S T A T I S T I C A L

AND

STATISTICAL-DYNAMICAL METHODS IN MEDIUM RANGE WEATHER FORECASTS

B Y

C. L E I T H

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

BOULDER, COLORADO, U.S.A.

INTRODUCTION

In this series of lectures we shall attempt to combine statistical and dynamical ideas about the behavior of the larger scales of motion in the atmosphere. The purely statistical methods of a few decades ago did not adequately treat the essentially nonlinear nature of the atmosphere and thus fell out of favor. They were replaced by the purely dynamical methods that culminated in the "deterministic" nonlinear numerical simulations of the atmosphere. But in recent years it has become clear that many aspects of the "deterministic" simulations rest on statistical grounds, that many important questions are statistical in nature, and thus that a full description requires a statistical hydrodynamics of the atmosphere.

There are two important developments in recent years that are contributing to a statistical hydrodynamics. One of these is the development of turbulence approximations for treating the statistical properties of solutions of the Navier-Stokes equations. The other is the recent work of Epstein (1969) and his students on stochastic dynamic prediction which has so far been applied to relatively simple models of the atmosphere. We shall be discussing both of these developments.

Because of the recent favoring of dynamical over statistical methods most of us are more familiar with equations of motion of the atmosphere than we are with the relevant statistical concepts. We shall begin then by reviewing some of the basic ideas of probability theory and statistical mechanics.

1. Probability and Statistics

1.1 Random Variables

We shall be dealing with random variables, random vectors, random functions, random vector fields, etc, and it is important at the outset to define precisely what is meant by the adjective "random." A random real variable u can take on real values at random but according to a precisely defined probability distribution. A simple example of a random variable is the height of Canadian children born during 1970. In this

example and in general we must bear in mind the population or ensemble in terms of which the adjective "random" is defined. Often the precise nature of an ensemble is not well known in practice, but we must at least conceive of the existence of an ensemble whenever we use the word "random."

In many cases in which we are considering random errors in measurement the ensemble is a hypothetical set of outcomes of many repetitions of the measurement process even if we are in reality only permitted to make a single measurement. In this application the ensemble describes the extent of our knowledge and ignorance about the measured quantity.

In the example given we may characterize the distribution of heights u in the population by defining the distribution function $P(u_0)$ as the fraction of the ensemble for which $u \leq u_0$. Obviously $0 \leq P(u) \leq 1$, and $P(u)$ is a nondecreasing function of u. For a large enough population we can often assume that the distribution function is continuous and define a probability density distribution $p(u)$ such that $p(u)du = dP(u)$ for any differential interval du. Clearly $p(u) \geq 0$ and $\int_{-\infty}^{\infty} p(u)\, du = 1$. The idea of probability has been introduced here in an inductive sense as a fraction of a population. Without overly concerning ourselves with philosophical arguments about the meaning of probability we shall say that $p(u)du = dP(u)$ is the probability that the random variable has a value in the interval du about u. This corresponds to the idea that of the individuals in the ensemble, in our example Canadian children born during 1970, each has an equal probability of random selection. Note also that, as in our example, the probability distribution may be time dependent.

Subject only to the integral and non-negative properties described, a probability density distribution can be quite general. A common example to which we shall be making frequent reference is the normal or Gaussian distribution. The normal distribution

$$p(u) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}(u-\mu)^2/\sigma^2} \tag{1.1}$$

is characterized by two parameters, the mean μ which is real and the the standard deviation σ which is real positive. The normal distribution has many simple analytical properties that make it an attractive choice

as an approximation especially when only two parameters of a real probability distribution are known.

We shall indicate the calculation of averages with respect to a probability distribution by the special angular brackets $< >$. Thus for any real function $f(u)$ of the random variable u we have

$$\langle f \rangle = \int_{-\infty}^{\infty} f(u)p(u)\ du.$$

The calculation of averages is a linear operation so that for any two constants a,b and for any two functions $f(u),g(u)$ we have $\langle af + bg \rangle = a\langle f \rangle + b\langle g \rangle$. Such an average is an ensemble average and, depending on the convergence of the integral, may or may not exist as a finite real number. It is not itself a random variable but rather is a constant. Special examples of such averages are the n-th moments $\langle u^n \rangle$ for $n = 0,1,2,3,\cdots$. Evidently for $n = 0$ we have $\langle 1 \rangle = 1$. For $n = 1$ we have the first moment $\langle u \rangle$ called the mean which for the normal distribution is $\langle u \rangle = \mu$. It is often convenient to translate the probability distribution by defining a new random variable $u - \langle u \rangle$ with vanishing mean $\langle u \rangle - \langle\langle u \rangle\rangle = 0$ and to speak of higher moments about the mean or central moments defined as $\langle (u - \langle u \rangle)^n \rangle$ for $n = 2,3,\cdots$. For $n = 2$ the second moment about the mean is called the variance which for the normal distribution is $\langle (u - \langle u \rangle)^2 \rangle = \sigma^2$. It is clear that the normal distribution is characterized in a simple way by its mean and its variance.

Of particular interest are the properties of the unit normal probability distribution with $\mu = 0$, $\sigma = 1$. For its normalization we note that

$$\int e^{-\frac{1}{2}x^2}\ dx \int e^{-\frac{1}{2}y^2}\ dy = \iint e^{-\frac{1}{2}r^2}\ dxdy = 2\pi \int_0^{\infty} e^{-\frac{1}{2}r^2}\ rdr = 2\pi \int_0^{\infty} e^{-z}\ dz = 2\pi \tag{1.3}$$

thus that

$$\frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}x^2}\ dx = 1 \tag{1.4}$$

We designate a unit normal random variable by a thus

$$p(a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2} \quad . \tag{1.5}$$

To compute the moments of a consider the definite integral

$$f(h) = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}a^2 h} \, da = h^{-\frac{1}{2}} \quad . \tag{1.6}$$

Multiple differentiation with respect to h gives us

$$f^{(n)}(h) = \frac{1}{\sqrt{2\pi}} \int (-\tfrac{1}{2}a^2)^n e^{-\frac{1}{2}a^2 h} \, da = (-1/2)(-3/2)\cdots(-(2n-1)/2)h^{-(2n+1)/2} \tag{1.7}$$

which we evaluate at h = 1 to see that

$$\langle a^{2n} \rangle = \frac{1}{\sqrt{2\pi}} \int a^{2n} e^{-\frac{1}{2}a^2} \, da = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1) \quad . \tag{1.8}$$

Odd moments vanish owing to the antisymmetric nature of the defining integrands.

The unit normal random variable a with

$$dP = p(a)da = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2} \, da \tag{1.9}$$

may be used to generate by an inhomogeneous linear transformation any normal random variable

$$u = \mu + \sigma a \tag{1.10}$$

since

$$a = (u-\mu)/\sigma \qquad da = du/\sigma$$

$$dP = p(u) \, du = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}(u-\mu)^2/\sigma^2} \, du \quad . \tag{1.11}$$

This is a convenient representation for computing moments of a normal random variable:

$$\langle u \rangle = \mu + \sigma\langle a \rangle = \mu$$

$$\langle u^2 \rangle = \mu^2 + 2\mu\sigma\langle a \rangle + \sigma^2\langle a^2 \rangle = \mu^2 + \sigma^2$$

$$\langle u^3 \rangle = \mu^3 + 3\mu^2\sigma\langle a \rangle + 3\mu\sigma^2\langle a^2 \rangle + \langle a^3 \rangle = \mu^3 + 3\mu\sigma^2$$

$$\langle u^4 \rangle = \mu^4 + 4\mu^3\sigma\langle a \rangle + 6\mu^2\sigma^2\langle a^2 \rangle + 4\mu\sigma^3\langle a^3 \rangle + \sigma^4\langle a^4 \rangle$$

$$= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \qquad (1.12)$$

In considering the random variable u that is generated by a transformation of the random variable a the question arises as to whether averages are to be computed with respect to the probability distribution of a or of u. In fact, the probability distributions are related in such a way that either may be used. This may be seen most easily by recognizing that an ensemble average can also be computed as

$$\langle f \rangle = \mathcal{N}^{-1} \sum_{m=1}^{\mathcal{N}} f^{(m)} \qquad (1.13)$$

for an ensemble of $\mathcal{N}$ members each with an attached value of $a^{(m)}$, $u^{(m)}$, or $f^{(m)}$. Here we have used the fundamental concept that the members of an ensemble are equally probable.

For a normal random variable all moments can be computed in terms of the first two. For random variables in general this is not so. For $n = 3,4,5,$ the difference between the actual nth moment and the value computed from an assumption of normality is called the cumulant. Thus for a normal distribution all cumulants vanish; the cumulants serve as a measure of the difference of a probability distribution from being normal.

1.2 Computer Generated Normal Random Variable

There is usually a routine available for generating random numbers x such that $p(x) = 1$ for $0 \leq x \leq 1$, $p(x) = 0$ otherwise. A good approximation for a unit normal random variable is

$$a = -6 + \sum_{i=1}^{12} x_i \tag{1.14}$$

i.e., the sum of -6 and 12 independent random numbers x. Use is being made here of the tendency for sums of independent random variables to approach a normal distribution when the number of variables becomes large.

For a general normal random variable we may, of course, set

$$u = \mu + \sigma a \quad . \tag{1.15}$$

1.3 Random Vectors

In two dimensions the random vector (u_1,u_2) will be characterized by the probability distribution $p(u_1,u_2) \geq 0$ such that

$$\iint p(u_1,u_2)du_1du_2 = 1 \quad . \tag{1.16}$$

We now have an ensemble distributed in a two-dimensional phase space Ω of vectors (u_1,u_2).

Note that although the vector components u_1 and u_2 are each random variables with probability distributions

$$\begin{aligned} p_1(u_1) &= \int p(u_1,u_2)du_2 \\ p_2(u_2) &= \int p(u_1,u_2)du_1 \end{aligned} \tag{1.17}$$

there is, in general, more information contained in $p(u_1,u_2)$ than is available from a knowledge of $p_1(u_1)$ and $p_2(u_2)$ alone.

The average of a function of phase $f(u_1,u_2)$ is defined by

$$\langle f \rangle = \iint f(u_1,u_2)p(u_1,u_2)du_1du_2, \tag{1.18}$$

and various moments of order $\alpha + \beta$ are given by $\langle u_1^{\alpha} u_2^{\beta} \rangle$.

In an arbitrary number N of dimensions we have an N-dimensional phase space of random vectors $(u_1,u_2,\cdots,u_N)$ with the average of a phase function defined by

$$f = \int\int\cdots\int f(u_1,u_2,\cdots,u_N)p(u_1,u_2,\cdots,u_N)du_1du_2\cdots du_N . \qquad (1.19)$$

The first moment $\langle u_i\rangle$ is a nonrandom vector. The second moment $M_{ij} = \langle u_iu_j\rangle$ is a nonrandom second rank tensor. Considered as a matrix it is obviously symmetric and also in fact non-negative definite since for any arbitrary constant vector x_i we have

$$M_{ij}x_ix_j = \langle u_iu_j\rangle x_ix_j = \langle(u_ix_i)(u_jx_j)\rangle = \langle(u_ix_i)^2\rangle \geq 0 . \qquad (1.20)$$

The matrix M has non-negative real eigenvalues. The matrix M is called the covariance matrix when the random vector u_i is such that $\langle u_i\rangle = 0$. To avoid pathologically singular cases we shall assume in general that covariance matrices are positive definite, nonsingular, and without zero eigenvalues.

Normal random vectors have multivariate normal probability distributions. We consider first the generalization to N dimensions of the unit normal random variable. This will be a random vector a_i with $\langle a_i\rangle = 0$, and with $M_{ij} = \langle a_ia_j\rangle = \delta_{ij}$, that is, the covariance matrix M is the identity matrix I. The probability distribution of the random vector $\{a_i\}$ is

$$p(a_1,a_2,\cdots,a_N) = \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}(a_1^2+a_2^2+\cdots+a_N^2)} \qquad (1.21)$$

As in one dimension we may generate any normal random vector by an inhomogeneous linear transformation

$$u_i = R_{ij}a_j + \mu_i \qquad (1.22)$$

with the matrix R being nonsingular, i.e. such that its inverse R^{-1} exists.

The probability distribution for u_i may be obtained from that for a_i. Let $S = R^{-1}$ so that

$$a_i = S_{ij}(u_j - \mu_j) \tag{1.23}$$

then

$$p(u_1,u_2,\cdots,u_N)du_1du_2\cdots du_N =$$

$$\frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}S_{ij}(u_j-\mu_j)S_{ik}(u_k-\mu_k)} \frac{\partial(a_1,a_2,\cdots,a_N)}{\partial(u_1,u_2,\cdots,u_N)} du_1du_2\cdots du_N \tag{1.24}$$

Thus we have

$$p(u_1,u_2,\cdots u_N) = \frac{1}{(2\pi)^{N/2}} |T|^{\frac{1}{2}} e^{-\frac{1}{2}T_{jk}(u_j-\mu_j)(u_k-\mu_k)} \tag{1.25}$$

where $|T|$ is the determinant of the matrix $T = S^*S$ where S^* is the (complex conjugate) transpose of S.

For this distribution the first two moments are

$$\langle u_i \rangle = R_{ij}\langle a_j \rangle + \mu_i = \mu_i$$

$$\langle u_iu_j \rangle = \langle (R_{ik}a_k + \mu_i)(R_{j\ell}a_\ell + \mu_j) \rangle = R_{ik}R_{j\ell}\langle a_ka_\ell \rangle + \mu_i\mu_j$$

$$= R_{ik}R_{jk} + \mu_i\mu_j = M_{ij} + \mu_i\mu_j \tag{1.26}$$

where the covariance matrix $M = RR^* = T^{-1}$.

To generate a normal random vector with given mean μ_i and covariance matrix M_{ij} we first compute the matrix R_{ij} such that $R^2 = M$. A positive definite matrix M has a unique positive definite square root R which we may compute in the following way.

We use throughout the coordinate system of the u_i's in which the matrix elements of M are M_{ij}. We can find for M a complete set of

orthonormal eigenvectors $e_i^{(k)}$, k=1,2,•••,N, with associated real eigenvalues $m^{(k)} > 0$. These are sometimes called the empirical orthogonal vectors of M. The decomposition of the identity relative to M is

$$I = \sum_{k=1}^{N} P^{(k)} \tag{1.27}$$

where the matrices $P^{(k)}$ are projections onto the eigendirections $e_i^{(k)}$ with matrix elements $P_{ij}^{(k)} = e_i^{(k)} e_j^{(k)}$. The spectral decomposition of M is

$$M = \sum_{k=1}^{N} m^{(k)} P^{(k)} \tag{1.28}$$

or in terms of matrix elements

$$M_{ij} = \sum_{k=1}^{N} m^{(k)} e_i^{(k)} e_j^{(k)} \quad . \tag{1.29}$$

By the spectral mapping theorem for any function f we have

$$f(M) = \sum_{k=1}^{N} f(m^{(k)}) P^{(k)} \tag{1.30}$$

thus, in particular,

$$R = M^{\frac{1}{2}} = \sum_{k=1}^{N} (m^{(k)})^{\frac{1}{2}} P^{(k)} \tag{1.31}$$

where we take the unique positive square root of each $m^{(k)} > 0$. The matrix elements of R are given by

$$R_{ij} = \sum_{k=1}^{N} (m^{(k)})^{\frac{1}{2}} e_i^{(k)} e_j^{(k)} \quad . \tag{1.32}$$

We may generate the random vector a_i by generating independent unit normal random variables for each component. The desired general normal random vector will be

$$u_i = R_{ij} a_j + \mu_i$$

$$= \sum_{k=1}^{N} (m^{(k)})^{\frac{1}{2}} (e_j^{(k)} a_j) e_i^{(k)} + \mu_i \quad . \tag{1.33}$$

The matrix M is real and thus symmetric and its eigenvectors have real components since they arise from the solutions of the real linear systems

$$(M - m^{(k)} I)\, e^{(k)} = 0 \quad . \tag{1.34}$$

Consequently $R = M^{\frac{1}{2}}$ will be real and the random vector u_i will be real. In fact any real function of M will be a real matrix.

The concept of a random function is an infinite-dimensional generalization of that of a random vector. For our purposes, however, all meteorological fields are, in practice, represented by a large but finite number of values, and we shall treat these values as components of a multi-dimensional vector in a meteorological phase space. In order to speak of such a vector as random we must consider an ensemble and a multivariate probability distribution to be defined in the phase space.

2. Dynamical Phase Space

2.1 Definition

The meteorological system with which we are dealing is, of course, evolving rapidly in time. Thus we are, in general, interested in an N-dimensional phase space Ω of vectors $(u_1,u_2,\cdots,u_N)$ in which there is defined at each point a fixed velocity vector

$$\dot{u}_\alpha = Q_\alpha(u_1,u_2,\cdots,u_N) \qquad \alpha = 1,2,\cdots,N \tag{2.1}$$

given by some dynamics equation. At a particular instant in time t we imagine an ensemble of dynamical systems each represented by a point in a cloud of points distributed through phase space according to the probability distribution

$$p(u_1,u_2,\cdots,u_N,t)$$

As each system evolves according to the dynamics equation, the corresponding phase point moves, the cloud as a whole will move, and its probability density distribution will change.

We deduce an equation for the change in the probability density distribution based on the conservation of phase points in the phase flow. At any point $\underset{\sim}{u}$ in Ω the flux of phase points will be $\dot{u}_\alpha p(u_1,u_2,\cdots,u_N,t)$, and the accumulation per unit N-volume will be

$$\begin{aligned}\partial p(u_1,u_2,\cdots,u_N,t)/\partial t &= -\sum_\alpha \partial[\dot{u}_\alpha(u_1,u_2,\cdots,u_N)p(u_1,u_2,\cdots,u_N,t)]/\partial u_\alpha \\ &= -\sum_\alpha \partial[Q_\alpha(u_1,u_2,\cdots,u_N)p(u_1,u_2,\cdots,u_N,t)]/\partial u_\alpha \,.\end{aligned} \tag{2.2}$$

This is a linear partial differential equation for the evolution of an ensemble probability density distribution as it is induced by specified dynamics. Its solution would answer all single-time statistical questions at any later time. The practical difficulty is that for N large it is difficult to compute such solutions exactly, and we must find some method of approximation.

In many cases the dynamics is such that

$$\sum_{\alpha} \partial \dot{u}_{\alpha} / \partial u_{\alpha} = 0 \tag{2.3}$$

that is, that the phase flow is nondivergent in its N-dimensional space. In these cases the probability evolution equation reduces to the Liouville equation

$$\partial p(u_1,u_2,\cdots,u_N,t)/\partial t = -\sum_{\alpha} \dot{u}_{\alpha}(u_1,u_2,\cdots,u_N)\, \partial p(u_1,u_2,\cdots,u_N,t)/\partial u_{\alpha} \tag{2.4}$$

Since in this case volume elements and phase points are conserved in the flow, the probability density remains constant along a phase path.

2.2 Time Dependence of Averages

Consider the phase function $f(\underset{\sim}{u},t)$ that may be a function of time as well as of position $\underset{\sim}{u}$ in the phase space Ω. Its average $\langle f \rangle$ is no longer a function of $\underset{\sim}{u}$, but is still a function of time explicitly through the time dependence of $f(\underset{\sim}{u},t)$ as well as implicitly through that of $p(\underset{\sim}{u},t)$. Let us compute the time derivative of $\langle f \rangle$.

$$d\langle f \rangle/dt = \int \partial\{f(\underset{\sim}{u},t)p(\underset{\sim}{u},t)\}/\partial t \, d\underset{\sim}{u} \tag{2.5}$$

The time derivative inside the integral is a partial derivative since the integrand unlike the integral is a function of $\underset{\sim}{u}$ as well as of t.

$$d\langle f \rangle/dt = \int p(\underset{\sim}{u},t)\partial f(\underset{\sim}{u},t)/\partial t \, d\underset{\sim}{u} + \int f(\underset{\sim}{u},t)\partial p(\underset{\sim}{u},t)/\partial t \, d\underset{\sim}{u} \, . \tag{2.6}$$

For the second integral we use the equation of conservation of probability to find

$$\begin{aligned} \int f(\underset{\sim}{u},t)\partial p(\underset{\sim}{u},t)/\partial t \, d\underset{\sim}{u} &= -\int f(\underset{\sim}{u},t) \sum_{\alpha} \partial\{\dot{u}_{\alpha}(\underset{\sim}{u})p(u,t)\}/\partial u_{\alpha} \, d\underset{\sim}{u} \\ &= -\int \sum_{\alpha} \partial\{f(\underset{\sim}{u},t)\dot{u}_{\alpha}(\underset{\sim}{u})p(\underset{\sim}{u},t)\}/\partial u_{\alpha} \, d\underset{\sim}{u} \\ &\quad + \int \sum_{\alpha} \dot{u}_{\alpha}(\underset{\sim}{u})\{\partial f(\underset{\sim}{u},t)/\partial u_{\alpha}\}p(\underset{\sim}{u},t) \, d\underset{\sim}{u} \, . \end{aligned} \tag{2.7}$$

The first of these two integrals vanishes as the integral of a divergence since $p \to 0$ as $u \to \infty$. Thus

$$d\langle f\rangle/dt = \int \{\partial f(\underset{\sim}{u},t)/\partial t + \sum_\alpha \dot{u}_\alpha(\underset{\sim}{u})\,\partial f(\underset{\sim}{u},t)/\partial u_\alpha\} p(\underset{\sim}{u},t)\, d\underset{\sim}{u}$$

$$= \langle \partial f(\underset{\sim}{u},t)/\partial t + \sum_\alpha \dot{u}_\alpha(\underset{\sim}{u})\,\partial f(\underset{\sim}{u},t)/\partial u_\alpha\rangle$$

$$= \langle \dot{f}(\underset{\sim}{u},t)\rangle \tag{2.8}$$

where $\dot{f}(\underset{\sim}{u},t)$ is the Lagrangian time derivative of $f(\underset{\sim}{u},t)$ following the phase flow in phase space. This notation is consistent with the use of $\dot{u}_\alpha$ as a velocity component in space since $\dot{u}_\alpha$ gives the time rate of change of the coordinate u_α with the phase flow.

$$du_\alpha/dt = \partial u_\alpha/\partial t + \sum_\beta \dot{u}_\beta \partial u_\alpha/\partial u_\beta$$

$$= 0 + \sum_\beta \dot{u}_\beta \delta_{\alpha\beta} = \dot{u}_\alpha \tag{2.9}$$

An alternate, more direct, derivation of the result $d\langle f\rangle/dt = \langle \dot{f}\rangle$ is based on the consideration of a myriad M phase points of the ensemble with $f = f_m$ being the value of f on point m. Then

$$\langle f\rangle = \frac{1}{M}\sum_{m=1}^{M} f_m \tag{2.10}$$

and

$$d\langle f\rangle/dt = \frac{1}{M}\sum_{m=1}^{M} df_m/dt = \langle \dot{f}\rangle \tag{2.11}$$

where df_m/dt is the rate of change of f attached to and moving with the point m. The evolving probability distribution in this picture is implicit in the evolving distribution of phase points.

2.3 Forecast Ensemble

Our principle concern in these lectures is with the optimal estimation of the present and future states of the atmosphere. The classical theory of optimal estimation (Gauss, 1809) is most succinctly

described in terms of a state or phase space for the system being considered. A discussion of the use of a phase space for atmospheric dynamics can be found in a paper by Gleeson (1970). The system of interest here is the whole atmosphere, but we always treat it, in fact, in some finite approximation in numerical models, and we shall consider therefore the large- but finite-dimensional phase space of a numerical model. The state of the model atmosphere at a particular time t is specified as a point u(t) in its phase space; the coordinates of the point are the values at that time of all prognostic variables in the model. Depending on the representation used, the phase space components may be made up of the values of fields of physical variables such as velocity, temperature, and humidity defined over an array of space mesh points or of the amplitudes of an orthogonal function expansion of these physical fields. The dimensionality of the model phase space is of the order of 250,000 in some present general circulation models.

The calculational cycle of a numerical model advances the state u(t) for a small time increment Δt to the new state $u(t+\Delta t)$, and the repetition of these cycles leads to a sequence of phase points which trace out a phase path. This process is completely deterministic; starting at the present from a given point the model generates a definite phase path into the future. If the initial state were known exactly and if the model exactly simulated the atmosphere, then the computed phase path would provide an exact forecast. But neither assumption is valid. We shall be interested first in these lectures in the internal error arising from the uncertainty in the determination of the initial state. We shall consider later the additional external error generated by the discrepancy between the dynamics of the model and that of the real atmosphere arising in part from the limited dimensionality of the model phase space. For studying internal error growth alone under the assumption that our model is perfect we assume that there exists a true phase path for the model that is not known but that we wish to estimate as closely as possible. Let us consider first the situation at an initial time.

Were we able to make a set of independent observations of a particular unknown true state u_0, we know that the errors associated

with our observational technique would lead in fact to a set of observed vectors u distributed about u_0. Since an observational technique in this discussion is intended to include an analysis method, the observational errors include interpolation and aliasing errors as well as instrumental errors. As the number of observations in the set approached infinity, the observed phase points would form a cloud of equally likely observations u about u_0. The density distribution of this cloud would give a probability distribution $p(u-u_0)$ for observational error. We shall assume that the observations have been corrected for bias and thus that the mean of the observations u is u_0. Although the error probability distribution $p(u-u_0)$ might be quite complicated in structure we are generally restricted to specifying it in terms of its first and second moments, in which case we cannot distinguish it from a Gaussian or normal distribution.

The practical situation is quite different from the hypothetical situation just described. Were we able, in fact, to make a large number n of independent observations u of a single true state u_0, the average of the n observations would represent a combined observation whose root-mean-square error would be smaller than that of individual observations by a factor $n^{-\frac{1}{2}}$. Instead we have, at a given time, a single observation which should, of course, include all available sources of information. The error distribution of this observation must be arrived at indirectly by a combination of theoretical analysis of the observation and analysis method and of experience acquired from observations of a long sequence of different true states. We shall assume that this has been done, that we know the observational technique to be free of bias, and that we know the second moments characterizing the error.

We are faced therefore with a situation in which we have been able to make a single imperfect observation u of the unknown true state u_0. We can only conclude that the true state u_0 is likely to be in the neighborhood of the observed state u with a probability distribution $p_0(u_0-u)$ obviously related to that of the observational error by $p_0(x) = p(-x)$. We may then reverse the earlier picture and consider a cloud of possible true states u_0 distributed about the observed

state u. If, further, the cloud is distributed with a density given by the probability distribution $p_0(u_0-u)$, then the points represent equally likely candidates to be the true state u_0. We cannot know the true state precisely, and the cloud represents the extent of our knowledge and ignorance given the observation. The mean position of the cloud is the observation u and represents the best estimate of the true state in the least-square-error sense. The second central moment is the covariance matrix for the ensemble of phase points and is the simplest representation of the size of the cloud and the degree of our uncertainty. The diagonal elements of the covariance matrix are variances which provide the usual mean-square measure of error or uncertainty in single components, but the off-diagonal elements are needed in order to specify correlations between the errors of pairs of components.

If we imagine each point in the initial cloud to be advanced along its phase path by the dynamics of the numerical model we see that the cloud of phase points will move through phase space. We know from predictability studies, which examine the divergence of pairs of paths in phase space, that in general the cloud will be distorted and its size as measured by its covariance matrix will increase with time. We expect further under an ergodic hypothesis that the cloud will tend, at times so late that all predictability has been lost, toward a stationary equilibrium distribution which represents the climate ensemble for the model. We shall refer, in general to the evolving cloud of phase points as the forecast ensemble in order to distinguish it from the climate ensemble toward which it tends.

The climate is, of course, more traditionally defined in terms of averages over phase points observed at regular time intervals during a hypothetically infinite evolution time. It is the ergodic hypothesis that we may equate such time averages to averages computed for a stationary climate ensemble. An important goal in the development of a numerical model to simulate the atmosphere is that the model climate agrees with that of the real atmosphere. This goal has been nearly reached for many present general circulation models in the sense that the time average statistical properties of the model agree with the

corresponding properties of the real atmosphere even though there may be differences in detailed evolution. For the purposes of this discussion we shall assume that we are dealing with such a climatologically realistic model with a known stationary climate ensemble.

So far as the internal dynamics of a model is concerned the evolving cloud of phase points at any later time continues to represent an ensemble of equally likely candidates to be the true state, and therefore the probability distribution of the forecast ensemble continues to represent the extent of our knowledge and ignorance. Specifically, the mean of the forecast ensemble continues to be the estimate of the true state of the atmosphere that is best in the least-square-error sense, and the covariance of the forecast ensemble serves as the simplest measure of uncertainty or error in that estimate.

Although the partial differential equations describing the evolution of the probability distribution in phase space are well known, the dimensionality of the phase space is so great that the direct integration of these equations is out of the question. It is necessary to make some approximation, and we may characterize a specific forecasting scheme in terms of the corresponding approximation. The stochastic dynamic forecasting scheme (Epstein, 1969) to be discussed in the next section is based on a moment expansion technique that gives equations for the evolution of the mean and covariance directly, and the approximation involved is a moment closure of which several have been tried. The Monte Carlo forecasting scheme to be discussed in Section 5 is based on a sampling approximation in which the forecast ensemble mean and covariance are estimated from a finite sample of phase points determined by computing m trajectories in phase space. A conventional single forecast is obtained by computing the trajectory that starts from the mean of the initial forecast ensemble and can be considered as an approximation to the equation for evolution of the mean that is closed by ignoring the second moment terms. Of course, a conventional forecast might also be considered as a special case of a Monte Carlo forecast with sample size m = 1.

A possibility exists that the true state may in a few particular cases lie very close to the mean of the initial forecast ensemble. Should they coincide, the conventional forecast would provide the true phase path and would therefore remain superior to any other forecasting scheme. In particular, stochastic dynamic or Monte Carlo forecasts of the mean of the forecast ensemble cannot attain such chance successes, but, as we shall see, they also avoid more frequent serious failures and provide a net improvement in average forecast skill. The value of those conventional forecasts which do turn out to be superior is limited somewhat by the fact that we can have no greater *a priori* confidence in them. In order to know that they were going to be more accurate, we should have had to know that their initial states were more accurate, but all such knowledge has already been used to determine the second moments of the initial ensemble.

In evaluating the average skill of a particular forecasting technique we shall determine the mean-square discrepancy between the true state and the forecast estimate. The averaging process involved here is over an ensemble of initial states which we naturally take to be the climate ensemble. Thus we are in this case interested in the average way in which the clouds of uncertainty grow for many repeated uses of the forecasting scheme. This kind of average growth of uncertainty for conventional single forecasts was treated by Leith and Kraichnan (1972) for a two-dimensional turbulence model. It is important to bear in mind the distinction between this climate average growth of error and the error growth associated with a particular forecast ensemble.

3. Stochastic-Dynamic Forecasts

3.1 Quadratically Nonlinear System

In dealing with various forecasting methods it is most useful to consider an abstract quadratically nonlinear forecast equation of the form encountered in spectral barotropic models. This is also the form encountered in spectral models of the Navier-Stokes equation and has been the subject of much study for two- and three-dimensional turbulence models. It is the quadratically nonlinear term in the forecast equation that causes the common difficulty in all of these systems.

An important class of turbulence approximations is based on the construction of stochastic models. We shall examine the essential structure of such models free of irrelevant kinematic complexity by considering a simple analog of the wavevector-space form of the Navier-Stokes equation in two dimensions. The analog will have two simple integrals, energy and enstrophy, but one may easily drop the enstrophy integral to have an analog for three-dimensional turbulence.

Our simple mechanical system will be characterized by a large but finite set of N real variables $u_\alpha(t)$ evolving according to

$$(d/dt+\nu_\alpha)u_\alpha(t) = \sum_{\beta,\gamma} A_{\alpha\beta\gamma}u_\beta(t)u_\gamma(t) + f_\alpha(t) \quad . \qquad (3.1)$$

The real constants ν_α simulate viscous damping when positive and prescribed mode instabilities when negative. The functions $f_\alpha(t)$ provide for specified external forcing. The nonlinear nature of the system is characterized by the interaction coefficients $A_{\alpha\beta\gamma} = A_{\alpha\gamma\beta}$. We shall assume that the mode indices are elements of an addition group such that $\alpha \dot{+} \alpha = 0$. We impose four conditions on the interaction coefficients:

I. $A_{\alpha\beta\gamma} = 0$, if $\alpha \dot{+} \beta \dot{+} \gamma \neq 0$

II. $A_{\alpha\beta\gamma} = 0$, if $\alpha = 0$ or $\beta = 0$ or $\gamma = 0$.

III. $A_{\alpha\beta\gamma} + A_{\beta\gamma\alpha} + A_{\gamma\alpha\beta} = 0$.

IV. There exists a set of real numbers m_α, not all equal, such that $m_\alpha A_{\alpha\beta\gamma} + m_\beta A_{\beta\gamma\alpha} + m_\gamma A_{\gamma\alpha\beta} = 0$.

The mode $\alpha = 0$ is special, and we remove it from the system by setting $f_0(t) = 0$ and $u_0(t) = 0$ as we may consistently according to Condition II. Conditions I and II together guarantee that there is no self-interaction, i.e., that $A_{\alpha\beta\gamma} = 0$ unless α, β and γ all differ. All of these conditions correspond to known properties of the two-dimensional Navier-Stokes equation.

3.2 Statistical Properties

We are not so much interested in individual solutions of Eq. (3.1) as we are in the statistical properties of ensembles of solutions. Thus, we consider the $u_\alpha(t)$ as random variables whose randomness is induced either by the initial conditions $u_\alpha(0)$ or by the random forcing $f_\alpha(t)$, or both. In the absence of the nonlinear interactions (3.1) reduces to a linear random equation whose statistical properties are well understood. To study the influence of the nonlinear term alone we consider the inviscid, force-free equation

$$du_\alpha(t)/dt = \dot{u}_\alpha(t) = \sum_{\beta,\gamma} A_{\alpha\beta\gamma} u_\beta(t) u_\gamma(t) \quad , \tag{3.2}$$

in which the randomness of $u_\alpha(t)$ arises only from that of the initial conditions $u_\alpha(0)$.

At any time t, including $t = 0$, the statistical properties of the ensemble are completely characterized by a probability distribution $P(u,t)$ defined on the points u of the N-dimensional phase space Ω of the system. In particular, any single-time moment is given by

$$\langle u_\alpha(t)\, u_\beta(t) \cdots u_\delta(t) \rangle = \int_\Omega u_\alpha u_\beta \cdots u_\delta P(u,t)\, du \ .$$

We shall assume that $P(u,t)$ is such that any moment $\langle u_\alpha(t) u_\beta(t) \cdots u_\delta(t) \rangle$ vanishes if $\alpha \dot{+} \beta \dot{+} \cdots \dot{+} \delta \neq 0$. By analogy to turbulence statistics we call this property (preserved in time by condition I on $A_{\alpha\beta\gamma}$) statistical homogeneity. In particular, we then have $\langle u_\alpha(t) \rangle = 0$, but $U_\alpha(t) = \langle u_\alpha(t) u_\alpha(t) \rangle \geq 0$.

The lack of self-interaction insures, in the phase space Ω, that

$$\sum_{\alpha} \partial \dot{u}_{\alpha}/\partial u_{\alpha} = 0 \quad ,$$

that the phase flow is nondivergent, and that a Liouville theorem holds for the evolution of the probability distribution, namely

$$\partial P(u,t)/\partial t = - \sum_{\alpha} \dot{u}_{\alpha} \partial P(u,t)/\partial u_{\alpha} = - \sum_{\alpha\beta\gamma} A_{\alpha\beta\gamma} u_{\beta} u_{\gamma} \partial P(u,t)/\partial u_{\alpha} \quad . \tag{3.3}$$

In principle the single-time statistical problem is completely solved by the solution of the linear partial differential equation (3.3); in practice for large N we must find some computable approximation to (3.3).

3.3 Moment Expansions

A traditional approach to the problem of the evolution of statistical properties has been to derive equations for the evolution of moments from the dynamics equation. For first moments we can average the dynamics equation to find

$$d\langle u_{\alpha} \rangle / dt = - \nu_{\alpha} \langle u_{\alpha} \rangle + \sum_{\beta,\gamma} A_{\alpha\beta\gamma} \langle u_{\beta} u_{\gamma} \rangle . \tag{3.4}$$

For second moments we find

$$\begin{aligned} d\langle u_{\beta} u_{\gamma} \rangle / dt &= \langle \dot{u}_{\beta} u_{\gamma} \rangle + \langle u_{\beta} \dot{u}_{\gamma} \rangle \\ &= -(\nu_{\beta} + \nu_{\gamma}) \langle u_{\beta} u_{\gamma} \rangle + \sum_{\delta,\varepsilon} A_{\beta\gamma\varepsilon} \langle u_{\delta} u_{\varepsilon} u_{\gamma} \rangle + \sum_{\delta,\varepsilon} A_{\gamma\delta\varepsilon} \langle u_{\beta} u_{\delta} u_{\varepsilon} \rangle \end{aligned} \tag{3.5}$$

and, in general, each equation for the rate of change of an nth moment will involve an (n+1)-st moment. The central problem of the statistical mechanics of such quadratically nonlinear systems has been the closure problem, that is, to find an approximation for (n+1)-st moments in terms of lower order moments in order to close this otherwise infinite sequence of equations.

3.4 Properties of the Nonlinear System

Before considering approximations we should list some known properties of the original Eqs. (3.2) and (3.3) that we may hope to preserve.

1) $P(u,t) \geq 0$ and $\int_{\Omega} P(u,t)\, du = 1$.

2) The total energy

$$E(t) = \frac{1}{2} \sum_{\alpha} u_{\alpha}(t) u_{\alpha}(t)$$

and the total enstrophy

$$G(t) = \frac{1}{2} \sum_{\alpha} m_{\alpha} u_{\alpha}(t) u_{\alpha}(t)$$

are integrals of the motion, for, by Condition III,

$$\dot{E} = \sum_{\alpha} \dot{u}_{\alpha} u_{\alpha} = \sum_{\alpha\beta\gamma} A_{\alpha\beta\gamma} u_{\alpha} u_{\beta} u_{\gamma} = 0 ,$$

and by Condition IV,

$$\dot{G} = \sum_{\alpha} m_{\alpha} \dot{u}_{\alpha} u_{\alpha} = \sum_{\alpha\beta\gamma} m_{\alpha} A_{\alpha\beta\gamma} u_{\alpha} u_{\beta} u_{\gamma} = 0 .$$

A third independent condition like III and IV would imply that all $A_{\alpha\beta\gamma}$ vanish.

3) Eq. (3.3) has stationary multivariate Gaussian solutions, corresponding to the generalized Boltzmann distribution for two temperatures θ and σ,

$$P(u) = M \exp(-E/\theta - G/\sigma)$$

$$= M \exp\left(- \frac{1}{2} \sum_{\alpha} u_{\alpha}^{2}/\theta - \frac{1}{2} \sum_{\alpha} m_{\alpha} u_{\alpha}^{2}/\sigma \right) . \tag{3.6}$$

Here M is a normalization factor determined by property 1).

4) More general multivariate Gaussian probability distributions are not stationary nor do they remain Gaussian, but we can compute exactly the initial change in moments. For the Gaussian distribution

$$P(u) = M \exp\left(- \frac{1}{2} \sum_{\alpha} u_{\alpha}^{2}/U_{\alpha} \right) ,$$

the first four moments are $\langle u_\alpha \rangle = 0$, $\langle u_\alpha u_\beta \rangle = U_\alpha \delta_{\alpha\beta}$, $\langle u_\alpha u_\beta u_\gamma \rangle = 0$, and

$$\langle u_\alpha u_\beta u_\gamma u_\delta \rangle = U_\alpha \delta_{\alpha\beta} U_\gamma \delta_{\gamma\delta} + U_\alpha \delta_{\alpha\gamma} U_\beta \delta_{\beta\delta} + U_\alpha \delta_{\alpha\delta} U_\beta \delta_{\beta\gamma} \quad .$$

The initial first and second time derivatives of the second moments are given by

$$\begin{aligned} \dot{U}_\alpha &= 2 \sum_{\beta,\gamma} A_{\alpha\beta\gamma} \langle u_\alpha u_\beta u_\gamma \rangle = 0 \\ \ddot{U}_\alpha &= 4 \sum_{\beta,\gamma} A_{\alpha\beta\gamma} (A_{\alpha\beta\gamma} U_\beta U_\gamma + A_{\beta\gamma\alpha} U_\gamma U_\alpha + A_{\gamma\alpha\beta} U_\alpha U_\beta) \end{aligned} \tag{3.7}$$

It is easy to check that if $U_\alpha = (\theta^{-1} + m_\alpha \sigma^{-1})^{-1}$, then $\ddot{U}_\alpha = 0$ as expected for the stationary Gaussian solutions (3.4).

5) A property of the system described by Eqs. (3.2) and (3.3) which is not known but which is a matter of conjecture is that after a long time the system forgets everything but the temperatures θ and σ and relaxes to the generalized Boltzmann distribution, that is, that $P(u,t)$ approaches the stationary Gaussian distribution (3.6). In the sense of uniform convergence of all moments this conjecture is clearly wrong since Eq. (3.2) is time reversible, but so far as the behavior of low-order moments is concerned it may be right. In the course of time a nonstationary Gaussian probability distribution $P(u,0)$ develops an increasingly fine non-Gaussian structure that never vanishes but that makes decreasing contribution to low-order moments while continuing to contribute significantly to some increasingly higher order moments. Even for the relaxation conjecture to be valid for the behavior of low-order moments, however, the system must be ergodic and mixing on surfaces in Ω with both E and G constant. It is easy to find systems with $N \approx 10$ that do not relax; a system with $N = 5$ that does has been found by Orszag (1970). It is observed that even for non-ergodic systems the final discrepancy of second moments from their Boltzmann values seems to be of the order of $1/N$. This encourages the hope that for increasing N the relaxation conjecture, if not exact, becomes increasingly valid for most physical systems. The pragmatic value of the conjecture is

that if it were false then the coarse statistical properties of the system would remember the indefinitely distant past and there would be very little that could usefully be said about them. We shall be optimistic and assume that the relaxation conjecture is true and that second moments approach in time the values that they would have in the stationary Gaussian distribution (3.6) with values of θ and σ determined by the total average energy and enstrophy in the system.

3.5 Stochastic Model

A stochastic model for the system evolving according to (3.2) is another system evolving according to a model equation replacing (3.2) and inducing an evolution of $P(u,t)$ that, first, agrees as closely as possible with that described by (3.3) and, second, is computable in practice. These two requirements tend, of course, to be incompatible. The principal advantage of a stochastic model is that it guarantees realizability, that is, that $P(u,t) \geq 0$. This is more difficult to do with closed moment expansions or with finite difference or truncated eigenfunction approximations to $P(u,t)$ and Eq. (3.3).

3.6 Stochastic-Dynamic Prediction

We consider the "true" state of the atmosphere to be characterized by a nonrandom true vector $\bar{u}_\alpha$. Our best estimate of the state of the atmosphere, however, is characterized by a random known vector u_α where the random differences $u_\alpha - \bar{u}_\alpha$ are the result of observation and analysis errors and inadequacies. An ensemble of independent estimates of the same $\bar{u}_\alpha$ would generate a probability distribution for u_α characterized, in part, by the variance $\langle (u_\alpha - \langle u_\alpha \rangle)^2 \rangle$. Hopefully our estimate is unbiased, $\langle u_\alpha \rangle = \bar{u}_\alpha$.

In fact, we observe a single u_α and try to make a forecast from this starting vector. In stochastic dynamic prediction an attempt is made to forecast the probability distribution of finding $\bar{u}_\alpha$ about u_α. For an unbiased estimate the covariance of $\bar{u}_\alpha - u_\alpha$ is that of $u_\alpha - \bar{u}_\alpha$ and $\sigma_{\alpha\beta} = \langle (u_\alpha - \bar{u}_\alpha)(u_\beta - \bar{u}_\beta) \rangle$. Our estimates of $\sigma_{\alpha\beta}$ at time 0 must be based on an analysis of errors associated with observation and analysis tempered by our knowledge of climatology.

For the quadratically nonlinear system with the equation of motion

$$\dot{u}_\alpha = \sum_{\rho,\sigma} A_{\alpha\rho\sigma} u_\rho u_\sigma - \sum_\rho B_{\alpha\rho} u_\rho + C_\alpha \tag{3.8}$$

having the usual conditions on $A_{\alpha\beta\gamma}$, with $B_{\alpha\beta}$ hermitian and positive definite for damping, and with C_α a forcing vector, we can carry out an expansion through second moments in terms of $\mu_\alpha = \langle u_\alpha \rangle$,

$$\sigma_{\alpha\beta} = \langle (u_\alpha - \mu_\alpha)(u_\beta - \mu_\beta) \rangle$$

$$\tau_{\alpha\beta\gamma} = \langle (u_\alpha - \mu_\alpha)(u_\beta - \mu_\beta)(u_\gamma - \mu_\gamma) \rangle \tag{3.9}$$

and find

$$\begin{aligned}
\dot{\mu}_\alpha &= \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\rho \mu_\sigma + \sigma_{\rho\sigma}) - \sum_\rho B_{\alpha\rho}\mu_\rho + C_\alpha \\
\dot{\sigma}_{\alpha\rho} &= \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\tau_{\beta\rho\sigma} + \mu_\rho \sigma_{\beta\sigma} + \mu_\sigma \sigma_{\beta\rho}) \\
&\quad + \sum_{\rho,\sigma} A_{\beta\rho\sigma}(\tau_{\alpha\rho\sigma} + \mu_\rho \sigma_{\alpha\sigma} + \mu_\sigma \sigma_{\alpha\rho}) - \sum_\rho B_{\alpha\rho}\sigma_{\beta\rho} - \sum_\rho B_{\beta\rho}\sigma_{\alpha\rho} .
\end{aligned} \tag{3.10}$$

These equations have an energy integral when $B_{\alpha\rho} = 0$, $C_\alpha = 0$

$$E = \frac{1}{2}\sum_\alpha (\mu_\alpha^2 + \sigma_{\alpha\alpha}) \tag{3.11}$$

$$\begin{aligned}
dE/dt &= \sum_\alpha \mu_\alpha \dot{\mu}_\alpha + \frac{1}{2}\sum_\alpha \dot{\sigma}_{\alpha\alpha} \\
&= \sum_{\alpha,\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\alpha \mu_\rho \mu_\sigma + \mu_\alpha \sigma_{\rho\sigma} + \mu_\rho \sigma_{\alpha\sigma} + \mu_\sigma \sigma_{\alpha\rho} + \tau_{\alpha\rho\sigma}) = 0.
\end{aligned}$$

Similarly we may establish the existence when appropriate of an enstrophy integral $G = \frac{1}{2}\sum_\alpha m_\alpha(\mu_\alpha^2 + \sigma_{\alpha\alpha})$.

Epstein (1969) closed by using the <u>third cumulant discard</u> approximation, $\tau_{\alpha\rho\sigma} = 0$, which preserves energy and enstrophy integrals. Unlike the fourth cumulant discard approximation the third cumultant discard approximation is realizable. The associated stochastic model equation is

$$du_\alpha/dt = \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\rho u_\sigma + \mu_\sigma u_\rho - \mu_\rho\mu_\sigma + \sigma_{\rho\sigma}) \tag{3.12}$$

which is random linear with nonrandom coefficients based on statistics as is typical of stochastic models. For the model, setting $v_\alpha = u_\alpha - \mu_\alpha$,

$$\begin{aligned} d\mu_\alpha/dt &= \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\rho\mu_\sigma + \sigma_{\rho\sigma}) \\ dv_\alpha/dt &= \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\rho u_\sigma + \mu_\sigma u_\rho - 2\mu_\rho\mu_\sigma) \\ &= \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\rho v_\sigma + \mu_\sigma v_\rho) \\ d\sigma_{\alpha\beta}/dt &= \sum_{\rho,\sigma} A_{\alpha\rho\sigma}(\mu_\rho\sigma_{\beta\sigma} + \mu_\sigma\sigma_{\beta\rho}) + \sum_{\rho,\sigma} A_{\beta\rho\sigma}(\mu_\rho\sigma_{\alpha\sigma} + \mu_\sigma\sigma_{\alpha\rho}) \end{aligned} \tag{3.13}$$

The linear terms, $-\sum_\rho B_{\alpha\rho}u_\rho - C_\alpha$, of course, present no problem.

As for the existence of equipartition solutions, we find for this model that any ensemble with $\mu_\alpha = 0$ and $\sigma_{\alpha\beta}$ diagonal is stationary. When applied to homogeneous turbulence the third cumulant discard approximation also discards nonlinear energy transfer between modes.

Epstein (1969) carried out prediction experiments with Lorenz's low order system, $d\omega/dt = 0$, $\omega = \nabla^2\psi$, of two-dimensional flow periodic in a rectangular box $(2\pi/k, 2\pi/\ell)$ with the vorticity given by 8 modes

$$\begin{aligned} \omega = {} & A_1 \cos \ell y + A_3 \sin \ell y + A_2 \cos kx + A_4 \sin kx \\ & + A_5 \cos kx \cos \ell y + A_7 \cos kx \sin \ell y \\ & + A_8 \sin kx \cos \ell y + A_6 \sin kx \sin \ell y \end{aligned} \tag{3.14}$$

reduced even further by setting $A_3 = A_4 = A_5 = A_7 = A_8 = 0$ to the set of dynamics equations

$$\begin{aligned} \dot{A}_1 &= -\frac{1}{2}\left(\alpha(\alpha^2+1)\right)^{-1} A_2A_6 \\ \dot{A}_2 &= \frac{1}{2}\alpha^3(\alpha^2+1)^{-1} A_1A_6 \\ \dot{A}_6 &= -(\alpha^2-1)\alpha^{-1} A_1A_2 \end{aligned} \tag{3.15}$$

where $\alpha = k/\ell$. This is an example of a quadratically nonlinear three component system that can be reduced to the canonical form with solutions given by the Jacobi elliptic functions $s = sn(t)$, $c = cn(t)$, $d = dn(t)$ satisfying the equations

$$\begin{aligned} \dot{s} &= cd \\ \dot{c} &= -sd \\ \dot{d} &= -k^2 sc \end{aligned} \tag{3.16}$$

where k is the modulus.

In these experiments Epstein made comparisons for 6 days of the results of a deterministic calculation with

$$du_\alpha/dt = \sum_{\rho,\sigma} A_{\alpha\rho\sigma} u_\rho u_\sigma \quad ,$$

stochastic calculations of $d\mu_\alpha/dt$, and $d\sigma_{\alpha\beta}/dt$ using the third cumulant discard approximation, and Monte Carlo calculations of an ensemble of 500 members satisfying the deterministic equations.

Fleming (1971a,b) extended these results by keeping third moments, closing with an eddy-damped quasinormal approximation, and experimented with a two-level model of 28 degrees of freedom.

Much of Fleming's work was related to the predictability problem. In the turbulence theory approaches to the predictability problem we examine the statistics of an ensemble of pairs with the ensemble means vanishing but the pairs starting close together. In the stochastic dynamic approach we examine an ensemble of not two but myriad members clustered about a nonzero mean. Thus we compute the change in uncertainty as measured by the covariance for specific initial conditions. The average of stochastic dynamic results over all initial conditions (distributed climatologically) should agree with the results of the turbulence theory approaches.

4. Theoretical Predictability Studies

Numerical weather prediction can never be exact owing to errors in the determination of the initial state and to external error sources arising from discrepancies between the dynamics of the numerical models and that of the real atmosphere. But in addition to the initial error and external error sources there is an internal error growth mechanism which imposes a fundamental limit on the predictability of the atmosphere. This internal error growth is not an artifact of the numerical model but is a consequence of the nonlinearity and inherent instability of atmospheric dynamics. Even a hypothetical model that treated all resolved scales of motion perfectly and for which was given an initial state as perfect as possible would have its forecast skill limited by the unavoidable errors in unresolved scales which would contaminate larger scales and lead to final destruction of skill in all scales. It is the internal error growth mechanism that has been the principal object of predictability studies in recent years.

There are three approaches (Lorenz, 1969a) to the predictability problem, all of which examine the way in which a pair of atmospheric evolutions diverge from each other with time when they differ slightly in their initial states. In one approach (Lorenz, 1969b) a search is made for close analogs in the historical record of atmospheric states to serve as an initial pair. Although this is the only method that deals directly with the real atmosphere, it was, unfortunately, found that the closest analogs are not very close nor would a longer record have helped very much. Lorenz's search for analog pairs was in a data base consisting of values of the heights of the 200-, 500-, and 850-mb pressure surfaces, twice daily for the 5 years 1963-1967 at a 1003-point checkerboard subset of the NMC grid. The closeness of analogs was measured in terms of mean square height differences over the 1003-point array. The best analogs found were still rather poor so that their divergence with time only measured the growth of errors which were initially already relatively large. For each large errors he found a doubling time of 8 days.

A simple quadratic law governing error growth that reflected the moderating influence of quadratically nonlinear terms for large error growth was found by Lorenz to fit his growth results well over the observed range. When used for extrapolation the quadratic law gave for small errors an rms doubling time of 2½ days. He estimated also that a data base of 140 winters would be needed in this approach in order to give a reasonable likelihood of finding analogs close enough to correspond to observational errors.

In a second approach the real atmosphere is replaced by any one of a number of numerical models simulating the atmosphere. Although one must now rely on the similitude of the model, it is easily possible to start with initial states differing slightly in arbitrary ways. Early "perfect model" predictability studies of this sort, as summarized by Smagorinsky (1969), found error growth rates with rms error doubling times of 5 days. Jastrow and Halem (1970) showed, however, that a refinement of resolution from an 800 km to a 400 km grid spacing shortened the rms error doubling time to about 3 days. This was borne out by calculations with the National Center for Atmospheric Research (NCAR) 5° model by Williamson and Kasahara (1971), which gave rms error doubling times of 2 to 3 days. The dependence of error growth rate on resolution seems to be a consequence of the fact that coarser mesh models incorrectly impede baroclinic processes, as shown by Miyakoda et al. (1971). This failing directly causes the slowing down of a baroclinic error growth mechanism and leads as well to a lower eddy kinetic energy level in the model with a correspondingly slower barotropic error growth process.

A limitation of concern in the numerical model approach to the predictability problem is the model scale truncation necessary for feasibility of computation. Models cannot resolve scales of motion smaller than the grid scale nor, therefore, their influence on error growth. A third approach to the predictability problem was described by Robinson (1967), who suggested that statistical theories of turbulence might be used to estimate the rate at which nonlinear processes transferred error from smaller to larger scales.

Using inertial range scaling arguments Robinson (1967) suggested that a particular scale of motion of wavenumber k in the atmosphere could not be predicted for longer than a characteristic eddy time which, assuming a -5/3 power spectrum, he took as

$$\tau_k = |\varepsilon|^{-1/3} k^{-2/3} . \tag{4.1}$$

Such an estimate seems unduly pessimistic since it denies that numerical models have any skill in describing the detailed nonlinear interactions between different wavenumbers. In a more detailed examination of this problem Lorenz (1969c) concluded that τ_k was a better measure of the time for transfer of error from wavenumber 2k to wavenumber k. Thus Lorenz imagined an error cascade process in which the predictability time was the sum of the individual octave transfer times from some initial error in some higher wavenumber. Even for infinitesimal error in infinitesimal scales at k = ∞ for a -5/3 power law the sum converges.

$$T_k = \sum_{n=0}^{\infty} \tau_{2^n k} = |\varepsilon|^{-1/3} k^{-2/3} (1 + 2^{-2/3} + 4^{-2/3} + \cdots)$$

$$= 2.8\tau_k . \tag{4.2}$$

The predictability time T_k of wavenumber k is then only a few times the local eddy time τ_k. More detailed calculations with the test-field model of turbulence (Leith and Kraichnan, 1973) give for a -5/3 power law range

$$T_k = 10\tau_k$$

in three dimensions and

$$T_k = 2.5\tau_k$$

in two dimensions. It is, however, clear from atmospheric observations (Wiin-Nielsen, 1967) that the relevant spectral power law is not -5/3, but is more nearly -3, consistent with the ideas of two-dimensional turbulence theory (Kraichnan, 1967; Leith, 1968). As was pointed out

by Lorenz (1969c) the situation is quite different for a -3 power law for in this case

$$\tau_k = \eta^{-1/3}$$

is independent of k, the predictability time increases by a fixed local eddy time for each additional octave that the error must cascade, and the sum diverges for an initial error at $k = \infty$. According to this estimate for each factor of two increase in resolution of the determination of initial conditions there will be gained $\tau_k = \eta^{-1/3} \approx 1$ day in the predictability of the atmosphere. This estimate has been born out in test-field model calculations (Leith and Kraichnan, 1972) in which the energy spectrum was assumed to approach a -3 power range above the planetary wavenumber 10.

Kraichnan (1970) pointed out that the predictability problem provided a fundamental test of the realiability of turbulence models. Although the turbulence models have treated the atmosphere as a simple two-dimensional fluid, the ease with which the rate of error transfer between scales can be determined makes such models useful. Predictability calculations have been carried out with the eddy-damped Markovian model (Leith, 1971) and with the test-field model (Kraichnan, 1971a, b; Leith and Kraichnan, 1972). These calculations have shown that scale truncation was not, in fact, a serious limitation of predictability studies with numerical models. The turbulence models predict an rms error doubling time of about 2 days, which is only slightly shorter than that predicted by numerical atmospheric models and by the analog study.

As a simple step toward studying predictability without the perfect model assumption, calculations have been carried out on the divergence of solutions of two numerical models differing only in their resolution or truncation properties. Williamson (1973) has done this for the 5° and 2½° versions of the NCAR model and finds a much more rapid error growth than under the perfect model assumption. The influence of truncation differences on predictability has been studied for a spectral model by Baer and Alyea (1974).

5. Monte Carlo Forecasts

5.1 Introduction

In the course of the statistical hydrodynamical studies of the predictability problem, it has become clear that single numerical forecasts do not provide the best estimate of the true state of the atmosphere in the classical least mean square sense. The most obvious demonstration of this statement is given by the fact that at late times mean-square numerical forecast errors approach twice the mean-square error of a forecast based on the climate mean. We shall describe in the next section in some detail a method for obtaining the best estimate from the forecast by using well-known statistical techniques variously referred to as linear regression or optimal filtering. We shall call such forecasts tempered although they might also be called regressed, filtered, or hedged. Tempered forecasts use climate information and approach the climate mean at times so late that all predictability is lost. The mean-square error of such forecasts therefore approaches that of a climate mean forecast. To provide the information required to predict the risks of large anomalies it is necessary to supplement such tempered forecasts with detailed information on mean-square error as provided by estimates of forecast variance.

Differing in kind from the conventional single forecast is the stochastic dynamic forecast (Epstein, 1969) in which inhomogeneous statistical hydrodynamical models are used to forecast directly mean and variance information. As with homogeneous turbulence models some closure approximation for the moment equations must be invoked. In principle, at least, stochastic dynamic forecasts should approach climate statistics at late times, and they are therefore tempered. Stochastic dynamic forecast models have been used to study predictability (Fleming, 1971a, b), and the forecast covariance information of such models has been used in experiments on optimal data assimilation (Epstein and Pitcher, 1972). Unfortunately, the amount of arithmetic required to carry out stochastic dynamic forecasts is overwhelming for systems with a reasonably large number of degrees of freedom N, since in addition to the expected N equations for evolution of the mean field,

there are, in general, N(N+1)/2 equations for the evolution of second moment quantities. Tests so far of stochastic dynamic methods have been for systems with N of the order of 100 or less; a reasonably detailed forecast model may require a value of N of the order of 100,000.

It is our principle purpose in this section to examine the theoretical skill of Monte Carlo approximations to the stochastic dynamic forecasting technique. For a Monte Carlo sample of size m the amount of arithmetic involved is proportional to mN, and if $m \approx 10$ is sufficient to gain most of the advantages of stochastic dynamic techniquest then we may hope to carry out such calculations with available computers.

5.2 Optimal Estimation

If, for any particular forecasting scheme, we have determined for many trials the climate average error statistics, we may determine from these whether the forecasting procedure leads to the best estimate of the true state in the least-square sense. If it does not, then by linear regression methods we may generate a final best forecast using the preliminary forecast as a predictor. For Monte Carlo forecasting schemes we shall see that as the sample size m approaches infinity the forecast based on the sample mean becomes best, but that for finite m a final regression step is needed and is of greater value the smaller the sample size m. The greatest benefit of linear regression is achieved for m = 1.

We shall represent any state of the model atmosphere by an N-dimensional state vector u which we shall treat as a column vector in matrix expressions. In order to simplify statistical expressions we shall choose the origin of the N-dimensional phase space to be the mean of the stationary climate ensemble. The components of a general state vector u will represent, therefore, anomalies from a climate mean. We shall always denote an average over the climate ensemble (and no other) by brackets < > so that for the climate random vector u we shall have $\langle u \rangle = 0$.

Let the true but unknown state of the model atmosphere be represented by u_0 and let $\hat{u}$ represent some estimate of u_0. We shall assume that $\hat{u}$ is an unbiased estimate of u_0 in the sense that over many estimation trials $\langle u_0 - \hat{u} \rangle = 0$ and therefore that $\langle \hat{u} \rangle = \langle u_0 \rangle = 0$. If $\hat{u}$ were not unbiased it could, of course, be made unbiased by adding the mean residual $\langle u_0 - \hat{u} \rangle$ to $\hat{u}$.

The climate average error of the estimate $\hat{u}$ is measured by the covariance matrix $\langle (u_0 - \hat{u})(u_0^* - \hat{u}^*) \rangle$ where the asterisk denotes in general a matrix transpose and here the transposed row vector. An estimate is best in the classical least squares sense if the residual $(u_0 - \hat{u})$ is statistically orthogonal to the estimate $\hat{u}$, in the sense that the matrix $\langle (u_0 - \hat{u})\hat{u}^* \rangle = 0$. If $\hat{u}$ is not a best estimate of u_0 we may use $\hat{u}$ as a predictor and find a regression matrix B such that the vector $B\hat{u}$ is a best estimate. In this case the residual $u_0 - B\hat{u}$ should be statistically orthogonal to the predictor $\hat{u}$ and we would have

$$\langle (u_0 - B\hat{u})\hat{u}^* B^* \rangle = 0 ,$$

which is satisfied for

$$B = \langle u_0 \hat{u}^* \rangle \langle \hat{u}\hat{u}^* \rangle^{-1} . \tag{5.1}$$

Of course, the determination of the regression matrix B requires knowledge of the climate statistical properties of the estimate $\hat{u}$. Note that if $\hat{u}$ is already a best estimate then $\langle u_0 \hat{u}^* \rangle = \langle \hat{u}\hat{u}^* \rangle$ and B = I, the identity matrix. In general the covariance of this new best estimate $B\hat{u}$ is readily determined for we have

$$\langle (u_0 - B\hat{u})(u_0^* - \hat{u}^* B^*) \rangle = \langle u_0 u_0^* \rangle - B\langle \hat{u} u_0^* \rangle ,$$

which can also be written in a number of other forms.

We shall, henceforth, reserve the caret notation to represent a best unbiased estimate $\hat{u}$ of the true state u_0. We shall denote the known fixed climate covariance by

$$U = \langle u_0 u_0^* \rangle \tag{5.2}$$

and the error covariance of the estimate $\hat{u}$ by

$$\hat{\Delta} = \langle (u_0 - \hat{u})(u_0^* - \hat{u}^*) \rangle . \tag{5.3}$$

A convenient measure of the knowledge of the true state u_0 that is provided by the best estimate $\hat{u}$ is given by the matrix

$$\hat{W} = \langle \hat{u} u_0^* \rangle = \langle \hat{u}\hat{u}^* \rangle = U - \hat{\Delta} . \tag{5.4}$$

Thus we have divided the total climate covariance into two parts, $U = \hat{W} + \hat{\Delta}$, with $\hat{W}$ measuring our knowledge and $\hat{\Delta}$ our ignorance. The matrices U, $\hat{W}$, and $\hat{\Delta}$ are all covariance matrices and thus are positive definite.

5.3 Monte Carlo Forecasting Scheme

The general procedure for carrying out a Monte Carlo forecast involves the generation of a finite sample of equally likely initial states, the advancing of each using model equations, and the determination using the computed sample at any later time of a best estimate of the true state as well as the uncertainty of this estimate. We shall assume that we are starting at time t = 0 with a best unbiased estimate $\hat{u}$ of $u_0(0)$. Such an estimate will be based on an optimal combination of our knowledge of the climate ensemble, of forecasts from previous times, and of new observations at time t = 0. The first two of these sources of knowledge are already combined in a tempered forecast; the optimal addition of new information from observations has been described in many recent papers (Gandin, 1963; Jones, 1965; Eddy, 1967; Petersen, 1968; Epstein and Pitcher, 1972; Rutherford, 1972). We shall also assume that we know the initial decomposition of U into the covariance matrices $\hat{\Delta}$ and $\hat{W}$.

In order to generate our initial sample we choose a finite set of vectors $u_i(0) = \hat{u} + r_i$, i=1,2,•••,m, with r_i being normal random vectors such that $\langle r_i \rangle = 0$, $\langle r_i u_0^* \rangle = 0$, $\langle r_i \hat{u}^* \rangle = 0$ and $\langle r_i r_j^* \rangle = \hat{\Delta}\delta_{ij}$. If the matrix $\hat{\Delta}$ is in diagonal form the generation of the random vectors r_i is straightforward; if not then it is probably easiest to transform to a

representation in which $\hat{\Delta}$ is diagonal, generate the sample, and transform back to the original representation.

The statistics of the initial sample are readily computed for we have $\langle u_i \rangle = 0$ and $\langle u_i u_j^* \rangle = \langle \hat{u}\hat{u}^* \rangle + \langle r_i r_j^* \rangle = \hat{W} + \hat{\Delta}\delta_{ij}$, that is $\langle u_i u_j^* \rangle = W + \hat{\Delta} = U$ if $i=j$, and $\langle u_i u_j^* \rangle = \hat{W}$ if $i \neq j$. We may also compute the statistical relations between the sample state $u_i(0)$ and the true state $u_0(0)$. We have, for each i,

$$\langle u_i u_0^* \rangle = \langle (\hat{u} + r_i) u_0^* \rangle = \langle \hat{u} u_0^* \rangle = \hat{W} \quad .$$

So far as its statistical properties are concerned the unknown state u_0 may be treated as another member of the set of u_i with then

$$\langle u_\alpha u_\beta^* \rangle = \hat{W} + \hat{\Delta}\delta_{\alpha\beta}, \quad \alpha,\beta = 0,1,2,\cdots,m \quad . \tag{5.5}$$

Inasmuch as u_0, u_1, u_2, $\cdots$, u_m are assumed to evolve with the same dynamics, this statistical equivalence persists for all later times.

The matrix $\hat{\Delta}$ measures the size of the cloud at time $t = 0$. We shall assume that the climate average predictability properties of the model have been determined in the sense that we know the evolution of the error covariance matrix $\Delta(t)$, with initial value $\Delta(0) = \hat{\Delta}$, that measures the climate average cloud size at any later time. Studies of predictability determine $\Delta(t)$ by examining the statistics of pairs of solutions and making use of the relation

$$\langle [u_1(t) - u_2(t)][u_1^*(t) - u_2^*(t)] \rangle = 2U - 2W(t) = 2\Delta(t) \quad , \tag{5.6}$$

where $W(t) = \langle u_1(t) u_2^*(t) \rangle = \langle u_2(t) u_1^*(t) \rangle$. At late times predictability becomes completely lost so that we have, as $t \to \infty$, $\Delta(t) \to U$, $W(t) \to 0$ and a pair of solutions becomes completely uncorrelated. In particular any single solution becomes uncorrelated with the true state and, as $t \to \infty$,

$$\langle [u_0(t) - u_i(t)][u_0^*(t) - u_i^*(t)] \rangle \to 2U \quad .$$

It is important here to emphasize that $\Delta(t)$ is a measure of climate

average cloud size growth; the aim of the Monte Carlo forecasting scheme is to give more detailed information in the form of an estimate of cloud size growth for each individual forecast ensemble.

We shall refer to the probability distribution of the forecast ensemble of phase points for a particular forecast period as the forecast probability distribution. In the Monte Carlo forecasting scheme we must estimate this forecast probability distribution at any time $t > 0$ from sample statistics of the set $u_i(t)$. We shall use an overbar to indicate sample averages.

The sample mean

$$\bar{u} = \frac{1}{m} \sum_{i=1}^{m} u_i$$

is an unbiased estimate of the mean of the forecast probability distribution, i.e., the mean of the cloud of points. It is also an unbiased estimate of the true state $u_0(t)$ since the climate ensemble is stationary and we have

$$\langle u_0 - \bar{u} \rangle = \langle u_0 \rangle - \frac{1}{m} \sum_{i=1}^{m} \langle u_i \rangle = 0 \quad .$$

The climate average error covariance matrix of $\bar{u}$ as an estimate of u_0 at any time is given by

$$\begin{aligned} \bar{\Delta} &= \langle (u_0 - \bar{u})(u_0^* - \bar{u}^*) \rangle \\ &= \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \langle (u_0 - u_i)(u_0^* - u_j^*) \rangle \\ &= \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} (U - 2W + W + \Delta\delta_{ij}) \\ &= (1 + m^{-1})\Delta \quad . \end{aligned} \tag{5.7}$$

A conventional single forecast takes $\hat{u}$ as the initial state. Let us designate the resulting forecast by $\tilde{u}(t)$. This forecast is not

quite equivalent statistically to a member of the sample since at time $t = 0$ we have $\langle\tilde{u}\tilde{u}^*\rangle = \langle\hat{u}\hat{u}^*\rangle = W$ rather than U. However, as with any single forecast, at late times the forecast becomes completely uncorrelated with the true state, thus as $t \to \infty$ we have $\langle\tilde{u}(t)u_0^*(t)\rangle \to 0$, and $\langle(u_0 - \tilde{u})(u_0^* - \tilde{u}^*)\rangle \to 2U$. By contrast we have for the error covariance of $\bar{u}$, as $t \to \infty$ and $\Delta \to U$,

$$\bar{\Delta} = \langle(u_0 - \bar{u})(u_0^* - \bar{u}^*)\rangle \to (1 + m^{-1})U \ , \qquad (5.8)$$

which shows that for $m > 1$ the sample mean $\bar{u}$ becomes in time a better estimate than the conventional single forecast $\tilde{u}$.

The sample covariance matrix

$$V = \frac{1}{m} \sum_{i=1}^{m} (u_i - \bar{u})(u_i^* - \hat{u}^*) \qquad (5.9)$$

is not an unbiased estimate of the forecast distribution covariance of which the climate mean is $\Delta(t)$. We have, in fact,

$$\begin{aligned} \langle V\rangle &= \frac{1}{m} \sum_{i=1}^{m} [\langle u_i u_i^*\rangle - \langle u_i \bar{u}^*\rangle - \langle \bar{u} u_i^*\rangle + \langle \overline{uu^*}\rangle] \\ &= \frac{1}{m} \sum_{i=1}^{m} \{U - 2[W(t) + m^{-1}\Delta(t)] + [W(t) + m^{-1}\Delta(t)]\} \\ &= \frac{m-1}{m} \Delta(t) \ , \end{aligned} \qquad (5.10)$$

as is to be expected from sampling theory. For an unbiased estimate of the forecast distribution covariance we should use instead

$$D = \frac{1}{m-1} \sum_{i=1}^{m} (u_i - \bar{u})(u_i^* - \bar{u}^*) \ , \qquad (5.11)$$

so that $\langle D\rangle = \Delta$.

In general the preference for $\bar{u}$ rather than $\tilde{u}$ as an estimate of u_0 arises from the nonlinear nature of the dynamics. The evolution of $\bar{u}$ takes into account in Monte Carlo approximation second moment terms,

such as Reynolds stresses and eddy transports of heat arising from uncertainty, that are ignored in the evolution of $\tilde{u}$. These are the terms that the stochastic dynamic forecasting methods (Epstein, 1969) forecast explicitly by equations for which a moment closure approximation is made. In the application of the Monte Carlo forecast covariance matrix D to the analysis problem we must recognize that although D is an unbiased estimate there remain sampling errors in the determination of D that can lead to errors in the relative weighting of forecast information and new observations. Since these sampling errors tend to decrease with increasing m the final choice of m should depend on their importance as compared to other errors in the overall forecast procedure, in particular, that error arising from external error growth. This aspect to Monte Carlo forecasts is beyond the scope of the theoretical analysis of this paper and will best be decided by experiments.

There are a number of basic questions to be answered by such experiments. The most important has to do with the relative importance of external and internal error sources. The use of a Monte Carlo sample of initial conditions indicates only the magnitude of internal errors. Additional random forcing terms have been used by Pitcher (1974) to simulate external error sources in stochastic dynamic forecasts, and such forcing terms are also applicable to Monte Carlo forecasts. If the external error sources for a given model are large compared to the internal error sources, then the value of detailed prediction of internal error growth based on the Monte Carlo sample estimate D is diminished, and more reliance should be placed on a general regression estimate that would take into account all error sources. It is to be hoped that experiments can serve to separate the internal and external sources so that not only can an efficient choice of m be made but also so that an analysis of purely external errors can lead to model improvement.

Although $\bar{u}$ is an unbiased estimate of u_0 which is, in time, an improvement over $\tilde{u}$, it still does not represent a best unbiased estimate of $\bar{u}_0$. We have

$$\begin{aligned} \langle (u_0 - \bar{u})\bar{u}^* \rangle &= W - [W + m^{-1}\Delta] \\ &= -m^{-1}\Delta \neq 0 , \end{aligned} \tag{5.12}$$

and thus $u_0 - \bar{u}$ is not statistically orthogonal to $\bar{u}$. We obtain a best unbiased estimate, $\hat{u} = B\bar{u}$, of u_0 by use of the regression matrix

$$B = \langle u_0 \bar{u}^* \rangle \langle \bar{u}\bar{u}^* \rangle^{-1}$$

$$= W(W + m^{-1}\Delta)^{-1} = (I + m^{-1}\Delta W^{-1})^{-1}, \tag{5.13}$$

which under our general assumption is a known function of time. A measure of the error of the best estimate $\hat{u}$ is given by the covariance matrix

$$\hat{\Delta} = \langle (u_0 - B\bar{u})(u_0^* - \bar{u}^* B^*) \rangle$$

$$= \langle u_0 u_0^* \rangle - B\langle \bar{u} u_0^* \rangle$$

$$= U - BW \quad . \tag{5.14}$$

Note that as $t \to \infty$, $W \to 0$, and $\Delta \to U$, we have $B \to 0$, $\hat{u} \to 0$, and $\hat{\Delta} \to U$. A Monte Carlo forecast with a final regression step tends toward the climatological forecast, $u = 0$, at late times even for finite m. Note also that as $m \to \infty$ we have $B \to I$, $\hat{\Delta} \to \Delta$, and the final regression step becomes unimportant. The $m \to \infty$ limit gives, of course, the ideal but unobtainable forecasting scheme which exactly reflects our state of knowledge and ignorance by exactly describing the evolution of the cloud of equally likely states.

The final regression step can be viewed as an optimal filter. It will tend in practice to diminish the smaller scale components for which the relative forecast error is greater.

The test-field turbulence model has been used to test the theoretical skill of Monte Carlo forecasts (Leith, 1974). Calculations have been made of the mean-square velocity errors both without regression $\overline{V}(m,t)$ and with a final regression step $\hat{V}(m,t)$ as a function of sample size m and forecast time t. These are naturally compared to the mean square velocity error V of a climate mean forecast.

The relative mean-square velocity errors are shown in Fig. 1 for $\overline{V}/F$ (dashed) and $\hat{V}/V$ (solid) for $m=1,2,8,\infty$ as a function of forecast time t. For $m = \infty$ the function $\overline{V}(\infty,t)$ and $\hat{V}(\infty,t)$ are the same and describe the errors of the ideal forecast scheme. The average error of traditional single forecasts without regression is essentially given by $\overline{V}(1,t) = 2\overline{V}(\infty,t)$. Note that even without the final regression step most of the benefit of a Monte Carlo forecast for improved accuracy of the mean is obtained with m = 8. It is clear, however, that regression for m = 1 leads to the largest reduction of error for the smallest cost and that the use of sample sizes m > 1 will have to be justified on the basis of the detailed knowledge obtained concerning the errors of a particular forecast.

The error growth results shown in Fig. 1 are based on an assumed initial error spectrum that corresponds to an observational accuracy and resolution that is much higher than is presently available. These results represent therefore a goal that hopefully may be attained in the 1980s by the use of satellite-based sensors. An indication of present error levels is shown in Fig. 1 by the plotted points which are based on reported root-mean-square vector wind errors at 500 mb of 15 and 17 knots at 12 and 24 hours, respectively, in April 1969 (National Meteorological Center, 1969) compared to an April rms deviation from the mean of 27 knots computed from climatological results of Oort and Rasmusson (1971). The forecast errors are computed from the difference of forecasts and station reports; likewise the climatological estimate of standard deviation is based on station reports, and the influence of smoothing through an analysis procedure is avoided. It is clear from Fig. 1 that there is much room for improvement.

The measure of forecast skill used throughout this section has been mean-square error, but there are serious questions as to whether such an objective measure corresponds to a subjective evaluation of what makes a good forecast. It is, of course, difficult to develop a theory of forecast skill without some objective measure of that skill. A large body of statistical theory is based on mean-square measures of error, and it would be most useful if such a measure could be devised that agreed

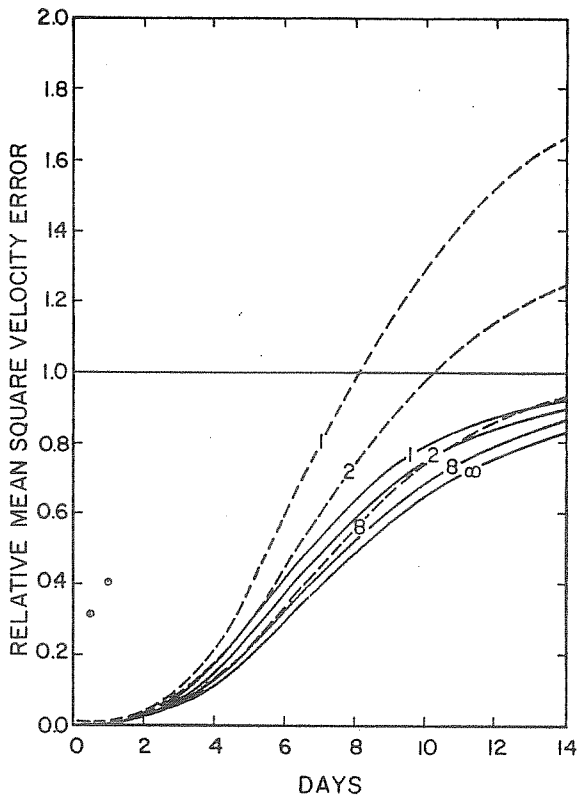

Fig. 1. Mean-square velocity error relative to that of a climate mean forecast without regression (dashed) and with regression (solid). The circle points show present forecast skill.

well with subjective evaluations. The mean-square wind error used in this section approximates somewhat the frequently used S_1 score (Teweles and Wobus, 1954) which is not itself based on mean-square error. There remains, however, considerable freedom in choosing other basic variables of the phase space through a linear or nonlinear transformation and thus in modifying the nature of the corresponding mean-square measure.

From a decision theoretical point of view the best measure of skill is that one which can be used to maximize expected gain or to minimize expected loss. Since the gain or loss as a function of the actual state of the atmosphere differs in different applications, the only universally useful information from forecasts would be a probability distribution of various outcomes. The best estimate of the mean and variance provides, of course, the lowest two moments which contribute much to the definition of such a probability distribution.

A subjective objection to any tempered forecast of the best estimate of the mean is that it is smoothed and does not therefore look "real." Such smoothing has been used to a lesser degree and accepted as necessary in analysis procedures so that evidently "real" looking weather maps should have some but not too much smoothing. An interesting possibility for overcoming this general objection is provided by the Monte Carlo forecasting procedure in that the forecast results might be displayed as a set of, say, m = 8 maps each of which looks "real" and is equally likely to be the true state so that the differences between them give a subjectively judged measure of the forecast uncertainty.

We conclude in general that a Monte Carlo forecasting procedure represents a practical, computable approximation to the stochastic dynamic forecasts proposed by Epstein (1969). Adequate accuracy should be obtained for the best mean estimate of the forecast field with sample sizes as small as 8. Improvement in skill is appreciable for Monte Carlo forecasts as compared to conventional single forecasts although much of this improvement comes from the filtered nature of the forecasts and is obtainable with a linear regression step applied to a single forecast.

The question of what sample size is adequate for the detailed determination of forecast error needed for optimal data assimilation has not

been decided by the present theoretical study and will require experiments with real data applications of the Monte Carlo procedure.

6. Regression Forecasts

6.1 Introduction

The use of pure statistical regression methods for forecasting is based on the observation that there exist significant time lagged correlations in atmospheric states and that knowledge of a present state can, from purely empirical considerations, provide information about future states. This general statistical problem of making a best estimate of an unobserved quantity on the basis of an observed quantity which is known to be correlated is of such general importance in analysis as well as in forecasting that it is worth a careful study.

6.2 Optimum Estimation Methods

Let us consider first the two-component random vector (x,y) for which is known, a priori, a normal climate ensemble characterized by $\langle x\rangle = 0$, $\langle y\rangle = 0$, $\langle x^2\rangle = X > 0$, $\langle y^2\rangle = Y > 0$, $\langle xy\rangle = W$, so that the covariance matrix is

$$S = \begin{pmatrix} X & W \\ W & Y \end{pmatrix} > 0 \quad . \tag{6.1}$$

When we write for a matrix $S > 0$ we mean that S is positive definite, i.e., that it is square, is symmetric, has real positive eigenvalues, is non-singular, and is invertible so that S^{-1} exists with $S^{-1} > 0$.

Now we make an observation that determines $y = y_0$. What is the best estimate of x given this new knowledge? We shall examine three methods of arriving at a single answer to this question.

LT method: Let us try to find a linear transformation (LT) that transforms from the random variables (x,y) to random variables (ε,y) such that $\langle\varepsilon\rangle = 0$, $\langle y\rangle = 0$, $\langle\varepsilon y\rangle = 0$, $\langle\varepsilon^2\rangle = E > 0$, $\langle y^2\rangle = Y > 0$, so that now the covariance matrix is

$$S' = \begin{pmatrix} E & 0 \\ 0 & Y \end{pmatrix} > 0 \quad . \tag{6.2}$$

Of course, y is unchanged and x is replaced by ε through the linear relation $x = by + \varepsilon$ or $\varepsilon = x - by$. To find b we note that

$$0 = \langle \varepsilon y \rangle = \langle xy \rangle - b\langle y^2 \rangle = W - bY$$

or $b = W/Y$. We can also readily compute

$$E = \langle \varepsilon^2 \rangle = X - 2bW + b^2Y$$

$$= X - 2W^2/Y + W^2/Y = X - W^2/Y$$

$$= X(1 - W^2/XY) \ . \qquad (6.3)$$

Since by the Schwarz inequality $W^2 \leq XY$, we have $X \geq E \geq 0$. Our best estimate of $\hat{x}$ is evidently $x = by_0$ with variance $\langle \varepsilon^2 \rangle = E$.

LS method: Let us try to find a linear estimating formula $\hat{x} = by$ that minimizes the climate ensemble average of

$$E = \langle (x - \hat{x})^2 \rangle \ . \qquad (6.4)$$

This is the method of least squares (LS). As function of b we have

$$E(b) = \langle x^2 \rangle - 2b \langle xy \rangle + b^2 \langle y^2 \rangle$$

$$= X - 2bW + b^2Y \qquad (6.5)$$

In order to minimize with respect to b we want

$$0 = dE(b)/db = -2W + 2bY$$

thus $b = W/Y$. Furthermore we have

$$d^2E(b)/db^2 = 2Y > 0$$

so that we have indeed a true minimum in E at $b = W/Y$. The value of E there is

$$E = X - W^2/Y + W^2/Y = x - W^2/Y$$

$$= X(1 - W^2/XY) \geq 0 \ . \qquad (6.6)$$

ML method: Let us try to find the most likely (ML) value of x given an observation of $y = y_0$. The inverse of the covariance matrix S is

$$S^{-1} = \Delta^{-1}\begin{pmatrix} Y & -W \\ -W & X \end{pmatrix} > 0 \qquad (6.7)$$

where $\Delta = |S| = (XY - W^2) > 0$. The climatic joint probability distribution of (x,y) is if normal

$$p(x,y) = \eta\, e^{-\frac{1}{2}\Delta^{-1}(Yx^2 - 2Wxy + Xy^2)} \qquad (6.8)$$

where η is some normalizing constant. Given $y = y_0$, the conditional probability distribution of x is

$$p(x|y_0) = \eta' e^{-\frac{1}{2}\Delta^{-1}(Yx^2 - 2Wxy_0 + Xy_0^2)} \quad . \qquad (6.9)$$

The most likely value of x is that one $x = \hat{x}$ which maximizes $p(x|y_0)$ or minimizes

$$L(x) = \Delta^{-1}(Yx^2 - 2Wxy_0 + Xy_0^2) \qquad (6.10)$$

Thus we want at $x = \hat{x}$

$$0 = dL(x)/dx = \Delta^{-1}(2Yx - 2Wy_0)$$

whence $\hat{x} = (W/Y)y_0$. This is a true minimum since

$$d^2L(x)/dx^2 = \Delta^{-1}(2Y) > 0 \quad .$$

In terms of x and $\hat{x}$ we have

$$\begin{aligned} p(x|y_0) &= \eta' e^{-\frac{1}{2}\Delta^{-1}(Yx^2 - 2Yx\hat{x} + (XY^2/W^2)\hat{x}^2)} \\ &= \eta' e^{-\frac{1}{2}Y\Delta^{-1}(x - \hat{x})^2} e^{-\frac{1}{2}YW^{-2}\hat{x}^2} \\ &= \eta'' e^{-\frac{1}{2}Y\Delta^{-1}(x - \hat{x})^2} \quad . \end{aligned} \qquad (6.11)$$

Note then that the conditional mean of x is also $\hat{x}$ and its variance is

$$E = \Delta/Y = X - W^2/Y = X(1- W^2/XY) \geq 0 \quad . \tag{6.12}$$

We have found by three methods that after an observation $y = y_0$ our best estimate of x changes from 0 to $\hat{x} = (W/Y)y_0$ with a reduction in variance from X to $X(1 - W^2/XY)$. The agreement between the three methods depends on the assumption that the climate ensemble has a normal probability distribution.

6.3 Regression Matrix

We shall next examine some generalizations of the LT method. The simplest generalization is to more than two dimensions. We let x be a vector to remain unobserved with $\langle x \rangle = 0$ and y a vector to be observed with $\langle y \rangle = 0$ before observation. We assume that we know before observation all second moments

$$\begin{aligned} X &= \langle xx^* \rangle > 0 \\ Y &= \langle yy^* \rangle > 0 \\ W &= \langle xy^* \rangle \\ W^* &= \langle yx^* \rangle \quad . \end{aligned} \tag{6.13}$$

Here x is a column vector and x^*, its transpose, is a row vector, each say with dimension m. In general x^*x is a scalar (an inner product), but xx^* is a m x m matrix. The matrices $X = X^*$, $Y = Y^*$ are positive definite. The dimension of y, say n, need not be the same as that of x. Thus W is a not necessarily square n x m matrix. The overall covariance matrix is, of course, the composite (m+n)x(m+n) matrix

$$S = \begin{pmatrix} X & W \\ W^* & Y \end{pmatrix} > 0 \tag{6.14}$$

We wish to find an m x n matrix B that effects a linear transformation to new random vectors ε, y related to x by

$$x = By + \varepsilon \tag{6.15}$$

and such that $\langle\varepsilon\rangle = 0$, $\langle\varepsilon y^*\rangle = 0$, and $\langle\varepsilon\varepsilon^*\rangle = E \geq 0$. We have then

$$W = \langle xy^*\rangle = B\langle yy^*\rangle = BY \tag{6.16}$$

thus we have $B = WY^{-1}$ with transpose $B^* = Y^{-1}W^*$. The matrix B is the regression matrix.

We readily determine E by computing

$$\begin{aligned} E &= \langle\varepsilon(x - By)^*\rangle \\ &= \langle\varepsilon x^*\rangle - \langle\varepsilon y^*\rangle B^* \\ &= \langle\varepsilon x^*\rangle \\ &= \langle(x - By)x^*\rangle \\ &= X - BW^* = X - WY^{-1}W^* \end{aligned} \tag{6.17}$$

We may also write

$$E = X - BYB^* \ .$$

Let us now assume that the correlation between x and y arises from a known linear relation $y = Ax$ with A a known $n \times m$ matrix. Then we have

$$\begin{aligned} Y &= \langle Ax(Ax)^*\rangle = AXA^* > 0 \\ W &= \langle x(Ax)^*\rangle = XA^* \\ W^* &= AX \end{aligned} \tag{6.18}$$

and the estimation matrices become

$$\begin{aligned} B &= WY^{-1} = XA^*(AXA^*)^{-1} \\ E &= X - WY^{-1}W^* = X - XA^*(AXA^*)^{-1}AX \ . \end{aligned} \tag{6.19}$$

If further A is a nonsingular square matrix so that A^{-1} exists then

$$B = XA^*(A^{*-1}X^{-1}A^{-1}) = A^{-1}$$

$$E = X - XA^*(A^{*-1}X^{-1}A^{-1})AX = X - X = 0 . \qquad (6.20)$$

In this case we have a precise determination of $x = A^{-1}y$ with vanishing variance and no dependence on X.

Suppose, finally, that the observations are subject to error η such that, in fact, the correlation between x and y arises from the relation

$$y = Ax + \eta \qquad (6.21)$$

where we assume that $\langle x\eta^* \rangle = 0$ and $\langle \eta\eta^* \rangle = N > 0$ is known. Now we have

$$\begin{aligned} Y &= AXA^* + N \\ W &= XA^* \\ B &= XA^*(AXA^* + N)^{-1} \\ E &= X - XA^*(AXA^* + N)^{-1}AX \end{aligned} \qquad (6.22)$$

For the special case that A is nonsingular we have

$$\begin{aligned} B &= ((AXA^* + N)(A^{*-1}X^{-1}))^{-1} \\ &= (A + NA^{*-1}X^{-1})^{-1} \\ E &= X(I - (I + X^{-1}A^{-1}NA^{*-1})^{-1}) \end{aligned} \qquad (6.23)$$

which owing to the noise η is no longer precise and does depend on X. Of course, for $N = 0$, we have $B = A^{-1}$, $E = 0$ as before.

To summarize, if the observations y are related to an unknown vector x by the relation

$$y = Ax + \eta \qquad (6.24)$$

where A and $X = \langle xx^* \rangle$ are known, and the random error of observation has $\langle \eta \rangle = 0$, $\langle \eta x^* \rangle = 0$ and $\langle \eta\eta^* \rangle = N$ known, then the best estimate for x is provided by

$$x = By + \varepsilon \qquad (6.25)$$

where $\langle y\varepsilon^* \rangle = 0$ and $\langle \varepsilon\varepsilon^* \rangle = E$ with

$$B = XA^*(AXA^* + N)^{-1}$$

$$E = X - XA^*(AXA^* + N)^{-1}AX \quad . \tag{6.26}$$

After an observation $y = y_0$, the best estimate of x is $\hat{x} = By_0$ with variance E.

The general expressions for B and E involve the inverse of the matrix $Y = AXA^* + N$ in y-space which is usually of smaller diemsnion than is x-space. In case x-space has the smaller dimension define $Z = X^{-1} + A^*N^{-1}A$ and note that

$$A^*N^{-1}Y = A^* + A^*N^{-1}AXA^* = ZXA^* \tag{6.27}$$

so that

$$Z^{-1}A^*N^{-1} = XA^*Y^{-1} = B \quad . \tag{6.28}$$

If N is easily inverted (as it is if it's diagonal) then it is better to use

$$B = Z^{-1}A^*N^{-1}$$

$$E = X - BYB^* \quad . \tag{6.29}$$

Example: The vector x represents coefficients in an orthogonal function expansion of a meteorological field. The vector y represents observations at a set of space points.

Example: The vector x represents a vertical temperature profile as given by values at 10 mb intervals. The vector y represents a set of observed radiances at a few different infrared wavenumbers (Strand and Westwater, 1968 a, b; Westwater and Strand, 1968; Rodgers, 1970).

Example: The vector x represents tomorrow's state of the atmosphere. The vector y represents today's observed state.

7. Statistical Assimilation of Observations

7.1 Statistically Optimum Analysis

Gandin's (1963) method of objective analysis makes extensive use of climate statistics in the form of two space point covariance functions which for homogeneous statistics depend only on the separation vector r.

$$U_{ij}(\underset{\sim}{r}) = \langle u_i(\underset{\sim}{x}) u_j(\underset{\sim}{x}+\underset{\sim}{r}) \rangle$$

For isotropic statistics we have the further simplification that $U_{ij}(\underset{\sim}{r}) = U_{ij}(\underset{\sim}{r})$ depends only on the separation distance $r = |\underset{\sim}{r}|$. As in the case of two time covariance functions for time series here too we have a Fourier transform relation

$$U_{ij}(r) = \int d\underset{\sim}{k} e^{i\underset{\sim}{k}\cdot\underset{\sim}{r}} U_{ij}(\underset{\sim}{k})$$

$$= \int d\underset{\sim}{k} e^{i\underset{\sim}{k}\cdot\underset{\sim}{r}} P_{ij}(\underset{\sim}{k}) U(k)$$

between the spectral function $U_{ij}(\underset{\sim}{k}) = P_{ij}(\underset{\sim}{k})U(k)$ and the space covariance function $U_{ij}(r)$. The space covariance function enters into a least squares method of analysis.

We consider a set of n predictors y_i, $i=1,2,\cdots,n$, which may be observations at station locations and a single predictand x which may be a value of the field at a mesh point location. We deal only with deviation fields so that all climatic means vanish. We wish to find the best linear prediction relation $\hat{x} = b_i y_i$ in the sense that we minimize

$$E = \langle (x - \hat{x})^2 \rangle = \langle x^2 \rangle - 2b_i \langle x y_i \rangle + b_i b_j \langle y_i y_j \rangle .$$

Let $Y_{ij} = \langle y_i y_j \rangle$, $W_i = \langle x y_i \rangle$ then we have

$$0 = \partial E / \partial b_i = -2W_i + 2b_j Y_{ij}$$

whence the coefficients b_j are given by

$$b_j = (Y^{-1})_{ji} W_i$$

Since $\partial^2 E/\partial b_i \partial b_j = 2Y_{ij} > 0$ is positive definite, b_j is at a true minimum. For any perturbation δb_i we would have $\delta E = \frac{1}{2}(\partial^2 E/\partial b_i \partial b_j)\delta b_i \delta b_j > 0$. For this b_i we have

$$b_i W_i = b_i b_j Y_{ij}$$

and thus we may compute E as

$$E = \langle x^2 \rangle - 2b_i W_i + b_i b_j Y_{ij} = \langle x^2 \rangle - b_i b_j Y_{ij} < \langle x^2 \rangle$$

In case of errors of observation which are

1) unbiased with variance σ^2
2) uncorrelated with observed fields
3) uncorrelated between stations

we replace the covariance matrix Y_{ij} of the fields at the observing stations with the covariance matrix $Y_{ij} + \sigma^2 \delta_{ij}$ of the fields with observational errors. The resulting optimum coefficients become

$$b_j = ((Y + \sigma^2 I)^{-1}{}_{ji} W_i \quad .$$

A simple example is that of a best estimate of field values in the neighborhood of a single observation. Here we have a single predictor y and the least square method reduces to that described earlier (with an interchange in the roles of x and y). For a homogeneous field we have

$$X = \langle x^2 \rangle = \langle y^2 \rangle = Y, \quad W = \langle xy \rangle$$

$$b = W/Y = \langle xy \rangle / \langle y^2 \rangle = r(x,y)$$

where r(x,y) is the correlation (normalized covariance) between x and y. We also have

$$E = X - b^2 Y = X(1 - W^2/XY) = X(1 - r^2) \leq X$$

A more complex example is that of interpolation between two points on a line. Consider the random function $f(\xi)$ with covariance function

$F(\rho) = \langle f(\xi)f(\xi+\rho)\rangle$ that has been observed at two points ξ and $\xi + \lambda$. An interpolation formula will give a prediction

$$f_p(\xi + x) = a(x)f(\xi) + b(x)f(\xi + \lambda)$$

at any other point x. A simple example is linear interpolation in which $a = (\lambda - x)/\lambda$, $b = x/\lambda$. But optimum interpolation will give in general a smaller variance

$$E = \langle (f_p(\xi + x) - f(\xi + x))^2 \rangle$$

$$E(x) = \langle (af(\xi) + bf(\xi + \lambda) - f(\xi + x))^2 \rangle$$

$$= a^2F(0) + b^2F(0) + F(0) + 2abF(\lambda) - 2aF(x) - 2bF(\lambda - x)$$

As usual we minimize E(x) with respect to a and b by setting

$$\partial E/\partial a = 2aF(0) + 2bF(\lambda) - 2F(x) = 0$$

$$\partial E/\partial b = 2bF(0) + 2aF(\lambda) - 2F(\lambda - x) = 0$$

to find the pair of linear equations

$$F(0)a + F(\lambda)b = F(x)$$

$$F(\lambda)a + F(0)b = F(\lambda - x)$$

with the solution

$$a(x) = \frac{F(0)F(x) - F(\lambda)F(\lambda - x)}{F(0)^2 - F(\lambda)^2}$$

$$b(x) = \frac{F(0)F(\lambda - x) - F(\lambda)F(x)}{F(0)^2 - F(\lambda)^2} \quad .$$

For this optimum interpolation formula we have

$$E(x) = F(0) - \frac{F(0)[F(x)^2 + F(\lambda - x)^2] - 2F(x)F(\lambda)F(\lambda - x)}{F(0)^2 - F(\lambda)^2}$$

Note that as expected for observations without errors, we have $E(0) = 0$ and $E(\lambda) = 0$.

7.2 Combining Two Independent Estimates

We consider two independent but imperfect observations of a fixed vector u in a space of dimension N. The estimates being imperfect can be treated as random vectors v, w. The estimates are assumed to be unbiased thus $\langle v \rangle = \langle w \rangle = u$. We define new random vectors $x = v - u$, $y = w - u$ so that $\langle x \rangle = \langle y \rangle = 0$. We shall assume that v, w hence also x, y are Gaussian, and thus the probability distributions are completely characterized by positive definite second moments X, Y which we assume to be known. We wish to find a positive definite weighting operator Q such that the new random vector $z = Qx + (I - Q)y$ will be a best estimate of 0 in the sense that the second moment Z will be a minimum. For then $Qv + (I - Q)w = u + z$ will be a best estimate of u. We also wish to determine Z.

For N = 1 the problem is well known. The weight Q should be inversely proportional to the variance X, $Q = (X^{-1} + Y^{-1})^{-1}X^{-1}$, and the variance of the combined estimate is $Z = (X^{-1} + Y^{-1})^{-1}$.

We introduce the characteristic function of vectors s

$$\theta_x(s) = \langle e^{i(x,s)} \rangle = \int e^{i(x,s)} p(x)\, dx = e^{-\frac{1}{2}(Xs,s)}$$

Note that for any linear transformation A

$$\theta_{Ax}(s) = \langle e^{i(Ax,s)} \rangle = \langle e^{i(x,A^*s)} \rangle = \theta_x(A^*s) = e^{-\frac{1}{2}(XA^*s,A^*s)}$$

$$= e^{-\frac{1}{2}(AXA^*s,s)}$$

and, for x, y independent,

$$\theta_{x+y}(s) = \langle e^{i(x+y,s)} \rangle = \langle e^{i(x,s)} e^{i(y,s)} \rangle = \langle e^{i(x,s)} \rangle \langle e^{i(y,s)} \rangle$$

$$= \theta_x(s)\theta_y(s)$$

thus for $z = Qx + (I - Q)y$, $0 \leq Q \leq I$,

$$\theta_z(s) = e^{-\frac{1}{2}(QXQ + (I - Q)Y(I - Q))s,s)}$$

$$Z = QXQ + (I - Q)Y(I - Q)$$

We assume first that X, Y commute, and thus Q, which is a function of X and Y, also commutes with X and Y.

We wish Z to be stationary relative to any positive definite commuting variation dQ in Q.

$$dZ = dQXQ - dQY(I-Q) + QXdQ - (I-Q)YdQ$$

$$= dQ((X+Y)Q - Y) + (Q(X+Y) - Y)dQ = 0$$

This condition will be satisfied for

$$Q = (X+Y)^{-1}Y = (X+Y)^{-1}(YXX^{-1}Y^{-1})Y = (X^{-1}Y^{-1}(X+Y))^{-1}X^{-1}$$

$$= (X^{-1} + Y^{-1})^{-1}X^{-1}$$

$$I-Q = (X^{-1} + Y^{-1})^{-1}Y^{-1}$$

$$Z = (X^{-1} + Y^{-1})^{-2}X^{-1} + (X^{-1} + Y^{-1})^{-2}Y^{-1} = (X^{-1} + Y^{-1})^{-1}$$

$$Q = ZX^{-1}$$

$$I-Q = ZY^{-1}$$

We know that Z is not only stationary but a minimum since

$$d^2Z = 2dQ(X+Y)dQ \geq 0$$

These expressions correspond to the well known results for N = 1, which is to be expected since if X, Y commute there is a unitary transformation that simultaneously diagonalizes all operators. In the diagonalizing coordinate system components are independent and can be combined one by one, each pair as for N = 1. The more interesting case then is when X, Y do not commute.

Since X is positive definite it has a positive definite square root $X^{\frac{1}{2}}$ with inverse $X^{-\frac{1}{2}}$. Make a transformation of coordinates

$$\bar{x} = X^{-\frac{1}{2}}x$$

$$\bar{y} = X^{-\frac{1}{2}}y \quad .$$

Then

$$\theta_{\underline{x}}(s) = e^{-\frac{1}{2}(x^{-\frac{1}{2}}XX^{-\frac{1}{2}}s,s)} = e^{-\frac{1}{2}(s,s)}$$

$$X = I$$

$$\theta_{\underline{y}}(s) = e^{-\frac{1}{2}(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}s,s)}$$

$$Y = X^{-\frac{1}{2}}YX^{-\frac{1}{2}}$$

But now X, Y commute, we can minimize $Z = (X^{-1}+Y^{-1})^{-1}$ by choosing

$$\bar{z} = \bar{Q}\bar{x} + (I-\bar{Q})\bar{y}$$

with $\bar{Q} = ZX^{-1} = Z$. Then

$$Z = X^{\frac{1}{2}}ZX^{\frac{1}{2}} = X^{\frac{1}{2}}(I + X^{\frac{1}{2}}Y^{-1}X^{\frac{1}{2}})^{-1}X^{\frac{1}{2}} = (X^{-\frac{1}{2}}(I + X^{\frac{1}{2}}Y^{-1}X^{\frac{1}{2}})X^{-\frac{1}{2}})^{-1}$$

$$= (X^{-1} + Y^{-1})^{-1}$$

and

$$z = X^{\frac{1}{2}}\bar{z} = X^{\frac{1}{2}}(\bar{Q}\bar{x} + (I-\bar{Q})\bar{y}) = X^{\frac{1}{2}}\bar{Q}X^{-\frac{1}{2}}x + X^{\frac{1}{2}}(I-\bar{Q})X^{-\frac{1}{2}}y$$

$$= Qx + (I-Q)y$$

where now

$$Q = X^{\frac{1}{2}}\bar{Q}X^{-\frac{1}{2}} = X^{\frac{1}{2}}ZX^{-\frac{1}{2}} = ZX^{-1}$$

$$I-Q = ZY^{-1}$$

so we still get the same answer when we are careful with the order of multiplication. Q now is not necessarily positive definite. There remains a question, however. Does minimizing Z with respect to $\bar{Q}$ also minimize Z with respect to Q? I rather doubt it, but this seems to be the only thing to do. The procedure is at least symmetric in X and Y.

A Bayesian might call X the prior variance, Y the observation variance, and Z the posterior variance. A stochastic dynamic forecaster and analyzer might call X the forecast variance, Y the observation variance, and Z the analyzed variance.

It has been assumed that X, Y, Z, etc. are positive definite non-singular, that is, that each estimate provides at least some measure of all degrees of freedom. The case where a measurement is incomplete should be readily available by a limiting process involving large variance for undetermined components, although then the noncommuting problem would seem to be more serious.

7.3 Assimilation of Observations into Numerical Weather Prediction Models

Numerical experiments

Review

Jastrow, 1972

Kasahara, 1972

Primitive equation models

Charney, Halem, Jastrow, 1969

Jastrow, Halem, 1970

Smagorinsky, Miyakoda, Strickler, 1970

Williamson, Kasahara, 1971

Gordon, Umscheid, Miyakoda, 1972

Baumhefner, Julian, 1972

Kasahara, Williamson, 1972

Økland, 1972

Mesinger, 1972

Rutherford, Asselin, 1972

High frequency damping with model

Morel, Lefevre, Rabreau, 1971

Mode analysis

Williamson, Dickinson, 1972

Dickinson, Williamson, 1972

Filtered models

Bengtsson, Gustavsson, 1971

Miyakoda, Talagrand, 1971

Talagrand, Miyakoda, 1971

- Dynamical constraints
  - Variational method
    - Sasaki, 1969
    - Thompson, 1969
    - Stephens, 1970
    - Lewis, Grayson, 1972
  - Algebraic method
    - Tadjbakhsh, 1969
    - Phillips, 1971
- Statistical methods
  - Objective analysis
    - Gandin, 1963
    - Jones, 1965a, b
    - Rutherford, 1972
  - Radiance analysis
    - Westwater and Strand, 1968
    - Rodgers, 1970
  - Stochastic dynamic methods
    - Epstein, 1969
    - Epstein and Pitcher, 1972
- Alternate representations
  - Empirical orthogonal functions
    - Craddock and Flood, 1969
    - Craddock and Flintoff, 1970
    - Rinne, 1971
  - Orthogonal polynomials
    - Dixon, Spackman, Jones, and Francis, 1972

BIBLIOGRAPHY FOR SECTION 7.3

Baumhefner, D. P., and P. R. Julian, 1972: The reference level problem: Its location and use in numerical weather predictions. J. Atmos. Sci., 29, 285-299.

Bengtsson, L., and N. Gustavsson, 1971: An experiment in the assimilation of data in dynamical analysis. Tellus, 23, 328-336.

Charney, J., M. Halem, and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci., 26, 1160-1163.

Craddock, J. M., and S. Flintoff, 1970: Eigenvector representations of Northern Hemisphere fields. Quart. J. R. Meteor. Soc., 96, 124-129.

_____ and C. R. Flood, 1969: Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere. Quart. J. R. Meteor. Soc., 95, 576-593.

Dickinson, R. E., and D. L. Williamson, 1972: Free oscillations of a discrete stratified fluid with application to numerical weather prediction. J. Atmos. Sci., 29, 623-640.

Dixon, R., E. A. Spackman, I. Jones, and A. Francis, 1972: The global analysis of meteorological data using orthogonal polynomial base functions. J. Atmos. Sci., 29, 609-622.

Epstein, E. S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-59.

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

Gandin, L. S., 1963: Objective analysis of meteorological fields. Gidromet. Izdat. Leningrad. (Translated 1965).

Gordon, C. T., L. Umscheid, Jr., and K. Miyakoda, 1972: Simulation experiments for determining wind data requirements in the tropics. J. Atmos. Sci., 29, 1064-1075.

Jastrow, R., 1972: Observing system simulation studies. GARP JOC-VII Report, Annex D.

Jastrow, R., and M. Halem, 1970: Simulation studies related to GARP. Bull. Amer. Meteor. Soc., 51, 490-513.

Jones, R. H., 1965a: Optimal estimation of initial conditions for numerical prediction. J. Atmos. Sci., 22, 658-663.

_____, 1965b: An experiment in non-linear prediction. J. Appl. Meteor., 4, 701-705.

Kasahara, A., 1972: Simulation experiments for meteorological observing systems for GARP. Bull. Amer. Meteor. Soc., 53, 252.

_____ and D. Williamson, 1972: Evaluation of tropical wind and reference pressure measurements: Numerical experiments for observing systems. Tellus, 24, 100-115.

Lewis, J. M., and T. H. Grayson, 1972: The adjustment of surface wind and pressure by Sasaki's variational matching technique. J. Appl. Meteor., 11, 586-597.

Mesinger, F., 1972: Computation of the wind by forced adjustment to the height field. J. Appl. Meteor., 11, 60-71.

Miyakoda, K., and O. Talagrand, 1971: The assimilation of past data in dynamical analysis. I. Tellus, 23, 310-317.

Morel, P., G. Lefevre, and G. Rabreau, 1971: On initialization and non-synoptic data assimilation. Tellus, 23, 197-206.

Økland, H., 1972: On the balance, initialization and data assimilation in primitive equation prediction models. J. Atmos. Sci., 29, 641-648.

Phillips, N. A., 1971: Ability of the Tadjbakhsh method to assimilate temperature data in a meteorological system. J. Atmos. Sci., 28, 1325-1329.

Rinne, J., 1971: Investigation of the forecasting error of a simple barotropic model with the aid of empirical orthogonal functions. Geophysica, 11, 185-239.

Rodgers, C. D., 1970: Remote sounding of the atmospheric temperature profile in the presence of cloud. Quart. J. R. Meteor. Soc., 96, 654-666.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. J. Atmos. Sci., 29, 809-815.

_____ and R. Asselin, 1972: Adjustment of the wind field to geopotential data in a primitive equations model. J. Atmos. Sci., 29, 1059-1063.

Sasaki, Y., 1969: Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. J. Meteor. Soc. of Japan, 47, 115-124.

Smagorinsky, J., K. Miyakoda, and R. F. Strickler, 1970: Relative importance of variables in initial conditions for dynamical weather prediction. Tellus, 22, 141-157.

Stephens, J. J., 1970: Variation initialization with the balance equation. J. Appl. Meteor., 9, 732-739.

Tadjbakhsh, I. G., 1969: Utilization of time dependent data in running solution of initial value problems. J. Appl. Meteor., 8, 389-391.

Talagrand, O., and K. Miyakoda, 1971: The assimilation of past data in dynamical analysis. II. Tellus, 23, 318-327.

Thompson, P. D., 1969: Reduction of analysis error through constraints of dynamical consistency. J. Appl. Meteor., 8, 738-742.

Westwater, E. R., and O. N. Strand, 1968: Statistical information content of radiation measurements used in indirect sensing. J. Atmos. Sci., 25, 750-758.

Williamson, D. L., and R. E. Dickinson, 1972: Periodic updating of meteorological variables. J. Atmos. Sci., 29, 190-193.

_____ and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating. J. Atmos. Sci., 28, 1313-1324.

8. Some Key Problems for Medium-Range Forecasting

8.1 False Regression

Much of the discussion of regression methods is based on statistical properties of the whole ensemble. But in practice we must determine a regression formula from a limited sample of available data called the dependent data set and then apply it to new data forming an independent data set. A fundamental limitation to regression methods is imposed by sampling errors. This has the practical consequence of limiting the number of predictors that can be used profitably (Lorenz, 1956). A complete theory of this limitation does not yet seem to have been developed, and the optimum number of predictors is usually decided by experiment. We shall examine the simplest version of this problem in which the predictors are uncorrelated with each other or the predictand, so that any nonzero regression coefficients determined by a sample must be false.

Consider first a two-dimensional random vector (x,p) with $\langle x\rangle = \langle p\rangle = 0$, $\langle x^2\rangle = \langle p^2\rangle = 1$, and $\langle xp\rangle = 0$. The correlation r is zero for the ensemble, but for finite samples of N vectors the sample correlation coefficient

$$\bar{r} = \sum_{n=1}^{N} (x_n - \bar{x}_n)(p_n - \bar{p}_n) \Big/ \left[\sum_{n=1}^{N} (x_n - \bar{x}_n)^2 \sum_{n=1}^{N} (p_n - \bar{p}_n)^2 \right]^{\frac{1}{2}}$$

will tend to cluster about zero with a probability density distribution (Cramer, 1946)

$$f_N(\bar{r}) = (1/\pi)(N - 2) \int_0^{\pi/2} \sin^{N-2} \theta \, d\theta \, (1 - \bar{r}^2)^{(N-4)/2}$$

The use of $\bar{r}$ as a regression coefficient would lead to an apparent reduction in variance given on the average by

$$\widetilde{1 - \bar{r}^2} = \int_{-1}^{1} (1 - \bar{r}^2) \, f_N(\bar{r}) \, d\bar{r}$$

$$= (2/\pi)(N - 2) \int_0^{\pi/2} \sin^{N-2} \theta \, d\theta \int_0^{\pi/2} \sin^{N-1} \theta \, d\theta$$

$$= 1 - 1/(N-1)$$

Here $\widetilde{\ }$ indicates an average taken over many samples.

There are two ways in which sampling errors enter into this false regression. In addition to the sample estimate of $\bar{r}$ differing in general from zero so also do the sample estimates of the means $\bar{x}$ and $\bar{p}$. On the basis of a given sample the regression formula would be

$$\hat{x} = \bar{x} + \bar{r}(p - \bar{p})$$

and when applied to the ensemble as a whole the mean square error of estimation would be

$$E = \langle(x - \hat{x})^2\rangle = \langle x^2\rangle - 2\langle x\hat{x}\rangle + \langle\hat{x}^2\rangle$$

$$= \langle x^2\rangle - 2\bar{r}\langle xp\rangle + (\bar{x} - \bar{r}\bar{p})^2 + \bar{r}^2\langle p^2\rangle$$

$$= 1 + \bar{r}^2 + (\bar{x} - \bar{r}\bar{p})^2 \quad .$$

The average of E over samples is

$$\tilde{E} = 1 + \tilde{\bar{r}^2} + \tilde{\bar{x}^2} - 2\tilde{\bar{r}\bar{x}\bar{p}} + \tilde{\bar{r}^2\bar{p}^2}$$

Since according to sampling theory central second moment quantities such as $\bar{r}$ are independent of first moment quantities such as $\bar{x}$ and $\bar{p}$, this can be written as

$$\tilde{E} = 1 + \tilde{\bar{r}^2} + \tilde{\bar{x}^2} + \tilde{\bar{r}^2}\tilde{\bar{p}^2}$$

which, for a sample of size N, is given by

$$\tilde{E} = 1 + 1(N-1) + 1/N + 1/(N-1)N = (N+1)/(N-1)$$

$$= 1 + 2/N + \cdots$$

Thus instead of the hoped for reduction in variance by a factor $1 - 1/(N-1)$ it has been increased by a factor $[1 + 1/(N-1)][1 + 1/N]$. Note that the sampling errors in $\bar{r}$ and in $\bar{x}$ each contribute error of order $1/N$, but the sampling error in $\bar{p}$ contributes only in higher order.

We examine now the case of many predictors by considering the unit normal $(k+1)$-dimensional random vector $(x, p_1, p_2, \cdots, p_k)$ with the p_i serving as predictors based on sample regression coefficients. If we ignore the (false) sample correlations between the p_i's we may treat p_j

as a predictor of the residual after all earlier predictors p_i, $i < j$, have been used and would estimate a final reduction in variance as

$$[1 - 1(N-1)]^k = 1 - k/N + \cdots$$

Based on a given sample the regression formula would be

$$\hat{x} = \bar{x} + \sum_i \bar{r}_i(p_i - \bar{p}_i)$$

so that the estimation error would be

$$E = \langle x^2 \rangle - 2 \sum_i \bar{r}_i \langle x p_i \rangle + (\bar{x} - \sum_i \bar{r}_i \bar{p}_i)^2 + \sum_i \bar{r}_i{}^2 \langle p^2 \rangle$$

$$= 1 + \sum_i \bar{r}_i{}^2 + (\bar{x} - \sum_i \bar{r}_i \bar{p}_i)^2 \quad .$$

The average of E over samples becomes

$$\tilde{E} = 1 + \sum_i \tilde{\bar{r}}_i{}^2 + \tilde{\bar{x}}^2 - 2 \sum_i \tilde{\bar{x}\bar{r}}_i \bar{p}_i + \sum_i \tilde{\bar{r}}_i{}^2 \tilde{\bar{p}}_i{}^2$$

$$= 1 + k/(N-1) + 1/N + k/(N-1)N$$

$$= (1 + 1/N)[1 + k/(N-1)] = 1 + (k+1)/N + \cdots$$

The sampling error in $\bar{x}$ has contributed an error of order 1/N and the sampling error in the k regression coefficients r_i have contributed an error of order k/N.

The time lag correlations in meteorological time series are such that the number of effectively independent samples is about N = 16 per season or about N = 500 in 30 seasons which is roughly the length of relevant data sets. The number of allowable predictors is therefore of order 10 rather than 100. The limitation imposed on regression methods by the limited data base is thus far more severe than that imposed by the computer power needed to do the matrix manipulation involved.

Clearly the optimum number of predictors is determined by a balance between their true value in reducing variance and their contribution to increased variance through sampling errors and associated false regression.

8.2 Empirical Linear Modification of a Nonlinear Dynamics Equation

Any numerical prediction model provides a theoretical prescription for the evolution of a phase path in its phase space of state vectors $\underset{\sim}{u} = (u_1, u_2, \cdots, u_N)$ by giving at each point $\underset{\sim}{u}$ a rate of change $\dot{\underset{\sim}{u}}$. In hydrodynamic models the advection and pressure terms are quadratically nonlinear, and we shall consider here only a dynamics equation of the form

$$\dot{u}_\alpha = \sum_{\beta,\gamma} A_{\alpha\beta\gamma} u_\beta u_\gamma + \sum_\beta B_{\alpha\beta} u_\beta + C_\alpha .$$

The nonlinear interaction coefficients $A_{\alpha\beta\gamma}$ are determined from fundamental principles presumed relevant such as, in barotropic models of the atmosphere, the conservation of vorticity. If, as in the case of the atmosphere, a large amount of data is available on the observed evolution of the dynamical system, then we may determine the coefficients $B_{\alpha\beta}$ and C_α empirically.

In order to do this, let us first distinguish the model from reality by letting the observed variable w_α correspond to the model variable u_α. We treat the variables w_α as forming a stationary multivariate time series (the climate) for which we have determined by statistical analysis the mean $\langle w_\alpha \rangle = m_\alpha$ and the covariance $\langle [w_\alpha(t+\tau) - m_\alpha][w_\beta(t) - m_\beta] \rangle = V_{\alpha\beta}(\tau)$ as a function of lag time τ. We need specifically for our purposes $V_{\alpha\beta}(0)$ and $\dot{V}_{\alpha\beta}(0) = \dot{V}_{\alpha\beta}(\tau)]_{\tau=0} = \langle \dot{w}_\alpha(t)[w_\beta(t) - m_\beta] \rangle$.

For any observed values $w_\alpha(t)$ we may also evaluate the nonlinear term

$$q_\alpha(t) = \sum_{\beta,\gamma} A_{\alpha\beta\gamma} w_\beta(t) w_\gamma(t)$$

and determine its mean $\langle q_\alpha \rangle = n_\alpha$ and cross covariance with w_β namely $\langle [q_\alpha(t+\tau) - n_\alpha][w_\beta(t) - m_\beta] \rangle = T_{\alpha\beta}(\tau)$ which we need specifically at $\tau = 0$.

For the empirical determinations of $B_{\alpha\beta}$ and C_α we first impose the condition that the mean be stationary. By taking the first moment of the dynamics equation we have

$$0 = n_\alpha + \sum_\beta B_{\alpha\beta} m_\beta + C_\alpha$$

thus

$$C_\alpha = -n_\alpha - \sum_\beta B_{\alpha\beta} m_\beta \quad ,$$

and the dynamics equation may be rewritten as

$$\dot{u}_\alpha = \sum_{\beta,\gamma} A_{\alpha\beta\gamma} u_\beta u_\gamma - n_\alpha + \sum_\beta B_{\alpha\beta}(u_\beta - m_\beta) \quad .$$

In order to evaluate the matrix $B_{\alpha\beta}$ we next impose the condition that the zero lag cross covariance of $\dot{u}_\alpha$ and u_β be as observed for $\dot{w}_\alpha$ and w_β. Thus we substitute w_α for u_α in the dynamics equation which we then multiply by $w_\beta - m_\beta$ and average to get

$$\dot{V}_{\alpha\beta}(0) = T_{\alpha\beta}(0) + \sum_\gamma B_{\alpha\gamma} V_{\gamma\beta}(0)$$

or, in matrix form, $\underset{\sim}{\dot{V}} = \underset{\sim}{T} + \underset{\sim}{B}\underset{\sim}{V}$. Since the covariance matrix $\underset{\sim}{V}$ is positive definite and nonsingular we may then solve for $\underset{\sim}{B} = (\underset{\sim}{\dot{V}} - \underset{\sim}{T})\underset{\sim}{V}^{-1}$ in terms of the empirically determined matrices $\underset{\sim}{V}$, $\underset{\sim}{\dot{V}}$, and $\underset{\sim}{T}$.

An alternate derivation of a linear correction term to the dynamics equation can be based on linear regression. Without the linear correction term involving the matrix $\underset{\sim}{B}$, the dynamics equation is

$$\dot{u}_\alpha = q_\alpha - n_\alpha$$

with

$$\langle \dot{u}_\alpha \rangle = \langle q_\alpha \rangle - n_\alpha = 0 \quad .$$

We wish to estimate the difference

$$\dot{w}_\alpha - \dot{u}_\alpha \hat{=} \sum_\gamma B_{\alpha\gamma}(w_\gamma - m_\gamma)$$

between observed and predicted rate of change using the variable $(w_\gamma - m_\gamma)$ as a predictor. Note that we have

$$\langle \dot{w}_\alpha - \dot{u}_\alpha \rangle = 0 \ , \ \langle w_\gamma - m_\gamma \rangle = 0$$

and

$$\langle (\dot{w}_\alpha - \dot{u}_\alpha)(w_\gamma - m_\gamma) \rangle = \dot{V}_{\alpha\gamma} - T_{\alpha\gamma} \quad ,$$

$$\langle (w_\alpha - m_\alpha)(w_\gamma - m_\gamma) \rangle = V_{\alpha\gamma} \quad .$$

The regression matrix $\underset{\sim}{B}$ is given in the usual way by

$$\underset{\sim}{B} = (\dot{\underset{\sim}{V}} - \underset{\sim}{T})\underset{\sim}{V}^{-1} ,$$

but this is the matrix $\underset{\sim}{B}$ already derived by imposing the condition that

$$\langle \dot{u}_\alpha u_\beta \rangle = \langle \dot{w}_\alpha w_\beta \rangle .$$

We may also show that the linear correction term involving $\underset{\sim}{B}$ guarantees the stationarity of the second moment

$$W_{\alpha\beta}(t) = \langle [w_\alpha(t) - m_\alpha][w_\beta(t) - m_\beta] \rangle .$$

We have

$$\begin{aligned} \dot{W}_{\alpha\beta}(t) &= \langle \dot{w}_\alpha(t)[w_\beta(t) - m_\beta] \rangle + \langle [w_\alpha(t) - m_\alpha]\dot{w}_\beta(t) \rangle \\ &= \langle \dot{u}_\alpha(t)[w_\beta(t) - m_\beta] \rangle + \langle [w_\alpha(t) - m_\alpha]\dot{u}_\beta(t) \rangle \\ &= T_{\alpha\beta}(0) + \sum_\delta B_{\alpha\delta} V_{\delta\beta}(0) + T_{\beta\alpha}(0) + \sum_\delta B_{\beta\delta} V_{\delta\alpha}(0) = 0 . \end{aligned}$$

An important consequence is that the climate energy spectrum is preserved by the model.

8.3 External Error Simulation

Any stochastic-dynamic or Monte Carlo forecasting scheme which is based on the perfect model assumption will provide an estimate of increasing forecast error variance based solely on the internal growth of initial errors. Considering the large external errors of present models such perfect model estimates will be far too small. This would lead in turn to an unduly heavy weight being given to forecast information relative to that of new observations.

In order to remedy this situation it would seem useful to seek for a simulation of external error in terms of a random external forcing introduced into the forecast model or at least into the equations for the evolution of variance. Such forcing terms should, of course, be based on an empirical study of real forecast errors for a model.

An amusing possibility exists that if a more accurate evolution of second moment information leads to a more accurate evolution of the mean in accordance with Epstein's ideas then the introduction of random forcing would actually improve the mean forecast skill as well as provide better error growth information. This possibility requires careful experimental investigation.

If random forcing terms are to be introduced into a forecast model it will be necessary to adjust the nonrandom damping terms in order to assure the preservation of climatologically realistic second moments.

8.4 Nonlinear Dispersive Waves

The statistical hydrodynamical treatment of the prediction and predictability problem discussed in earlier sections has been largely based on experience with the simpler statistical hydrodynamics of turbulent flows. Much of the present theoretical work in turbulence deals with statistically homogeneous and isotropic situations, and the underlying Navier-Stokes equations involve no linear dispersion. The stochastic models say little or nothing about the frequently observed intermittent or local nature of turbulence nor about the possibility of local structure in the flow.

During recent years there has been an increasing mathematical interest in nonlinear wave phenomena which are observed and predicted to occur in simple systems in which both nonlinearity and linear dispersion play a role. In such systems can sometimes be found rather stable and completely predictable local structures called "solitons" in which the tendency for linear dispersion is balanced by nonlinear terms.

Rotational waves in the atmosphere are dispersive and the possibility that "Rossby solitons" might exist in the atmosphere was pointed out by Long (1964) and by Benney (1966). If they do then the whole analysis of the predictability of the atmosphere should be reexamined. Present numerical models should also be examined for numerical dispersion errors that might interfere with the required balance. Although there is still a large gap between the simplicity of systems showing soliton behavior and the complexity of the atmosphere, some of the qualitative behavior of the atmosphere is perhaps more naturally described in terms of local structures than in terms of harmonic components.

References

Baer, F., and F. N. Alyea, 1974: Predictability and spectral truncation. J. Atmos. Sci., 31, 920-931.

Benney, D. J., 1966: Long non-linear waves in fluid flows. J. Math. and Phys., 45, 5263.

Cramer, H., 1946: Mathematical Methods of Statistics. Princeton University Press, Section 29.7.

Eddy, A., 1967: Statistical objective analysis of scalar data fields. J. Appl. Meteor., 6, 597-609.

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

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

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.

Gandin, L. S., 1963: Objective Analysis of Meteorological Fields. Hydrometeor. Publ. House, Leningrad. Israel Program for Scientific Translations, 1965.

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.

Gleeson, T. A., 1970: Statistical-dynamical predictions. J. Appl. Meteor., 9, 333-344.

Jastrow, R., and M. Halem, 1970: Simulation studies related to GARP. Bull. Amer. Meteor. Soc., 51, 490-513.

Jones, R. H., 1965: Optimal estimation of initial conditions for numerical predictions. J. Atmos. Sci., 22, 658-663.

Kraichnan, R. H., 1967: Inertial ranges in two-dimensional turbulence. Phys. Fluids, 10, 1417-1423.

_____ 1970: Instability in fully developed turbulence. Phys. Fluids, 13, 569575.

_____ 1971a: An almost-Markovian Galilean-invariant turbulence model. J. Fluid Mech., 47, 513-524.

_____ 1971b: Inertial range transfer in two- and three-dimensional turbulence. J. Fluid Mech., 47, 525-535.

Leith, C. E., 1968: Diffusion approximation for two-dimensional turbulence. Phys. Fluids, 11, 671-673.

_____ 1971: Atmospheric predictability and two-dimensional turbulence. J. Atmos. Sci., 28, 145-161.

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

_____ and R. H. Kraichnan, 1972: Predictability of turbulent flows. J. Atmos. Sci., 29, 1041-1058.

Long, R. R., 1964: Solitary waves in the westerlies. J. Atmos. Sci., 21, 197-200.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Statistical Forecasting Project. Massachusetts Institute of Technology, Department of Meteorology.

_____ 1969a: Three approaches to atmospheric predictability. Bull. Amer. Meteor. Soc., 50, 345-349.

_____ 1969b: Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci., 26, 636-646.

_____ 1969c: The predictability of a flow which possesses many scales of motion. Tellus, 21, 289-307.

Miyakoda, K., R. F. Strickler, C. J. Nappo, P. L. Baker, and G. D. Hembree, 1971: The effect of horizontal grid resolution in an atmospheric circulation model. J. Atmos. Sci., 28, 481-499.

National Meteorological Center, 1970: Numerical Weather Prediction Activities, Second Half of 1969. U.S. Dept. of Commerce Publication.

Orszag, S. A., 1970: private communication (to appear in the Statistical Theory of Turbulence).

Petersen, D. P., 1968: On the concept and implementation of sequential analysis for linear random fields. Tellus, 20, 673-686.

Pitcher, E. J., 1974: Stochastic Dynamic Prediction using Atmospheric Data. Ph. D. Dissertation, University of Michigan.

Robinson, G. D., 1967: Some current projects for global meteorological observation and experiment. Quart. J. Roy. Meteor. Soc., 93, 409-418.

Rodgers, C. D., 1970: Remote sounding of the atmospheric temperature profile in the presence of cloud. Quart. J. Roy. Meteor. Soc., 96, 654-666.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. J. Atmos. Sci., 29, 809-815.

Smagorinsky, J., 1969: Problems and promises of deterministic extended range forecasting. Bull. Amer. Meteor. Soc., 50, 286-311.

Strand, O. N., and E. R. Westwater, 1968a: Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind. J. Assoc. Comp. Mach., 15, 100-115.

_____ and _____ 1968b: Minimum-rms estimation of the numerical solution of a Fredholm integral equation of the first kind. Siam J. Numerical Anal., 5, 287-295.

Teweles, S., and H. B. Wobus, 1954: Verification of prognostic charts. Bull. Amer. Meteor. Soc., 35, 455-463.

Westwater, E. R., and O. N. Strand, 1968: Statistical information content of radiation measurements used in indirect sensing. J. Atmos. Sci., 25, 750-758.

Williamson, D. L., 1973: The effect of forecast error accumulation on four-dimensional data assimilation. J. Atmos. Sci., 30, 537-543.

_____ and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating. J. Atmos. Sci., 28, 1313-1324.

Wiin-Nielsen, A., 1967: On the annual variation and spectral distribution of atmospheric energy. Tellus, 19, 540-559.