Design of a Monte Carlo Forecasting Procedure

Eric J. Pitcher[1]

National Center for Atmospheric Research[2]

Boulder, Colorado 80307

1 On leave from the Division of Meteorology and Physical Oceanography, University of Miami.

2. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

1. Introduction

Principally two reasons may be given for designing a forecasting scheme which, along with the conventional prediction of meteorological fields, produces a state-dependent estimate of the reliability of those predictions. In the first place, this permits the identification of those flow regimes which have more (or less) predictable consequences than others. Such additional information would be a valuable aid to the users of such forecasts. Secondly, a prediction of forecast error valid at analysis time would allow a quantitative blending of the forecast with the new observations. Stochastic dynamic forecasts have the potential of achieving this objective although the penalty to be paid is a substantial one in additional computing resources. However, Monte Carlo approximations to stochastic dynamic forecasts may be possible with present resources.

Fig. 1 is an example of the type of information produced by a stochastic dynamic forecast. This was made using a hemispheric barotropic spectral model with 105 real spectral coefficients (rhomboidal truncation at wavenumber 10). The standard deviation field is a measure of confidence that may be placed in the geopotential height field and, as expected, is rather inhomogeneous. Such a display presents only a small part of the total information contained in such forecasts because the technique permits the prediction of the whole covariance structure as well. All of these second moment statistics may be utilized in depicting the uncertainty information in various ways (Epstein and Fleming, 1971), as well as forming part of an integrated analysis

procedure (Epstein and Pitcher, 1972). Fig. 1c is the standard deviation of the subsequent analysis and is less than that of the forecast or observation (not shown) alone.

2. Formulation of stochastic dynamics

Consider a physical system of N parameters x_i ($i = 1,2,...N$). The state of the system at any instant may be represented by a point in an N-dimensional state or phase space, where the coordinates of that point are the values taken on by each of the x_i. Subsequent evolution of the system traces out a path in phase space. Whenever the initial state is subject to uncertainty, its specification is limited to a hypothetical ensemble of phase points distributed according to some assumed probability distribution. The size of this ensemble reflects our uncertainties in the observation and analysis of the initial state of the system. Of particular interest is the prediction of the ensemble mean, as this is the best estimate of the true state based on the criterion of minimum mean-square error, and also the ensemble variance, which becomes a measure of the uncertainty in the mean. The formalism for achieving this consists of ensemble averaging the equations governing the system. To the writer's knowledge, stochastic dynamic prediction, as discussed in the meteorological literature, has been limited to quadratically nonlinear systems. The consequence of this restriction is considered later. As is typical of such systems, the prognostic equation for each x_i may be put in the form

$$dx_i/dt = \dot{x}_i = \sum_{j,k} a_{ijk}\, x_j x_k - \sum_j b_{ij}\, x_j + c_i \,, \tag{2.1}$$

where a_{ijk}, b_{ij} and c_i are constants. Epstein (1969) derived prognostic equations for the mean $\mu_i = E(x_i)$ and covariance $\sigma_{ij} = E(x_i - \mu_i)(x_j - \mu_j)$, where E denotes an ensemble average. The results are

$$\dot{\mu}_i = \sum_{j,k} a_{ijk}(\mu_j\mu_k + \sigma_{jk}) - \sum_j b_{ij}\, \mu_j + c_i \,, \tag{2.2}$$

$$\begin{aligned}\dot{\sigma}_{ij} = \sum_{k,\ell} [&a_{ik\ell}(\mu_k\, \sigma_{j\ell} + \mu_\ell\, \sigma_{jk} + \tau_{jk\ell}) \\ &+ a_{jk\ell}(\mu_k\, \sigma_{i\ell} + \mu_\ell\, \sigma_{ik} + \tau_{ik\ell})] \\ &- \sum_k (b_{ik}\, \sigma_{jk} + b_{jk}\, \sigma_{ik}) \,,\end{aligned} \tag{2.3}$$

where $\tau_{ijk} = E(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)$. Eqs. (2.2) and (2.3) are generally referred to as the stochastic dynamic equations.

Except for the study of low order systems (Fleming, 1971a, 1971b, 1973), the approach just outlined has some rather severe limitations if application to current numerical weather prediction models is contemplated. Foremost among these is the excessive computation time required over a conventional forecast for a system with a reasonably large number of x_i. In an N-parameter system a total of $N(N + 1)/2$ second moment quantities exists, and this represents an increase in computation time of roughly N orders of magnitude. This of course is intolerable in models where N is the order of 10^5. No doubt an approximation in the

form of a subset of (2.3) could be used in which not all covariance quantities were predicted, but it is not immediately clear how this might be achieved. Secondly, the above formulation is limited to quadratically nonlinear systems and, while this appears to be the dominant nonlinearity in atmospheric prediction models, a reformulation taking into account higher order nonlinearities, as are present in certain local physical processes such as radiative exchanges for example, would pose some severe mathematical difficulties and an extra computational burden. Thirdly, stochastic dynamic prediction models ought to be spectral rather than grid point in order to reduce the total number of covariance quantities, but the programming of such models to take advantage of the spectral transform technique of Eliasen et al. (1970) is by no means obvious. Finally, Eqs. (2.2) and (2.3) are an unclosed system and cannot be solved as they stand because of the appearance of third moments τ_{ijk}. Merely discarding third moments is one closure which has been used successfully by the writer (1977) in making short range forecasts, but its performance has not been tested with forecasts that exceed one week.

3. Monte Carlo forecasting

Many of the practical difficulties associated with the solution of the stochastic dynamic equations may be substantially alleviated, if not eliminated entirely, by utilizing a Monte Carlo approach. Leith (1974) has advocated such an alternative and investigated the theoretical skill of the Monte Carlo procedure by application to a two-dimensional

turbulence model. Instead of predicting the moments of the forecast ensemble directly, a sample of points is chosen at random from this ensemble initially, a forecast is made starting from each phase point, and then the moments are estimated by averaging over the statistical sample. Of course such a calculation could be repeated with a different sample, thereby obtaining a different set of statistical estimates. The hope, however, is that we may choose, within the practical limits of computer resources, a sample size m sufficiently large to produce stable statistics. The total computation involved is proportional to mN, but if $m << N$ we may capture the advantages of stochastic dynamic prediction without the penalty of excessive arithmetic. Moreover, the remaining limitations of the stochastic dynamic system of equations discussed previously are not present with the Monte Carlo procedure. However, the problem of moment closure is replaced by the sampling problem in deciding what value to choose for m. This must be done essentially by experimentation.

The first step in carrying out a Monte Carlo forecast is the generation of a sample of initial states chosen at random from a hypothetical ensemble of phase points. We assume the existence of an unbiased initial state vector $\underline{\hat{x}}$ which might be an optimal blend of all the information available. This state should also be compatible with the intended forecast model in satisfying any necessary initialization requirements. For primitive-equation models this could be a linear (Flattery, 1970; Williamson, 1976) or nonlinear (Machenhauer, 1977; Baer, 1977; Daley, 1979) normal mode initialization, for example. The initial sample may now be generated by choosing $\underline{x}_i = \underline{\hat{x}} + \underline{\varepsilon}_i$ $(i = 1,2...m)$.

While the $\underline{\varepsilon}_i$ are chosen at random, they must be done so according to some definite probability distribution, and in view of our limited knowledge of the statistical nature of analysis error, it is convenient to take $\underline{\varepsilon}$ to be normally distributed with zero mean and covariance matrix $\underline{C} = E(\underline{\varepsilon}\ \underline{\varepsilon}')$, which is presumed known. The prime denotes a vector transpose. $\underline{C}$ is derived from the error properties of the observational and analysis procedure. Nonuniformity in the diagonal elements of $\underline{C}$ reflects the inhomogeneity of analysis errors, whereas the off-diagonal elements reveal the spatial correlations in analysis error, both horizontal and vertical. This covariance structure should be accounted for in generating the $\underline{\varepsilon}_i$ since the evolution of the forecast ensemble depends not only upon its initial size, as given by σ_{ii}, but also upon its initial shape, as measured by $\sigma_{ij}(i \neq j)$. Evidently then, the generation of the initial sample is a nontrivial matter as the matrix $\underline{C}$ is not, in general, diagonal. Transforming to a representation in which $\underline{C}$ is diagonal, generating the sample, and then transforming back to the original representation is possible, in principle, though in practice such a task would be computationally burdensome. Usually the initial sample will be generated in some approximate fashion, and one such procedure is discussed in the next section.

Once the initial sample has been generated, the model integrations can be performed. However, in the case of primitive-equation models, each sample member cannot serve directly as an initial condition since there is no assurance that the wind and mass fields are dynamically balanced. This is a consequence of the fact that not all dimensions in

phase space are independent of one another. To the extent that a dynamical balance exists between some of the elements of $\underline{\hat{x}}$, each element of $\underline{\hat{x}}$ may not be perturbed at random without upsetting this balance. Consequently, an initialization step must be performed to suppress spurious gravitational oscillations which would otherwise appear during the course of the model integrations. Whether a relatively simple linear initialization (say) will suffice can perhaps best be answered by experiments.

The subsequent evolution of this sample produces, in the case of a perfect model, a set of equally likely representations of the true state of the atmosphere. The sample variance reflects the uncertainty in the sample mean as an estimate of the true state, and shows the characteristic predictability error growth arising from nonlinear interactions. This internal growth of error does not take into account model imperfections. As such, the predictions of error will be unrealistically low in the absence of some parameterization of external error growth stemming from model inadequacies. Such a parameterization has been considered by the writer (1977) for a stochastic dynamic forecast, wherein random forcing terms were added to the dynamical equations. The simplicity of the basic forecast model required rather large random forcing terms and so reduced somewhat the advantage of state-dependent error growth. However, with every model improvement comes the expectation that less reliance will have to be placed on artificial prescriptions of external error growth. Nonetheless, this will remain a continuing problem whenever realistic error predictions are desired. Its solution may lie

in a more detailed investigation of the nature of forecast error itself. On the other hand, predictions of sample variance based only upon internal error growth may be useful, even now, in making relative assessments of regional forecast reliability.

4. Application to a barotropic model

We turn now to the consideration of a relatively straightforward manner of generating an initial sample that possesses a statistical structure reflecting both inhomogeneity and spatial correlations among analysis errors. The covariance structure of $\underline{\varepsilon}$ is taken to be of the form

$$E[\varepsilon(\lambda_1,\mu_1)\ \varepsilon(\lambda_2,\mu_2)] = r(\lambda_1,\mu_1,\lambda_2,\mu_2) \quad , \tag{4.1}$$

where $\varepsilon(\lambda,\mu)$ denotes the error at a geographic point, λ is the longitude, $\mu = \sin\phi$, and ϕ is the latitude. Choosing spherical harmonics as basis functions, we represent ε as follows:

$$\varepsilon = \sum_{m,n} \left(A_n^m \cos m\lambda + B_n^m \sin m\lambda\right) P_n^m(\mu)$$

$$= \sum_{m,n} C_n^m \cos\left(m\lambda - \theta_n^m\right) P_n^m(\mu) \quad , \tag{4.2}$$

where $A_n^m = C_n^m \cos\theta_n^m$ and for $m > 0$ $B_n^m = C_n^m \sin\theta_n^m$. If θ_n^m and C_n^o are assigned at random, then (4.2) permits the calculation of a random error field. The covariance structure of (4.2) may be easily evaluated and assumes the form:

$$E(\varepsilon_1\varepsilon_2) = \sum_n E\left(C_n^{o\,2}\right) P_n^o(\mu_1)\, P_n^o(\mu_2)$$

$$+ \frac{1}{2} \sum_{m,n} C_n^{m\,2} \cos m(\lambda_1 - \lambda_2)\, P_n^m(\mu_1)\, P_n^m(\mu_2) \quad , \quad (4.3)$$

where in the derivation of (4.3) all of the θ_n^m and C_n^o are assumed to be uncorrelated. This assumption is relaxed subsequently.

A degree of arbitrariness may be exercised in assigning $E\left(C_n^{o\,2}\right)$ and $C_n^{m\,2}$. One choice is to relate these to the climatological variance spectrum by taking

$$\left.\begin{aligned} E\left(C_n^{o\,2}\right) &= k\, f_1(n)\, \gamma_n^{o\,2} \\ C_n^{m\,2} &= k\, f_2(n)\, \gamma_n^{m\,2} \end{aligned}\right\} \quad , \quad (4.4)$$

where $\gamma_n^{m\,2}$ is the natural variance, and k is a constant chosen such that $E(\varepsilon^2)$ as calculated from (4.3) matches the variance of analysis error. The presence of the functions $f_1(n)$ and $f_2(n)$ introduces relative scale dependence in the analysis error. In the following $f_1(n) = f_2(n) = n$, and this implies a linear increase with wavenumber of relative error variance.

Given in Fig. 2 is a normalized plot of (4.3) for $\phi = 45°$. While this is probably not an unreasonable representation of the autocorrelation of analysis error, the statistical nature of this quantity is not well known; no doubt it depends upon the nature of the particular analysis scheme. Once the shape of this function has been established, however, the C_n^m in (4.3) may be reassigned for a more realistic representation.

From (4.3) it is apparent that the variance of analysis error is constant for a given latitude. Nonuniformities in $E(\epsilon^2)$ may be accounted for most easily by summing (4.2) for a given set of θ_n^m and C_n^o, and then selectively enhancing the error field in those regions where the analysis errors are known to be greater. This is equivalent to permitting correlations between the θ_n^m in (4.2) although the procedure just described is much easier to carry out in practice. The modified error field may then be used to perturb $\underline{\hat{x}}$ in physical space, or transformed back to the wavenumber domain if the elements of $\underline{\hat{x}}$ are spectral coefficients. The spatial autocorrelations are not substantially affected by this additional step.

A series of five-day Monte Carlo forecasts has been made using the above statistical model for initial error. The dynamical component is a spectral barotropic model with triangular truncation at wavenumber 18. The error growth properties are given in Figs. 3 and 4. The average root-mean-square forecast error of the sample mean is given in Fig. 3 for a three-month winter period. A choice of m = 10 was found to give relatively stable statistical estimates of second moment quantities. The growth of the sample standard deviation averaged over these 18 winter cases is also shown.

In Fig. 4 is presented one half of the mean-square forecast error as a function of wavenumber. Also shown is the spectrum of sample variance labeled in forecast days. The difference between the respective sets of curves is indicative of the external error growth for this model. It is clear that a Monte Carlo prediction of internal error

growth is a rather unrealistically low estimate of the actual error for this model. While existing operational forecast models would fare somewhat better in predictions of both the mean and variance, the improvements would not be dramatic.

5. Outlook

Realization of the full potential of Monte Carlo forecasts must await further model improvements, and perhaps a more complete understanding of the effects of initial error on model evolution. Somerville (1979) has shown that a consideration of both of these aspects of the problem results in an improved prediction of ultra-long waves.

Since steady advances are expected in computer technology, further research on Monte Carlo forecasting is warranted in anticipation of the day that a full statistical-hydrodynamical treatment of the atmospheric prediction problem may be possible.

REFERENCES

Baer, F., 1977: Adjustment of initial conditions required to suppress gravity oscillations in nonlinear flows. Beitr. Phys. Atmos., 50, 350-366.

Daley, R., 1979: The application of nonlinear normal mode initialization to an operational forecast model. Atmosphere-Ocean, 17, 97-124.

Eliasen, E., B. Machenauer and E. Rassmussen, 1970: On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields. Report No. 2, Institute for Theoretical Meteorology, University of Copenhagen, 35 pp.

Epstein, E.S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-759.

-----, and R.J. Fleming, 1971: Depicting stochastic dynamic forecasts. J. Atmos. Sci., 28, 500-511.

-----, and E.J. Pitcher, 1972: Stochastic analysis of meteorological fields. J. Atmos. Sci., 29, 244-257.

Flattery, T.W., 1970: Spectral models for global analysis and forecasting. Proc. Sixth AWS Technical Exchange Conference, Air Weather Service Technical Report 242, 42-53.

Fleming, R.J., 1971a: On stochastic dynamic prediction: I. The energetics of uncertainty and the question of closure. Mon. Wea. Rev., 99, 851-872.

-----, 1971b: On stochastic dynamic prediction: II. Predictability and utility. Mon. Wea. Rev., 99, 927-938.

-----, 1973: Gravitational adjustment in phase space. J. Appl. Meteor., 12, 1114-1128.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. Mon. Wea. Rev., 102, 409-418.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization. Beitr. Phys. Atmos., 50, 253-271.

Pitcher, E.J., 1977: Application of stochastic dynamic prediction to real data. J. Atmos. Sci., 34, 3-21.

Somerville, R.C.J., 1979: Predictability and prediction of ultra-long planetary waves. Preprint Volume of AMS Fourth Conference on Numerical Weather Prediction, 182-185.

Williamson, D.L., 1976: Normal mode initialization procedure applied to forecasts with the global shallow water equations. Mon. Wea. Rev., 104, 195-206.

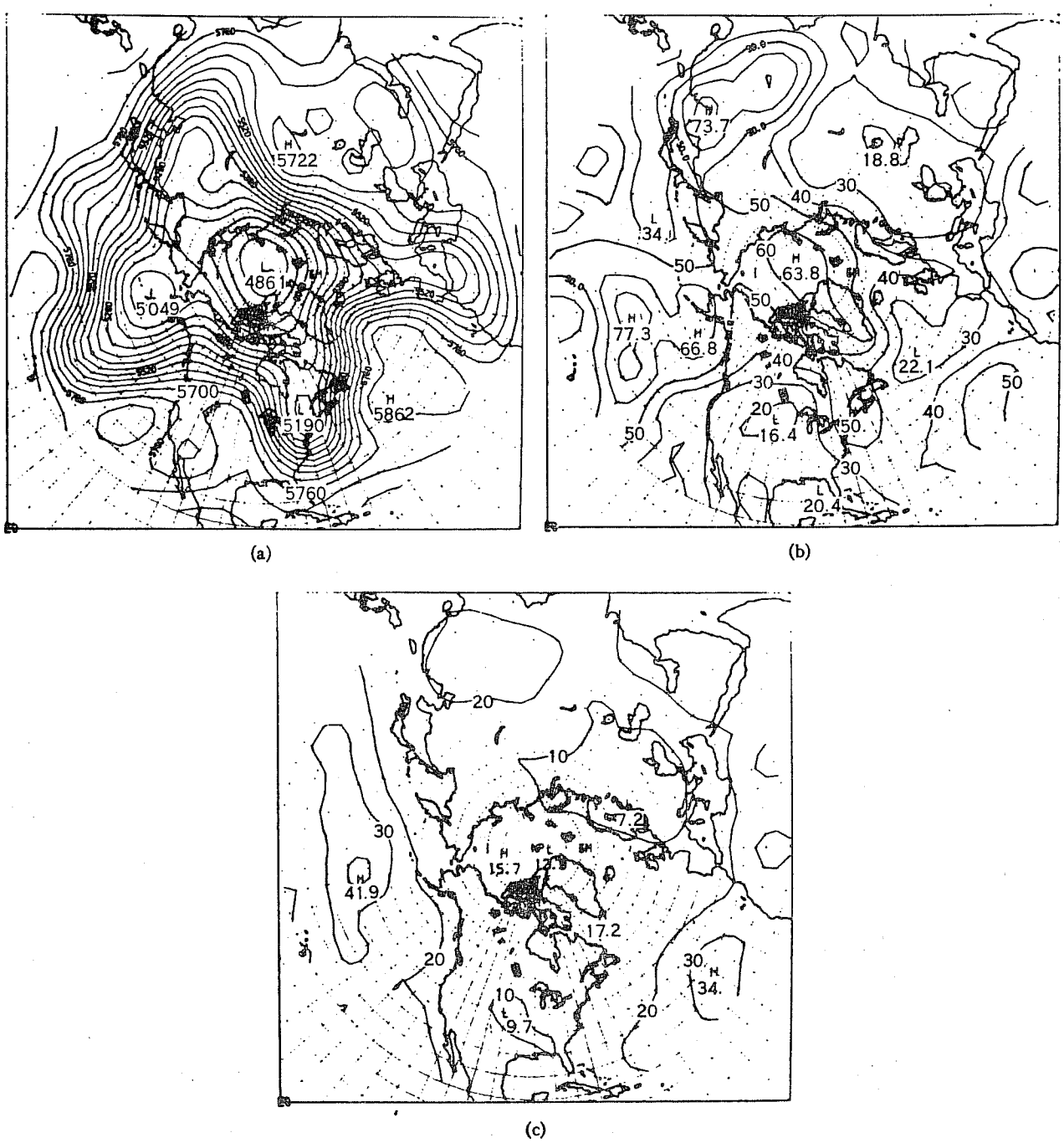

Fig. 1. Height (a) and standard deviation (b), both in meters, of the 500 mb surface from a one-day stochastic dynamic forecast. The standard deviation as derived from a subsequent stochastic analysis is shown in (c) (from Pitcher, 1977).

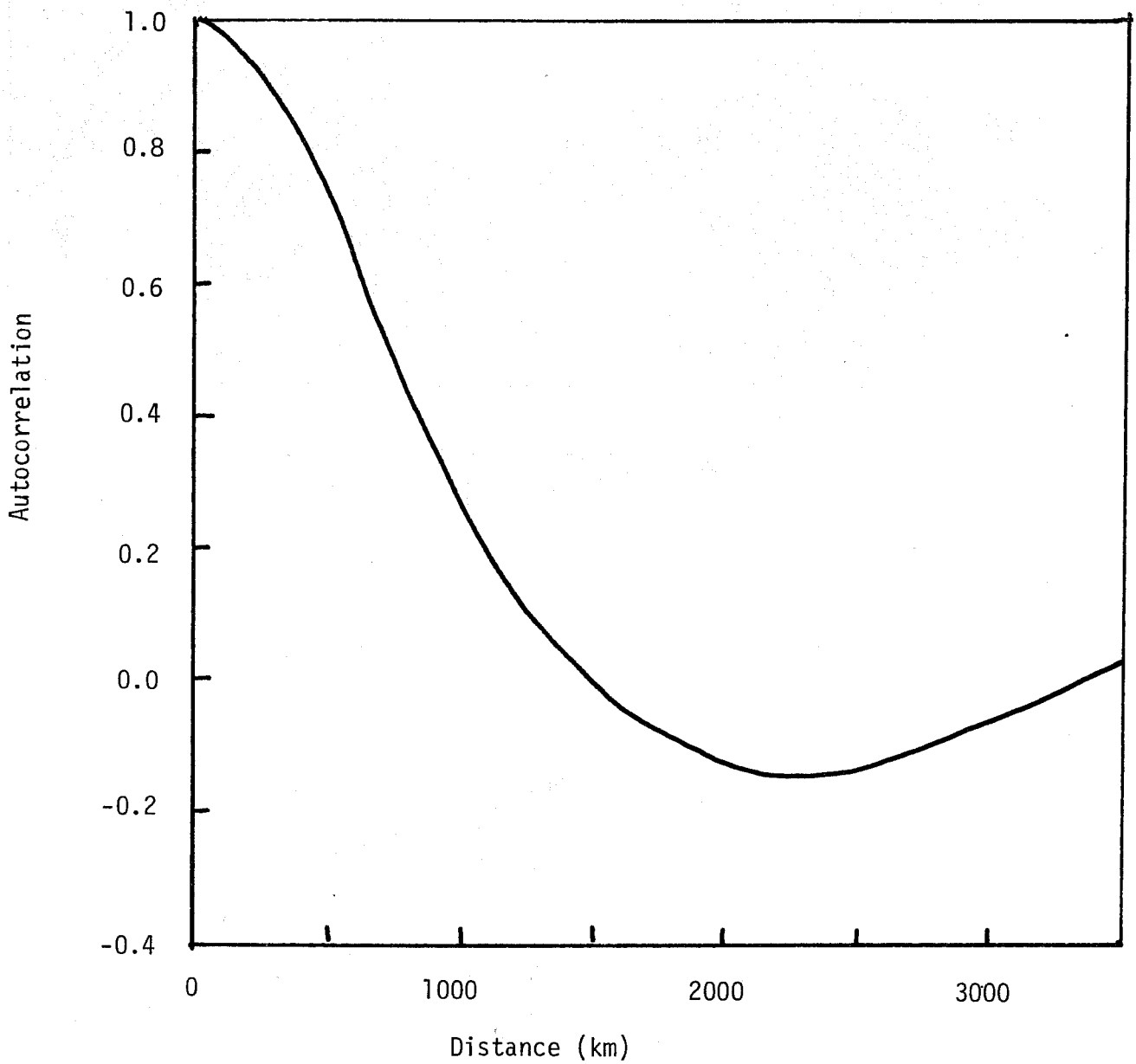

Fig. 2. Model of autocorrelation of analysis error.

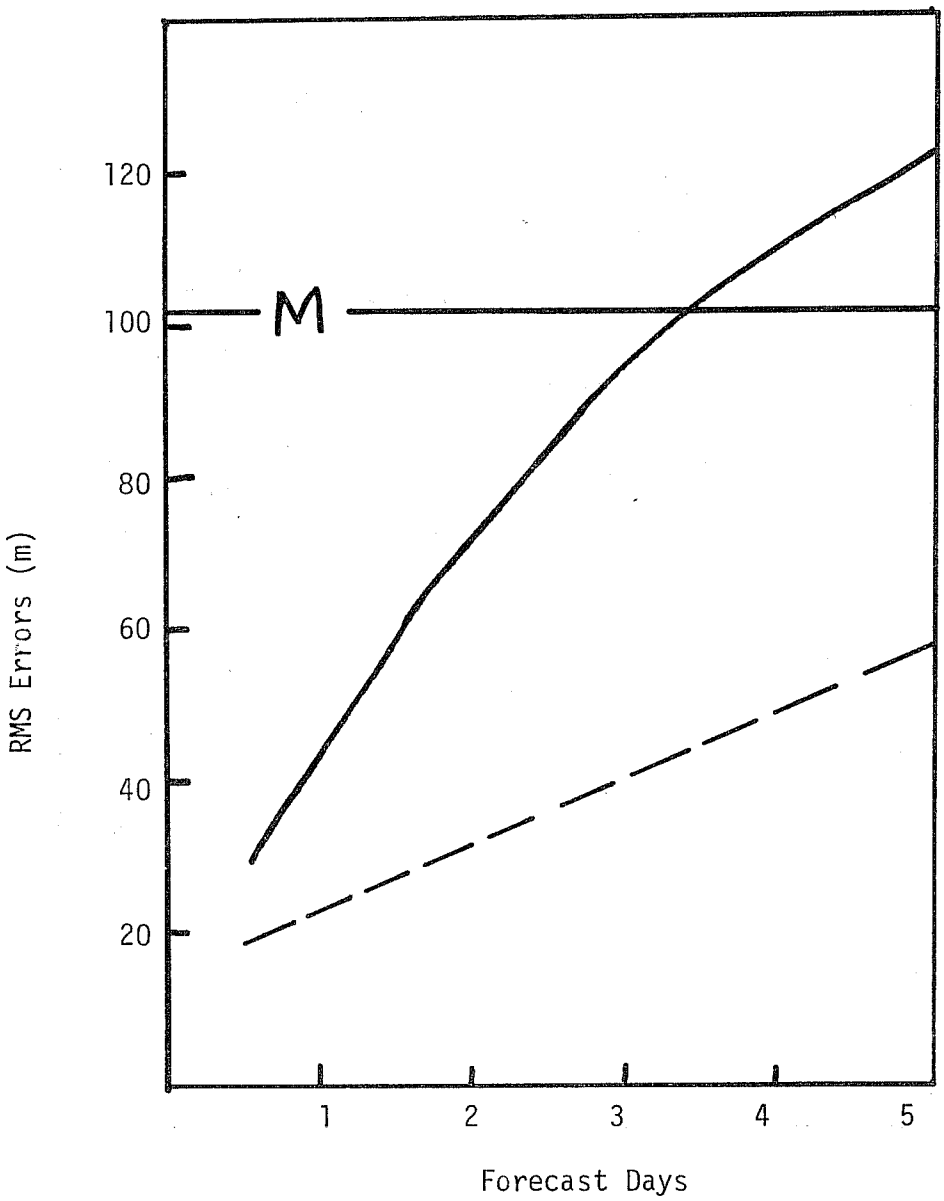

Fig. 3. RMS error of sample mean (solid line), and sample standard deviation (dashed line) for 18 Monte Carlo forecasts. The error of a forecast based on climatology is designated by M.

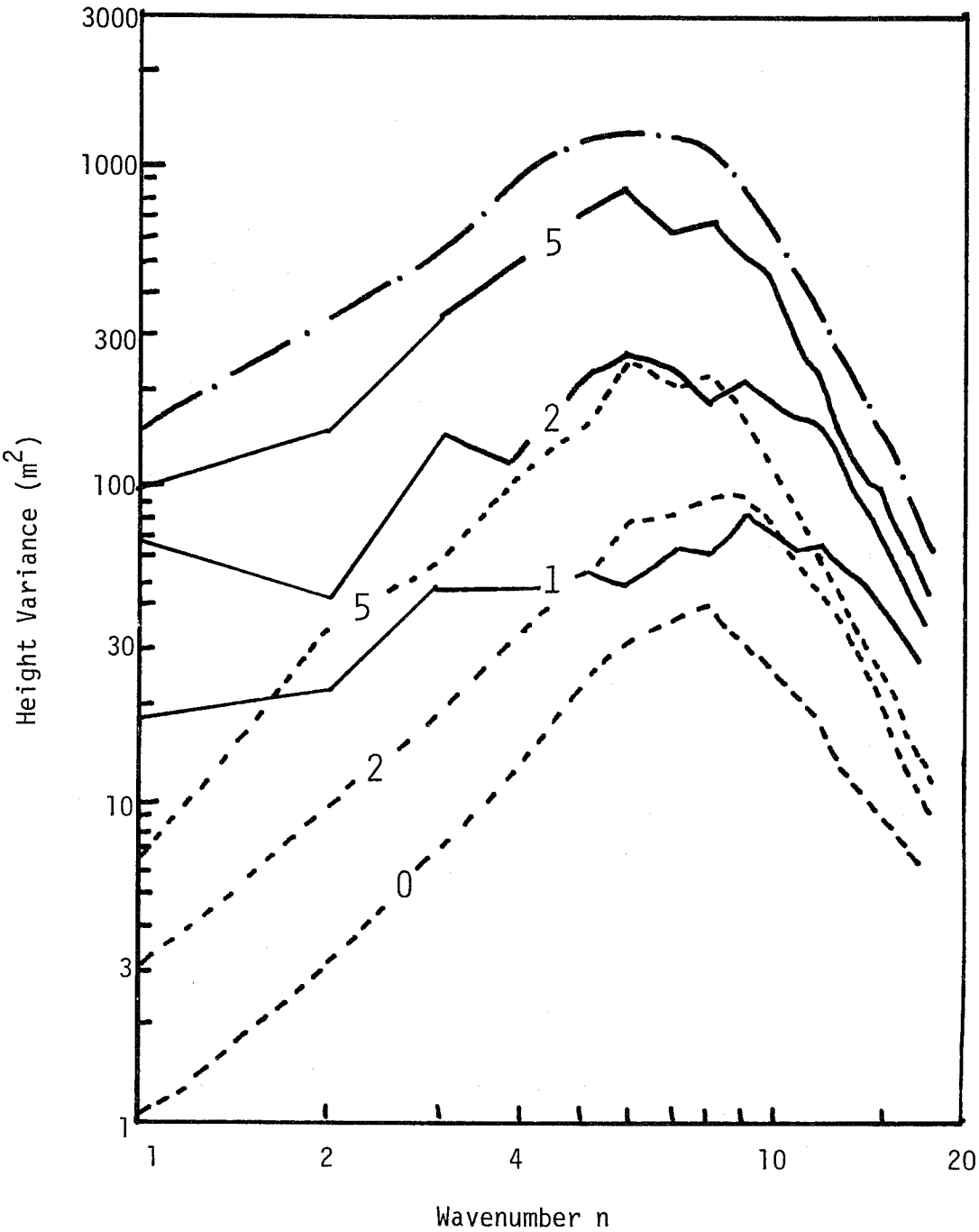

Fig. 4. One half mean-square error of sample mean (solid line), and sample variance (dashed line) for 18 Monte Carlo forecasts. Climatological variance is given by the upper dash-dot line.