STATISTICAL METHODS FOR THE VERIFICATION OF LONG AND SHORT RANGE FORECASTS

C. E. Leith*
Courant Institute of Mathematical Sciences
New York University, New York, USA

1. INTRODUCTION

Weather forecasts are often accurate but never exact. Forecasting must be treated in part therefore as a random process with errors distributed according to some probability distribution. It is important for us to quantify these error distributions not only as a guide to users of the forecasts but also as a measure of progress in the development of improved forecasting methods. A general theory of observation and prediction error was first formulated by Gauss (1809) for orbit calculations in celestial mechanics and has since been applied in many fields of science and technology.

In these lecture notes I shall show how these general ideas have been applied to an analysis of forecasting error. I have two general goals. The first is to develop a calculus of error variance from which a simple forecast error budget may be constructed. This is done in Sections 3 and 4. The second is to provide the basis for judging the significance of any change in error statistics in tests of proposed improvements in an observing and forecasting system. This is done in Section 5. In Section 2 I summarize the general definitions of random variables and vectors which I apply to random errors in Section 3.

2. Random variables and vectors

2.1 Random scalars

In common usage the adjective random means uncertain, unknown, or unpredictable, and a random variable would be considered in some way ill-defined. In probability theory a random variable is defined as a variable which takes on different values with a specified proba-

*On leave from the National Center for Atmospheric Research which is sponsored by the National Science Foundation.

bility. Thus associated with a real random variable x is a probability function P(x) such that the probability that $x \leq x_0$ is $P(x_0)$. If P(x) is differentiable then there is an associated probability density distribution p(x) such that $dP(x) = p(x)dx$. Without much loss in generality I shall consider that this is the case in these notes.

A random variable may be characterized by a hypothetical ensemble of an infinite number of possible and equally likely members each labeled by a value of x. The probability density p(x) is then a measure of the number density of members at x when they are sorted and distributed along an x-axis.

A probability density p(x) must clearly be normalized,

$$\int_{-\infty}^{\infty} p(x)dx = P(\infty) = 1 \tag{2.1}$$

and everywhere non-negative,

$$p(x) \geq 0. \tag{2.2}$$

but is otherwise unconstrained. For an arbitrary function f(x) of a random variable, which attaches to a member of the ensemble labeled by x a functional value f, we denote the average of f over the ensemble as

$$\langle f \rangle = \int_{-\infty}^{\infty} f(x)p(x)dx \tag{2.3}$$

Only if the integral converges does the average of f exist as a finite number.

Moments are defined as averages of powers, thus the nth moment is given by

$$\langle x^n \rangle = \int_{-\infty}^{\infty} x^n p(x)dx. \tag{2.4}$$

The zeroth moment is identically equal to unity by the normalization condition (2.1). We shall be primarily interested in the first and second moments. The first moment $\langle x \rangle$ is called the mean of the distribution p(x). Averages such as the mean are not random variables but sharp variables or constants independent of x. As for the zeroth moment, the average of a sharp variable is equal to itself. The averaging operation is clearly linear, so that if we define a new random variable as

$$y = x - \langle x \rangle \tag{2.5}$$

its mean will vanish since

$$\langle y \rangle = \langle x \rangle - \langle\langle x \rangle\rangle = \langle x \rangle - \langle x \rangle = 0. \tag{2.6}$$

Higher order central moments or moments about the mean are defined in terms of this shifted variable. The variance, X, is the second central moment,

$$X = \langle (x-\langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \tag{2.7}$$

I shall generally denote the variance as here by capitalization.

In his analysis Gauss (1820) introduced for simplicity the normal (or Gaussian) distribution

$$p(x) = (2\pi)^{-1/2} X^{-1/2} \exp\left[-(\tfrac{1}{2})X^{-1}(x-\langle x \rangle)^2\right] \tag{2.8}$$

as a plausible distribution for random errors. It depends on only two parameters $\langle x \rangle$ and X which are, as indicated, its mean and variance. We shall follow the lead of Gauss and describe a theory of errors in terms of first and second moments only. We shall also assume that we are dealing with normal probability distributions.

2.2 Random vectors

A random vector is characterized by an ensemble of vectors $\underline{x}$ in an n-dimensional vector space with an associated probability density distribution which is nonnegative

$$p(\underline{x}) \geq 0. \tag{2.9}$$

and normalized

$$\int p(\underline{x})\, d\underline{x} = 1 \tag{2.10}$$

A random vector is generally more than a collection of random components. If the components are independent random variables with distributions $p_1(x_1), p_2(x_2), \ldots p_n(x_n)$ then

$$p(\underline{x}) = p_1(x_1)p_2(x_2)\ldots p_n(x_n), \tag{2.11}$$

but in general $p(\underline{x})$ can not be so factored and induces probabilistic relations between the components.

Averages and moments are defined as generalizations of their

definition for random variables. In particular, the mean is a vector

$$\langle \underline{x} \rangle = \int \underline{x}\, p(\underline{x})\, d\underline{x} \tag{2.12}$$

For notational convenience a random vector $\underline{x}$ will be treated as a column vector in component form with $\underline{x}^*$ indicating the transposed row vector. Then $\underline{x}\underline{x}^*$, in usual matrix multiplication notation, is an n×n square matrix. The covariance matrix is defined as the central second moment

$$\underline{X} = \langle (\underline{x}-\langle \underline{x} \rangle)(\underline{x}-\langle \underline{x} \rangle)^* \rangle = \langle \underline{x}\underline{x}^* \rangle - \langle \underline{x} \rangle \langle \underline{x}^* \rangle \tag{2.13}$$

Its diagonal elements are variances of the associated components; its nondiagonal elements are covariances of the associated pairs of components. An important property of a covariance matrix is that it is non-negative definite, which means that for any (sharp) vector $\underline{a}$ the scalar, $\underline{a}^* \underline{X}\, \underline{a}$ is nonnegative. To see this let $\underline{y} = \underline{x} - \langle \underline{x} \rangle$ so that, since $\underline{a}^* \underline{y} = \underline{y}^* \underline{a}$ is a random scalar, we have

$$\underline{a}^* \underline{X}\, \underline{a} = \underline{a}^* \langle \underline{y}\underline{y}^* \rangle \underline{a} = \langle (\underline{a}^* \underline{y})(\underline{y}^* \underline{a}) \rangle = \langle (\underline{a}^* \underline{y})^2 \rangle \geq 0. \tag{2.14}$$

By a suitable linear transformation of coordinates the covariance matrix $\underline{X}$ may be diagonalized, thus displaying its eigenvalues as variances on the diagonal of a matrix whose non-diagonal covariance elements vanish. These eigenvalues are nonnegative, but some might vanish. The associated eigenvector would in that case be nonrandom or sharp. We shall assume that such a degeneracy does not occur in our applications and thus that $\underline{X}$ is positive definite, i.e. that $\underline{a}^* \underline{X}\, \underline{a}$ is strictly positive. We shall then feel free to invert $\underline{X}$ at will.

By letting the dimensionality of the random vectors become infinite it is possible to generalize the definitions to apply to random functions or field. We shall not do this since the meterological fields of interest for weather forecasting must be represented for computing purposes in terms of a large but finite member of degrees of freedom. But this serves as a reminder that we are interested in random vectors whose components may consist of all the spectral components in a prediction model and may have a dimension of order 10^5.

3. Probabilistic measures of error

3.1 Scalar error

Let us suppose that we measure using an imperfect technique a

physical quantity, whose true value, is x_t. Owing to imperfections the measured value x_m will differ from x_t by a measurement error $e = x_m - x_t$. We consider the measurement to select one member out of an ensemble of equally likely possible errors, and thus the error e is a random variable with some probability density p(e). We shall assume that the first two moments of p(e) are known from experience such as provided by repeated independent applications of the measurement technique to a known standard. The first moment $\langle e \rangle$ is the mean error or bias which by recalibration we may assume to vanish. The second central moment, $E = \langle e^2 \rangle$ is the mean square error or error variance which, following Gauss, we take to be the primary probabilistic measure of error. Clearly the smaller is E the more accurate is the measurement technique. We shall define the accuracy $A = E^{-1}$ as the inverse of E.

The random error e induces randomness in the measured value

$$x_m = x_t + e \tag{3.1}$$

even though x_t is sharp. Over an infinite ensemble of measurements the mean of x_m is

$$\langle x_m \rangle = x_t + \langle e \rangle = x_t \tag{3.2}$$

and its variance is

$$E + \langle e \rangle^2 = E. \tag{3.3}$$

Suppose in addition that x_t is a random variable with mean 0 and variance X_t and that the error e is independent of x_t. Then the variance X_m of measured values is given by

$$X_m = \langle x_m^2 \rangle = \langle (x_t+e)^2 \rangle = X_t + E \geq X_t \tag{3.4}$$

since, in this case, $\langle x_t e \rangle = 0$. Note that this result is unchanged by a change in sign of e. Similarly if there are two independent errors e_1 and e_2 in a measurement then

$$x_m = x_t + e_1 + e_2 \tag{3.5}$$

and

$$X_m = X_t + E_1 + E_2 \tag{3.6}$$

It is this additive property of independent sources of variance that makes variance such a useful simple measure of error.

Another common error measure is the standard error or root mean square (rms) error defined in general as $\langle e^2 \rangle^{1/2}$. This is a measure of the width of the error probability distribution in the same physical units as x_m. The standard error does not, however, have the additive property of error variance.

Accuracy can be additive also. Consider two measurements of x_t made by separate, independent, and perhaps differing measurement techniques giving the pair of results

$$\begin{aligned} x_{m1} &= x_t + e_1 \\ x_{m2} &= x_t + e_2 \end{aligned} \tag{3.7}$$

We assume that each measurement is unbiased, but that they may have different error variances E_1 and E_2. One expects that some weighted combination of the two measured values

$$x_m = qx_{m1} + (1-q)x_{m2} \tag{3.8}$$

should be more accurate than either by itself. This is, in fact, the case in a probabilistic sense. Clearly the combined error

$$e = qe_1 + (1-q)e_2 \tag{3.9}$$

has a variance

$$E = \langle \left[qe_1 + (1-q)e_2 \right]^2 \rangle = q^2 E_1 + (1-q)^2 E_2 \tag{3.10}$$

which may be minimized by finding a value of q such that

$$dE/dq = 2qE_1 - 2(1-q)E_2 = 0 \tag{3.11}$$

Eq. (3.11) is satisfied by

$$q = \frac{E_2}{E_1+E_2} = \frac{A_1}{A_1+A_2}, \quad 1 - q = \frac{A_2}{A_1+A_2} \tag{3.12}$$

where $A_1 = E_1^{-1}$, $A_2 = E_2^{-1}$ are the respective accuracies. Note that the weights are proportional to the corresponding accuracies and that this is indeed a minimum in E since

$$d^2E/dq^2 = 2(E_1+E_2) > 0. \tag{3.13}$$

The value of E at the minimum is given by

$$\begin{aligned} E &= (A_1+A_2)^{-2}\left[A_1^2E_1 + A_2^2E_2\right] \\ &= (A_1+A_2)^{-1} \end{aligned} \tag{3.14}$$

The final accuracy

$$A = E^{-1} = A_1 + A_2 \tag{3.15}$$

is the sum of the two contributing accuracies and therefore greater than either.

A special case of some interest is for an independent repetition of the same measurement technique whence $E_2 = E_1$, $A_2 = A_1$, and the two weights are equally 1/2. Then $A = 2A_1$ and $E = \frac{1}{2}E_1$. This can be extended to many repetitions, thus for n independent measurements the best estimate of x_t is an equally weighted average of the x_{mi}'s and its accuracy is n times greater than that of a single measurement.

In summary, independent sources of error lead to additive error variances, but independent sources of information lead to additive accuracies.

3.2 Error vectors

For an imperfect measurement of a vector $\underline{x}_t$ we have

$$\underline{x}_m = \underline{x}_t + \underline{e} \tag{3.16}$$

where $\underline{e}$ is a random error vector. The extension of error measures to random vectors is fairly straightforward. We may again without much loss of generality set the mean error, now a vector, to zero, thus $\langle \underline{e} \rangle = \underline{0}$, where $\underline{0}$ is a vector all of whose components are zero. The error is then characterized by a positive definite error covariance matrix

$$\underline{E} = \langle \underline{e}\underline{e}^* \rangle \tag{3.17}$$

whose matrix inverse

$$\underline{A} = \underline{E}^{-1} \tag{3.18}$$

serves as a positive definite accuracy matrix.

In the case that all error covariance and accuracy matrices under discussion commute they all have a common set of eigenvectors and can all be diagonalized by the same transformation. Each error eigenmode then is independent of all others, and the scalar algebraic manipulations of adding error variance or accuracies carry through separately.

In particular, two independent sources of error with covariance matrices $\underline{E}_1$ and $\underline{E}_2$ contribute a total error with covariance matrix $\underline{E} = \underline{E}_1 + \underline{E}_2$. Two independent sources of information with accuracy matrices $\underline{A}_1$ and $\underline{A}_2$ provide a total accuracy $\underline{A} = \underline{A}_1 + \underline{A}_2$ when combined with matrix weights

$$\underline{Q}_1 = \underline{A}^{-1}\,\underline{A}_1,\ \underline{Q}_2 = \underline{A}^{-1}\,\underline{A}_2 \tag{3.19}$$

so that

$$\underline{x} = \underline{A}^{-1}\,\underline{A}_1\underline{x}_1 + \underline{A}^{-1}\,\underline{A}_2\underline{x}_2 \tag{3.20}$$

In case the matrices $\underline{E}_1$ and $\underline{E}_2$ do not commute the summing of error covariances remains valid, but the solution to the problem of combining information is not so clear. The use of Eqns (3.19) and (3.20) appears however, to remain plausible (Leith 1975).

In the case that $\underline{x}_t$ itself is random, with say $\langle \underline{x}_t \rangle = 0$ and covariance matrix $\underline{X}_t = \langle \underline{x}_t \underline{x}_t^* \rangle$, then this knowledge can serve as an independent source of information to improve the estimate $\underline{x}_m$. We seek, in this case, a regression matrix $\underline{R}$ such that

$$\underline{x}_t = \underline{R}\,\underline{x}_m + \underline{f} \tag{3.21}$$

This is a sort of inversion of Eqn (3.16). Here

$$\hat{\underline{x}}_t = \underline{R}\,\underline{x}_m \tag{3.22}$$

is a best estimate of $\underline{x}_t$ with error covariance

$$\underline{F} = \langle \underline{f}\,\underline{f}^* \rangle \tag{3.23}$$

if $\underline{R}$ is chosen such that

$$\langle \underline{f}\,\underline{x}_m^* \rangle = \underline{0} \tag{3.24}$$

for then $\underline{x}_m$ provides no further information about $\underline{f}$. We have from

Eqn (3.21)

$$\langle \underline{x}_t \, \underline{x}_m^* \rangle = \underline{R} \langle \underline{x}_m \, \underline{x}_m^* \rangle = \underline{R} \, \underline{X}_m \tag{3.25}$$

and from Eqn (3.16)

$$\langle \underline{x}_t \, \underline{x}_m^* \rangle = \langle \underline{x}_t \, \underline{x}_t^* \rangle = \underline{X}_t \tag{3.26}$$

so that with the use of Eqn (3.4) we find

$$\begin{aligned} \underline{R} &= \underline{X}_t \, \underline{X}_m^{-1} = \underline{X}_t (\underline{X}_t + \underline{E})^{-1} \\ &= [\, (\underline{X}_t + \underline{E}) \underline{X}_t^{-1}]^{-1} = [\, \underline{I} + \underline{E} \, \underline{X}_t^{-1}]^{-1} \\ &= [\underline{E} (\underline{E}^{-1} + \underline{X}_t^{-1})]^{-1} = [\underline{E}^{-1} + \underline{X}_t^{-1}]^{-1} \underline{E}^{-1} \\ &= (\underline{A} + \underline{A}_t)^{-1} \underline{A} \end{aligned} \tag{3.27}$$

where $\underline{A} = \underline{E}^{-1}$ and $\underline{A}_t = \underline{X}_t^{-1}$.
The final error covariance is

$$\begin{aligned} \underline{F} &= \langle \underline{f} (\underline{x}_t - \underline{R} \, \underline{x}_m)^* \rangle = \langle \underline{f} \, \underline{x}_t^* \rangle \\ &= \langle (\underline{x}_t - \underline{R} \, \underline{x}_m) \underline{x}_t^* \rangle = \underline{X}_t - \underline{R} \langle \underline{x}_m \, \underline{x}_t^* \rangle \\ &= \underline{X}_t - \underline{R} \, \underline{X}_t = [\underline{I} - (\underline{A} + \underline{A}_t)^{-1} \underline{A}] \underline{X}_t \\ &= (\underline{A} + \underline{A}_t)^{-1} \, \underline{A}_t \, \underline{X}_t = (\underline{A} + \underline{A}_t)^{-1} \end{aligned} \tag{3.28}$$

Standard regression analysis leads thus to the expected result. $\underline{R}$ is the normalized weight for $\underline{x}_m$ in the combination, $\underline{I}-\underline{R}$ is the normalized weight for $\langle \underline{x}_t \rangle = \underline{0}$, and the final accuracy is $\underline{A} + \underline{A}_t$. No assumption has been made that $\underline{E}$ and $\underline{X}_t$ or $\underline{A}$ and $\underline{A}_t$ commute.

3. Forecasting errors

I have described the error analysis in terms of measurement error, but it is equally valid for forecasting error. In that case $\underline{x}_m = \underline{x}_f$ represents the forecast atmosphere state vector in a model and $\underline{x}_t$ represents the state vector in the model that most closely describes the true state of the atmosphere at that time. The error ensemble is based hypothetically on an infinite number of forecasts for an infinite ensemble of atmospheric states. This ensemble of

states $\underline{x}_t$ can be considered as the climate ensemble and it is convenient to set $<\underline{x}_t> = 0$ by defining the state vector components as anomalies from the climate mean. The regression analysis that utilizes information about the climate covariance matrix $\underline{X}_t$ is an example of "statistically optimal analysis" when applied to observations or of "optimal filtering" when applied to forecasts.

One practical problem in utilizing random error vector analysis becomes clear immediately. The number of different components in an error covariance matrix $\underline{E}$ is N(N+1)/2 where N is the dimension of the model state vector. For N of order 10^5 this is a tremendous amount of information. In practice therefore one usually reduces the measure of error back to a scalar variance quantity by some combination of selection and averaging over diagonal elements. Such a procedure can be strictly justified only if all covariance matrices involved are multiples of a common matrix. In the next section I shall develop simple error budget equations based on this kind of a reduction.

4. Simple prediction error budget equations

4.1 Introduction

It is known from many theoretical predictability studies that even for a perfect prediction model any initial error, no matter how small in scale and amplitude, will progressively contaminate increasingly larger scales and grow with an rms error doubling time of about 2 1/2 days. Prediction models, however, are not perfect and introduce an additional source of error growth. A simple error-growth equation (Leith, 1978) has been used to describe the initial growth of small errors and to distinguish model-induced error sources from the inherent growth of initial analysis errors. Although the equation is based on rather crude assumptions, it seems to provide a consistent fit to observed error growths during the first day or so of prediction. It has recently been applied with some success to the operational ECMWF prediction model by Bengtsson (1981).

In this section the equation will be summarized and extended to the calculation of the error budget for a standard data assimilation procedure. This extension provides an estimate of the relative impact of model errors on the equilibrium error level of the final analysis with assimilation. The saturation effects of climate variance on error growth at late times will also be discussed.

4.2 Error variance

In dealing with error budgets, it is far more natural to use

mean square error or error variance E rather than the commonly used root mean square (rms) error. If, for example, a particular determination is afflicted by two independent errors with variances E_1 and E_2, then the resulting error variance is the simple sum $E = E_1 + E_2$. If, on the contrary, two independent determinations with error variances E_1 and E_2 are combined into a better final determination, then the inverses, which measure accuracy or information content, are summed; $E^{-1} = E_1^{-1} + E_2^{-1}$. We shall use both of these general statistical principles which were developed in Section 3.

The use of a single number E to describe the error variance of an atmospheric state ignores details of error distributions over space or over spatial scales. This is less serious for compositing error than for compositing information, but in either case an implicit assumption is made that all distributions are similar.

4.3 Error growth equation

The simple error growth equation (Leith, 1978) is

$$\dot{E} = \alpha E + S \tag{4.1}$$

The term αE describes the inherent tendency for error to grow owing to the unstable nature of atmospheric dynamics. The rms error doubling time of 2 1/2 days given by predictability theory translates into an error variance doubling time of 1.25 days and a value of $\alpha = 0.5545 \text{ day}^{-1}$. The term S describes the model error source rate, which is model dependent and can be empirically determined by fitting observed error growth values.

Analysis error variance, which includes observation errors, will be denoted by E_0 and provides an initial value for the integration of Eq. (4.1) with the result

$$E(t) = E_0 + (E_0 + S/\alpha)[\exp(\alpha t) - 1] \tag{4.2}$$

It is convenient to replace the time variable t with the pseudo-time variable

$$\tau = \frac{1}{\alpha}[\exp(\alpha t) - 1] \tag{4.3}$$

in terms of which the error growth is linear. The perceived error variance involves a verification against a later analysis, and this contributes an additional term E_0 under the simple assumption that the verifying analysis has independent errors. Thus, for $\tau > 0$, we

have

$$E_p(\tau) = 2E_0 + (\alpha E_0 + S)\ \tau \qquad (4.4)$$

For a particular model a linear empirical fit to a plot of perceived values of E_p against values of $\tau > 0$ for a day or so determines first E_0 from the intercept and then S from the slope. Greater confidence is achieved by, at the same time, fitting values of $E_p(\tau)$ for a null model, namely, those of persistence forecasts. The intercept should be the same but the null model slope is greater and determines a value S_0. The ratio S/S_0 is a dimensionless figure of merit for a model. It must be remembered in making the linear empirical fit that Eq.(4.1) includes no effects of saturation for large errors, thus that smaller errors at shorter times should be more heavily weighted.

It is assumed that the foregoing determination of E_0 is without any benefit of assimilation methods. Thus, E_0 is the error variance of an analysis which may use climate but not a model prediction as a source of a priori information. The benefits of assimilation will be examined next.

4.4 Data assimilation

The basic idea of data assimilation is to combine information from a new set of observations with the a priori information about the state of the atmosphere available from a short-range prediction valid at the new observing time. In this way information from earlier observations is carried forward, although somewhat degraded, to provide an independent source of information to be added to that newly acquired. It is straightforward to compute an error budget for the assimilation process by using Eq. (4.1) between observation times and the general principle for compositing information at observation times.

Let now τ be the fixed pseudo-time interval of the assimilation cycle, and let E_n be the error variance after data assimilation at the nth cycle. According to Eq. (4.1), prediction over the pseudo-time interval τ leads to a prediction error variance

$$E_{\tau,n} = E_n\ (1 + \alpha\tau) + S\tau \qquad (4.5)$$

The introduction of new observations with error variance E_0 will lead to a new value E_{n+1} according to the general principle by which information is composited, thus

$$E_{n+1}^{-1} = E_0^{-1} + E_{\tau,n}^{-1} \tag{4.6}$$

To cast the problem in dimensionless form, let $\varepsilon_n = E_n/E_0$, $\sigma = S/\alpha E_0$, and $\beta = 1 + \alpha\tau$. Then Eqs. (4.5) and (4.6) may be combined to give the iterative expression

$$\varepsilon_{n+1} = [\, 1 + \{\beta\, \varepsilon_n + \sigma\alpha\tau\}^{-1}\,]^{-1} \tag{4.7}$$

As n increases, an equilibrium level

$$\varepsilon = \lim_{n\to\infty} \varepsilon_n = \lim_{n\to\infty} \varepsilon_{n+1} \tag{4.8}$$

is reached which is the factor by which data assimilation reduces the observational error variance E_0. It is easy to deduce from Eqs. (4.7) and (4.8) that $\in$ must satisfy the quadratic equation

$$\varepsilon^2 + \eta(\sigma - 1)\, \varepsilon - \eta\sigma = 0 \tag{4.9}$$

where $\eta = \alpha\tau/\beta$ is a dimensionless parameter depending only on the assimilation time interval. The relevant root of Eq. (4.9) is given by

$$\varepsilon = [\, \eta\sigma + \{\eta(\sigma - 1)/2\}^2\,]^{1/2} - \{\eta(\sigma - 1)/2\}. \tag{4.10}$$

and is displayed in Fig. 1 as a function of σ for values of η corresponding to assimilation time intervals of 0.25 day and 0.50 day and for $\alpha = 0.5545$ day^{-1}.

Error growth results from an early GISS research model (Druyan, 1974) were fitted by Eq. (4.4) both for 500 mb height errors and velocity errors (Leith, 1978). For height errors, the resulting values are $E_0 = 200$ m^2, $S = 900$ m^2 day^{-1} and thus $\sigma = 8.25$. For velocity errors, the fitting parameters are $E_0 = 15$ m^2 sec^{-2}, $S/\alpha = 17.3$ m^2 sec^{-2} day^{-1} and thus $\sigma = 2.1$. More recently, Bengtsson (1981) reports for the ECMWF operational prediction model 5oo mb height error values of $E_0 = 150$ m^2 and $S = 400$ m^2 day^{-1} with $\sigma = 4.8$. It is not clear, however, whether E_0 in this case reflects already the benefits of assimilation.

Fig. 1 shows quantitatively how a decrease in model error sources leads to an improved equilibrium error variance. The greatest benefits appear to accrue when $\sigma = S/\alpha E_0$ is reduced to less than 1. It also appears that velocity errors may be decreased by assimilation more than are height errors.

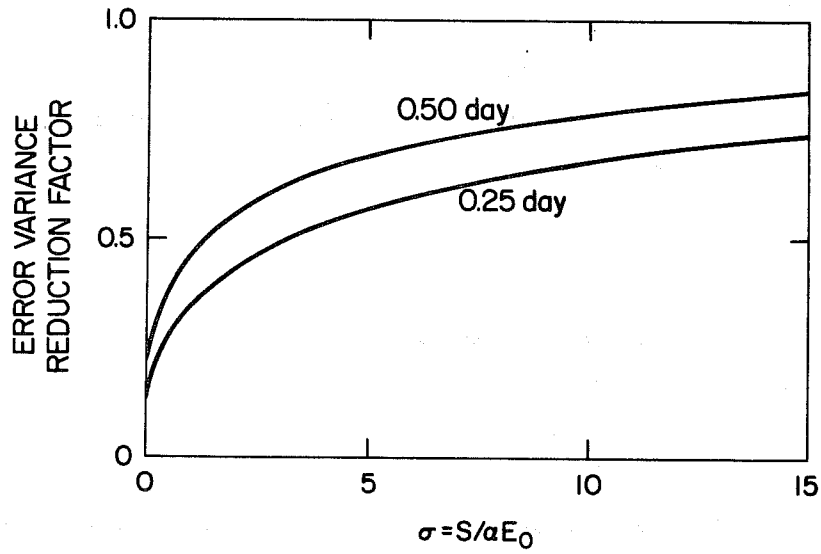

Fig. 1. Assimilation equilibrium error variance reduction ε vs. dimensionless model error source rate σ. Inherent error growth rate parameter $\alpha = 0.5545$ day^{-1}.

4.5 Error growth saturation

It is a well known phenomenon that if either rms error or error variance for forecasts are plotted as a function of forecast period the curves approach asymptotically an upper level related to the climate variance. The reasons for this are clear for at late time there is vanishing correlation between the forecast anomaly $\underline{x}_f$ and the true anomaly $\underline{x}_t$. In this limit the error covariance matrix becomes

$$\underline{E} \to \langle (\underline{x}_f - \underline{x}_t)(\underline{x}_f - \underline{x}_t)^* \rangle$$

$$= \langle \underline{x}_f \, \underline{x}_f^* \rangle + \langle \underline{x}_t \, \underline{x}_t^* \rangle$$

$$= \underline{X}_f + \underline{X}_t \qquad (4.11)$$

and any derived scalar variance will likewise give

$$E \to X_f + X_t. \qquad (4.12)$$

If the climate generated by a forecasting model when run for a long time is the same as the true climate then $X_f = X_t$ and $E \to 2X_t$ or for rms error $E^{1/2} \to 2^{1/2} X_t^{1/2}$.

Clearly in this case in which knowledge of the climate has not entered into the forecasting process the forecast error at long range reaches a value greater than that of a climate mean forecast, namely $\underset{\sim}{x}_f = \underset{\sim}{0}$, for which the error is only X_t. As described in Section 3, however, it is possible to introduce climate as an additional source of information by regression filtering and then saturation of forecast error variance occurs at a lower level with $E \to X_t$.

This large difference in saturation levels provides a warning that error variance can depend on any filtering or smoothing that has been carried out advertently or not. This suggests that the most reliable error variance for verification purposes is the irreducible value resulting from optimal regression filtering.

No saturation effects have been included in the simple error budget equations so far derived. For empirical fitting to initial error growth curves and for analysis of the assimilation process this has not been important. Perhaps the simplest way to add the effects of saturation is to add climate information or accuracy to that of the model forecast. Eq (4.2) when written in terms of the dimensionless pseudo-time $\bar{\tau} = \alpha\tau$ and the dimensionless parameter $\sigma = S/\alpha E$ becomes

$$E(\tau) = E_0[\,1 + (1+\sigma)\bar{\tau}\,] \,. \tag{4.13}$$

The associated model accuracy equation becomes

$$A(\tau) = A_0\,[\,1 + (1+\sigma)\bar{\tau}\,]^{-1} \tag{4.14}$$

which approaches zero as $\bar{\tau} \to \infty$. Eq (4.14) may be modified for saturation effects, in the case that regression filtering is used, to give

$$A(\tau) = A_0[\,1 + (1+\sigma)\bar{\tau}\,]^{-1} + A_t \tag{4.15}$$

whence Eq (4.13) is modified to become

$$E(\tau) = \{E_0^{-1}[\,1 + (1+\sigma)\bar{\tau}\,]^{-1} + X_t^{-1}\}^{-1} \tag{4.16}$$

Note that $E(\tau) \to X_t$ as $\bar{\tau} \to \infty$.

4.6 Anomaly correlation

The correlation between forecast and observed anomalies is a frequently used measure of forecasting skill which is related to the various measures already discussed. Consider first the anomaly co-

variance matrix

$$\underline{C} = \langle \underline{x}_f \, \underline{x}_t^* \rangle \tag{4.17}$$

An expansion of the error covariance matrix $\underline{E}$ leads to

$$\underline{E} = \langle (\underline{x}_f - \underline{x}_t)(\underline{x}_f - \underline{x}_t)^* \rangle$$

$$= \underline{X}_f + \underline{X}_t - [\underline{C} + \underline{C}^*]$$

A scalar covariance measure based solely on diagonal elements becomes then

$$C = \frac{1}{2}[X_f + X_t - E] \tag{4.19}$$

or

$$C = X_t \left[1 - \frac{1}{2}(E/X_t)\right] \tag{4.20}$$

when $X_f = X_t$. In this case the anomaly correlation is the bracketed term in Eq (4.20) namely

$$r = 1 - \frac{1}{2}(E/X_t) \tag{4.21}$$

As $t \to \infty$, $E \to 2X_t$ and $r \to 0$.

5. Statistical sampling fluctuations

5.1 Introduction

The discussion so far has been probabilistic in nature in that first and second moments of all probability density distributions have been presumed to be known. In practice, however, we must estimate these by computing error statistics from a finite sample of forecasts. These estimates are subject to statistical sampling fluctuations that occur for finite samples even if the infinite ensemble from which they are drawn remains unchanged. These sampling errors tend to mask any small change in ensemble properties that we are trying to detect. This is a familiar problem for the analysis of climate sensitivity experiments in which one tries to detect small changes in model climate properties in response to changes in the model structures. We consider here the analogous problem for forecast sensitivity experiments. I will first summarize results from sampling theory. These are discussed in more detail in many standard

texts such as that of Cramer (1945).

5.2 Statistical sampling theory

Let the random variable x have mean $\langle x \rangle = 0$ and variance $\langle x^2 \rangle = X$. We draw a sample of n elements $(x_1, x_2, \ldots, x_n)$ from the infinite ensemble with each member equally likely to be chosen. The sample average

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{5.1}$$

should approach the ensemble mean $\langle x \rangle = 0$ as n increases and the sample becomes increasingly representative. As a sum of random variables $\bar{x}$ is itself a random variable whose distribution is determined by an infinite number of repetitions of the sampling process. We can compute the first two moments of $\bar{x}$ and find

$$\langle \bar{x} \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle x_i \rangle = 0$$

$$\bar{X} = \langle \bar{x}^2 \rangle = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle x_i x_j \rangle = \frac{1}{n} X \tag{5.2}$$

Thus $\bar{x}$ has the same mean as x but its variance is less by a factor 1/n. As $n \to \infty$ we see that $\bar{X} \to 0$, and it is in this sense that $\bar{x} \to 0$.

An unbiased estimate of X is provided by the sample statistic

$$v = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{5.3}$$

since

$$\langle v \rangle = \frac{1}{n-1} \sum_{i=1}^{n} \langle (x_i - \bar{x})^2 \rangle$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left[X - \frac{2}{n} X + \frac{1}{n} X \right] = X \tag{5.4}$$

its second moment is

$$\langle v^2 \rangle = \frac{1}{(n-1)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle (x_i - \bar{x})^2 (x_j - \bar{x})^2 \rangle \tag{5.5}$$

which for a normal distribution of x can be evaluated as

$$\langle v^2 \rangle = X^2 + \frac{2}{n} X^2 \tag{5.6}$$

The variance of v is then given by

$$\langle v^2 \rangle - \langle v \rangle^2 = \frac{2}{n} X^2 \tag{5.7}$$

which also tends to zero as $n \to \infty$.

In summary although $\bar{x}$ of Eq (5.1) and v of Eq (5.3) are unbiased estimates of $\langle x \rangle$ and X they have associated sampling error variances of $\frac{1}{n}X$ and $\frac{2}{n} X^2$ respectively. The corresponding rms sampling errors are, of course, the square root of these, and if an observed change in $\bar{x}$ or v is small in magnitude compared to these, we may doubt its statistical significance as indicating a change in $\langle x \rangle$ or X.

In application to forecast sensitivity experiments $\langle x \rangle$ becomes the forecast bias $\langle x_f \rangle$ and X the error variance E. We have for convenience assumed that $\langle x_f \rangle = 0$, but in fact the climate mean of a model usually differs somewhat from that of the atmosphere and thus of the initial states. The model forecasts tend to drift toward the model climate mean in a few days introducing a climate mean drift bias. Experiments on model improvements designed to decrease this bias must be interpreted in the light of sampling errors of the mean which have a variance E/n.

We are of course, also interested in changes that will reduce E as estimated by v, but here we must face sampling errors of the variance with, in turn, a variance of $(2/n)E^2$ or standard error of $(2/n)^{1/2}$ E. The significance of any change δ E will depend on the dimensionless ratio

$$w = (n/2)^{1/2} \, (\delta E/E) \tag{5.8}$$

Eq (5.8) may be combined with Eq (4.16) to determine the best forecast range for determining, say, the effect of a change in the model error source rate S. If we let

$$\epsilon_c = X_t/E_0 \tag{5.9}$$

and

$$\bar{\epsilon} = E/E_0 \tag{5.10}$$

then Eq (4.16) may be written in dimensionless form as

$$\bar{\epsilon}(\tau) = \{ 1 + (1+\sigma)\bar{\tau}]^{-1} + \epsilon_c^{-1}\}^{-1} \qquad (5.11)$$

According to Eq (5.8) the detectability of a change in σ is proportional to

$$\kappa(\tau) = \frac{1}{\bar{\epsilon}} \frac{\partial \bar{\epsilon}}{\partial \sigma} = \bar{\epsilon}\, \lambda^{-2}\bar{\tau} \qquad (5.12)$$

where $\lambda = 1 + (1+\sigma)\bar{\tau}$.

The quantity κ approaches 0 as $\bar{\tau} \to 0$ where the effect of model errors has not yet been felt and as $\bar{\tau} \to \infty$ where climate variance dominates the error variance. In between it has a maximum at $\bar{\tau} = \bar{\tau}_m$ where

$$\bar{\tau}_m = \frac{(1+\epsilon_c)1/2}{1 + \sigma} \qquad (5.13)$$

Evaluation of Eq. (5.13) for typical values such as $\sigma = 6$ and $\epsilon_c = 48$ gives a value $\bar{\tau}_m = 1$ and an optimal forecast time interval $t_m = 1.25$ days.

The anomaly correlation measure r defined by Eq. (4.21) is also subject to sampling errors. The correlation coefficient r is bounded with $-1 \leq r \leq 1$, and its sampling probability distribution is a quite complicated function of the ensemble correlation ρ, the sample correlation r, and n. This distribution is considerably simplified by the transformation of variables introduced by Fisher (1941)

$$\zeta = \frac{1}{2} \log \frac{1+\rho}{1-\rho} \qquad (5.9)$$

$$z = \frac{1}{2} \log \frac{1+r}{1-r} \qquad (5.10)$$

in terms of which the distribution of z is more nearly normal with moments

$$\langle z \rangle = \zeta + \frac{\rho}{2(n-1)} \qquad (5.11)$$

$$Z = \langle z^2 \rangle - \langle z \rangle^2 = \frac{1}{n-3} \qquad (5.12)$$

to lowest order in 1/n.

5.3 Effective sample size

Sampling theory is based on knowledge of the number n of independently drawn members of the infinite ensemble. A special problem

arises in applications to atmospheric statistics where time and space correlations can decrease the effective sample size. For climate studies based on time averages over a time interval T this has led to the definition of an effectively independent sampling time T_0 given (Leith, 1973) as

$$T_0 = \int_{-\infty}^{\infty} R(\tau)d\tau$$

where $R(\tau)$ is a characteristic time-lagged correlation function. For many atmospheric variables T_0 is of the order of a week. The effective sample size is then given asymptotically for large T as $n = T/T_0$. This analysis was appropriate for estimation of a mean. For estimation of a variance a better time is

$$T_0' = \int_{-\infty}^{\infty} R^2(\tau)d\tau \approx T_0/2$$

The number n is greater by about a factor of 2 for variance than for mean estimates which tends to cancel the factor 2 appearing in Eq (5.7).

Similar arguments are appropriate for spatial correlations when statistics are generated by averaging over space. By analogy one defines an effectively independent averaging area as

$$A_0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(x,y)dx\,dy$$

for first moments and

$$A_0' = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R^2(x,y)dx\,dy \approx A_0/2$$

for second moments. The approximations for T_0' and A_0' are based on observations that time-lagged correlations are nearly decaying exponentials while space correlations are nearly Gaussian.

6. Conclusion

These lecture notes have been primarily concerned with the more theoretical aspects of the subject and have ignored many practical difficulties in forecast verification. I hope that the following key ideas can serve as a guide for future work.

1) With appropriate definitions error and accuracy can be treated as additive (Section 3)

2) Efforts to construct error budget equations can sharpen our understanding of the total forecasting system (Section 4)

3) Our knowledge of the state of the atmosphere at any time depends to a considerable extent on the forecasting model used to bring forward past information (Section 4.4).

4) We must not be misled by statistical sampling fluctuation when testing possible improvements in a forecasting technique (Section 5)

References

Bengtsson, L. 1981 Medium range weather forecasting at ECMWF and remaining problems. Extended abstract, Symposium on Current Problems of Weather Prediction, Vienna, June, 1981.

Cramer, H. 1945. Mathematical Methods of Statistics. Princeton University, Press, chaps 27-29.

Druyan, L.M. 1974. Short-range forecasts with the GISS model of the global atmosphere. Mon. Weather Rev., 102, 269-279.

Fisher, R.A. 1941 Statistical Methods for Research Workers. Eigth ed., Edinburgh and London

Gauss, K.F. 1809 Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections. C.H. Davis Transl., republished by Dover Publ., 1963.

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

Leith, C.E. 1973 The standard error of time-averaged estimates of climatic means. J. Appl. Meteor., 12, 1066-1069.

Leith, C.E. 1975 Statistical-dynamical forecasting methods. In Seminars on Scientific Foundation of Medium Range Weather Forecasts, Reading, ECMWF.