. The last term vanishes and the first term is independent of ensemble averaging, which is denoted by $[F_i] = \Sigma_{i=1}^{M} F_i/M$.

First, with vanishing field variable, $X = 0$, (3.1) is confined to the forecast ensemble and the first two terms describe the ensemble variance or spread:

$$[F_i^2] - [F_i]^2 = s_M^2. \tag{3.1a}$$

Next, let the field variable X be the verification $X(t)$ of the forecast, then the first term describes the ensemble averaged squared error of the individual forecasts $F_i(r)$, $[e_i^2] = [(F_i - X)^2]$ where $e_i(r) = F_i(r) - X(t)$; the second term denotes the squared error of the ensemble-mean forecast $[F_i(r)]$, $e_M^2 = ([F_i] - X)^2$; and the last term is the spread or dispersion of the individual forecasts from the ensemble mean, $s_M^2 = [(F_i - [F_i])^2]$, which is a forecast variance. This quantifies the average improvement of the ensemble-mean forecast, $[F_i]$, over the individual members in terms of the mean squared error:

$$[e_i^2] = e_M^2 + s_M^2. \tag{3.1b}$$

This improvement is achieved by removing parts of the variance, namely s_M^2, from the forecast error $[e_i^2]$. 'Although true in a least squares sense, this is somewhat misleading since part of the reduction in error variance is caused simply by the reduction in anomaly intensity (smoothing) resulting from formation of an ensemble mean. Smoothing a forecast does not by itself improve the signal-to-noise ratio. The real benefit lies in the fact that, compared with an individual forecast, the ensemble mean is a better estimate of the true state' (Murphy 1988).

Finally, corresponding to the ensemble mean of the individual squared errors, $[e_i^2] = \Sigma_{i=1}^{M} (F_i - X)^2/M$, one defines the ensemble mean of the squared distances between all non-identical pairs of individual forecasts, $[d_i^2] = \Sigma_i \Sigma_j (F_i - F_j)^2/M(M - 1)$. Substituting F_j for X in (3.1), summing over the whole ensemble and dividing by $M - 1$, one obtains

$$[d_i^2] = \frac{2M \cdot s_M^2}{M - 1}. \tag{3.1c}$$

(*a*) *Perfect model/ensemble environment*

A perfect ensemble ($r = 0$) and perfect model ($r > 0$) hypothesis can be introduced:

(i) The perfect ensemble hypothesis ($r = 0$) assumes that the ensemble members are chosen such that initially the spread amongst them is representative of the initial analysis error. Then $[e_i^2] = [d_i^2] = 2M s_M^2/(M - 1)$ at zero lead time.

(ii) The perfect model hypothesis ($r > 0$) assumes that during the time evolution the growth of the mean distance between the members of the ensemble d_i^2 is equal to the

average growth of the internal deterministic errors. Then $[e_i^2] = [d_i^2] = 2Ms_M^2/(M-1)$ holds also for increasing lead time $r > 0$. Now, combination with (3.1b) and (3.1c) yields the forecast error of the ensemble mean, e_M^2, related to the mean error of the individual forecasts and the ensemble variance (spread);

$$\left.\begin{aligned} e_M^2 &= [e_i]^2 \frac{M+1}{2M} \sim \frac{[e_i^2]}{2} \\ e_M^2 &= s_M^2 \frac{M+1}{M-1} \sim s_M^2 \text{ for large } M. \end{aligned}\right\} \tag{3.2}$$

That is, for sufficiently large ensemble size M, one obtains (from 3.2) the theoretical perfect model/perfect ensemble limit for the skill of an ensemble forecast (Leith 1974): the mean squared error of an ensemble forecast is half the average mean squared error of the individual members of the ensemble (first goal). Furthermore, there is a linear relation between forecast error e_M^2 of the ensemble mean and the ensemble spread s_M^2, (second goal). Note that (time) averaging $\langle\rangle$ over the forecast samples will be employed to analyse (later in this section) the statistics of predictability experiments in the imperfect model/ensemble environment.

(*b*) *Imperfect model/ensemble*

Practical forecasts are based on imperfect models and the related ensembles are also imperfect as their initial spread is not a direct measure of the analysis error. In the following we simulate the imperfect model–imperfect ensemble environment by the persistence model and a set of M time lagged (unweighted) persistence forecasts as the imperfect ensemble:

$$[F(r+i)] = \sum_{i=0}^{M-1} F(r+i)/M. \tag{3.3}$$

The forecasts are verified at time t and issued at $t-(r+i)$; here the square brackets, [ ], define the average over the lagged forecast ensemble; note that the counting index i commences at $i = 0$ to include the latest forecast $F(r)$ in the ensemble. Furthermore, the M-lagged ensemble mean persistence can be reformulated as a persistence plus equally weighted trend model, $[F(r+i)] = F(r) + \{F(r+1) - F(r)\}/M \ldots + \{F(r+M-1) - F(r)\}/M$ (see section 6 for error–spread relation).

(*c*) *Error budget*

The error budget is determined by the time or sample average $\langle\rangle$ of the squared forecast errors $E_M(r) = \langle e_M^2\rangle = \langle\{X(t) - [F(r+i)]\}^2\rangle = \langle X^2\rangle + \langle[F(r+i)]^2\rangle - 2\langle X[F(r+i)]\rangle$. Leaving these three terms in the same order we obtain (appendix A):

$$E_M(r) = s_x^2\left(1 + \frac{1+a}{M(1-a)} - \frac{2a(1-a^M)}{M^2(1-a)^2} - \frac{2a^r(1-a^M)}{M(1-a)}\right) \tag{3.4}$$

and the growth rate of the error variance, $\mathrm{d}E_M(r)/\mathrm{d}r = E_{Mr}$:

$$E_{Mr}(r) = s_x^2 \ln\left(\frac{1}{a}\right)\frac{a^r(1-a^M)}{M(1-a)}. \tag{3.5}$$

The related distribution densities of the error and the squared error, g and f, can easily be derived in analogy to those of the individual persistence forecasts (2.5) using the random variable $y = (X-F)/\sqrt{E_M(r)}$. Figures 3(a) and (b) show the error budget for

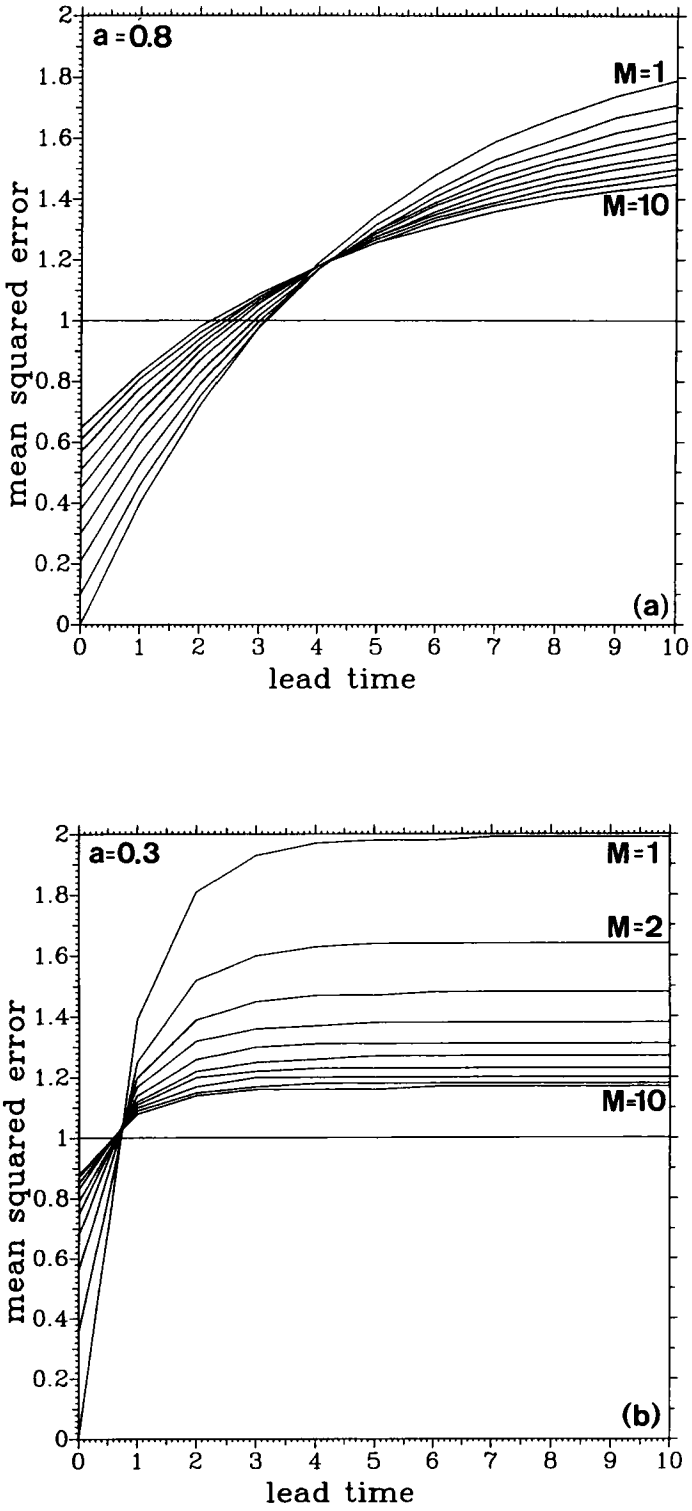


Figure 3. Ensemble-mean error budget, $E_M(r)$. The error variance of ensemble averaged lagged persistence forecasts for varying ensemble size $M = 1$ to 10 in red-noise atmospheres with the autocorrelation $a = 0.8$ (a) and $a = 0.3$ (b).

the individual and the ensemble averaged forecasts ($M = 1$ to 10) evolving with lead time, r, for red noise with the autocorrelation $a = 0.8$ and 0.3, respectively. The following results are of interest:

(i) Forecasts of fixed ensemble size M realize small (large) initial errors $E_M(r = 0)$, and small (large) growth rates $E_{Mr}(r = 0)$ in red-noise regimes of large (small) autocorrelations, a, or time scales, τ, because these regimes are associated with small (large) stochastic white-noise forcing (compare Fig. 4(a) with 4(b)). This behaviour will also be discussed in section 5 for an ensemble of $M = 2$ lagged persistence forecasts and in connection with the regime related error budgets of NWP predictability experiments.

(ii) At small lead times r, and large autocorrelation values a, the ensemble-mean forecast error variance, E_M, grows with ensemble size M. However, if the lead time is sufficiently large so that the forecasts enter the region of saturation, the error variances E_M decrease with growing ensemble size M.

(iii) Before reaching the predictability limit, unweighted lagged ensemble-mean forecasts appear to be always worse than the single forecast, in some general agreement with NWP experiments (Hoffman and Kalnay 1983; Dalcher *et al.* 1988; Tracton *et al.* 1989; Brankovic *et al.* 1990). This only holds in the unconditional sample average and up to a lead time of about ten days (that is near the limit of predictability (Tracton *et al.* 1989)). However, the skill of ensemble-mean forecasts depends on initial conditions, ensemble size and red-noise regime, so that ensembles improve over the latest single forecasts under favourable conditions (section 5). Here it should be noted that optimally weighted lagged ensemble forecasts (with recent forecasts having larger weights) improve in skill over the latest ensemble member even unconditionally. However, for skill prediction by ensemble spread, the forecast spread correlates more strongly with the uniformly weighted ensemble forecast skill than with the optimally weighted ensemble forecast skill (Palmer and Tibaldi 1988).

(iv) The predictability limit T, given by $E_M(r = T) = s_x^2$:

$$T = \frac{\ln 2}{\ln(1/a)} - \frac{\ln\left(\dfrac{1+a}{1-a^M} - \dfrac{2a}{M(1-a)}\right)}{\ln(1/a).} \tag{3.6}$$

For both $M = 1$ and $M = 2$ the limit of predictability is $T = \ln 2/\ln(1/a)$; for increasing ensemble size, $M \to \infty$, the ensemble-mean forecasts approach the limit of predictability, $T = \{\ln 2/(a + a)\}/\ln(1/a)$ at $E_M = s_x^2$, while the initial error of the mean, $[F_i(r = 0)]$, also approaches the climate variance: $E_M(r = 0) \to s_x^2$ for $M \to \infty$.

A final comment on the predictability limit and its initial error dependence may be relevant for NWP predictability analyses. It can be observed in Figs. 3(a) and (b) that the limit of predictability depends on the threshold of error variance chosen for its definition (here we used the climate variance, s_x^2). A value larger than the climate variance, $E_M(r = T) > s_x^2$, does not only lead to different numbers but it may also reverse the relation between the decreasing predictability limit and the increasing initial error. This holds independently of whether the variability of the initial error, $E_M(r = 0)$, is due to changing the ensemble size M or due to the changing autocorrelation a. Increasing the predictability threshold from s_x^2 to As_x^2 (with $1 < A < 2$) can lead to a qualitatively different behaviour of the associated predictability limit T_A given by $E_M(r = T_A) = As_x^2$. Increasing initial errors $E_M(r = 0)$ can, therefore, lead to increasing predictability limits, T_A. This is plausible, because initial, $E_M(r = 0)$, and saturation error

variance, $E_M(r \to \infty)$, depend in an opposite way on the autocorrelation regime. Thus the limit T at higher (smaller) predictability threshold values is influenced by the saturation (initial) error; in extreme cases the predictability limit may not even be reached (that is $T \to \infty$). One may apply this result to the real atmosphere, where regimes of large time-scales τ, or high autocorrelations a, occur in more or less irregular alternation with regimes of the opposite character. If regimes of shorter time-scale are also related to smaller saturation values $E_M(r \to \infty)$, and vice versa, a relatively large constant and not regime-related predictability threshold might lead to an overestimation of the average predictability limit.

(*d*) *Anomaly correlation*

The anomaly correlation coefficient serves as another measure of skill. For an individual persistence forecast, $F_1(r) = X(t - r)$, the anomaly correlation between forecast and verification is, $A_1(r) = \langle X(t)\, X(t - r)\rangle / s_x^2 = a^r$, and the forecasts by an ensemble mean, $[F(r + i)] = \Sigma_{i=0}^{M-1} X\{(t - (r + i)\}/M$ of M members yields the anomaly correlation (see appendix A):

$$A_M(r) = \frac{\langle X(t)\,[F(r + i)]\rangle}{s_x\,\{\langle [F(r + i)]^2\rangle\}^{1/2}} = \frac{a^r(1 - a^M)}{\{M(1 - a^2) - 2a(1 - a^M)\}^{1/2}} \tag{3.7}$$

where $s_F^2 = \langle [F(r + i)]^2\rangle - \langle [F(r + i)]\rangle^2 = s_x^2[(1 + a)/(M(1 - a)) - 2a(1 - a^M)/(M^2(1 - a)^2)]$; for $M = 1$, $A_1 = a^r$. From a practical point of view it is important to consider whether the forecast by the latest ensemble member (or control forecast, Brankovic *et al.* (1990)) is superior to the ensemble mean. As in their NWP model experiments we observe that 'the unweighted lagged-average forecast gives no significant advantage over the single (deterministic) forecast from the latest initialization date'. For lagged persistence forecasts this is easily deduced from the ratio between anomaly correlation of the latest, A_1, and the ensemble-mean forecasts, A_M:

$$\frac{A_1}{A_M} = \frac{\{M(1 - a^2) - 2a(1 - a^M)\}^{1/2}}{1 - a^M} > 1 \text{ for } M > 1. \tag{3.8}$$

Note that the latest member of the lagged persistence ensemble gives the best short-term forecast only in the average, because the latest persistence forecast initialized by an extreme anomaly is not necessarily better than the related ensemble-mean prediction. Thus forecast errors depending on the magnitude of the anomaly at the initial condition will be analysed in the next section.

4. Conditional, systematic and non-systematic errors

Error growth has been observed to depend on the state from which the forecast is initiated. Recent general circulation model (GCM) experiments (Molteni and Tibaldi 1990; Molteni *et al.* 1990) have shown that error distributions may tend towards a bimodal density of the error variance if the initial state in phase space is situated near the boundary of two weather-regime basins. That is the time evolution of the forecast error statistics can be employed as a diagnostic tool to extract information on the dynamical properties of the atmospheric phase space. In this sense conditional error (variance) distributions are a useful extension of the error-budget analysis, which has not been considered in the previous section on ensemble forecasting where only time averages, $\langle\rangle$, over all initial anomalies of the red-noise regimes were considered.

The error variance conditional on the initial anomaly, X_0, from which the forecast starts, can be derived analytically (see appendices A and B):

$$E_M(r|X_0) = \langle X_0^2 \rangle \left(a^r - \frac{1-a^M}{(1-a)M} \right)^2 + \\ + s_x^2 \left(1 - a^{2r} + \frac{1+a}{(1-a)M} + \frac{1-a^{2M}}{(1-a)^2M^2} - \frac{2(1-a^M)(1+a)}{(1-a)^2M^2} \right). \quad (4.1)$$

Note that formally the average is to be taken over all realizations of the red-noise atmosphere given the same starting point X_0, but the stochastic forcing remaining random. This conditional formulation will be of further use when analysing the spread–skill relation (section 6) depending on the initial anomalies, X_0.

Figures 4(a) and (b) show the mean squared errors $E_M(r|X_0)$, and Figs. 4(c) and (d) the skill, $1 - E_M(r|X_0)/E_1(r|X_0)$, conditional on the initial anomalies, X_0, for varying ensemble size M in the red-noise regime $a = 0.8$. The following results substantiate the results of section 3. The smallest error is attained when persistence forecasts commence at situations of small anomalies about the climate mean. This result is independent of the autocorrelation, a, and has also been found in the real atmosphere (van den Dool 1989) when analysing past-weather analogues; it may also hold for NWP forecasts. Given small anomalies the individual persistence forecasts are better than ensemble forecasts for short lead times, $r = 1$ (Fig. 4(a)). However, the ensemble forecasts improve over the individual forecasts at larger lead times (Fig. 4(b) for $r = 6$) when the conditional anomaly, X_0, is large. Note that this occurs near the predictability limit. However, at smaller lead times one observes an additional constraint by the ensemble size M. That is short-term forecasts with too large an ensemble M do not gain skill (over the predictions by a smaller ensemble) at certain conditions of a, M, r and X_0. Figures 4(c) and (d) show the skill of the ensemble forecasts taking the individual persistence forecast as reference, $M^* = 1$, in the red-noise flow $a = 0.8$ for lead times $r = 1$ (and 6). One observes that for sufficiently large initial anomalies, X_0, ensemble-mean forecasts can be superior to the latest individual forecast which, however, depends on the regime a, the ensemble size M, and the lead time r.

For unconditional situations, that is averaged over all conditions $\langle X_0^2 \rangle = \langle X^2 \rangle = s_x^2$, the error variance (3.4) is recovered. Then (4.1) can be used to distinguish between the systematic $SE_M(r)$ and random or non-systematic contribution $RE_M(r)$ to the total error $E_M(r)$:

$$\left.\begin{aligned} SE_M &= s_x^2 \left(a^r - \frac{1-a^M}{(1-a)M} \right)^2 \\ \frac{dSE_M}{dr} &= 2s_x^2 \left(a^r - \frac{1-a^M}{(1-a)M} \right) a^r \ln \frac{1}{a} \\ RE_M &= s_x^2 \left(1 - a^{2r} + \frac{1+a}{(1-a)M} + \frac{1-a^{2M}}{(1-a)^2M^2} - \frac{2(1-a^M)(1+a)}{(1-a)^2M^2} \right) \\ \frac{dRE_M}{dr} &= 2s_x^2 a^{2r} \ln \frac{1}{a} \end{aligned}\right\} \quad (4.2)$$

and the respective growth rates. Figures 5 and 6 display the following structures of the systematic and the random contributions to the total error budget (4.2) changing with ensemble size M and lead time r in dependence of the regime a:

(i) For lead times $r > 0$ the mean systematic error is generally smaller than the non-systematic contribution. Furthermore, the systematic errors SE (non-systematic errors

RE) attain minima (maxima) at the ensemble sizes M, whose magnitudes and positions are related to the autocorrelation regime. The larger the autocorrelation (Fig. 5(d)) the smaller (larger) the ensemble size for $SE = \min.$ ($RE = \max.$). Furthermore, the non-systematic error variance magnitudes at the extrema decrease with increasing autocorrelation (Fig. 5(e)). The initial ($r = 0$) systematic error deviates from the structures observed for $r > 0$; it increases continuously with decreasing autocorrelation a and growing ensemble size (Fig. 5(a)).

(ii) The optimal ensemble size M^* generates ensemble-mean forecast errors that are almost completely determined by the non-systematic error contributions (Fig. 5). That is the systematic error vanishes, and model deficiencies have the least influence on the error growth that affects the error–spread relation and the skill forecasting.

(iii) At a fixed ensemble size M the systematic errors can pass through a minimum when increasing the lead time from $r = 0$, so that a 'return of skill' (Fig. 6(d)) may be possible under favourable conditions. The initial anomaly X_0 needs to be larger than a standard deviation of the fluctuations of the weather regime, and the ensemble should be of sufficient size to warrant a substantial decrease of the systematic over the increasing random-error contribution. However, this return of skill with negative systematic error growth rates commences at lead time $r = 0$. A similar behaviour can be observed in NWP predictability experiments (Tracton *et al.* 1989) occurring at larger lead times. The reason may be similar. A change from a weather regime with relatively poor model performance (or large systematic error) to a regime of opposing characteristics may be realized by the model as a large anomaly, so that its systematic error returns to smaller values before rising again. This is observed in situations when the circulation moves to a state that is closer to the lagged average than to the initial condition. In the atmosphere it occurs under seemingly fortuitous circumstances, when similar anomalous situations are recurrent; this is not due to the model's ability to simulate the relevant circulation. There is no return of skill for the two-lagged ensemble forecasts, because the systematic error growth is positive for all $0 < a < 1$ and the systematic error remains constant for $r = 0$ and $r = 1$ (see Fig. 6(b)).

(iv) The growth rate of the non-systematic error, $\mathrm{d}RE(r)/\mathrm{d}r = 2s_x^2 a^{2s} \ln(1/a)$, depends only on the autocorrelation regime a, and the lead time r, and not on the ensemble size M. This is plausible because the non-systematic forecast error represents the perfect model conditions, when initially close trajectories diverge at a rate which is determined solely by the internal dynamics represented by the white-noise forcing in the red-noise atmosphere.

5. Two-lagged persistence forecasts: a parametrization

The error budgets of individual persistence forecasts are a poor parallel to common predictability analyses. There is no initial error and the forecast-verification trajectory pair to be analysed is degenerated to a single trajectory analysis. In this section a forecast model is introduced which is generated by a model of similar simplicity as the individual persistence forecast, but is closer to the real situation as it allows for initial errors. It may also serve as an illustrative example for an ensemble mean of lagged forecasts. Dalcher *et al.* (1988) experienced that an ensemble of only two members provided a good prediction of the skill of the individual forecasts 'for regions which are not large enough to contain many mid-latitude cyclones (say, the size of North America or Europe)'. Let F_1 and F_2 be a forecast pair commencing one time lag apart and running over a forecast range r in parallel to be verified at the same time t, then:

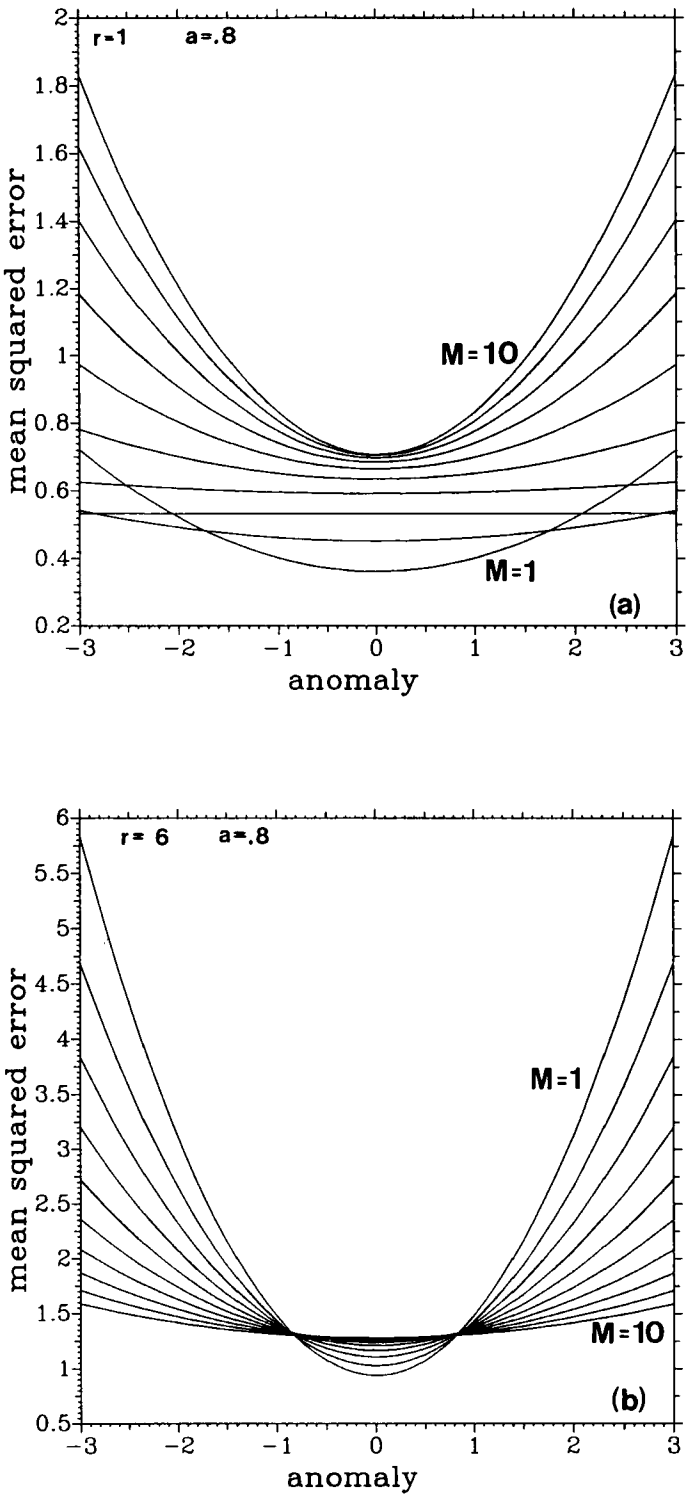
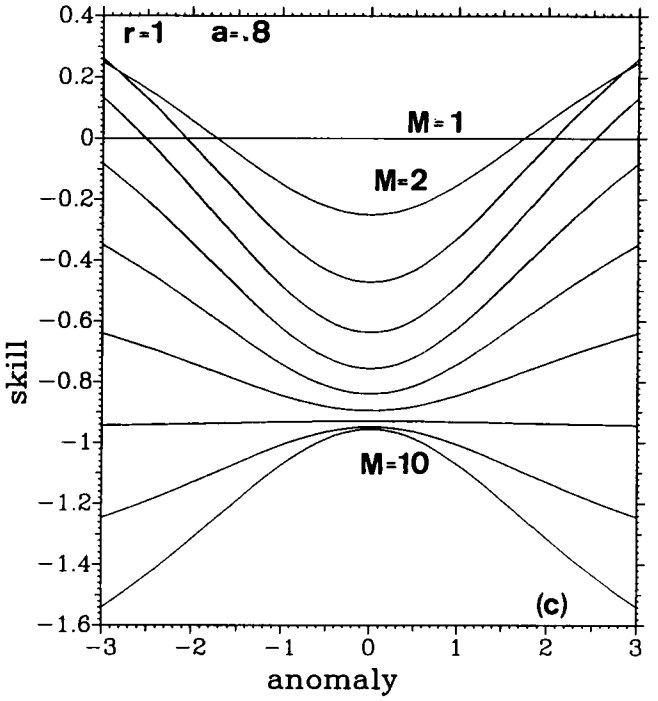


Figure 4. Ensemble-forecast errors, $E_M(r|X_0)$, conditional on initial anomalies, X_0 at $r = 0$, in the red-noise autocorrelation regime $a = 0.8$ for the ensemble size varying from $M = 1$ to 10 at lead time $r = 1$ (a) and $r = 6$ (b). For completeness, the skill is also shown for $r = 1$ (c) and $r = 6$ (d).



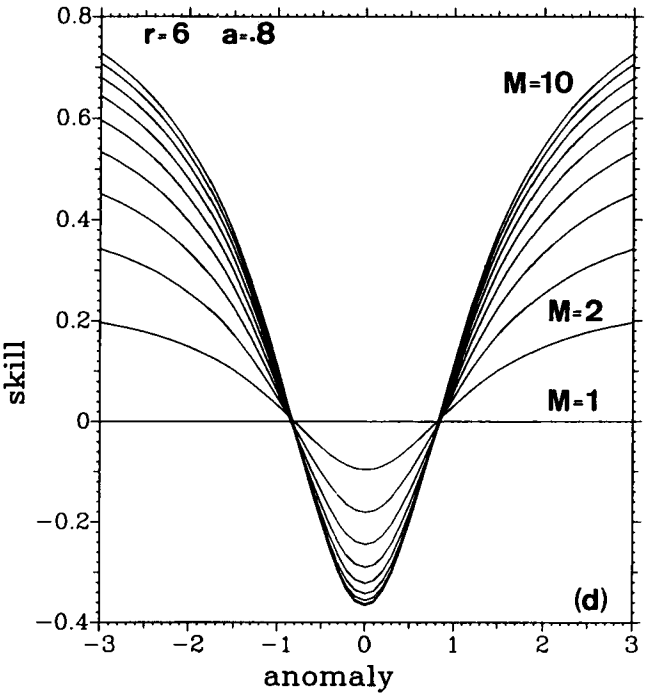


Figure 4. Continued.

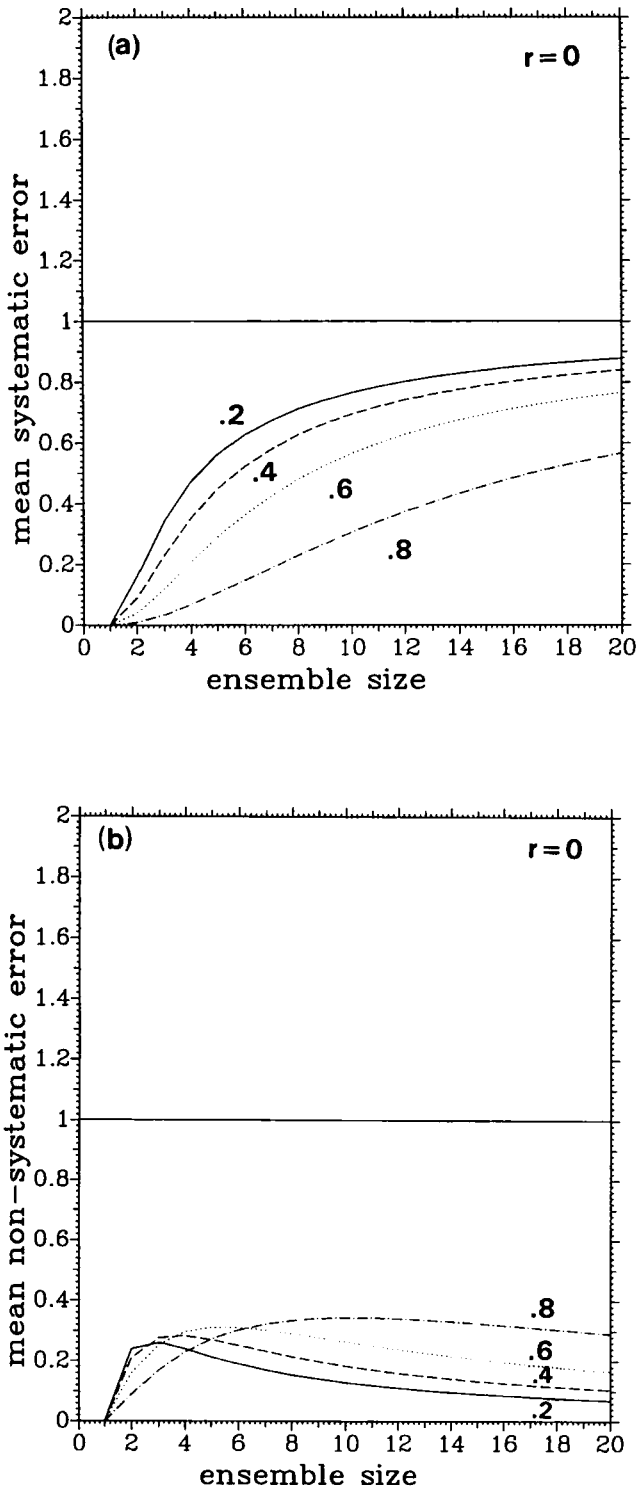
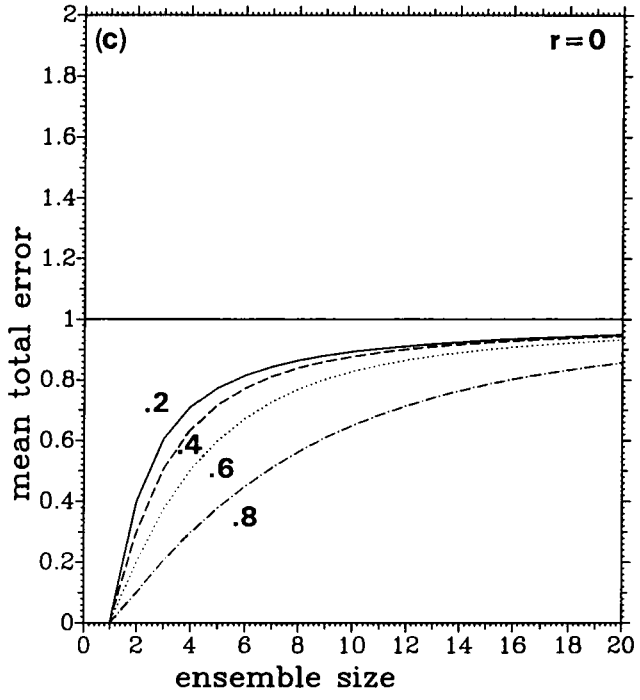


Figure 5. Ensemble-mean predictions: (a) and (d) systematics, (b) and (e) non-systematic and (c) and (f) total error variance changing with ensemble size at different lead times r ((a), (b) and (c): $r = 0$; (d), (e) and (f): $r = 1$) and for various autocorrelation regimes $a = 0.2, 0.4, 0.6, 0.8$.



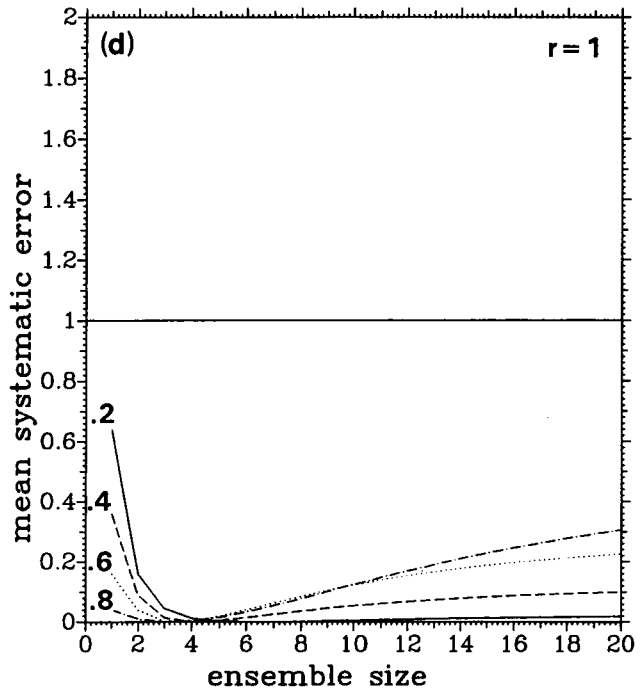
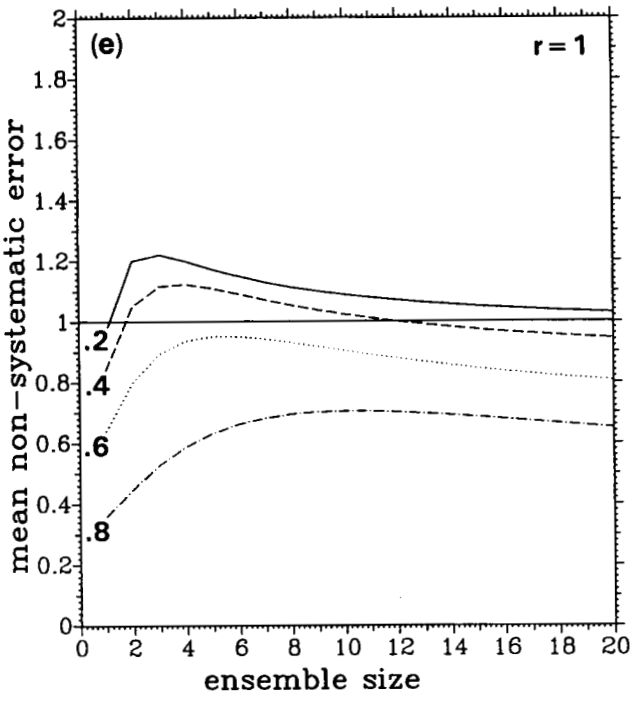


Figure 5. Continued.



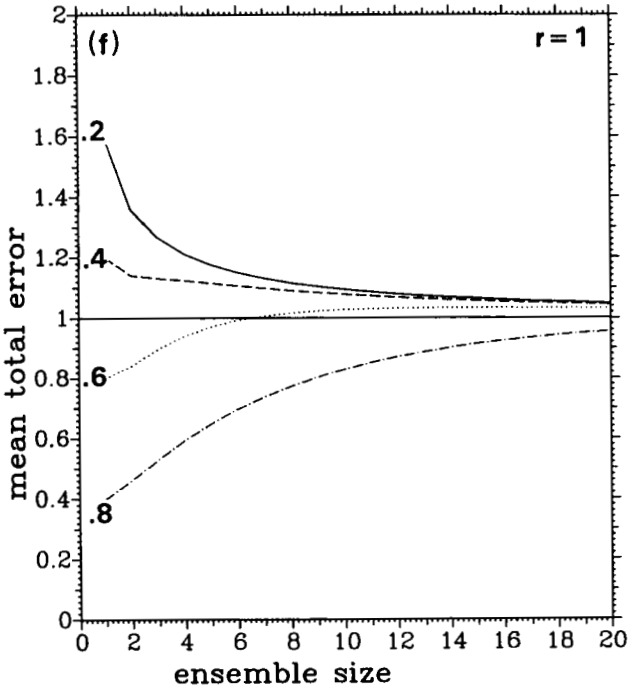


Figure 5. Continued.

$$\left.\begin{aligned} F_1(r) &= X(t-r) = X(t_0) \\ F_2(r) &= X\{t-(r+1)\} = X(t_0-1). \end{aligned}\right\} \tag{5.1a}$$

The lagged forecast pair can be combined to an ensemble averaged, [ ], forecast $F(r)$. It consists of $M = 2$ members:

$$F = [F_i] = (F_1 + F_2)/2 = F_1 + (F_2 - F_1)/2. \tag{5.1b}$$

This forecast can also be interpreted as a persistence plus half-trend prediction. In long-range forecasting, for example, a persistence plus analogue–trend scheme is successfully used (Livezey *et al.* 1990). Because of its initial errors this scheme may be more suitable for analysing error budgets than the individual persistence forecasts. In particular, the predictability limits of both forecasts are the same (see below).

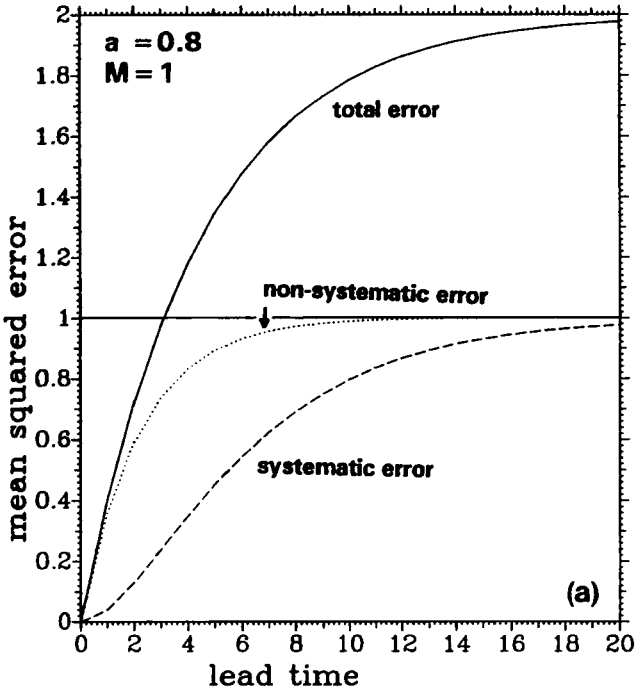
(*a*) *Error budget*

Setting $M = 2$ in (4.6) gives the error budget of ensemble-mean forecasts of a pair of two-lagged consecutive persistence predictions, $E_2(r) = \langle\{X(t) - F(t)\}^2\rangle$, and its rate of change, $\mathrm{d}E/\mathrm{d}r$, varying with lead time:

$$\left.\begin{aligned} E_2(r|X_0) &= \langle X_0^2\rangle\left\{a^r - \frac{1+a}{2}\right\}^2 + s_x^2\left\{1 - a^{2r} + \frac{1-a^2}{4}\right\} \\ E_{2r}(r|X_0) &= 2\langle X_0^2\rangle\left\{\frac{1+a}{2} - a^r\right\}a^r \ln\frac{1}{a} + 2s_x^2\, a^{2r} \ln\frac{1}{a}. \end{aligned}\right\} \tag{5.2}$$

For unconditional forecasts, $\langle X_0^2\rangle = s_x^2$, the error variance is $E_2(r) = s_x^2\{1 + (0.5 - a^r)(1-a)\}$. The magnitudes of the initial error $E_2(r=0)$, and the saturation error $E_2(r\to\infty)$ are bounded: $0 < E_2(r=0) = s_x^2(1-a)/2 < s_x^2$ and $1.5s_x^2 < \mathrm{E}(\mathrm{r}\to\infty) = s_x^2\{1.5 + a/2\} < 2s_x^2$. Their values depend on the autocorrelation regime a in an inverse sense; for increasing time-scale τ (or a) the initial error is reduced whereas the saturation level is enhanced and vice versa. To complete the error budget the Gaussian distributions of the error and the chi-squared distributions of the squared error are deduced in analogy to the individual persistence forecast; that is, replacing the forecast error variance, E, in (2.5) by the lead time r dependent error variance, $E_2(r)$, and using the normalized error random variable $y = (X - F)/\sqrt{E_2(r)}$. From this information the quantiles of the distribution can be calculated. Figure 7(a) presents the error budget and the quantiles of the error distribution. Figure 7(b) exhibits the sensitivity on changing autocorrelation. These measures show a behaviour which appears to be similar to the individual persistence-forecast model. There are, however, differences. At zero lead time the initial error does not vanish, and for large lead times a saturation value, $E(r\to\infty)$, is reached that is smaller than that attained by individual persistence predictions (due to ensemble averaging). In addition, the limit of predictability T of the two-lagged forecast is the same as for the individual persistence prediction (2.7). At this limit the systematic error attains the value $SE(r=T) = a^2 s_x^2/4$, the non-systematic error is $RE = 1 - SE$.

The rate of error growth, $\mathrm{d}E/\mathrm{d}r = E_r$, gives more details. The initial rate of unconditional error growth, $E_{2r}(r=0) = s_x^2(1+a)\ln(1/a)$ increases with increasing magnitude of the mean initial error, $E_2(r=0) = s_x^2(1-a)/2$, which depends on the weather regime a. For $= 0.8$ (or 0.3) the initial growth is 0.4 (or 1.6) per unit time step, which is smaller than for individual forecasts; note, however, the mean initial error of 0.1 (or 0.35) s_x^2. As weather regimes of large (small) autocorrelation $a = (1 - 2E_2(r=0)/s_x^2)$ or time-scale τ are related to small (large) white-noise fluctuations $s_z^2 = s_x^2(1-a^2)$, forecasts made in these regimes realize small (large) initial errors, $E_2(r=0)$ and small (large)



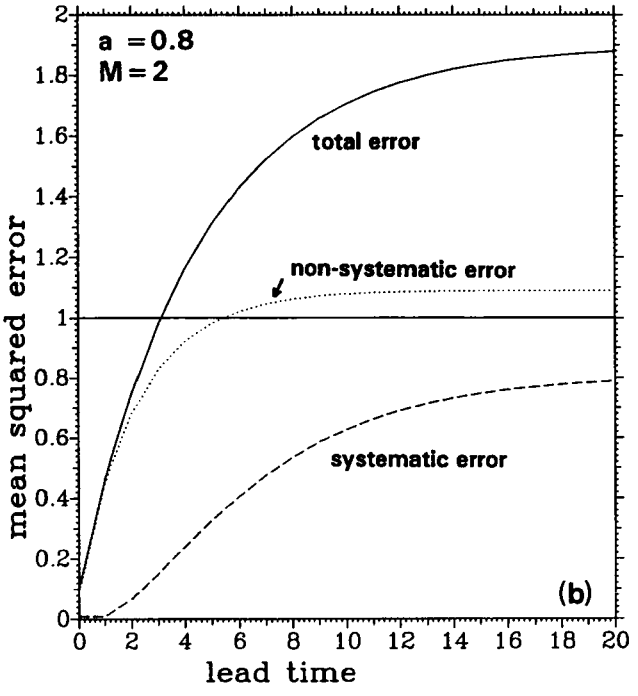
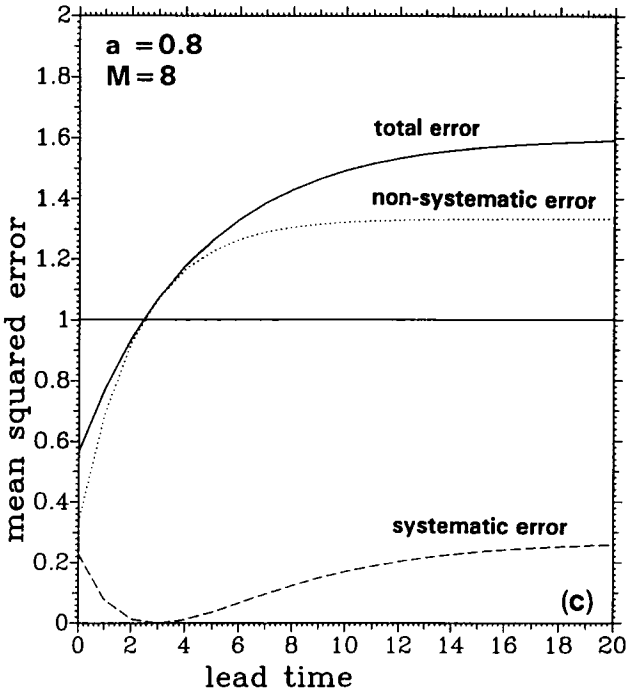


Figure 6. Ensemble-mean predictions: Systematic, non-systematic and total error variance changing with lead time r for different ensemble sizes, (a) $M = 1$, (b) $M = 2$, (c) $M = 8$, in the autocorrelation regime $a = 0.8$. Examples for the 'return of skill' phenomenon are shown in (d).



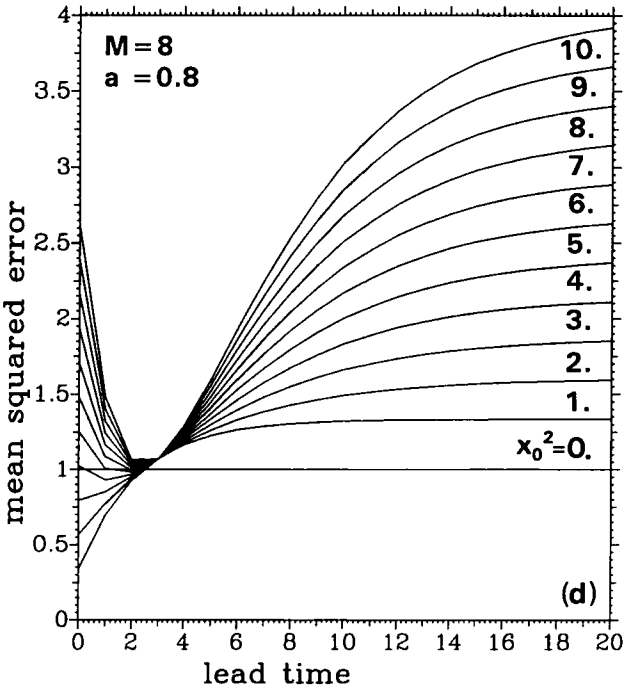


Figure 6. Continued.

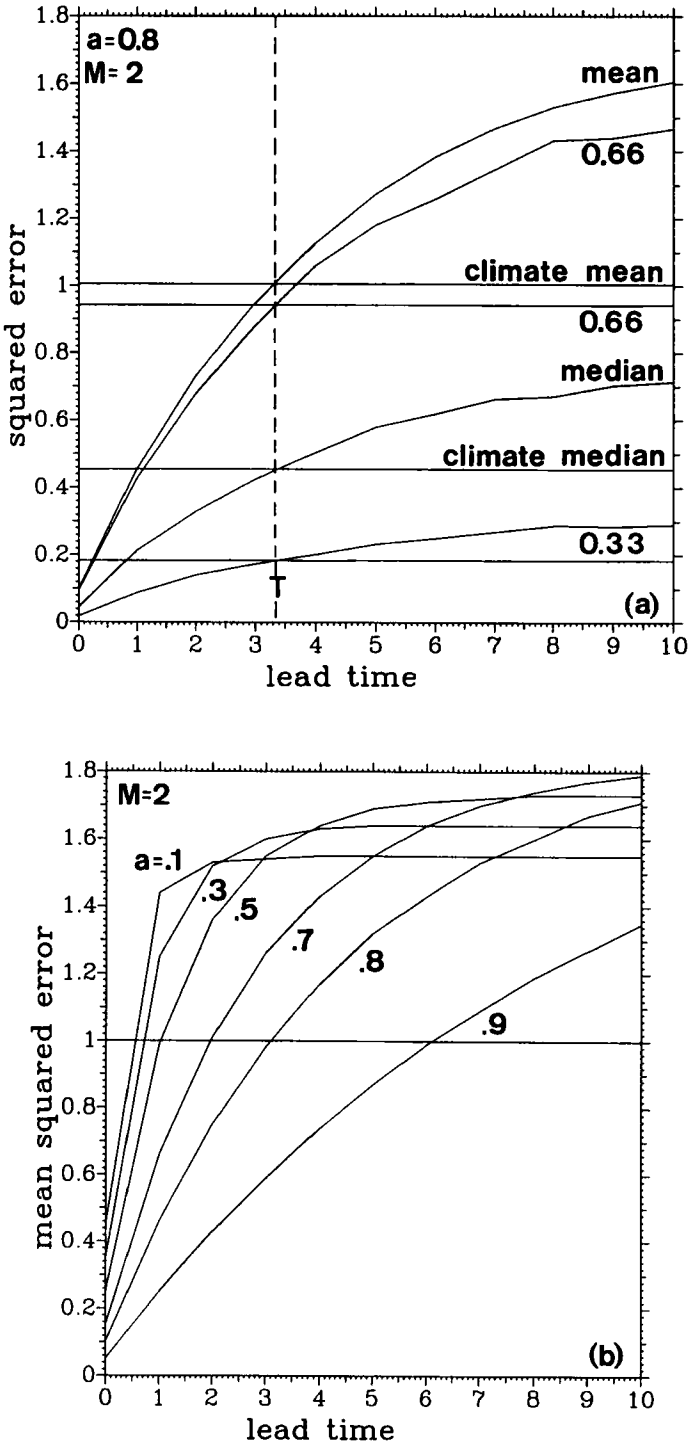


Figure 7. Error budget of the average of two-lagged persistence forecasts in a red-noise atmosphere (autocorrelation $a = 0.8$). (a) The time evolution of the error variance, the median and the upper and lower terciles (0.66; 0.33) of the squared error distribution. The same is shown for the climate-mean forecasts (horizontal lines). The predictability limit T, is also indicated. (b) The lead-time-dependent error variance for various red-noise atmosheres: $a = 0.1$ to 0.9; note the associated initial errors.

initial error growth rates, $E_{2r}(r=0)$; and vice versa. In the limit $a \rightarrow 1$, infinitely large values of $E_{2r}(r=0)$ are approached. Such error-growth behaviour is reported from both NWP predictability experiments and historical-weather analogue analyses (Horel and Roads 1988, Fig. 9; Chen 1989, Figs. 7 and 9; Toth 1991, Fig. 5). One can only guess about the cause of the phenomenon. The structure of the same magnitude of error, its wave number of geographical distribution, might be different with various initial errors. Indeed, the magnitude of the initial error, $E_2(r=0)$, is associated with the weather regimes in the substitute red-noise atmosphere where the autocorrelation, a, or the integral time-scale, τ, characterize the dynamics (Fig. 7(b)) in the following sense: the larger the autocorrelation, the smaller the stochastic forcing, s_z^2, the smaller the initial error, $E_2(r=0)$, and, subsequently, the smaller the initial rate of error growth, $E_{2r}(r=0)$, and vice versa. As described in section 4 this holds for $M>1$. These considerations lead directly to a parametrization of the error budget to derive the upper bound of the NWP predictability limit (Chen 1989).

These qualitative results of NWP perfect-model predictability experiments are complemented by the imperfect model-verification (persistence – red-noise) system. Separation of the systematic from the non-systematic error variances and their associated growth rates (first and second term on the right-hand side of (5.2), setting $\langle X_0^2 \rangle = s_x^2$, normalized by the respective saturation values $E_2(r \rightarrow \infty)$) shows that the non-systematic error contributions exhibit qualitatively the same sensitive dependence on the weather regimes as revealed by the perfect NWP experiments. The initial non-systematic error (and its growth rate), which characterize the 'dynamics' unaffected by model errors, increase with decreasing integral time-scale of the weather regime; however, the initial growth of the systematic error remains unchanged (or reverses sign for $M>2$) while the initial error grows with decreasing time-scale. In comparison, Chen (1989, Fig. 5) shows that the small-scale dynamics of NWP models (with wave numbers >18) are related to larger initial errors and growth rates than the large-scale processes. In the non-systematic (and total) error budget of persistance in red noise, the small (scale) integral time-scale is associated with relatively strong random forcing which leads to larger initial errors and growth rates compared with processes of large integral time-scale (see Schubert and Suarez (1989) for error-budget modelling of the combined effect).

(*b*) *Parametrization*

Error-budget models like that of the constrained population growth (Lorenz 1969; for a review see Stroe and Royer (1993)) are constructed to estimate upper and lower bounds of the predictability limied and to determine error doubling times of infinitesimally small errors. These error-budget models are fitted to internal and external predictability experiments by NWP models, using finite perturbed initial conditions, time-lagged forecasts (Hoffman and Kalnay 1983) or observed past-weather analogues (Lorenz 1969). Such experiments provide a set of (relatively large) initial errors, $E(r=0)$, which are associated with a set of predictability limits, T. Now, the upper bound of the predictability limit is defined by a hypothetical limit reached under the assumption of vanishing initial error: $E(r=0) \rightarrow 0$. Based on NWP and past-weather-analogue predictability experiments, a linear extrapolation of the limit of predictability, T, towards zero initial error is suggestive, although very small initial errors may happen to be associated with anomalously large predictability which would not fit such a linear (but a geometric) relation. Chen (1989, Fig. 8) has introduced this parametrization of the error budget, examining NWP predictability experiments; his analysis has been further substantiated

by Toth (1991, Fig. 2) using past-weather analogues. The linear initial error versus predictability limit parametrization has been developed, because the commonly used Verhulst-type error-growth models (adopted from population dynamics) tend to underestimate predictability (Toth 1991) when its parameters are fitted to all error data irrespective of lead time. This linear predictability/initial-error parametrization will be discussed using the persistence plus half-trend forecast.

Figure 8 displays the predictability versus initial-error relation (using (3.4) and (3.6)) for the lagged persistence forecasts using unit lag and $M = 2$. Assume autocorrelations (representing weather regimes) to be uniformly distributed over the a-interval $(0, A)$ and to remain unchanged during the prediction until the limit of predictability has been reached, $r > T = \ln 2/\ln(1/a)$, then an average limit of predictability can be defined:

$$T^* = A^{-1} \int_0^A T \mathrm{d}a' = -A^{-1} \ln 2 \, \mathrm{Li}(A)$$

with the logarithm-integral, $\mathrm{Li}(A)$ (Gradshteyn and Ryzhik 1980). Now regime averaged predictability limits T^* can be calculated; for $0 < a < A = 0.9$ one obtains the average predictability limit $T^* \sim 2.5$ unit time steps. This analytically determined value compared favourably with the linear extrapolation (see Fig. 8) as suggested by the parametrization scheme. The coincidence is not surprising; the datasets, from which the parametrization has been originally deduced, are taken from both NWP and past-weather-analogue predictability experiments, which cover only relatively large initial errors (Chen 1989; Toth 1991). It is equivalent to not letting the autocorrelation regime reach the upper limit $a \to 1$, where the stochastic forcing becomes relatively small (compared with the deterministic part) and likewise the initial errors. This is the case in regimes of large time-scale, small stochastic fluctuations, $0 < E_2(r =) < 0.15 s_x^2$, and large predictability T, where the initial error/predictability-limit relation is far from linear.

A final comment on the interpretation of error budgets is in order before proceeding to the error–spread relation. Within a fixed red-noise regime, a = constant, small (large) error magnitudes, $E(r)$, are related with large (small) growth rates $E_r(r)$, because the error growth rate decreases when the error magnitude approaches saturation. On the other hand, however, the predictability experiments with lagged persistence forecasts and also NWP models realize regime dependence of the error budget (that is, the autocorrelation, a, is no longer fixed), so that the initial errors exhibit an inverse behaviour as explained above. Small (large) initial errors are related to small (large) initial error growth rates because, in regimes with large (small) autocorrelation times, the white-noise level is relatively small (large) and subsequently the initial forecast errors, $E(r = 0)$, and their growth rates, $E_r(r = 0)$, tend to be smaller (larger). Therefore, realistic error-growth models such as those taken from population dynamics need to discriminate between weather regimes and, possibly, between the initial states within these regimes (which is certainly true for 'chaotic' regimes).

The results of this section (and section 4) can be summarized in a qualitative sense. If the intensity of the anomalous fluctuations of the deterministic part, $s_x^2 a^2$, is enhanced, the white-noise forcing $s_z^2 = s_x^2(1 - a^2)$ decreases, and vice versa. Consequently, the initial error $E_2(r = 0)$ and its initial growth rate $E_{2r}(r = 0)$ become smaller, as they are dominated by the relatively small random noise, s_z^2. However, the saturation level $E_2(r \to \infty)$ rises, because the inherently longer memory (and time-scale) of the anomalous fluctuations and their enhanced intensity, $s_x^2 a^2$, are dominated by the deterministic part of the system. This leads to a rise of the predictability limit, T (provided the selected predictability threshold is sufficiently small). A similar behaviour is observed in the real atmosphere (van den Dool and Saha 1990); more than 50% of the total variance of the

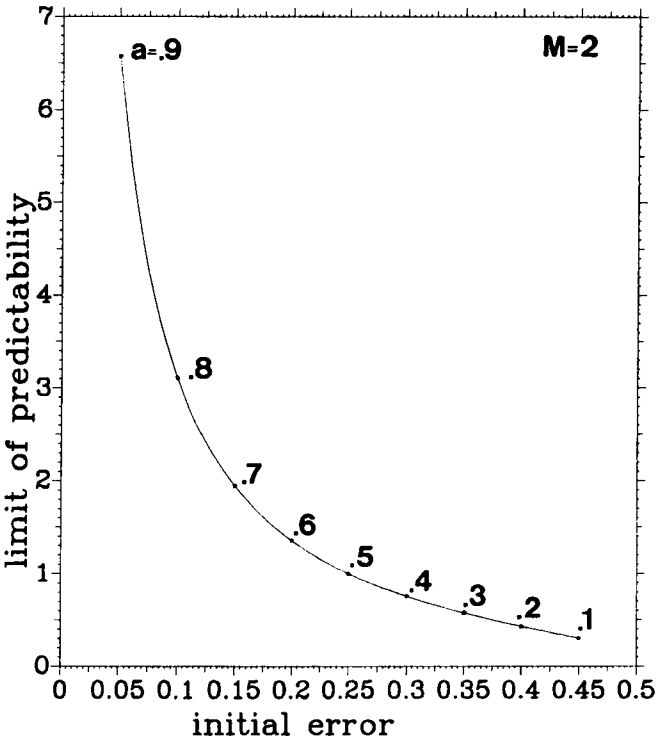


Figure 8. The predictability limit, T, changing with the magnitude of the initial error $E(r = 0)$ due to changing red-noise regime, $a = (1 - 2E_2(r = 0)/s_x^2)^{1/R}$. The related autocorrelation values, a, are denoted on the T-graph; the initial error axis is labelled in multiples of the unit climate variance, $s_x^2 = 1$.

500 mb height fields is found at periods of 18 days or longer associated with phenomena characterized by a long lifetime or large autocorrelation coefficient. These low-frequency regimes can be predicted over longer time-scales than their high-frequency counterparts.

6. Ensemble dispersion and forecast error

The time evolution of two initially close trajectories in phase space (say a forecast and its verification in a perfect-model experiment) provides information on the sensitivity of the system's dependence on initial conditions and its internal predictability. In the perfect-model environment a perfect ensemble of forecasts can be used to determine the predictability (measured in terms of the squared error of the ensemble-mean forecast) by the ensemble variance. In the imperfect model/ensemble environment no such relation exists for individual ensemble-mean forecasts; firstly because the error is influenced not only by the initial conditions but also by systematic model deficiencies and, secondly, because the ensemble is imperfect since its members do not necessarily provide the correct distribution about the unknown true initial state. Therefore, such an error–spread relation may exist only in a statistical sense, depending on many parameters like the weather regime characterized by the red-noise parameter a, the magnitude of the initial

anomaly, X_0, and on the lead time r. The predictability experiment with an ensemble of lagged persistence forecasts may be considered as a first step towards the more comprehensive analyses of the imperfect model/ensemble hypothesis.

(a) *Spread*

The ensemble variance (dispersion or spread) is defined (Eq. (3.1a)) by the ensemble average over the individual lagged forecasts that comprise the ensemble, $s_M^2 = [(F(r+i) - [F(r+i])^2] = [F(r+i)^2] - [F(r+i)]^2$. Subsequently, the (time or sample) average over all forecasts leads to a sample-averaged spread

$$S_M = \langle s_M^2 \rangle = \langle [F(r+i)^2] \rangle - \langle [F(r+i)]^2 \rangle$$

which can be derived conditional on the (value of the latest) initial anomaly, X_0, from which the ensemble-averaged forecast (latest member of), $[F(r+i)]$, starts (see appendix A):

$$S_M(X_0) = \langle X_0^2 \rangle \left(\frac{1-a^{2M}}{(1-a^2)M} - \left(\frac{1-a^M}{(1-a)M} \right)^2 \right) + \\ + s_x^2 \left(1 - \frac{1-a^{2M}}{(1-a^2)M} - \left[\frac{1+a}{(1-a)M} + \frac{1-a^{2M}}{(1-a)^2M^2} - \frac{2(1-a^M)(1+a)}{(1-a)^2M^2} \right] \right). \quad (6.1)$$

Note that for persistence forecasts the spread does not evolve with lead time. As the error budget allows an interpretation based on its systematic and random components, the ensemble variance or spread may also be interpreted in a similar manner. Again, the conditional spread is composed of two terms, the first describing the systematic and the second the non-systematic or random contribution. In the climate average over all initial anomalies $\langle X_0^2 \rangle$ is to be replaced by the variance s_x^2. For unconditional situations, $\langle X_0^2 \rangle = s_x^2$, so that the time-averaged spread is given by the sum of the systematic and non-systematic parts:

$$S_M = s_x^2 \left(1 - \frac{1+a}{(1-a)M} + \frac{2a(1-a^M)}{(1-a^2)M^2} \right). \quad (6.2)$$

Adding (6.2) and (3.4) gives the ensemble mean of M sample-averaged individual forecasts (see 6.5) which, for large lead times r, tends towards the mean squared distance of two randomly chosen (independent) weather states. Indeed, in the sample average $\langle\,\rangle$, the improvement of the squared error of the ensemble mean forecast, $\langle e_M^2(r) \rangle$, over the ensemble mean error of the individual forecasts $\langle [e(r+i)^2] \rangle = \langle \Sigma_{i=0}^{M-1} e^2(r+i)/M \rangle$ with $e_i = e(r+i) = F(r+i) - X(t)$, is given by the (sample averaged) spread S_M (see also Brankovic *et al.* 1990):

$$\langle [e(r-i)^2] \rangle = [E_1(r+i)] = E_M(r) + S_M(r). \quad (6.3)$$

Furthermore, the spread S_M is related to the squared distances between all pairs of individual forecasts: $D_M = \langle [d_i^2] \rangle = \langle \Sigma_i^M \Sigma_j^M (F_i - F_j)^2 / \{M(M-1)\} \rangle = 2MS_M/(M-1)$ (Eq. (3.1c); Brankovic *et al.* 1990).

The contributions of the systematic and non-systematic spread (6.1) to the total (6.2) depend on the ensemble size, and vary with the autocorrelation time-scale. Let $\langle X_0^2 \rangle = s_x^2$ be unity, both contributions to the total spread change with ensemble size and weather regime, M and a, respectively (Fig. 9). The structures are similar to the initial error ($r = 0$) but the systematic and random contributions are reversed; while the non-

systematic error grows continuously with ensemble size M and the systematic contribution has a maximum at finite M, the random spread increases monotonically and the systematic spread shows a maximum.

(*b*) *Error–spread relation*

In a perfect model/ensemble environment the squared error of individual ensemble-mean forecasts, e_M^2, is linked to the ensemble spread, s_M^2; that is, $e_M^2 \sim s_M^2$ holds for individual realizations. For the imperfect model/ensemble environment the error–spread relation is analysed in terms of the statistics provided by the predictability experiment, using samples of time-lagged ensemble-mean forecasts. The analysis of lagged persistence forecasts in the red-noise atmosphere proceeds in two steps. (i) The statistics of a sample of individual ensemble forecasts (with ensemble size M and lead time r) in a given climate or weather regime is evaluated. This may lead to practical aspects of error prediction. (ii) Functional relationships between the sample averaged conditional/unconditional squared forecast errors and their related spreads are derived.

(i) Scatter diagrams, Figs. 10(a) and (b), show a sample of (1000 from a total of 10000) squared errors, e_M^2, of ensemble-mean forecasts for lead time $r = 1$ versus their related spreads, s_M^2, in the red-noise regime $a = 0.8$; the error–spread regression lines estimated from the total sample represent correlations of 0.31 (0.14) for $M = 8$ (2) ensemble members. This analysis is extended to evaluate error–spread correlations in dependence of the ensemble size M and the forecast range r:

$$\langle (e_M^2 - E_M)\,(s_M^2 - S_M)\rangle / \{\langle (e_M^2 - E_M)^2\rangle \langle (s_M^2 - S_M)^2\rangle\}^{1/2}. \tag{6.4}$$

To search for an optimal M' for skill predictions, a set of error–spread correlation coefficients is plotted against the ensemble size M for lead times varying from $r = 1$ through 5 (Fig. 10(c)). The points $M = 8$ and 2 on the $r = 1$ line correspond to the two examples discussed above.

For a given weather regime and at a fixed ensemble size M, the average error–spread correlation decreases with increasing lead time. For a given forecast range r one observes ensemble size-dependent correlations indicating a size M', for which the error–spread correlation is maximal and optimal for skill prediction; for example, the optimal ensemble size $M' = 8$ is associated with a maximum error–spread correlation of 0.31. These optimal sizes M' tend to increase with growing lead time.

The occurrence of an optimal ensemble size M' for the error–spread correlation may be interpreted as follows. Realizing that there is an optimal ensemble size M^* (section 4) at which the systematic error contribution vanishes, then the total error is almost completely determined by the non-systematic part, which describes the internal dynamics (that is, the stochastic forcing) in terms of a perfect model. These non-systematic errors grow owing to the stochastic nature of the dynamics, even if initial conditions were perfectly known ($M = 1$, section 2). Although the perfect model–ensemble hypothesis is best met in the neighbourhood of M^* (when the systematic error attains a minimum and tends towards zero) the related error–spread correlation is expected to be optimal but not perfect, because imperfect (that is lagged) ensemble members are used. An independent analysis of all nine possible correlations (not shown) between the individual realizations of the systematic, random and total error (at $r = 1$) and the respective spread (obtained from 10000 'forecast experiments') support the interpretation. The structure of the ensemble size-dependent correlation between error and spread, the position and the magnitude of its maximum, is found only in the non-systematic contributions.

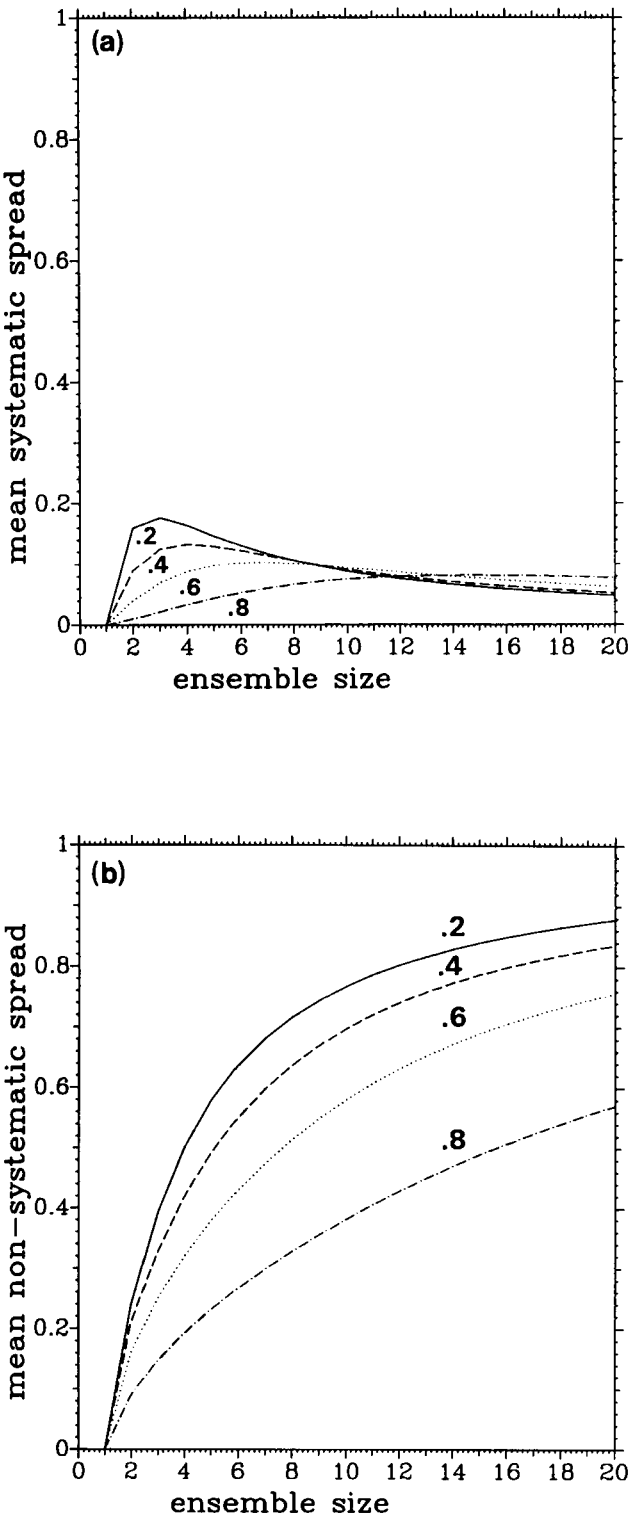


Figure 9. Ensemble spread: (a) systematic, (b) non-systematic and (c) total ensemble variance for various autocorrelation regimes $a = 0.2, 0.6, 0.4, 0.2$.

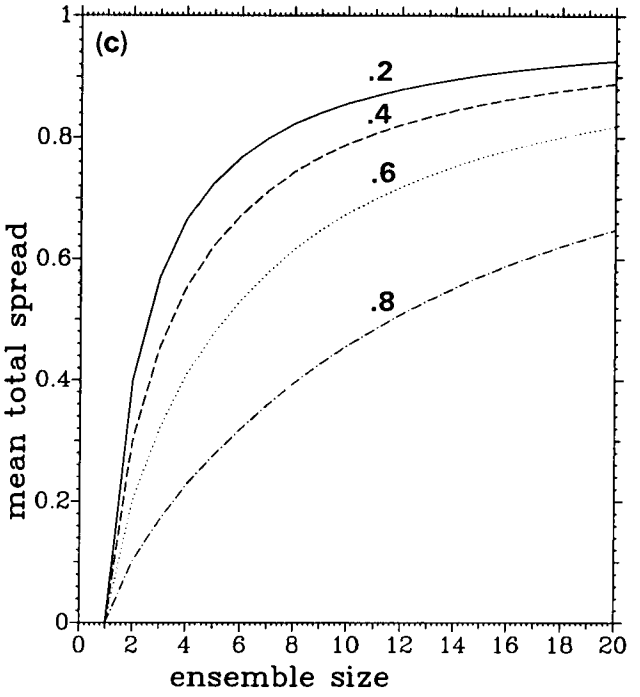


Figure 9. Continued.

These results may be compared with NWP predictability experiments in an imperfect model environment (Dalcher and Kalnay 1987; Brankovic *et al.* 1990, but see also Murphy (1990) and Tracton *et al.* (1989)). In the imperfect NWP model environment the (hemispheric and regional) spread did not turn out as a very good predictor of skill in the extended range; it has been argued that smaller regions (and fewer synoptic disturbances) might improve the error–spread relation, but this has also not been substantiated by the analysis of Brankovic *et al.* (1990), most likely because NWP forecast experiments are influenced by a large variability of the climate and weather regimes and the initial states and, in particular, by the systematic error of the model.

In the perfect-model environment the correlation between the mean-square spread and the forecast error has been determined by Barker (1991) using a sample of 120 cases from a two-layer GCM. This is one of the few studies based on a large sample. The error–spread correlation obtained is 0.35 (0.58) for lead time $r = 1$ (12 days), explaining about 10% (30%) of the variance, which may have little practical use. Barker (1991), however, notes that the Monte Carlo generated ensembles have the deficiency that their initial spread is always generated in the same way; it is not related to the initial error, although it should be associated with the confidence in the initial analysis to obtain a perfect initial spread–initial analysis error correlation. In an imperfect (operational) environment this is difficult to achieve notwithstanding the systematic error growth even at short lead times. Many earlier investigations use small samples, leading to considerable variability in the correlations. Furthermore, Dalcher and Kalnay (1987), Murphy (1988), Tracton *et al.* (1989) used anomaly correlation measures that can be misleading in cases (Palmer and Tibaldi 1988). Additionally, forecast error–spread correlations are much greater when forecasts from different seasons include the seasonal variations in spread

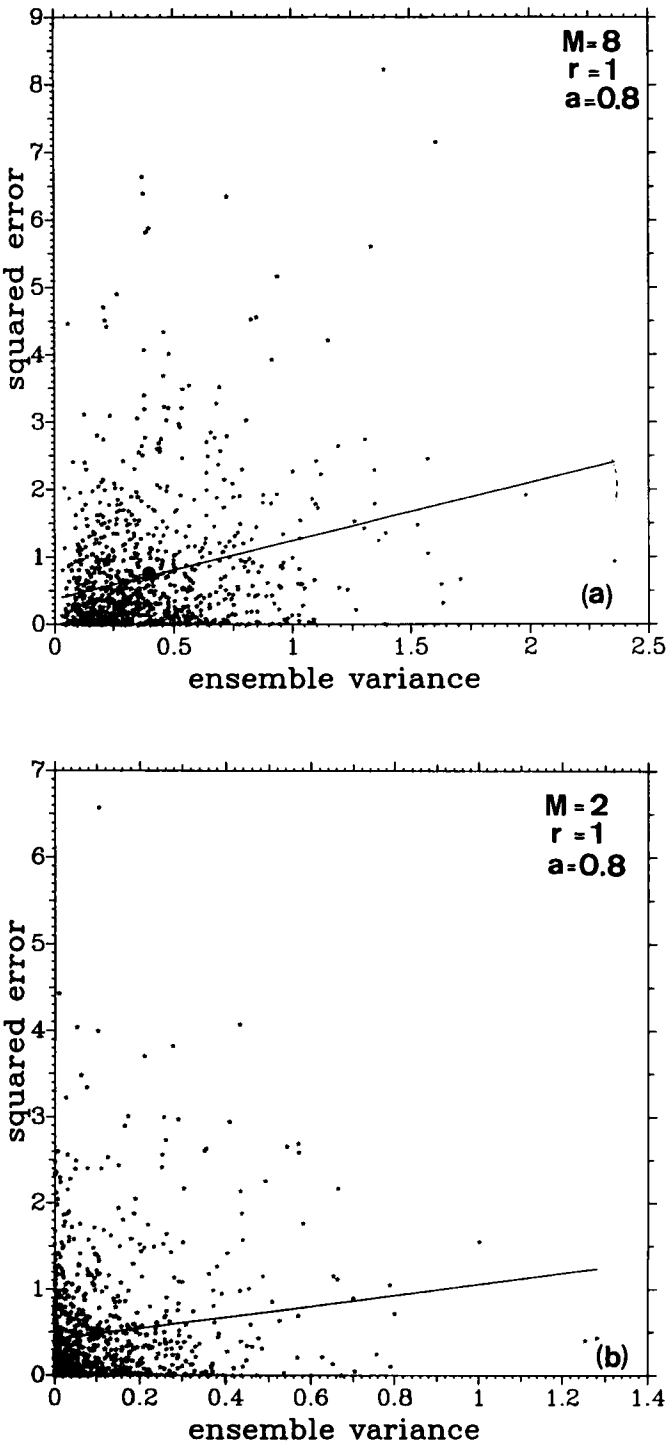


Figure 10. (a) and (b) Scatter diagrams of squared errors versus ensemble variance (spread) of a sample of 1000 individual ensemble mean forecasts of lead time $r = 1$ in a red-noise atmosphere with autocorrelation $a = 0.8$ and ensemble size $M = 8$ (a) and $M = 2$ (b). The linear regression is also shown. (c) Error–spread correlation of ensemble mean forecasts in the red-noise regime $a = 0.8$ ($= 0.3$), versus the ensemble size M for lead times increasing from $r = 1, \ldots, 5$. Note that there appears to be an optimal ensemble size depending on lead time and regime.

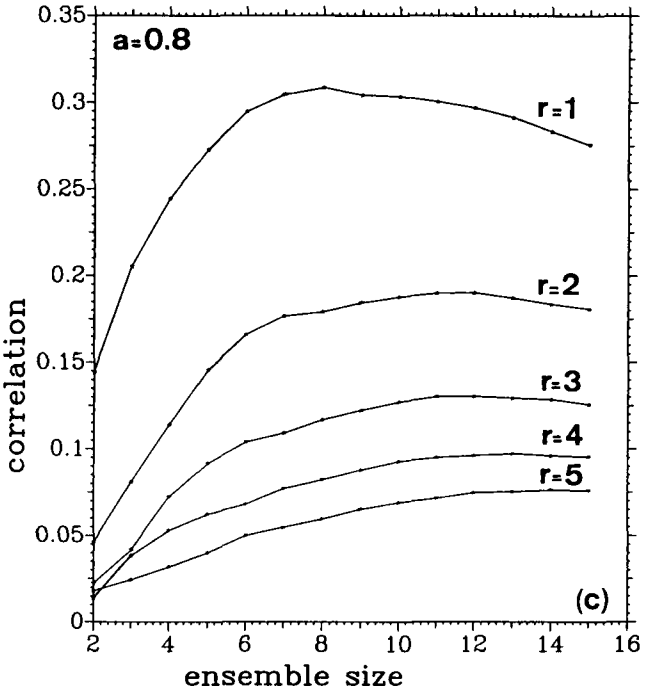


Figure 10. Continued.

and error (Brankovic *et al.* 1990). Finally, the extremely high correlations quoted from van den Dool (1989, based on weather analogues) appear to be created by a regression not weighted by the number of the cases. Therefore, it appears plausible that Monte Carlo generated ensembles may not be optimal to determine the error–spread relation but dynamically conditioned ensemble members (see Mureau *et al.* 1993). In this sense ensemble forecasts may be improved by two steps. First a control prediction may be constructed in terms of a lagged average forecast with an ensemble size which minimizes the systematic error contribution. Then this optimal control forecast may be applied to make ensemble forecasts, using, for example, dynamically conditioned perturbations (see Mureau *et al.* 1993) of the optimal control instead of Monte Carlo generated or lagged ensembles.

(ii) Error and spread are now analysed to determine functional relationships between both sample-averaged error and spread to extend the analysis of the individual realizations.

Firstly both the unconditional sample-averaged squared error (error variance) and the sample-averaged spread, $E_M(r)$ and S_M (Eqs. (3.4) and (6.2)) are added to provide a regime (a = constant) averaged error–spread relation. This is basically a sample-averaged version of (3.1b):

$$E_M = -S_M + 2s_x^2\left(1 - \frac{a^r(1-a^M)}{M(1-a)}\right). \tag{6.5}$$

Two examples of (6.5), $M = 8$ (and 2) for $a = 0.8$ and $r = 1$, correspond to the sample-averaged squared errors $E_M = \langle e_M^2 \rangle$ and spreads $S_M = \langle s_M^2 \rangle$ of the individual ensemble-mean forecasts shown in Figs. 11(a) and (b). Note that such a sample-averaged error–

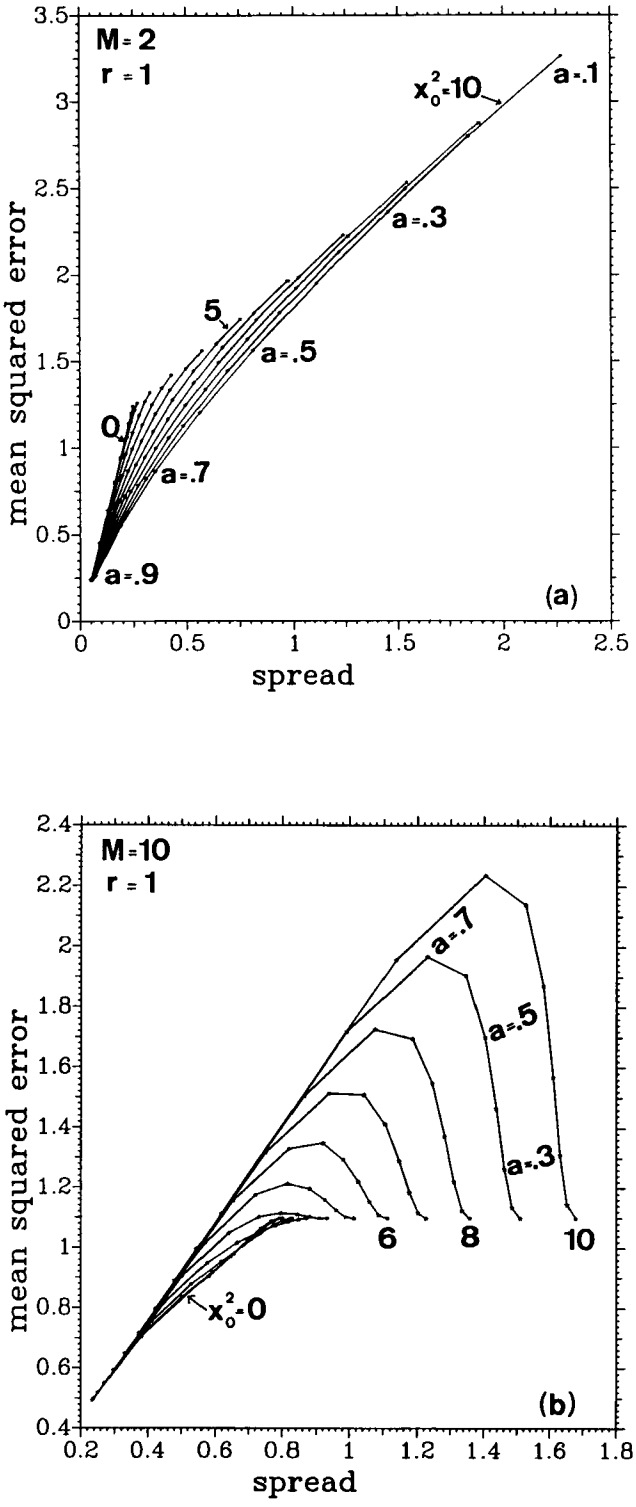


Figure 11. Error versus spread diagrams depending on the initial conditions, $\langle X_0^2 \rangle = 0, 1, \ldots 10$ and changing with the red-noise regime, autocorrelation a (dots). Individual diagrams are presented for lead time $r = 1$ associated with the ensemble size $M = 2$ (a) and $M = 10$ (b).

spread relation is not suitable for 'practical' forecasts of the error of individual predictions. It merely reveals the properties of a sample-averaged ensemble statistics (3.1b) applied to forecasts in an imperfect-model environment. At large lead times, $E_M(r \to \infty) + S_M = 2$. That is saturation error and spread add to the mean squared error, $2s_x^2$, of the individual forecasts that comprise the ensemble; at zero lead time, for example, one obtains $E_M(r = 0) + S_M = 2s_x^2\{1 - (1 - a^M)/(M(1 - a)\}$. In combination with the perfect ensemble hypothesis (3.2), $E_M(r = 0) = S_M(M + 1)/(M - 1)$, one can show that, in the sample or climate average, there is no perfect ensemble except the singular solution $M = \infty$.

Secondly, conditional sample averages are analysed. Both the conditional sample-averaged squared error (error variance) and the spread, $E_M(X_0|r)$ and $S_M(X_0)$, (Eqs. (6.1) and (4.1)) are combined by eliminating the magnitude of the initial condition X_0 to yield a linear relation of the form

$$E_M(X_0|r) = bS_M(X_0) + B \tag{6.5}$$

whose slope b and offset B depend on the lead time r, the ensemble size M and the weather regime a; b is the quotient of the factors attached to the initial conditions $\langle X_0^2 \rangle$ in (6.1) and (4.1). This linear relation can be traced in the displays of Figs. 11(a) and (b), showing the rather complex structure of the conditional sample averages of error and spread in $E_M(X_0)$) versus $S_M(X_0)$ diagrams based on (4.1) and (6.1). The results are presented for isolines of constant initial anomalies $\langle X_0^2 \rangle$, and autocorrelation coefficients a, given the lead time $r = 1$ and the ensemble sizes $M = 8$ (Fig. 11 (a)) and $M = 2$ (Fig. 11(b)). The unconditional error–spread relation is defined by the initial condition $\langle X_0^2 \rangle = \langle X^2 \rangle = s_x^2 = 1$; the non-systematic error is obtained by $\langle X_0^2 \rangle = 0$.

The non-systematic error–spread relation is represented by the line, defined by $\langle X_0^2 \rangle = 0$. Adding the systematic error–spread relation leads to the imaginary straight lines which follow constant autocorrelations a = constant.

For common initial conditions, that is $\langle X_0^2 \rangle < 2$ to $3\, s_x^2$, the conditional *sample averaged* error–spread relation is dominated by the non-systematic contribution (that is near $\langle X_0^2 \rangle = 0$). This, at least partially, supports the error–spread correlations of *individual* forecasts, which are also dominated by the non-systematic fluctuations. Furthermore, the larger the ensemble size the wider the spread interval to be covered by the related average error.

For large initial conditions the systematic contributions dominate the error–spread relation. The magnitudes of the initial anomalies set a limit to the spread; the dependence on the regime, a, defines the slope of the error–spread relation, b, modulated by ensemble size M and lead time r.

Finally, the persistence predictability experiment with lagged ensemble-mean forecasts has demonstrated that both initial values and weather regimes need to be considered for the evaluation of ensemble-mean predictions in the imperfect-model environment. In particular, for both skill improvement and skill prediction, the predictability experiments need to be statistically analysed by conditional sampling.

7. Concluding remarks

Predictability experiments are performed by a univariate persistence model associated with ensembles of time-lagged forecasts. The predictions are made in a red-noise atmosphere so that predictability can be analysed analytically for an imperfect model/ensemble environment. In this sense these predictability experiments have some aspects

in common with the complex problem of weather predictability posed by NWP model forecasts in the real atmosphere, so that it is not surprising to obtain results that are, at least qualitatively, comparable with NWP predictability studies. As the dominating weather regimes are relatively well parametrized by the autocorrelation time-scale and the noise level of a red atmosphere, a and s_z^2, their influence on the error budget can be studied. For lagged persistence ensemble forecasts, the systematic and non-systematic error budget and derived measures of predictability (initial and saturation error, error growth rates, limit of predictability, error distributions, etc.) can be determined analytically. There is no need to fit models of population-growth dynamics to the results in order to obtain a simple and interpretable format. Besides the traditional error-budget approach with its derived properties and sample statistics, other predictability aspects have been discussed which may be explored further in terms of the two goals, improvement and prediction of skill.

(i) *Skill improvement.* In an imperfect-model environment the sample statistics (of error and spread) of predictability experiments depend sensitively both on initial anomalies and the particular weather regime. That is although unweighted lagged average forecasts do not give advantage over the latest single forecast in the unconditional sample mean (Fig. 3), this is not the case for conditional samples (Fig. 4) that depend on the magnitude of the initial anomalies, the autocorrelation regime and the ensemble size. This suggests that, for example, it would be more effective to use lagged averages when initial perturbations are larger than the climate standard deviation (depending on the lead time and the autocorrelation regime). This may be further explored in practical weather forecasting and needs to be incorporated in error-growth models.

(ii) *Skill prediction.* In the perfect model/ensemble environment the optimal ensemble size to achieve the two goals (improvement and prediction of skill) is assumed to be the same. However, in an imperfect model/ensemble environment this need not necessarily be the case. For skill improvement (first goal) there are indications that the ensemble-mean forecasts are, in the average, not sufficiently superior to the latest individuals of the ensemble, suggesting an optimal size $M = 1$; optimal weighting, however, can provide ensemble-averaged forecasts that are better than the latest member. For skill prediction (second goal) the lagged persistence forecasts suggest an optimal ensemble size M, which is close to the ensemble size associated with minimum systematic error (Figs. 5(d) and 10(c)). For example, for observed atmospheric autocorrelations, $a = 0.8$, corresponding to a five-day integral time-scale (Dole and Gordon 1983, appendix A) an optimal number of three to four lagged forecasts minimizes the systematic error of the average prediction at short lead times, and a slightly larger number leads to a maximum error–spread correlation.

Recent developments in skill prediction indicate that deterministic NWP models can provide more suitable (than random or lagged) ensemble members. Within the linear error-growth range non-normal mode disturbances or singular vectors can be derived with the adjoint-model technique (Molteni and Palmer 1993, see also Lorenz (1965) and Lacarra and Talagrand (1988)) which leads to a maximum ensemble dispersion and provides an estimate of the maximum possible error. Another efficient approach is to find the maximum growing modes by breeding growing-mode perturbations (Toth and Kalnay 1992). Now, the systematic error has been assumed to be part of the cause of the observed relatively small correlations between ensemble spreads and forecast errors; especially in the extended range where systematic errors become as important as errors due to uncertainties in the initial conditions (Barker 1991). In this sense a lagged average forecast (with a systematic error minimizing ensemble size) may still be useful because

it could provide a suitable control forecast from which subsequent ensemble predictions can be determined by applying the dynamically conditioned perturbations (breeding or adjoints) for skill prediction.

ACKNOWLEDGEMENT

Discussions with Drs L. Smith, E. Kalnay, T. Palmer and M. Deque are gratefully acknowledged. Thanks are due to Dr S. Schubert and another referee for their constructive comments. Part of the research is related to a Bundesminister für Forschung und Technologie grant on climate variability.

APPENDIX A

Conditional error variance

Two averaging operators are introduced: the average of an ensemble of forecasts, $[F]$, and the average over a sample of forecast experiments (sample, time or climate average), $\langle\rangle$. Then the error variance, E_M, and the dispersion (spread), S_M, of the ensemble-mean forecasts, $[F] = [F(r+i)] = \Sigma_{i=0}^{M-1} F(r+i)/M$ can be expressed as a sum of terms, which are individually deduced:

$$E_M(r|X_0) = \langle (X(t_0+r) - [F(r+i)]^2) \rangle = \langle X^2 \rangle + \langle [F^2] \rangle - 2\langle X[F] \rangle$$

$$S_M(X_0) = \langle ([F(r+i) - [F(r+i)])^2 \rangle = \langle [F^2] \rangle - \langle [F]^2 \rangle = \langle s_M^2 \rangle$$

where the ensemble variance of the individual forecasts, $s_M^2 = [F(r+i)^2] - [F(r+i)]^2$, is given in (3.1a). The verification is denoted by $X(t_0+r)$, the conditional initial value by $X_0 = X(t_0)$, and the members of the forecast ensemble by $F(r+i)$. The statistics are analysed in terms of a (time) average over all forecast samples, $\langle\rangle$, all of which are conditional on the same initial value X_0. For convenience we use the notation $s_x^2 = \langle X^2 \rangle$; the initial value, X_0, is not affected by the sample averaging, $X_0 = \langle X_0 \rangle$; z and w are introduced to differentiate between noises that are characterized by the same index in the verification and ensemble-forecast building mode:

Verification ($r > 0$):

$$X(t_0+r) = a^r X(t_0) + a^{r-1} z_1 + \ldots + z_r$$

Individual forecasts ($0 < i < M-1$):

$$F(r+i) = X(t_0 - i) = a^i X(t_0) + a^{i-1} w_1 + \ldots + w_{i-1}$$

Conditional state ($r = 0$):

$$F(r=0) = X(t_0) = X_0.$$

Now the following statistics can directly be derived:

Sample variance of verification:

$$\langle X^2(t_0+r) \rangle = a^{2r} \langle X_0^2 \rangle + \langle z_i^2 \rangle \{a^{2(r-1)} + \ldots + a^0\} = a^{2r} \langle X_0^2 \rangle + \langle z_i^2 \rangle (1 - a^{2r})/(1 - a^2)$$

$$= a^{2r} \langle X_0^2 \rangle + \langle X^2 \rangle (1 - a^{2r}).$$

Ensemble-mean forecast:

$$[F(r+i)] = \{X_0(1 + a + \ldots + a^{M-1}) + w_1(1 + a + \ldots + a^{M-2}) + \ldots + w_{M-1}\}/M$$
$$= \{X_0(1 - a^M) + w_1(1 - a^{M-1}) + \ldots + w_{M-1}(1 - a)\}/\{M(1 - a)\}.$$

Sample average of squared ensemble-mean forecasts:

$$\langle [F(r+i)]^2 \rangle = \{\langle X_0^2 \rangle (1 - a^M)^2 + \langle w^2 \rangle [(1 - 2a^{M-1} + a^{2(M-1)}) + \ldots +$$
$$+ (1 - 2a + a^2)]\}/\{M(1 - a)\}^2$$
$$= \langle X_0^2 \rangle (1 - a^M)^2/(1 - a)^2 + \langle X^2 \rangle [(1 - a^2)/\{M(1 - a^2)\} +$$
$$+ (1 - a^{2M})/\{M^2(1 - a)^2\} - 2(1 - a^M)(1 + a)/\{M^2(1 - a)^2\}$$

because $\langle z^2 \rangle = \langle w^2 \rangle = \langle X^2 \rangle (1 - a^2)$. For $\langle X_0^2 \rangle = \langle X^2 \rangle = s_x^2$, the unconditional sample averages are attained, so that we obtain $s_F^2 = \langle X^2 \rangle [(1 + a)/\{(1 - a)M\} - 2a(1 - a^M)/\{M^2(1 - a^2)\}$. This is identical to the sample or climate variance of the ensemble-mean forecast (see the following term).

Sample variance of ensemble-mean forecasts (s_F^2):
$\langle ([F(r+i)] - \langle [F(r+i)] \rangle)^2 \rangle = \langle [F(r+i)]^2 \rangle - \langle [F(r+i)] \rangle^2 = \langle [F(r+i)]^2 \rangle$, because for unbiased persistence forecasts $\langle [F(r+i)] \rangle = 0$. For $\langle X_0^2 \rangle = \langle X^2 \rangle = s_x^2$ this is identical with the sample or climate variance of the ensemble-mean forecasts (for $M = 1$, $s_F^2 = s_x^2$); this will be used in deriving the anomaly correlation:

$$s_F^2 = \langle [F(r+i)]^2 \rangle = s_x^2[(1 + a)/\{M(1 - a)\} - 2a(1 - a^M)/\{M^2(1 - a)^2\}].$$

Sample average of the mean of squared forecasts:

$$\langle [F(r+i)^2] \rangle = \{\langle X_0^2 \rangle + (a^2 \langle X_0^2 \rangle + \langle w^2 \rangle) + \ldots + (a^{2(M-1)} \langle X_0^2 \rangle +$$
$$+ a^{2(M-2)} \langle w^2 \rangle + \ldots + \langle w^2 \rangle)\}/M$$
$$= \{\langle X_0^2 \rangle (1 - a^{2M}) + \langle w^2 \rangle [(1 - a^{2(M-1)}) + \ldots + (1 - a^2)]\}/\{M(1 - a^2)\}$$
$$= \{\langle X_0^2 \rangle (1 - a^M)^2/(1 - a^2) + \langle X^2 \rangle [M - (1 - a^{2M})/(1 - a^2)]\}/M$$

because $\langle w^2 \rangle = \langle X_0^2 \rangle (1 - a^2)$. For $\langle X_0^2 \rangle = \langle X^2 \rangle = s_x^2$, the unconditional sample averages are attained.

Anomaly covariance:

$$\mathrm{ACOV} = \langle X(t_0 + r)[F(r+i)] \rangle = a^r \langle X_0^2 \rangle (1 - a^M)/\{M(1 - a)\}$$

where the correlations between the independent noise fluctuations, z and w, vanish.

The appropriate terms can now be combined to provide the conditional forecast-error variance and ensemble spread (4.1 and 6.1):

$$E_M(r|X_0) = \langle X_0^2 \rangle \{a^{2r} - 2a^r[(1 - a^M)/\{(1 - a)M\}] + [(1 - a^M)/\{(1 - a)M\}]\} +$$
$$+ s_x^2\{1 - a^{2r} + [(1 + a)/\{(1 - a)M\} + (1 - a^{2M}/\{(1 - a)^2M^2\} -$$
$$- 2(1\, a^M)(1 + a)/\{(1 - a)^2M^2\}]\} \qquad (4.1)$$

$$S_M(X_0) = \langle X_0^2 \rangle \{(1 - a^{2M})/\{(1 - a^2)M\} - (1 - a^M)^2/\{(1 - a)^2M^2\}\} +$$
$$+ s_x^2\{1 - (1 - a^{2M})/\{(1 - a^2)M\} - [(1 + a)/\{(1 - a)M\} + (1 - a^{2M})/\{(1\, a)^2M^2\} -$$
$$- 2(1 - a^M)(1 + a)/\{(1 - a)^2M^2\}]\}. \qquad (6.1)$$

Setting $\langle X_0^2\rangle = \langle X^2\rangle$ one obtains the climate or sample-averaged (unconditional) values (3.4) and (6.2). The individual and two-lagged persistence forecasts are derived for $M = 1$ or 2, respectively. The anomaly correlation is also easily deduced from the anomaly covariance, setting $\langle X_0^2\rangle = \langle X^2\rangle = s_x^2$ and realizing the variance of the ensemble-mean forecasts $s_F^2 = \langle [F(r+i)]^2\rangle$.

Forecast agreement: For completeness, the forecast agreement, FA, is introduced here as a measure of the ensemble dispersion related to anomaly correlations. It is defined after Tracton *et al.* (1989) as the average anomaly correlation between the latest (base or control) prediction and the remaining $M-1$ members of the ensemble:

$$FA = \frac{\langle [F(r)\,F(r+i)]\rangle}{s_x^2} = \sum_{i=1}^{M-1} \frac{x(t-r)\,x\{t-(r+i)\}}{(M-1)s_x^2}$$

$$= \sum_{i=1}^{M-1} \frac{a^i}{(M-1)} = \frac{a(1-a^{M-1})}{(M-1)\,(1-a)}.$$

If one includes the latest forecast in the forecast agreement, then $i = 0$ is included and $FA^* = (1-a^M)/\{(1-a)M\}$. Note that FA^* enters both the systematic and random contribution of the spread.

Appendix B

Conditional distribution density of the squared error e^2 at X_0

The conditional distribution density $f(e|X_0)$ of the error e at the fixed state X_0 can be deduced from the joint bivariate Gaussian distribution density, $f(e, X_0)$ in (B.1), of the error e with zero mean and variance E_M (denoted by E in the following) divided by the marginal density $f(X_0) = (2\pi s_x^2)^{-1/2} \exp\{-(2s_x^2)^{-1} X_0^2\}$:

$$\begin{aligned} f(e|X_0) = f(e, X_0)/f(X_0) &= \{2\pi\, E(1-c^2)\}^{-1/2} \exp\{-\{2(1-c^2)\}^{-1} \times \\ &\times [X_0^2/s_x^2 - 2cX_0 e/s_x E^{1/2} + e^2/E] + (2s_x^2)^{-1} X_0^2\} \\ &= \{2\pi E(1-c^2)\}^{-1/2} \exp\{-\{2(1-c^2)\}^{-1} \times \\ &\times [-2cX_0 e/(s_x E^{1/2}) + e^2/E] - c^2\{2(1-c^2)s_x^2\}^{-1} X_0^2\}. \end{aligned} \tag{B.1}$$

The probability distribution of the squared error, $e^2 < l$, smaller than a threshold l is given by integration between square root limits, $\mathrm{Prob}(e^2 < l) = H(l) = {}_{-\sqrt{l}}\int^{\sqrt{l}} f(e|X_0)\mathrm{d}e$, with the density $h(l|X_0) = \mathrm{d}H(l)/\mathrm{d}l$. Applying the Leibniz rule of differentiation for variable boundaries, $\mathrm{d}(\int^{g(y)} f(y)\mathrm{d}y)/\mathrm{d}y = f(g(y))\mathrm{d}g/\mathrm{d}y + \int^{g(y)}(\mathrm{d}f/\mathrm{d}y)\mathrm{d}y$, and noting that $\mathrm{d}l^{1/2}/\mathrm{d}l = l^{-1/2}/2$, leads to

$$h(l|X_0) = (2\sqrt{l})^{-1}\,[f(\sqrt{l}|X_0) + f(-\sqrt{l}|X_0)]. \tag{B.2}$$

Finally we derive the expectation of the squared error $l = e^2$ at fixed X_0 by

$$E(l|X_0) = {}_{l=0}\int^{\infty} l h(l|X_0) = {}_{l=0}\int^{\infty} 1/2\,\sqrt{l}[f(\sqrt{l}|X_0) + f(-\sqrt{l}|X_0)]\,\mathrm{d}l. \tag{B.3}$$

Substituting $z = \sqrt{l}$ with $\mathrm{d}l = 2z\mathrm{d}z$ yields $E(l|X) = {}_{z=0}\int^{\infty} z^2[f(z|X_0) + f(-z|X_0)]\mathrm{d}z$, and with (B.1) the expectation of the squared error is

$$E(l|X_0) = \{2\pi E(1-c^2)\}^{-1/2} \exp\{-c^2\{2(1-c^2)\,s_x^2\}^{-1} X_0^2\} \times$$
$$\times \left[{}_{z=0}\int^{\infty} z^2 \exp\{-\{2(1-c^2)\}^{-1}\,[-2cX_0 z/(s_x E^{1/2}) + z^2/E]\}\mathrm{d}z + \right.$$
$$\left. + {}_{z=0}\int^{\infty} z^2 \exp\{-\{2(1-c^2)\}^{-1}\,[2cX_0 z/(s_x E^{1/2}) + z^2/E]\}\mathrm{d}z\right]$$
$$= \{2\pi E(1-c^2)\}^{-1/2} \exp\{-c^2\{2(1-c^2)s_x^2\}^{-1} X_0^2\}\,[A]. \qquad \text{(B.4)}$$

Both integrals can be solved (Gradshteyn and Ryzhik 1980, p. 338):

$${}_0\int^{\infty} X^2 \exp\{-\mu X^2 - 2\nu X\}\mathrm{d}X$$
$$= -\nu/(2\mu^2) + (\pi/\mu^5)^{1/2}\,(2\nu^2+\mu)/4 \exp(\nu^2/\mu)[1 + \mathrm{erf}(\nu/\sqrt{\mu})]$$

for $[|\arg \nu| < \pi/2,\ Re\ \mu > 0]$. Since μ and the absolute value of ν are identical in both integrals, but ν changes sign in the second one. Noting that $\mathrm{erf}(-x) = -\mathrm{erf}(x)$ their sum is $A = {}_0\int^{\infty} \ldots + {}_0\int^{\infty} \ldots = 2(\pi/\mu^5)^{1/2}\,(2\nu^2+\mu)/4 \exp\{\nu^2/\mu\}$. With $\mu = 1/\{2(1-c^2)E\}$ and $\nu = cX_0/\{2(1-c^2)s_x E^{1/2}\}$,the term A substituted in (B.4) yields:

$$A = E\{2\pi(1-c^2)E\}^{1/2}\,[\{c^2X_0^2 + (1-c^2)s_x^2\}/s_x^2] \exp\{c^2\{2(1-c^2)s_x^2\}^{-1}X_0^2\}. \qquad \text{(B.5)}$$

Replacing A in (B.4) by (B.5) finally leads to the mean squared error at fixed X_0:

$$E(l|X_0) = (E/s_x^2)\,[c^2 X_0^2 + (1-c^2)s_x^2]. \qquad \text{(B.6)}$$

Replacing in Eq. (B.6) the correlation $c = A_M(r)$ in (3.7) and E by E_M in (3.4) leads after some transformations to (4.1).

References

Barker, T. W. 1991 The relationship between spread and forecast error in extended-range forecasts. *J. Climate*, **4**, 733–742

Brankiovic, C., Palmer, T. N., Molteni, F., Tibaldi, S. and Cubasch, U. 1990 Extended range prediction with ECMWF models: Time lagged ensemble forecasting. *Q. J. R. Meteorol. Soc.*, **116**, 867–912

Chen, W. Y. 1989 Estimate of dynamical predictability from NMC DERF experiments. *Mon. Weather Rev.*, **117**, 1227–1236

Dalcher, A. and Kalnay, E. 1987 Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39A** 474–491

Dalcher, A., Kalnay, E. and Hoffman, R. 1988 Medium range lagged average forecasts. *Mon. Weather Rev.*, **116**, 402–416

Deque, M. 1988 The probabilistic formulation: A way to deal with ensemble forecasts. *Annales Geophysicae*, **6**, 217–224

Dole, R. and Gordon, N. 1983 Persistent anomalies of the extratropical northern hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics. *Mon. Weather Rev.*, **111**, 1567–1586

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

Fraedrich, K. and Leslie, L. M. 1989 A minimal model for the short term prediction of rainfall in the tropics. *Weather and Forecasting*, **3**, 243–246

Gleeson, T. A. 1970 Statistical-dynamical prediction. *J. Appl. Meteorol.*, **9**, 333–344

Gradshteyn, I. S. and Ryzhik, I. M. 1980 *Table of integrals, series and products*. Academic Press

Gutzler, D. S. and Mo, K. C. 1983 Autocorrelation of northern hemisphere geopotential heights. *Mon. Weather Rev.*, **111**, 155–164

Hayashi, Y. 1986 Statistical interpretation of ensemble-time mean predictability. *J. Meteorol Soc. Japan*, **67**, 164–181

Hoffman, R. and Kalnay, E. 1983 Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118

Horel, J. D. 1985 Persistence of the 500 mb height field during northern hemisphere winter. *Mon. Weather Rev.*, **113**, 2030–2042

Horel, J. D. and Roads, J. O.	1988	Sensitivity of regional predictability to flow characteristics. *J. Geophys. Res.*, **93**, 11005–11014
Kalnay, E. and Dalcher, A.	1987	Forecasting forecast skill. *Mon. Weather Rev.*, **115**, 349–356
Lacarra, J.-F. and Talagrand, O.	1988	Short-range evolution of small perturbations in a barotropic model, *Tellus*, **40A**, 81–95
Legras, B. and Ghil, M.	1985	Persistent anomalies, blocking and variations in atmospheric predictabiilty. *J. Atmos. Sci.*, **42**, 433–471
Leith, C. C.	1974	Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.*, **102**, 409–418
Livezey, R. E., Barnston, A. G. and Neumeister, B. K.	1990	Mixed analog/persistence prediction of seasonal mean temperatures for the USA. *Int. J. Climatol.*, **10**, 329–340
Lorenz, E. N.	1965	A study of predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333
	1969	Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646
	1975	Climatic predictability. Pp. 132–136 in The physical basis of climate modelling. WMO, GARP Publication Series, 16
	1982	Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513
Molteni, F. and Palmer, T. N.	1993	Predictability and non-model finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc.*, **119**, 269–298
Molteni, F. and Tibaldi, S.	1990	Regimes in the wintertime circulation over northern extratropics. II: Consequences on dynamical predictability. *Q. J. R. Meteorol. Soc.*, **116**, 1263–1288
Molteni, F., Tibaldi, S. and Palmer, T. N.	1990	Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Q. J. R. Meteorol. Soc.*, **116**, 31–67
Munk, W.	1960	Smoothing and persistence. *J. Meteorol.*, **17**, 92–93
Mureau, R., Molteni, F. and Palmer, T. N.	1993	Ensemble prediction using dynamically conditioned perturbations. *Q. J. R. Meteorol. Soc.*, **119**, 299–323
Murphy, J. M.	1988	The impact of ensemble forecasts on predictability. *Q. J. R. Meteorol. Soc.*, **114**, 463–493
	1990	Assessment of the practical utility of extended range ensemble forecasts. *Q. J. R. Meteorol. Soc.*, **116**, 89–125
Palmer, T. N.	1993	Extended-range atmospheric prediction and the Lorenz model. *Bull. Am. Meteorol. Soc.*, **74**, 49–65
Palmer, T. N. and Tibaldi, S	1988	On the prediction of forecast skill. *Mon. Weather Rev.*, **116**, 2453–2480
Reinhold, B. B. and Pierrehumbert, R. T.	1982	Dynamics of weather regimes: quasi-stationary waves and blocking. *J. Atmos. Sci.*, **110**, 1105–1145
Roads, J.	1987	Estimate of errors in lagged time average numerical weather prediction. *Tellus*, **39A**, 492–499
	1988	Lagged average predictions in a predictability experiment. *J. Atmos. Sci.*, **45**, 147–162
Saha, S. and van den Dool, H. M.	1988	A measure of the practical limit of predictability. *Mon. Weather Rev.*, **116**, 2522–2526
Schubert, S. D. and Suarez, M. J.	1992	Dynamical predictability in a simple general circulation model: average error growth. *J. Atmos. Sci.*, **46**, 353–370
Schubert, S. D., Suarez, M. J. and Schemm, J.-K.	1992	Persistence and predictability in a perfect model. *J. Atmos. Sci.*, **49**, 256–269
Seidman, A. N.	1981	Average techniques in long range weather forecasting. *Mon. Weather Rev.*, **109**, 1367–1379
Stroe, R. and Royer, J. F.	1993	Comparison of different error growth formulas and predictability estimation numerical extended range forecasts. *Annales Geophysicae*, **11**, 296–316
Tennekes, H.	1991	'Karl Popper and the accountability of numerical forcasting'. Pp. 22–28 in New developments in predictability. ECMWF Workshop
Toth, Z.	1991	Estimation of atmospheric predictability by circulation analogs. *Mon. Weather Rev.*, **119**, 65–72
Toth, Z. and Kalnay, E.	1992	'Growing modes of the atmosphere obtained by "breeding" and their application to ensemble forecasting'. Abstracts of the Second international conference on modelling of global climate change and variability. Hamburg, 7–11 September 1992

Tracton, M. S., Mo, K., Chen, W., Kalnay, E., Kistler, R. and White, G. 1989 Dynamical extended range forecasting (DERF) at the National Meteorological Center. *Mon. Weather Rev.*, **117**, 1604–1635

Trenberth, K. 1984 Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelation. *Mon. Weather Rev.*, **112**, 2359–2368

1985 Persistence of daily geopotential heights over the southern hemisphere. *Mon. Weather Rev.*, **113**, 38–53

van den Dool, H. M. 1989 A new look at forecasting through analogues. *Mon. Weather Rev.*, **117**, 2230–2247

van den Dool, H. M. and Chervin, R. 1986 A comparison of month to month persistence of anomalies in a general circulation model and in the earth's atmosphere. *J. Atmos. Sci.*, **43**, 1454–1466

van den Dool, H. M. and Saha, S. 1990 Frequency dependence in forecast skill. *Mon. Weather Rev.*, **118**, 128–137