APPLICATIONS OF ESTIMATION THEORY TO NUMERICAL WEATHER PREDICTION

M. Ghil*, S. Cohn, J. Tavantzis**, K. Bube, and E. Isaacson

Courant Institute of Mathematical Sciences
New York University, New York, New York 10012

Abstract

Numerical weather prediction (NWP) is an initial-value problem for a system of nonlinear partial differential equations (PDEs) in which the initial values are known only incompletely and inaccurately. Data at initial time can be supplemented, however, by observations of the system distributed over a time interval preceding it. Estimation theory has been successful in approaching such problems for models governed by systems of ordinary differential equations and of linear PDEs. We develop methods of sequential estimation for NWP.

A model exhibiting many features of large-scale atmospheric flow important in NWP is the one governed by the shallow-fluid equations. We study first the estimation problem for a linearized version of these equations. The vector of observations corresponds to the different atmospheric quantities measured and space-time patterns associated with conventional and satellite-borne meteorological observing systems. A discrete Kalman-Bucy (K-B) filter is applied to a finite-difference version of the equations, which simulates the numerical models used in NWP.

* Also Laboratory for Atmospheric Sciences,
NASA Goddard Space Flight Center, Greenbelt, MD 20771

** Department of Mathematics, New Jersey Institute of Technology,
Newark, N.J. 07102

The specific character of the equations' dynamics gives rise to the necessity of modifying the usual K-B filter. The modification consists in eliminating the high-frequency inertia-gravity waves which would otherwise be generated by the insertion of observational data. The modified filtering procedure developed here combines in an optimal way dynamic initialization (i.e., elimination of fast waves) and four-dimensional (space-time) assimilation of observational data, two procedures which traditionally have been carried out separately in NWP. Comparisons between the modified filter and the standard K-B filter have been made.

The matrix of weighting coefficients, or filter, applied to the observational corrections of state variables converges rapidly to an asymptotic, constant matrix. Using realistic values of observational noise and system noise, this convergence has been shown to occur in numerical experiments with the linear system studied; it has also been analyzed theoretically in a simplified, scalar case. The relatively rapid convergence of the filter in our simulations leads us to expect that the filter will be efficiently computable for operational NWP models and real observation patterns.

Our program calls for the study of the asymptotic filter's dependence on observation patterns, noise levels, and the system's dynamics. Furthermore, the covariance matrices of system noise and observational noise will be determined from the data themselves in the process of sequential estimation, rather than be assigned predetermined, heuristic values. Finally, the estimation procedure will be extended to the full, nonlinear shallow-fluid equations.

CONTENTS

1. Introduction
2. A Review of the State-Space Approach to Estimation
   2.1 Statistical considerations in estimation: a simple illustration
   2.2 Estimation for stochastic-dynamic systems
   2.3 General remarks on the Kalman-Bucy (K-B) filter
3. Estimation for the Shallow-Water Equations
   3.1 The equations
   3.2 Discretization
   3.3 The modified K-B filter
   3.4 Observational pattern and choice of parameters
   3.5 Previous work
4. Results
   4.1 Estimation for a perfect, noise-free model
   4.2 Estimation in the presence of system noise
   4.3 Theoretical analysis of the scalar case
5. Concluding Remarks

References

Appendix A. List of Major Symbols

1. INTRODUCTION

One of the main reasons we cannot tell what the weather will be tomorrow is that we do not know what the weather is today. In other words, numerical weather prediction (NWP) is an initial-value problem for which initial data are not available in sufficient quantity and with sufficient accuracy.

Numerical forecasts are produced now routinely on a daily basis by a number of weather services in different countries. The models used in NWP are discretized versions of the partial differential equations (PDEs) governing large-scale atmospheric flow. The discretization is performed by finite differencing, finite element or spectral representations. The number of degrees of freedom of the discretized models is typically of the order of 10^5-10^6. The spatial domain of the models is the entire globe or at least an entire hemisphere.

A large number of observations is made by the conventional, ground-based meteorological network, coordinated by the World Weather Watch (WWW). They consist of point values of temperature, humidity, pressure and horizontal velocity. These observations, of the order of 10^5 in number, are produced at the so-called synoptic times, 0000 GMT and 1200 GMT. It is customary therefore to choose a synoptic time as initial time for a numerical forecast. Conventional observations are insufficient in number in order to determine

the initial state of the model atmosphere. Furthermore, they are very unevenly distributed in space, being concentrated over the continents of the Northern Hemisphere, and much sparser over the oceans and over the Southern Hemisphere (Fig. 1).

A number of additional observations are made at the so-called subsynoptic times, 0600 GMT and 1800 GMT. A still larger number of observations, exceeding by now that given by the WWW network, is gathered in an essentially time-continuous manner from polar-orbiting satellites and other non-conventional measuring platforms (Fleming et al., 1979 a,b, and references therein). All these observations together form a rather bewildering array by their uneven distribution in space and time, as well as by their different error characteristics.

In order to obtain the best possible estimate of the model state at the initial instant, NWP centers use the data available over a time interval preceding that instant. The most common procedure to use such data, called updating, was suggested by Charney et al. (1969). The model is provided the best available data at some preceding instant, e.g., 24 h or 48 h earlier, and is integrated forward in time. Additional data replace the model values as they become available: the model is updated. When the model integration reaches the initial instant for the next scheduled forecast, its guess of the initial state is blended with the data available at that instant to produce the desired estimate. Thus the model itself is used to assimilate the data

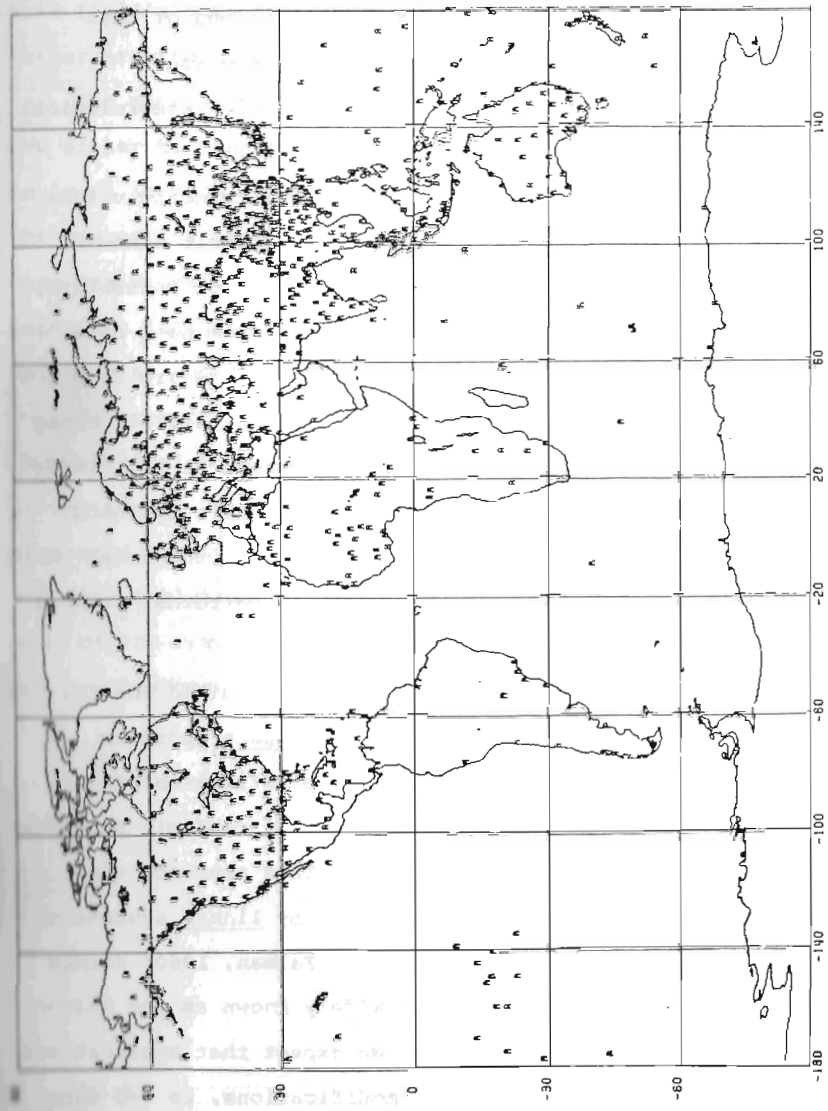


Fig. 1 The global network of ground-launched radiosondes, as of 19 February 1976. This conventional network of upper-air measurements is today supplemented by numerous airborne and space-borne observing systems.

available up to the initial instant. Variations of this procedure, as well as other procedures for the four-dimensional (4-D), space-time assimilation of meteorological observations are reviewed in Bengtsson (1975).

The blending of observations and model forecast values has been made most recently using in an explicit manner the error structure of the data (Phillips, 1976; Rutherford, 1972). This error structure is determined from past data and the resulting linear regression coefficients are computed once and for all, their constant values being used all along the assimilation cycle of the model (McPherson et al., 1979; Schlatter et al., 1976). A modification of this approach, which combines more intimately the dynamics of the model with the time-continuously changing observation patterns, appears in Ghil et al. (1979).

Clearly, a mathematical framework well suited for the 4-D assimilation problem of NWP is the state-space approach of estimation theory. It deals with the estimation of stochastic processes which are generated by randomly perturbed differential or difference equations. This approach was first formulated for processes governed by linear systems of ordinary differential equations (ODEs: Kalman, 1960; Kalman and Bucy, 1961); its results are widely known as the Kalman or the Kalman-Bucy (K-B) filter. We expect that applications of the K-B filter, with suitable modifications, to 4-D data assimilation will provide additional physical insight into

this field's outstanding problems and eventually lead to better practical algorithms for solving them.

The purpose of the present article is to apply the K-B filter to the linearized shallow-water equations and to learn as much as possible from this application about the properties of the filter relevant to operational 4-D data assimilation. This linear system already contains important features of the equations used in operational NWP models, and our application should be instructive.

We present a brief review of the K-B filter in Section 2. The dynamical model and observing pattern studied are presented in Section 3, along with the modification of the K-B filter suggested by the system's dynamics. Numerical results and some analytical ones follow in Section 4. A discussion of the results, comparison with operational practice and conclusions follow in Section 5.

2. A REVIEW OF THE STATE-SPACE APPROACH TO ESTIMATION

The intent of this section is to familiarize the reader with the ideas and methods of sequential estimation. Systematic, more or less rigorous expositions of the theory are in existence for the interested reader (Bucy and Joseph, 1968; Curtain and Pritchard, 1978; Davis, 1977; Gelb, 1974; Jazwinski, 1970). Here we shall stay on the purely formal, and hopefully intuitive, level.

2.1 Statistical Considerations in Estimation: A Simple Illustration

Given a quantity x, suppose that two independent measurements of this quantity, x_1 and x_2, are available. For instance x could be the temperature in a room and x_1 and x_2 the readings of two thermometers placed in the room. In the absence of any additional information about x, it is natural to seek an estimate of x, $\hat{x}$ say, as a linear function of x_1 and x_2,

$$\hat{x} = \alpha_1 x_1 + \alpha_2 x_2 . \qquad (2.1a)$$

The function itself is called an estimator; the estimate is its value.

We wish to determine α_1 and α_2 so that the estimate $\hat{x}$ will be optimal in some sense. The conditions to achieve such optimality depend on the nature of the measurements.

Let us assume first of all that there are no systematic errors in the measurements x_1 and x_2, i.e. that if we repeat our measurements many times then their average would equal the true value x. In terms of the expectation operator E, this is written as

$$E(x_1 - x) = E(x_2 - x) = 0 ,$$

and we say the measurements are unbiased. It is natural to require that the estimate $\hat{x}$ also be unbiased:

$$E(\hat{x} - x) = 0 ;$$

this requirement is equivalent to

$$\alpha_1 + \alpha_2 = 1, \qquad (2.1b)$$

so that (2.1a) becomes

$$\hat{x} = x_1 + \alpha_2(x_2 - x_1) . \qquad (2.1c)$$

Next it is assumed that the measurement errors $x_1 - x$ and $x_2 - x$ are uncorrelated:

$$E(x_1 - x)(x_2 - x) = 0 ,$$

and that their variances σ_1^2 and σ_2^2,

$$\sigma_1^2 = E(x_1 - x)^2 , \quad \sigma_2^2 = E(x_2 - x)^2$$

are known from previous measurements, viz. from instrument calibration. Then the variance of the estimation error, $\hat{\sigma}^2 = E(\hat{x} - x)^2$, is given by

$$\hat{\sigma}^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 . \qquad (2.2)$$

Suppose for the moment that in addition to satisfying Eq. (2.1b), α_1 and α_2 are nonnegative but otherwise arbitrary; the linear combination in Eq. (2.1a) is then said to be convex. Convexity will imply that $\hat{x}$ always lies between x_1 and x_2, and furthermore that

$$\hat{\sigma}^2 \leq \max \{\sigma_1^2, \sigma_2^2\} . \qquad (2.3a)$$

While property (2.3a) is reassuring, one should be able to do better. We expect to be able to achieve

$$\hat{\sigma}^2 \leq \min \{\sigma_1^2, \sigma_2^2\} \; ; \qquad (2.3b)$$

otherwise there would be no point in making more than one measurement. The assumed knowledge of σ_1^2 and σ_2^2 will in fact yield (2.3b); it is accomplished by our optimality requirement.

This requirement can now be formulated precisely: $\hat{x}$ should be a minimum variance estimate, which in our case means that $\hat{\sigma}^2$ in Eq. (2.2) should be minimized with respect to α_1 and α_2, subject to (2.1b). The resulting optimal weights are

$$\alpha_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\hat{\sigma}^2}{\sigma_1^2} \qquad (2.4a)$$

and

$$\alpha_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{\hat{\sigma}^2}{\sigma_2^2} \; , \qquad (2.4b)$$

where the optimal error variance $\hat{\sigma}$ is given by

$$\hat{\sigma}^{-2} = \sigma_1^{-2} + \sigma_2^{-2} \; . \qquad (2.4c)$$

Notice that α_1 and α_2 are nonnegative, so that the optimal estimator,

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2 \; ,$$

is, in fact, convex. Moreover, the optimal weights (2.4a,b) satisfy the intuitive requirement that they should reflect our relative confidence in x_1 and x_2 : if σ_1 is smaller than σ_2, for example, then x_1 is weighted more heavily.

Notice also that formula (2.4c) for the optimal error variance immediately implies property (2.3b). In particular, when $\sigma_1 = \sigma_2 = \sigma$, one obtains $\hat{\sigma} = \sigma/\sqrt{2}$, which generalizes to $\hat{\sigma} = \sigma/\sqrt{N}$ for N independent measurements of equal variance.

2.2 Estimation for Stochastic-Dynamic Systems

The purpose of sequential estimation theory for dynamic systems is to extend the simple ideas outlined above to the case in which the quantity x of interest evolves in time according to a given (system of ordinary or partial) differential or difference equation(s). In this case, x_1 will represent the state of the system as determined from previous observations (measurements), while x_2 represents observations at the current time.

To stress the analogy, let us consider here a system of randomly perturbed difference equations for the state vector $\underset{\sim}{x}$:

$$\underset{\sim}{x}_{k+1} = \Psi \underset{\sim}{x}_k + \underset{\sim}{\xi}_k \; , \quad k = 0,1,2,\ldots \; ; \qquad (2.5a)$$

$\underset{\sim}{x}$ and $\underset{\sim}{\xi}$ have dimension n, and Ψ is a constant n×n matrix. (See Appendix A for a list of recurring symbols.) In our application, $\underset{\sim}{x}_k$ will stand for the meteorological variables at time k at the grid points of a global atmospheric prediction model; Ψ stands for the finite-difference operator which advances the variables by one time step. The random

vector sequence $\{\underset{\sim}{\xi}_k:\ k = 0,1,2,\ldots\}$ is assumed to be a white noise sequence with mean zero and covariance matrix Q,

$$E\underset{\sim}{\xi}_k = 0 \ , \quad E\underset{\sim}{\xi}_k\underset{\sim}{\xi}_\ell^T = Q\delta_{k\ell} \ . \tag{2.5b,c}$$

The transpose of a vector or matrix $(\ \ )$ is indicated by $(\ \ )^T$ and $\delta_{k\ell}$ is the Kronecker delta, $\delta_{k\ell} = 0$ if $k \neq \ell$ and $\delta_{k\ell} = 1$ if $k = \ell$. The white noise $\underset{\sim}{\xi}$ represents dynamical and physical processes not described by the model Ψ, especially smaller-scale phenomena not resolved by the grid.

Suppose for the moment that an initial unbiased estimate $\hat{\underset{\sim}{x}}_0 = E\underset{\sim}{x}_0$ is available from observations at time zero, and that no further observations are available at later times. In this case the best estimate of $\underset{\sim}{x}_k$, $\hat{\underset{\sim}{x}}_k$ say, would be the mean of $\underset{\sim}{x}_k$, $\hat{\underset{\sim}{x}}_k = E\underset{\sim}{x}_k$, and is computed according to (2.5), by the recursion

$$\hat{\underset{\sim}{x}}_{k+1} = \Psi\hat{\underset{\sim}{x}}_k \ , \quad \hat{\underset{\sim}{x}}_0 = E\underset{\sim}{x}_0 \ ;$$

this estimate can be improved if further observations become available.

We have seen in the introduction that the description of the atmospheric state at a given synoptic time from observations at that time is entirely inadequate and that we are interested in using observations at other times as well. This situation is modeled by assuming a continuous stream of observations

$$\underset{\sim}{z}_k = H\underset{\sim}{x}_k + \underset{\sim}{\zeta}_k \ , \quad k = 1,2,3,\ldots \ ; \tag{2.6a}$$

$\underset{\sim}{\zeta}_k$ models observational errors and is also assumed to be a white noise sequence,

$$E\underset{\sim}{\zeta}_k = 0, \quad E\underset{\sim}{\zeta}_k\underset{\sim}{\zeta}_\ell^T = R\delta_{k\ell} \ , \tag{2.6b,c}$$

uncorrelated with $\underset{\sim}{\xi}$,

$$E\underset{\sim}{\zeta}_k\underset{\sim}{\xi}_\ell^T = 0 \ . \tag{2.6d}$$

The zero-mean assumptions (2.5b, 2.6b) are made for convenience only; they are not essential for the theory we describe here.

The observation vectors $\underset{\sim}{z}_k$ have dimension $p \leq n$, and H is a $p \times n$ matrix. This formulation describes in particular the meteorological situation in which the observations at any time k are incomplete; measurements are made over some areas (continents) and not over others (oceans); at a given location some variables are measured and not others, e.g., geostationary satellites determine winds but not temperatures. Finally, some functional of the variables may be observed rather than a variable itself, e.g., polar-orbiting satellites measure radiances, which depend on vertical temperature profiles. Thus the entries of H need not be only zero or one, i.e., H is not necessarily a permutation matrix.

In fact, H, as well as Ψ, Q, and R, need not be constant in time; they are only taken constant here for simplicity. In particular, the rank of H, i.e., the effective dimension of $\underset{\sim}{z}_k$, can change from one time to the next: the largest number of observations are available at synoptic times, with fewer observations provided at subsynoptic times, and even fewer in between, at intermediate times.

The linearity assumption that Ψ and H are independent of $\underset{\sim}{x}$ is, however, important for the theory we shall use here. This assumption is essential in guaranteeing the optimality of the filtering algorithm (2.11) below. Extensions to the nonlinear case will be discussed in Sec. 5.

Having described our stochastic-dynamical model (Ψ,H),

Eqs. (2.5, 2.6), we are now in a position to make the connection with (2.1-2.4). Given an estimate $\hat{\underset{\sim}{x}}_k(+)$ based on all the observations up to and including the time k, the best prediction at time k+1, $\hat{\underset{\sim}{x}}_{k+1}(-)$, is simply

$$\hat{\underset{\sim}{x}}_{k+1}(-) = \Psi\hat{\underset{\sim}{x}}_k(+) \ ; \qquad (2.7)$$

$\hat{\underset{\sim}{x}}_{k+1}(-)$ will be the analogue of x_1 in Section 2.1. The analogue of x_2 is the actual observation $\underset{\sim}{z}_{k+1}$ at time k+1. We wish to combine $\hat{\underset{\sim}{x}}_{k+1}(-)$ and $\underset{\sim}{z}_{k+1}$ in order to obtain an estimate $\hat{\underset{\sim}{x}}_{k+1}(+)$, which we require to be: (a) linear, (b) unbiased, and (c) optimal in some suitable sense.

In Sec. 2.1, we discussed the simple illustrative example of estimating the room temperature x from the readings x_1 and x_2 of two thermometers. In that situation, the requirements (a) and (b) lead to formulae (2.1). The optimality condition was expressed in (2.4), which yields the minimum $\hat{\sigma}$ among all linear, unbiased estimators.

For our dynamic system (Ψ,H), requirement (a) leads to the formula

$$\hat{\underset{\sim}{x}}_{k+1}(+) = L_{k+1}\hat{\underset{\sim}{x}}_{k+1}(-) + K_{k+1}\underset{\sim}{z}_{k+1} \ , \qquad (2.8a)$$

which is analogous to (2.1a), while requirement (b) leads to

$$L_{k+1} = I - K_{k+1}H \ , \qquad (2.8b)$$

which is analogous to (2.1b). Our estimator is therefore of the form

$$\hat{\underset{\sim}{x}}_{k+1}(+) = \hat{\underset{\sim}{x}}_{k+1}(-) + K_{k+1}(\underset{\sim}{z}_{k+1} - H\hat{\underset{\sim}{x}}_{k+1}(-)) \ ; \qquad (2.9)$$

the analogy with (2.1c) is obvious. It remains for the gain matrix K_{k+1} to reflect the relative uncertainties in $\underset{\sim}{z}_{k+1}$ and $\hat{\underset{\sim}{x}}_{k+1}(-)$. This will be achieved by imposing the optimality requirement (c), and will yield formulae analogous to (2.4).

In order to formulate the optimality criterion, we define first the estimation error covariance matrices

$$P_{k+1}(\pm) = E(\hat{\underset{\sim}{x}}_{k+1}(\pm) - \underset{\sim}{x}_{k+1})(\hat{\underset{\sim}{x}}_{k+1}(\pm) - \underset{\sim}{x}_{k+1})^T \ ;$$

$P_{k+1}(-)$ is the counterpart of σ_1^2 and $P_{k+1}(+)$ is that of $\hat{\sigma}^2$. Using Eqs. (2.5) and (2.7), one finds that $P_k(+)$ is advanced by one time step to $P_{k+1}(-)$ according to

$$P_{k+1}(-) = \Psi P_k(+)\Psi^T + Q \ ; \qquad (2.10a)$$

the derivation of (2.10a) depends on the fact that

$$E\underset{\sim}{\xi}_k(\hat{\underset{\sim}{x}}_k - \underset{\sim}{x}_k)^T = 0$$

(cf. (2.5c)). Eqs. (2.6) and (2.9) imply that, in the presence of observations $\underset{\sim}{z}_{k+1}$, $P_{k+1}(+)$ is found from $P_{k+1}(-)$ by the formula

$$P_{k+1}(+) = (I - K_{k+1}H)P_{k+1}(-)(I - K_{k+1}H)^T + K_{k+1}RK_{k+1}^T \ ; \qquad (2.10b)$$

in the total absence of observations at time k+1, we have instead $\hat{\underset{\sim}{x}}_{k+1}(+) = \hat{\underset{\sim}{x}}_{k+1}(-)$ and $P_{k+1}(+) = P_{k+1}(-)$.

The estimation error covariance matrices explicitly contain all relevant information about the error structure of the current estimate.

The statistics of all errors committed up to and including time k are accumulated in $P_k(+)$. Formula (2.10a) shows how this information is advanced to the next time step. For example, this equation determines how the presumably small errors committed over a continent, or other data-rich region, propagate over an ocean, or other data-sparse region. Equation (2.10b) then determines precisely the extent to which the estimate is improved by the new observations.

We have defined the estimation error covariance matrices and considered their changes in time. We are ready now to derive the optimal gain matrix by imposing the optimality requirement (c): it is required that $\hat{\underset{\sim}{x}}_{k+1}(+)$ be a minimum variance estimate in the sense that

$$J = E(\hat{\underset{\sim}{x}}_{k+1}(+) - \underset{\sim}{x}_{k+1})^T S(\hat{\underset{\sim}{x}}_{k+1}(+) - \underset{\sim}{x}_{k+1})$$

be minimized with respect to each element of K_{k+1}, for all symmetric, positive definite matrices S. In particular, for S = I, we see from the definition of $P_{k+1}(+)$ that the trace, or sum of the diagonal elements, of $P_{k+1}(+)$ is to be minimized. The trace of $P_{k+1}(+)$ is the expected mean-square (m-s) estimation error.

We recall that a symmetric matrix S, $S^T = S$, is positive definite if, for any vector $\underset{\sim}{x} \neq 0$, the scalar product $\underset{\sim}{x}^T S \underset{\sim}{x} > 0$. Since every such matrix has a factorization $S = C^T C$, one finds that in general

$$J = \text{trace } C P_{k+1}(+) C^T .$$

Using $P_{k+1}(+)$ from Eq. (2.10b) and setting the derivative of J with respect to each element of K_{k+1} equal to zero, one finds that a unique, absolute minimum is attained at

$$K_{k+1} = P_{k+1}(-)\, H^T (H P_{k+1}(-) H^T + R)^{-1} .$$

This result is valid independently of C, and hence is the same for all possible positive definite matrices S. In other words, K_{k+1} above minimizes simultaneously all reasonable measures, or norms, of the expected estimation error.

The formula above gives the so-called Kalman-Bucy (K-B) optimal gain matrix, or filter. Substituting this into (2.10b) yields the optimal error covariance matrix

$$P_{k+1}(+) = (I - K_{k+1} H)\, P_{k+1}(-) .$$

Assuming the availability of an unbiased initial estimate

$$\hat{\underset{\sim}{x}}_0 = \hat{\underset{\sim}{x}}_0(+) = E\underset{\sim}{x}_0$$

and an initial estimation error covariance matrix

$$P_0 = P_0(+) = E(\hat{\underset{\sim}{x}}_0 - \underset{\sim}{x}_0)(\hat{\underset{\sim}{x}}_0 - \underset{\sim}{x}_0)^T ,$$

the description of the Kalman filtering algorithm is now complete: for k = 0,1,2,..., one computes in order

$$\hat{\underset{\sim}{x}}_{k+1}(-) = \Psi \hat{\underset{\sim}{x}}_k(+) , \qquad (2.11a)$$

$$P_{k+1}(-) = \Psi P_k(+) \Psi^T + Q , \qquad (2.11b)$$

$$K_{k+1} = P_{k+1}(-)\, H^T (H P_{k+1}(-) H^T + R)^{-1} , \qquad (2.11c)$$

$$P_{k+1}(+) = (I - K_{k+1}H)\, P_{k+1}(-) , \qquad (2.11d)$$

$$\hat{x}_{k+1}(+) = \hat{x}_{k+1}(-) + K_{k+1}(z_{k+1} - H\hat{x}_{k+1}(-)) . \qquad (2.11e)$$

In the absence of observations at time k+1, Eqs. (2.11c,d,e) are replaced by

$$K_{k+1} = 0 , \qquad (2.11c')$$

$$P_{k+1}(+) = P_{k+1}(-) , \qquad (2.11d')$$

$$\hat{x}_{k+1}(+) = \hat{x}_{k+1}(-) . \qquad (2.11e')$$

Actually, the gain matrix sequence $\{K_k : k = 1,2,3,\ldots\}$ may be precomputed once and for all, i.e., for all realizations of the state and noise processes, x_k , ξ_k, ζ_k. Indeed, (2.11b,c,d) do not depend on the estimates (2.11a,e). This is a result of the assumed linearity of the model (Ψ,H).

To complete the analogy with our earlier illustrative example, notice that Eqs. (2.11c,d) can be rewritten as

$$P_{k+1}^{-1}(+) = P_{k+1}^{-1}(-) + H^T R^{-1} H , \qquad (2.12a)$$

$$K_{k+1} = P_{k+1}(+) H^T R^{-1} \qquad (2.12b)$$

In our analogy, $(P_{k+1}(-), R, P_{k+1}(+), K_{k+1})$ correspond to $(\sigma_1^2, \sigma_2^2, \hat{\sigma}^2, \alpha_2)$; Eqs. (2.12a,b) are analogous to (2.4c) and (2.4b), respectively.

Some intuitively appealing results follow, as in the simple example. For instance, Eq. (2.12a) implies that

$$P_{k+1}(+) \leq P_{k+1}(-) ;$$

the matrix inequality $A \leq B$ means that $C = B - A$ is nonnegative semidefinite, $x^T C x \geq 0$ for all x. Eq. (2.12b) implies that if R is small (large) then the observations z_{k+1}

are weighted more (less) heavily than the predictions $\hat{x}_{k+1}(-)$.

2.3 General Remarks on the K-B Filter.

Before proceeding with the description of our dynamical and observational model (Ψ,H), a number of theoretical remarks are in order. Recall first that $\hat{x}_{k+1}(+)$ was chosen to be the optimal, minimum variance, unbiased estimate among all estimators of the form (2.8a). It is not clear that in computing $\hat{x}_{k+1}(+)$ all past observations z_j , $j = 1,2,\ldots,k$, have been fully utilized. It can be shown, however (e.g. Jazwinski, 1970, Sec. 7.3), that our estimate is in fact the optimum unbiased estimate among all estimators which are linear combinations of all the available data

$$\hat{x}_{k+1}(+) = A_0 \hat{x}_0 + \sum_{j=1}^{k+1} A_j z_j . \qquad (2.13)$$

This wider optimality is due to assumptions (2.5c, 2.6c,d) that system errors and observational errors are uncorrelated in time. It can sometimes still be achieved without these assumptions (Jazwinski, 1970, Sec. 7.3, Examples 7.5-7.7). Carrying the estimation error covariance matrices along in the computation makes it possible for the filtering algorithm to be sequential, or recursive: each observation is discarded as soon as it is processed. This sequential nature of the estimation makes the algorithm conceptually simple, as well as having great practical advantages. It is one of the major reasons for the broad applicability of Kalman filtering.

Another important feature of the filtering algorithm (2.11) is the fact that only first-order statistics, i.e., means, and second-order statistics, i.e., covariance matrices, of the

vector processes of interest are involved. In other words, it suffices to know these statistics for the system noise $\underset{\sim}{\xi}$, observational noise $\underset{\sim}{\zeta}$, and initial error $\hat{\underset{\sim}{x}}_0 - \underset{\sim}{x}_0$. These will provide the estimate $\hat{\underset{\sim}{x}}$ as well as its error covariance P at all future times. This property is important since in practice it is usually difficult to obtain even this much information about the random errors one wishes to filter; it is well nigh impossible to obtain more, i.e., to prescribe higher-order statistics. Moreover, the first-order and second-order statistics of the error processes can be determined adaptively, i.e., by the filtering algorithm itself (Chin, 1979, and references therein).

Gaussian processes are in fact completely determined by their first and second-order statistics. Furthermore, the Central Limit Theorem (e.g. Parzen, 1960, Sec. 8.5 and 10.4) states, in its various forms, that the superposition of a large number of random effects is approximately Gaussian, regardless of the distribution of the individual effects. It is reasonable, therefore, to expect our errors, which come from a large number of sources, to be approximately Gaussian. It follows that retention of only first and second-order statistics in the filtering algorithm (2.11) should be rather satisfactory.

Actually, when $\underset{\sim}{\xi}$, $\underset{\sim}{\zeta}$, and $\underset{\sim}{x}_0$ are Gaussian, it is known (e.g. Jazwinski, 1970, Sec. 5.2) that the best possible nonlinear estimate of $\underset{\sim}{x}$, i.e., one which might depend nonlinearly on all the observations, is still our linear estimate $\hat{\underset{\sim}{x}}$.

The sequential nature of the filter implies in particular that, in the absence of further observations at times $k \geq N$, the best prediction is simply given by (2.11a, 2.11e'):

$$\hat{\underset{\sim}{x}}_{k+1} = \Psi \hat{\underset{\sim}{x}}_k , \qquad k = N, N+1, N+2, \ldots, \qquad (2.14a)$$

$$\hat{\underset{\sim}{x}}_N = \hat{\underset{\sim}{x}}_N(+) . \qquad (2.14b)$$

The covariance of this predictor is then given by (2.11b, 2.11d'). This corresponds roughly to what is done in operational practice: an initial state $\hat{\underset{\sim}{x}}_N$ is determined by 4-D data assimilation from all observations up to and including the synoptic time of interest. Then the forecast model is integrated forward in time from the initial state obtained, without further use of the data. We intend therefore to study only the assimilation and initialization problem, over an assimilation interval $k = 0,1,2,\ldots,N$.

The framework of estimation theory can provide also insight into the nature of the system noise $(\underset{\sim}{\xi},Q)$ and ways to its determination. It could lead to improvements in modeling and hence forecasting, by helping to pinpoint deterministic components of $\underset{\sim}{\xi}$. This, however, is not our purpose here. With these remarks, we turn to the description of the dynamic system to which the theory outlined in this section will be applied.

3. ESTIMATION FOR THE SHALLOW-WATER EQUATIONS

3.1 The Equations

The shallow-water equations are a simple system whose solutions exhibit some of the properties of large-scale atmospheric flow. They have certain important characteristics in common with the more complicated, three-dimensional systems currently used in NWP models.

We shall study here a linear, spatially one-dimensional version of the equations, written in cartesian coordinates for a plane tangent to the Earth at latitude θ_0:

$$u_t + Uu_x + \phi_x - fv = 0 \ , \qquad (3.1a)$$

$$v_t + Uv_x + fu = 0 \ , \qquad (3.1b)$$

$$\phi_t + U\phi_x + \Phi u_x - fUv = 0 \ . \qquad (3.1c)$$

The coordinate x points eastward, in the zonal direction, along the circle of latitude $\theta = \theta_0$, while y points northward, in the meridional direction; u and v are velocity components in the x and y directions, $f = 2\Omega \sin\theta_0$ is the Coriolis parameter, with Ω the angular velocity of the Earth. The geopotential $\phi = gh$ measures the deviation of the height $h + H$ of the free surface from its equilibrium value H, with $\Phi = gH$, U is a constant zonal mean flow velocity. All quantities are independent of y.

These equations are derived from the full, nonlinear shallow-water equations on a tangent plane,

$$u_t + uu_x + vu_y + \phi_x - fv = 0 \ , \qquad (3.2a)$$

$$v_t + uv_x + vv_y + \phi_y + fu = 0 \ , \qquad (3.2b)$$

$$\phi_t + u\phi_x + v\phi_y + \phi(u_x+v_y) = 0 \ , \qquad (3.2c)$$

by linearization around a solution ($u = U$, $v = 0$, $\phi = \Phi$) satisfying $fU = -\phi_y = \text{const}$. We assume in this derivation that the perturbation quantities, i.e., the deviations from equilibrium values of (u,v,ϕ), do not depend on y.

It is advantageous to work with a constant-coefficient system, as long as the basic phenomena of interest are not obscured by this simplification. Hence f, U and Φ in (3.1) are taken to be constants. The variation with latitude of the Coriolis parameter f, however, has an important effect on planetary flows. This so-called β-effect, $\beta = f_y(\theta_0) \neq 0$, is equivalent to the effect of bottom topography, $\phi_y \neq 0$, in the tangent-plane approximation (Pedlosky, 1979, Sec. 3.17 and Ch. 6). The term $(-fUv)$ in (3.1c) introduces this effect into the solutions of the system considered, without sacrificing the simplicity of constant coefficients.

All the solutions $\underset{\sim}{W} = (u,v,\phi)$ of (3.1) can be expressed as a superposition of plane waves:

$$\underset{\sim}{W}(x,t) = \sum_{\ell} \underset{\sim}{w}_{\ell} \, e^{i\ell(x-c_\ell t)} \ , \qquad (3.3a)$$

where ℓ is the wave number. For each ℓ, the speed of the individual waves, c_ℓ, is given by the dispersion relation

$$\ell^2(U - c_\ell)\{\Phi - (U - c_\ell)^2\} - f^2 c_\ell = 0 \ . \qquad (3.3b)$$

This is a cubic equation for c_ℓ , which has three <u>real</u> roots, $c_\ell^{(m)}$, m = 1,2,3. It turns out that (3.1) has two types of solutions: slow waves, corresponding to

$$c_\ell^{(1)} \cong U - \frac{f^2}{\ell^2\Phi + f^2} U \ , \qquad (3.4a)$$

and fast waves, with

$$c_\ell^{(2,3)} \cong U \pm \sqrt{\Phi + f^2/\ell^2} + \frac{f^2}{2(\ell^2\Phi + f^2)} U \ . \qquad (3.4b)$$

In the absence of the β-like term (-fUv) in (3.1c), the last term in (3.4a), as well as in (3.4b), would not be present. The expressions in (3.4) are actually exact to first order in the small nondimensional parameter $U/\sqrt{\Phi}$.

The slow waves are the meteorologically important ones, which correspond to the slow traveling of planetary waves on which synoptic weather systems are superimposed. Their speed is comparable to that of the mean zonal current, U = O(10 m/sec), and they <u>retrogress</u>: their propagation relative to the mean flow is westward. These slow, retrogressing waves are named after <u>Rossby</u>, and they are an important feature of mid-latitude atmospheric dynamics. Their frequency is always smaller than $f = O(10^{-4}\text{sec}^{-1})$.

The speed of the fast waves is dominated by the second term in (3.4b), being $O(10^2\text{m/sec})$. They are called <u>inertia-gravity waves</u>, since they are the familiar gravity waves of shallow-water theory, for which $c_\ell = U \pm \sqrt{\Phi}$,

modified by the presence of the Coriolis force. Their dispersion and dissipation plays a role in the mechanism of geostrophic adjustment, which maintains the atmosphere in a <u>quasi-geostrophic</u> state, in which the Coriolis force nearly balances the pressure-gradient force. But they carry very little energy at any given time, and appear mostly as higher frequency oscillations superimposed on the meteorologically significant ones, i.e., as <u>meteorological noise</u>.

An important problem in NWP is the <u>filtering</u> of the fast waves in large-scale numerical forecasts, in order to prevent their spurious growth to amplitudes larger than those found in the atmosphere. The need for this filtering is different from that discussed in the previous section: it stems from the two time scales of the deterministic motion itself, rather than from the presence of extraneous, random noise perturbing the deterministic motion. In particular, the fast waves can be eliminated or reduced at the initial time of the forecast by an <u>initialization</u> procedure (Bengtsson, 1975; Leith, 1980; and references therein). Once such an initialization has been performed, the fast waves will not grow excessively over periods of time comparable to the evolution time of the slow waves (Browning <u>et al</u>., 1979; Bube and Ghil, 1980).

Clearly, the two problems of: a) determining a solution to the forecast equations from a continuous stream of noisy data (4-D data assimilation) and b) rendering this solution as free of fast waves as possible (initialization) are related. The connection was stressed and a first step in the direction of their joint solution taken in Ghil (1980) and Leith (1980), among others.

We shall present in the sequel a systematic way of combining the two aspects of filtering within our framework. This will involve a modification of the standard K-B filter outlined in the previous section. Before proceeding with this modification, we shall discretize the equations. This corresponds to what is done in operational practice and will bring the problem to the form (2.5a).

3.2 Discretization

The discretization chosen for (3.1) is in terms of finite differences. Finite-difference models are still the most widely used in NWP. They also facilitate somewhat the assimilation of observations made in irregular patterns. Spectral models, on the other hand, have certain advantages with regard to the initialization aspect of our problem. An analysis similar to the present one should be easy to reproduce for spectral, finite-element or hybrid models.

The grid

$$x^j = j\Delta x, \quad j = 1,2,\ldots,M, \quad t_k = k\Delta t, \quad k = 0,1,2,\ldots \tag{3.5a}$$

is introduced, with $\underset{\sim}{W}{}_k^j$ approximating $\underset{\sim}{W}(j\Delta x, k\Delta t)$. Then the state vector $\underset{\sim}{x}_k$ of Sec. 2 corresponds simply to $\underset{\sim}{W}{}_k^*$, with

$$\underset{\sim}{x}_k = (u_k^1, v_k^1, \phi_k^1, u_k^2, v_k^2, \ldots, \phi_k^M)^T , \tag{3.5b}$$

so that $n = 3M$.

We used the Richtmyer two-step formulation of the Lax-Wendroff scheme (Richtmyer and Morton, 1967, Sec. 12.7 and 13.4). Let A and B denote the matrices of system (3.1),

$$\underset{\sim}{W}_t = A\underset{\sim}{W}_x + B\underset{\sim}{W} .$$

The scheme can then be written as

$$\begin{aligned} \underset{\sim}{W}{}_{k+1/2}^{j+1/2} &= \underset{\sim}{W}{}_k^{j+1/2} + \frac{\Delta t}{2}\,\underset{\sim}{W}_t\Big|_k^{j+1/2} \\ &= \frac{1}{2}\,(I + \frac{\Delta t}{2}\,B)(\underset{\sim}{W}{}_k^j + \underset{\sim}{W}{}_k^{j+1}) + \frac{\Delta t}{2\Delta x}\,A(\underset{\sim}{W}{}_k^{j+1} - \underset{\sim}{W}{}_k^j) , \end{aligned} \tag{3.6a}$$

$$\begin{aligned} \underset{\sim}{W}{}_{k+1}^{j} &= \underset{\sim}{W}{}_k^j + (\Delta t)\underset{\sim}{W}_t\Big|_{k+1/2}^{j} \\ &= \underset{\sim}{W}{}_k^j + \frac{\Delta t}{\Delta x}\,A(\underset{\sim}{W}{}_{k+1/2}^{j+1/2} - \underset{\sim}{W}{}_{k+1/2}^{j-1/2}) + \frac{\Delta t}{2}\,B(\underset{\sim}{W}{}_{k+1/2}^{j-1/2} + \underset{\sim}{W}{}_{k+1/2}^{j+1/2}) \end{aligned} \tag{3.6b}$$

Eqs. (3.6) define the dynamics Ψ of our system for the state vector $\underset{\sim}{x}_k$ which is given by (3.5) in terms of $\underset{\sim}{W}_k$,

$$\underset{\sim}{x}_{k+1} = \Psi \underset{\sim}{x}_k \ .$$

The discrete system (3.6) has the same type of slow and fast plane wave solutions as the continuous system (3.1). Their dispersive properties are similar. The Lax-Wendroff (L-W) scheme was used because its numerical dissipation is very useful in our simple model in order to simulate the physical dissipation mechanisms active in the atmosphere. Such mechanisms are also present in more complex NWP models, and they are essential for geostrophic adjustment to occur.

The plane waves of the continuous system (3.1) are better approximated by those of the Richtmyer two-step version of the L-W scheme (3.6) than by those of the standard, one-step version. In particular, in the absence of the β-term, the slow, quasi-geostrophic wave of (3.6) satisfies $u_k^j \equiv 0$, as that of (3.1) satisfies $u(x,t) \equiv 0$. This turned out to be a useful check on the departure of solutions $\underset{\sim}{W}_k^j$ from geostrophy.

3.3 The Modified K-B Filter

In the absence of any constraints on the dynamics, it is clear from Sec. 2 that the K-B filter corresponds to the solution of an optimization problem: obtaining a minimum variance estimator for the system (Ψ, H). We have seen in Sec. 3.1 that, in the application of interest here, it is desirable to select among the solutions of the discrete evolution operator Ψ defined by (3.5, 3.6) a special subset —

solutions with fast components which are vanishing or small. Obtaining a best estimate in the presence of such a constraint corresponds to a constrained optimization problem. We shall study in this subsection an appropriate modification of the standard K-B filter described in Sec. 2.

All solutions of (3.6) can be represented as a superposition of plane waves, similar to (3.3a). For the purposes of this discussion it is actually more convenient to think of $\underset{\sim}{x}_k$ in its physical interpretation $\underset{\sim}{W}_k$, so that we write

$$\underset{\sim}{W}(j\Delta x, k\Delta t) = \sum_{\ell,m} \left(\lambda_\ell^{(m)}\right)^k \underset{\sim}{w}_\ell^{(m)} \exp\{i\ell(j\Delta x - c_\ell^{(m)} k\Delta t)\}. \quad (3.7a)$$

For each wave number ℓ, there is a slow wave $\underset{\sim}{w}_\ell^{(1)} e^{i\ell j\Delta x}$ with speed $c_\ell^{(1)}$, and two fast waves $\underset{\sim}{w}_\ell^{(2,3)} e^{i\ell j\Delta x}$ with speeds $c_\ell^{(2,3)}$ of opposite signs. The decay factors $\lambda_\ell^{(m)}$, $|\lambda_\ell^{(m)}| \leq 1$, are present due to the dissipation in the difference scheme (3.6).

Denote by R (for Rossby waves), the span of the real parts of all compound n-vectors, $n = 3M$, of the form

$$\begin{pmatrix} \underset{\sim}{w}_\ell^{(m)} e^{i\ell\Delta x} \\ \underset{\sim}{w}_\ell^{(m)} e^{i\ell 2\Delta x} \\ \vdots \\ \underset{\sim}{w}_\ell^{(m)} e^{i\ell M\Delta x} \end{pmatrix} \quad (3.7b)$$

for $m = 1$, as ℓ ranges over all possible wave numbers. This is the slow wave space. The fast wave space is denoted by G (for gravity waves), and is defined as the span of the real parts of all n-vectors of the form (3.7b) for $m = 2,3$, as ℓ again ranges over all possible wave numbers.

The spaces $\mathcal{R}$ and $\mathcal{G}$ are subspaces of Euclidean n-space E^n, and they span E^n, $\mathcal{R} \oplus \mathcal{G} = E^n$. Furthermore, $\mathcal{R}$ and $\mathcal{G}$ are invariant under the matrix Ψ, defined by (3.6), i.e., they are invariant subspaces of the system's dynamics. Indeed, the eigenvectors of Ψ are precisely the set of all vectors of the form (3.7b). Hence any vector $\underset{\sim}{x}$ in $\mathcal{R}$ will be advanced by the unperturbed system (3.6) to a vector $\Psi\underset{\sim}{x}$ in $\mathcal{R}$ after a time step Δt. Similarly, a vector $\underset{\sim}{y}$ in $\mathcal{G}$ will evolve to $\Psi\underset{\sim}{y}$ in $\mathcal{G}$. Notice, however, that $\mathcal{R}$ and $\mathcal{G}$ are not orthogonal to each other.

Given any n-vector $\underset{\sim}{x}$, there is a unique vector $\underset{\sim}{y}$ in $\mathcal{R}$ which is closest to $\underset{\sim}{x}$ in the sense that $\|\underset{\sim}{y} - \underset{\sim}{x}\|^2 = (\underset{\sim}{y}-\underset{\sim}{x})^T(\underset{\sim}{y}-\underset{\sim}{x})$ is minimized. This vector $\underset{\sim}{y}$ is called the orthogonal projection of $\underset{\sim}{x}$ onto $\mathcal{R}$, and is denoted by $\underset{\sim}{y} = \Pi\underset{\sim}{x}$. The orthogonal projection operator Π is a symmetric matrix, $\Pi^T = \Pi$, and satisfies $\Pi^2 = \Pi$. The orthogonal projection $\underset{\sim}{y} = \Pi\underset{\sim}{x}$ is found in O(n log n) arithmetic operations by performing three Fast Fourier Transforms (FFTs) on the vector $\underset{\sim}{x}$ (one for each component $\underset{\sim}{u}$, $\underset{\sim}{v}$, $\underset{\sim}{\phi}$), then multiplying by a block diagonal matrix comprised of M 3×3 blocks, then performing three inverse FFTs.

We assume in the sequel that

$$\hat{\underset{\sim}{x}}_0 = \Pi\,\hat{\underset{\sim}{x}}_0 \,, \tag{3.8}$$

i.e., that initialization has been performed and $\hat{\underset{\sim}{x}}_0$ lies in $\mathcal{R}$ already. These concepts are discussed and similar notation is used in Leith (1980).

Having defined the operator Π, we are now ready to describe the modified filter. Let

$$\underset{\sim}{\eta}_k = \underset{\sim}{z}_k - H\,\hat{\underset{\sim}{x}}_k(-) \tag{3.9a}$$

denote the innovation vector at time k. This vector carries the new information contained in the observations at time k; $\hat{\underset{\sim}{x}}_k(-)$ carries all the previously accumulated information, as propagated by the dynamic system.

The filtering step (2.11e) is now written

$$\hat{\underset{\sim}{x}}_{k+1}(+) = \hat{\underset{\sim}{x}}_{k+1}(-) + K_{k+1}\underset{\sim}{\eta}_{k+1} \,. \tag{3.9b}$$

What is desired is that, for all k, $\hat{\underset{\sim}{x}}_{k+1}(+)$ lies in $\mathcal{R}$. It is clear by inspection of Eqs. (2.11 a,e,e') that, under assumption (3.8), this will be the case provided that, for all k, $K_{k+1}\underset{\sim}{\eta}_{k+1}$ lies in $\mathcal{R}$. As the observation vector $\underset{\sim}{z}_{k+1}$ is a noisy perturbation of the noisy true state $\underset{\sim}{x}_{k+1}$, cf. Eqs. (2.5a, 2.6a), the correction vector $K_{k+1}\underset{\sim}{\eta}_{k+1}$ does not, in general, lie in $\mathcal{R}$.

What we seek, then, is a modified filter K^*_{k+1}, possibly depending on $\underset{\sim}{\eta}_{k+1}$, which has the property that $K^*_{k+1}\underset{\sim}{\eta}_{k+1}$ does lie in $\mathcal{R}$. This filter is found by minimizing trace $P_{k+1}(+)$, as before, but subject now to the constraint that the correction vector lie in $\mathcal{R}$. A Lagrange multiplier method was used to solve this constrained optimization problem. The result is simply that

$$K^*_{k+1} = \Pi\,K_{k+1} \,, \tag{3.10a}$$

independently of $\underset{\sim}{\eta}_{k+1}$, where K_{k+1} is the standard Kalman-Bucy

filter. For the modified filter, therefore, we replace Eq. (2.11e) by

$$\hat{x}_{k+1}(+) = \hat{x}_{k+1}(-) + \Pi K_{k+1}\eta_{k+1} \cdot \qquad (3.10b)$$

Since this filter is no longer optimal, we must also replace (2.11 d) by (2.10 b), in which K_{k+1} is replaced by ΠK_{k+1}.

In the sequel we shall call the modified algorithm the Π-filter, while the standard algorithm will be called the K-filter. We shall see in Sec. 4 that the Π-filter produces optimal estimates of the slow waves at the expense of estimation errors only slightly larger than the fast-wave contaminated estimates produced by the K-filter.

3.4 Observational Pattern and Choice of Parameters.

In the present article we shall restrict ourselves to the study of a "classical" observational pattern, corresponding to the conventional meteorological upper-air network: all quantities (u,v, ϕ) are observed over "land", and none over the "ocean". This is only meant to serve as an illustration of K-B filtering in a meteorologically familiar situation. Clearly, the power of this approach lies in its ability to handle observations which are arbitrarily distributed in space and time.

Our physical domain is an interval of the x-axis of length 2L. It is meant to correspond to a circle of latitude near 45° N. Hence $f = 10^{-4}\text{sec}^{-1}$ and L = 14000 km. The actual distribution of land and ocean at this latitude has been simplified to be 2-periodic, so that in each interval of length L, half is covered by ocean (Pacific or Atlantic), and half by land (North America or Eurasia). It is reasonable to consider, therefore, only 2-periodic solutions of (3.1), and consequently our computational domain is of length L. We consider the left half of the computational L-domain to be covered by land, and the right half to be covered by ocean, so that our observation matrix is

$$H = (I\ 0) .$$

The mean flow about which (3.2) is linearized was taken to have U = 20 m/s and $\Phi = 3\times10^4 m^2/s^2$. The value of U is typical for mid-tropospheric flow at this latitude; Φ corresponds to an equivalent depth for a homogeneous atmosphere of $H \cong 3$ km, which gives realistic phase speeds for inertia-gravity waves. The slow waves in the solution of (3.1) which we wish to estimate will travel across the fundamental L-domain of one continent and one ocean in a time of approximately L/U; cf. (3.4a), the actual time, for our choice of L, U, Φ and ℓ, is roughly 12 days.

For a single wave number ℓ, initial data for the continuous system (3.1) which lead only to slow waves are, to first order in the small parameter $U/\sqrt{\Phi}$,

$$\phi(x,0) = \phi_0 \sin \ell x \ , \tag{3.11a}$$

$$u(x,0) = \frac{\ell^2 U}{\ell^2\Phi + f^2}\,\phi(x,0) = \frac{\ell^2 U\phi_0}{\ell^2\Phi + f^2} \sin \ell x \ , \tag{3.11b}$$

$$v(x,0) = \frac{1}{f}\,\phi_x(x,0) = \frac{\ell\phi_0}{f} \cos \ell x \ . \tag{3.11c}$$

The solution $\underset{\sim}{W}(x,t)$ of (3.1) with initial data $\underset{\sim}{W}(x,0) = \underset{\sim}{W}_0(x)$ given by (3.11) is, to first order in $U/\sqrt{\Phi}$, $\underset{\sim}{W}(x,t) = \underset{\sim}{W}_0(x - c_\ell^{(1)} t)$, with $c_\ell^{(1)}$ given by (3.4a).

We chose initial data corresponding to a single Rossby wave of wave length L/2, i.e., $\ell = 4\pi/L$, and amplitude $\phi_0 = 2.5 \times 10^3 m^2/s^2$. The latter is in accordance with a typical ridge-to-trough difference of 500 m in the height of the 500 mb pressure surface (Palmén and Newton, 1969, Sec. 6.6). It follows that $\phi_0/\Phi = 1/12$, which partially justifies the linearization of (3.2). It follows also that the amplitude of $v(x,0)$, $v_{max} = \ell\phi_0/f$, is roughly equal to U, a realistic value. Note, however, that $u_{max} = \ell^2 U\phi_0/(\ell^2\Phi+f^2)$ is relatively small, $u_{max} \cong 0.053\ v_{max}$, for our choice of parameters.

Initial data for the discrete estimation problem, $\hat{\underset{\sim}{x}}_0(-)$, were computed by evaluating (3.11) at the grid points $x^j = j\Delta x$, while $\hat{\underset{\sim}{x}}_0(+) = \Pi\ \hat{\underset{\sim}{x}}_0(-)$, in accordance with (3.8). The projection is desirable because the slow waves of (3.6) are slightly different from those of (3.1). To obtain $\underset{\sim}{x}_0$, random

errors are added to $\hat{\underset{\sim}{x}}_0(+)$ (cf. Eqs. (3.12 b,c) below).

Due to the linearity of the problem, the gain matrices K_{k+1}, or $\Pi\ K_{k+1}$, are independent of both the "true" atmospheric state $\underset{\sim}{x}$ and the estimated state $\hat{\underset{\sim}{x}}$: Eqs. (2.11a,e) are decoupled from Eqs. (2.11b,c,d). In particular, the gain matrices are independent of the initial data $\underset{\sim}{x}_0$, $\hat{\underset{\sim}{x}}_0(+)$. Thus the choice of initial data (3.11) has been made only for orientation purposes, and similar results will obtain for any initial estimate satisfying (3.8).

The discretization used the minimum number of grid points which would resolve our wave, namely 16, for the L-domain of interest. This left a computational problem (2.11) of easily manageable size, and it was deemed sufficient for a preliminary illustration of the method. One numerical experiment was performed with a total of 32 points for the computational L-domain; such a spatial resolution of O(400 km) is close to that used in operational NWP models. The results were quite similar to those of the comparable experiment with 16 points.

With a resolution of $\Delta x = L/16$, the stability criterion for the difference scheme (Richtmyer and Morton, 1967, Sec. 12.7) imposes a limit of approximately 47 min on the time step, when using (3.4b) for the maximum wave speed. A much more stringent limit on Δt, closer to the value of Δt used in most primitive-equation NWP models, in which inertia-gravity waves are present, would be imposed if our spatial resolution Δx were closer to the resolution of such models. We actually took Δt = 30 min. This results in two time steps per hour, or 24 steps per synoptic period.

Observations are made over the 8 grid points located on "land" at every synoptic time, that is twice per day. The observation error covariance matrix R is taken diagonal, with equal entries at all grid points. The values assigned to the diagonal entries of R are based on data from McPherson et al. (1979, Table 2). The standard deviation of conventional temperature observations used there is 1°C. This can be converted, based on the customary hydrostatic assumption, to a 500 mb level geopotential error of approximately 200 m^2/s^2. This value results, for our choice of ϕ_0, in an error of about 0.1 ϕ_0. A corresponding 10% error in the wind components is roughly 2 m/s; this is slightly larger than the value of 1.5 m/s used by McPherson et al. (1979). We took the standard deviation in observations of ϕ to be 200 m^2/s^2, and that in observations of u and v to be 2 m/s. Relative errors in all observations are thus about 10%.

The system noise covariance matrix Q is taken to be the sum of a geostrophic and an ageostrophic part

$$Q = \Pi\, D_1^2 \Pi^T + (I - \Pi)\, D_2^2 (I - \Pi)^T , \qquad (3.12a)$$

with D_1 and D_2 diagonal. The choice of D_1 and D_2 involves further dynamical considerations, and will be discussed after we describe first the experiments with a perfect, noise-free model.

The initial error covariance matrix P_0 also has the form

$$P_0 = \Pi\, D_3^2 \Pi^T + (I - \Pi)\, D_4^2 (I - \Pi)^T , \qquad (3.12b)$$

which results from the assumption that geostrophic and ageostrophic errors are uncorrelated. This assumption is only made for convenience, and because of lack of information on the cross-correlations of the two types of errors; it can be easily removed as further information becomes available. Let D be the diagonal matrix with the elements $(v_{max}, v_{max}, \phi_0)$ repeated on the diagonal; then

$$D_3 = 0.4\, D, \quad D_4 = 0.1\, D . \qquad (3.12c)$$

For simplicity, we have taken the initial error covariances (cf. (3.12 b,c)) to be uniform over the entire L-domain. This choice results in initial errors which are much larger than the observational errors over land, although they are approximately equal to the expected mean forecast errors over the ocean. This uniform distribution of initial errors makes it easier to visualize the initial error reduction by synoptic information over land and the propagation of information from land to ocean, as well as the effect of high errors over the ocean

propagating inland. Eventually, we shall be interested in the asymptotic state which corresponds to a stationary, continuous assimilation cycle; hence, the choice of initial errors will ultimately be immaterial.

3.5 Previous Work

The realities of 4-D data assimilation have suggested to practitioners, as well as theoreticians, ideas related to those presented here. A number of authors have preceded us on this ground, and we shall mention their work at least briefly.

Jones (1965a,b) seems to have been the first to bring the formalism of K-B filtering to the attention of the meteorological community. Jones (1965a) is an excellent compendium of the formulae for the discrete-time filter, including results on the asymptotic filter. Jones (1965b) is an attempt at nonlinear filtering for a single, scalar quantity; an improvement over direct insertion obtains when statistical ideas are included.

Petersen (1968, 1970, 1973a,b,c, 1976) offers the most comprehensive treatment of estimation ideas in the meteorological literature. He applied these ideas to a linear form of the quasi-geostrophic barotropic potential vorticity equation (Petersen, 1976), in which fast waves do not appear. The estimation is carried out in terms of spectral transforms and the dynamics incorporated in the

form of Green's functions. This approach to estimation does not seem to extend easily to nonlinear dynamics. Its implementation in linear cases is also hampered by not exploiting the sequential aspect of K-B filtering, nor the use of asymptotic filter values.

A variational approach to updating, which bears certain similarities to sequential estimation, appears in Tadjbakhsh (1969) and Phillips (1971). It was also implemented for real satellite sounding data by Ghil and Mosebach (1978). This approach, however, does not include explicitly the statistical optimality considerations of K-B filtering.

Miyakoda and Talagrand (1971) discussed the possibility of blending forecasts from previous synoptic times with current observations, by averaging. They analyzed this possiblity for the linear, one-dimensional vorticity equation and carried out numerical experiments for the same equation in its nonlinear, two-dimensional form. They did not use a sequential filter or statistically determined weights. Still their work showed the importance of using past observations, as carried forward by the dynamical model itself, in obtaining a better estimate for the current state of the atmosphere.

Phillips (1976) developed an estimation procedure for a one-dimensional, linear, two-level quasi-geostrophic model. The model uses fairly realistic flow quantities, observational patterns and error variances. This procedure, like those of Petersen (1976) and of Miyakoda and Talagrand

(1971) is not sequential; it requires, in particular, the specification of the actual solutions' statistics, rather than solely observational and forecast error covariances. Since the model only admits slow waves, no modifications were necessary to eliminate the fast waves present in operational, primitive-equation models.

Phillips' work stressed the importance of statistical concepts in 4-D data assimilation and had a considerable influence in their operational implementation. We hope in turn that our results will lead to a better theoretical understanding of the interaction of statistics and dynamics in meteorological data assimilation and also help practitioners in optimizing further its operational applications.

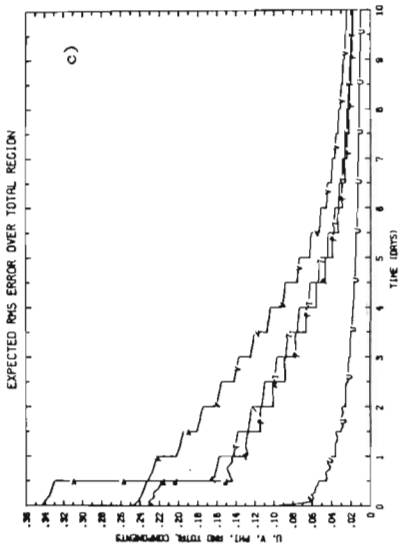
4. RESULTS

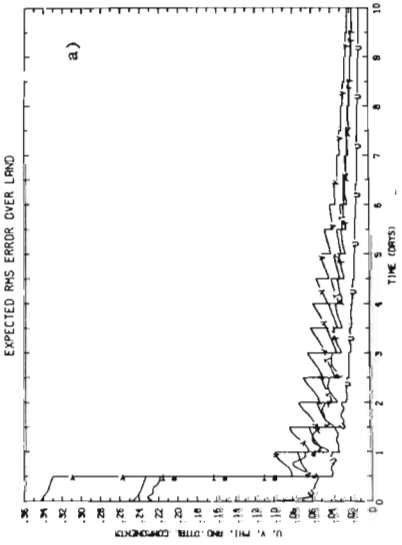
We wish to stress here again that these results are preliminary. They present an illustration of our estimation approach for a very simple model with some nontrivial features of operational NWP models. The main feature of interest is the presence in the model of travelling large-scale waves with different speeds. Optimal estimates of the slow waves are obtained, while eliminating the fast waves.

4.1 Estimation for a Perfect, Noise-Free Model

We study first the way in which the K-filter, using partial observations over land, reduces the initial error in a system without noise, Q = 0. Fig. 2 shows the components of the expected root-mean-square (rms) estimation error(trace P_k)$^{1/2}$, over a 10 day numerical experiment, or run. Fig. 2a shows the expected rms error over "land", Fig. 2b over "the ocean", and Fig. 2c over the entire L-domain. The individual curves correspond to the errors in u(U), v(V), ϕ(P), and the total error (T).

Obviously, sharp error reduction occurs at the observation times over land (Fig. 2a). The more interesting fact is that noticeable error reduction occurs at synoptic times also over the ocean (Fig. 2b). The latter reduction is due to the corrections applied to ocean grid points by the filter. In this, the K-filter acts like current objective analysis schemes, in particular those using linear regression: the new information





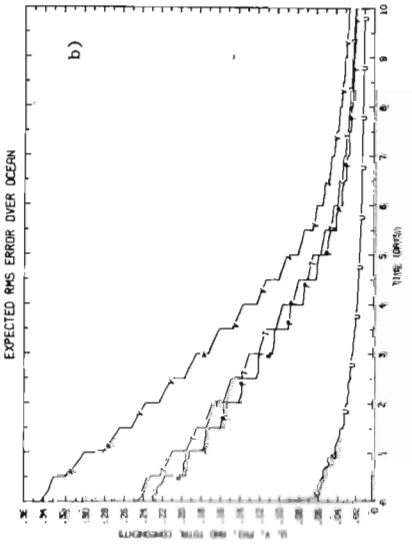


Fig. 2
The components of the total expected rms error (Erms), (trace: $P_k)^{1/2}$, in the estimation of solutions to the stochastic-dynamic system (Ψ,H), with Ψ given by (3.6) and H = (I 0). System noise is absent, Q = 0. The filter used is the standard K-B filter (2.11) for the model.

a) Erms over land; b) Erms over the ocean; c) Erms over the entire L-domain

In each one of the figures, each curve represents one component of the total Erms error. The curves labelled U, V, and P represent the u component, v component and ϕ component, respectively. They are found by summing the diagonal elements of P_k which correspond to u, v, and ϕ, respectively, dividing by the number of terms in the sum, and then taking the square root. In a) the summation extends over land points only, in b) over ocean points only, and in c) over the entire L-domain. The vertical axis is scaled in such a way that 1.0 corresponds to an Erms error of vmax for the U and V curves, and of ϕ_0 for the P curve. The observational error level is 0.089 for the U and V curves, and 0.080 for the P curve. The curves labelled T represent the total Erms error over each region. Each T curve is a weighted average of the corresponding U, V, and P curves, with the weights chosen in such a way that the T curve measures the error in the total energy $u^2 + v^2 + \phi^2/\phi$, conserved by the system (3.1). The observational noise level for the T curve is then 0.088. Notice the immediate error decrease over land and the gradual decrease over the ocean. The total estimation error tends to zero.

from observations over land is spread out over adjacent ocean areas at the observation time itself. The difference between our approach and conventional schemes will become more apparent in the next subsection, inasmuch as the K-filter is capable of discerning between data-sparse ocean areas upstream and downstream from data-rich land.

Even in the present case of a noise-free model, Q = 0, the effects of advection of information are noticeable. In between observations, the error over land grows (Fig. 2a). This is due to the advection of error from over the ocean. The total error (Fig. 2c) between observations decreases. This decrease is due to the dissipation in the model, which is conservative when Q = 0, except for numerical dissipation. The error over the ocean (Fig. 2b), however, decreases considerably more than the total, due to the advection of information from land.

The expected rms error over land falls below the observational noise level at every observation; over the ocean, this happens after approximately 4-5 days. In our noise-free model, the total expected rms error eventually decays to zero: no information is lost and the observational errors can be eliminated entirely by repeating the observations as long as necessary for the expected estimation error to become negligible. This is a result of the fact that our system is completely observable (Bucy and Joseph, 1968, Ch. 3 and Ch. 5).

Components of the actual rms error in the estimation of an individual realization of our system, with given initial data x_0 and observation error $\{\zeta_k: k = 1,2,\ldots\}$, appear in Fig. 3a. The corresponding components of relative m-s error, $\|x_k - \hat{x}_k\|^2/\text{trace } P_k$, where $\|y\|$ is the length of the vector y are given in Fig. 3b. For a different random choice of x_0 and sequence ζ_k, the same plots appear in Figs. 3c, 3d.

It is easy to see that the rms error in an individual realization decreases to the observational noise level in about the same time as its expected value (compare Fig. 3a and 3c with Fig. 2c). After that time, however, it continues to fluctuate near zero, rather than decaying monotonically to zero. Our experimental results (Figs. 3b,d) show that the spread of individual estimation errors about their expected value is not too large. It would be interesting, in general, to know a priori how large this spread is expected to be, and we intend to study this question further.

We show in Fig. 4 the actual time histories at a number of grid points for the estimated solution corresponding to the realization in Figs. 3a,b; $\hat{u}$ is shown in Fig. 4a, $\hat{v}$ in Fig. 4b and $\hat{\phi}$ in Fig. 4c. We chose to show a point on the West Coast of the continent, labeled SF (for San Francisco), one on the East Coast, labeled NY (for New York), and one in the middle of the ocean, labeled HA (for Hawaii). Note that "Tokyo" = "New York" by periodicity.

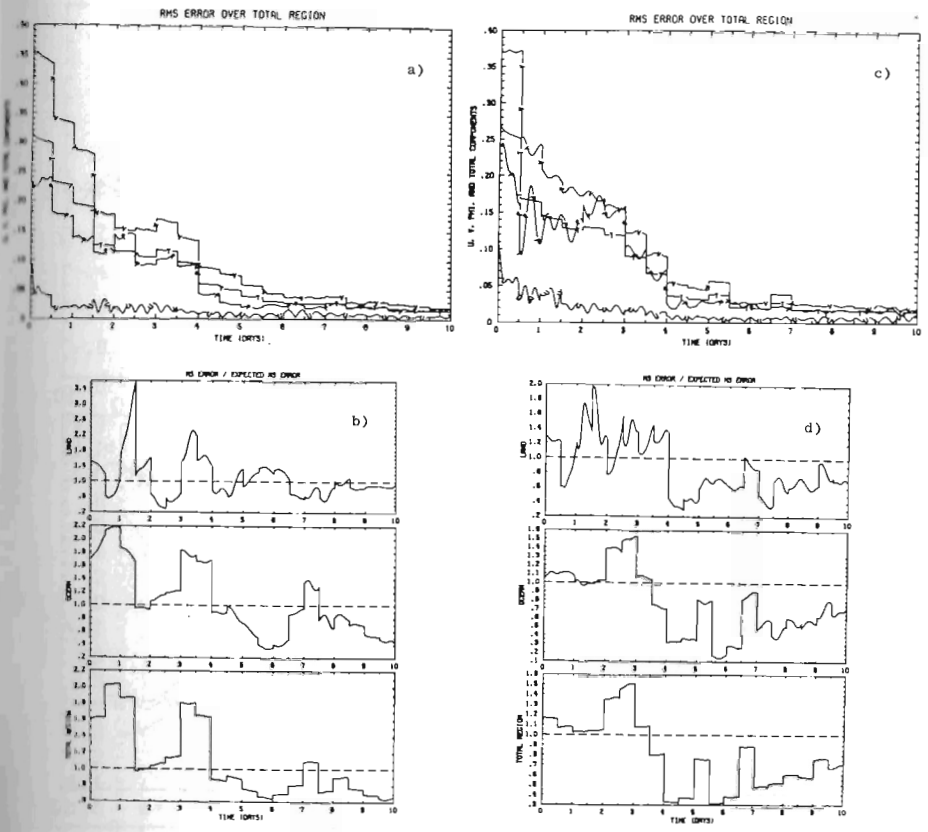


Fig. 3 Components of the actual root-mean-square (rms) estimation error and the relative mean-square (m-s) estimation error, for two realizations of the experiment whose Erms errors appear in Fig. 2. The two realizations correspond to two different random choices of initial condition x_0 and observational noise ζ_k, sampled from the same respective probability distributions, with covariance matrices P_o and R, respectively.

a) The rms errors for a realization of the run whose Erms errors are given in Fig. 2. The curves have the same labels, and are plotted on the same scale, as in Fig. 2.

b) Relative m-s errors for the same realization. The three panels give the ratio of total m-s error over land, ocean, and the entire region, respectively, to the corresponding expected m-s error.

c) Same as Fig. 3a, for a different realization.

d) Same as Fig. 3b, for the run in Fig. 3c. Notice the difference between Fig. 3a and Fig. 3c: the actual rms estimation error in two realizations of the filtering process can be quite different. Figs. 3b and 3d show the variation of the relative m-s error: equality of absolute m-s error and expected m-s error corresponds to a value of 1 in these figures.

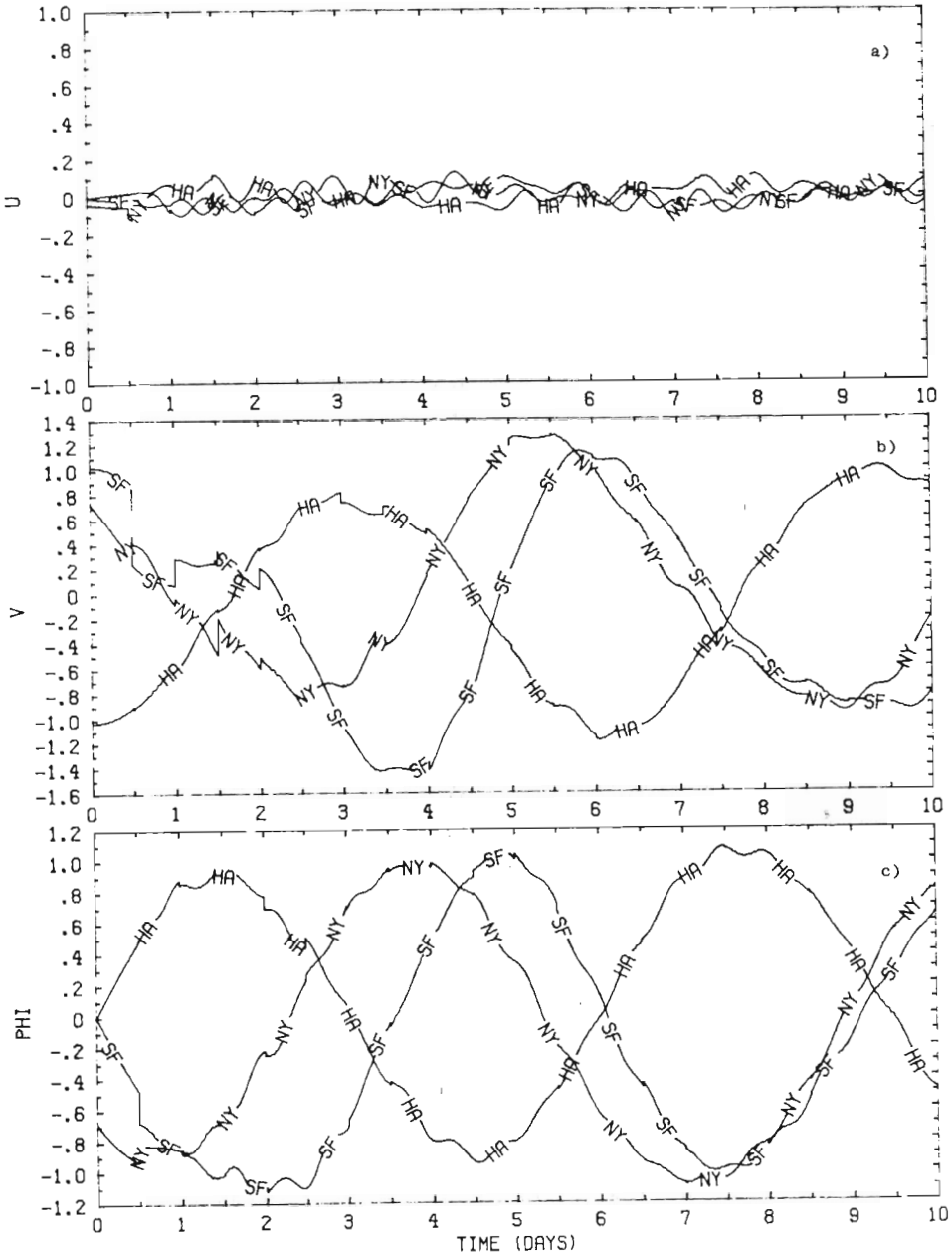


Fig. 4 Time history of the estimated solution without system noise, at three locations, labeled HA (for Hawaii, a mid-ocean location), SF (for San Francisco, a West Coast location), and NY (for New York, an East Coast location).

a) u-component of velocity; b) v-component of velocity,

c) ϕ, the geopotential

Notice the slow waves with a period of approximately 6 days, upon which are superimposed smaller, fast waves with a period of approximately one-half day.

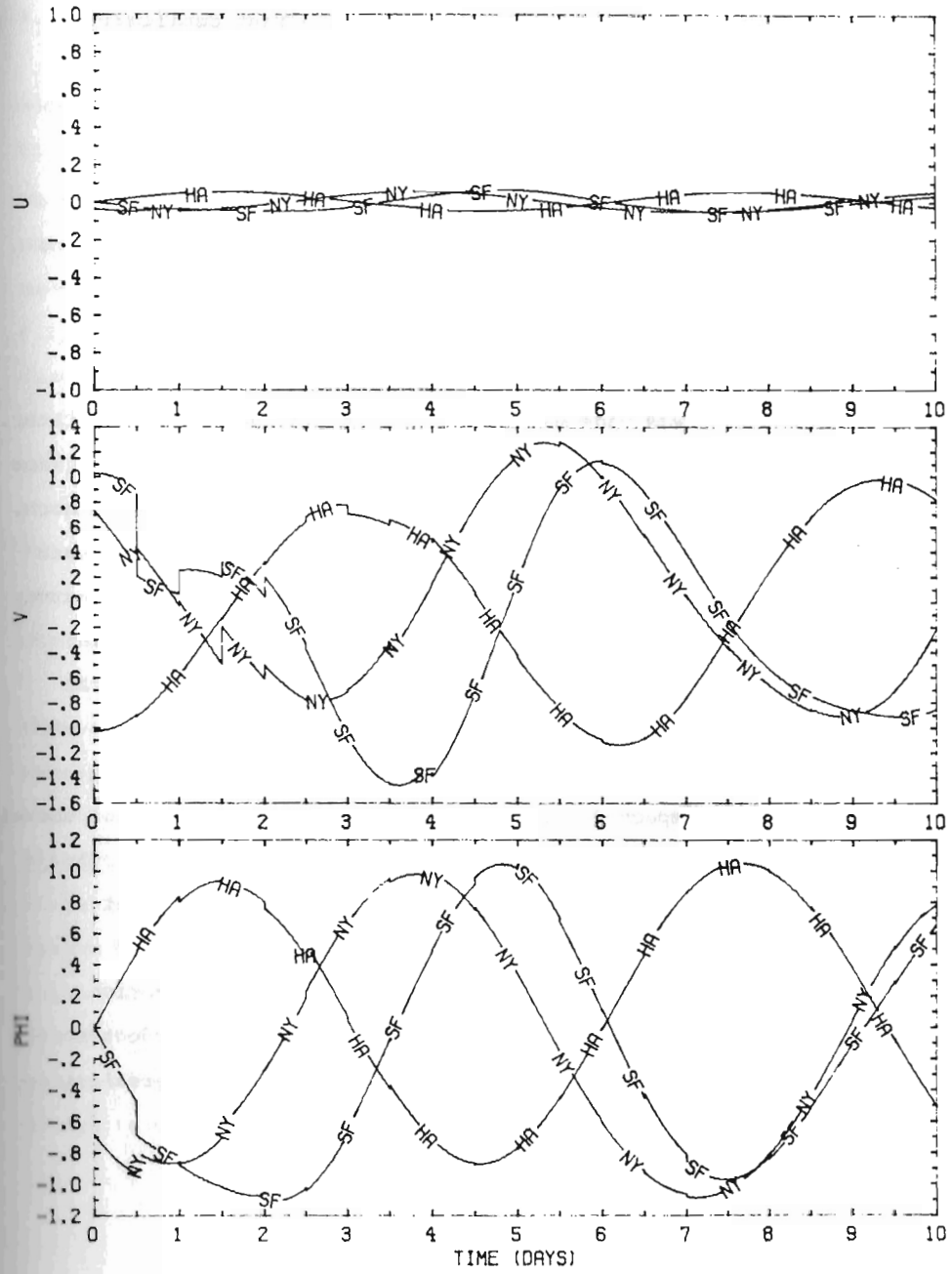


Fig. 5 Same as Fig. 4, only for a run using our modified K-B filter, the Π-filter. The fast waves have been completely eliminated and the estimated solution is slowly varying.

We see that for each curve, small, fast oscillations are superimposed on a smooth, slowly varying wave pattern. The fast oscillations are caused by the ageostrophic component of the initial error, cf. (3.12 b,c), and of the observations. They are especially apparent in the u-component of the estimated solutions. These oscillations are partially damped between synoptic times by the dissipativity of the model.

A run identical in every other respect to that in Fig. 4 was made using the Π-filter instead of the K-filter. The time histories of the corresponding estimates at the same points are shown in Figs. 5a,b,c. They are perfectly smooth, except for the jumps due to observations, which are rather large in the SF and NY curves and very small in the HA curve. In other runs without the β-term (not shown) one had in particular $\hat{u} \equiv 0$, to within machine accuracy. Notice the periodicity of approximately 6 days, due to the passage of the 2-wave we are estimating.

The expected error reduction for the Π-filter (not shown) was only slightly smaller than that for the K-filter. Thus a slowly varying estimated solution was obtained without sacrificing the optimality of the estimate.

We have thus studied error reduction, information propagation and filtering of fast waves in the perfect model. We shall turn our attention presently to the more realistic model with system noise in it.

4.2 Estimation in the Presence of System Noise

In the previous subsection we saw that, for a perfect model, the estimation error covariance matrix $P_k(+)$ tends to zero, and hence so does K_k (Eq. (2.12 b)). We shall study now the case in which, due to the simultaneous presence of system noise Q and observation noise R, the gain matrix K will tend to a nonzero, asymptotic constant value.

4.2.1 Modeling of system noise. We shall discuss at the beginning our formulation of Q, cf. (3.12a), in particular the choice of the diagonal matrices D_1 and D_2, left open in Sec. 3.4. This choice has to reflect the error growth properties of NWP models. The overwhelming dynamical consideration in this context is the inherent unpredictability of the atmosphere.

Numerous studies (cf. Lorenz, 1969, and references therein) have shown that realistic models of the atmosphere, subject to a small random perturbation, will evolve in a finite amount of time to a state which is statistically independent of the corresponding unperturbed state. This amount of time depends upon the scales of motion of interest, and for synoptic scale motions is about two weeks. Unpredictability is a decidedly nonlinear effect: perturbations at any given scale of motion are nonlinearly fed into all the scales and eventually grow enough to completely contaminate the state.

At that time, forecast error growth levels off (Phillips, 1976). Energy conservation implies in fact that such leveling off must occur in a nonlinear realistic model of the atmosphere.

In our linear model, the estimated state $\hat{\underset{\sim}{x}}$ is governed, in the absence of observations, by

$$\hat{\underset{\sim}{x}}_{k+1} = \Psi \hat{\underset{\sim}{x}}_k \quad , \qquad k = 0,1,2,..., \tag{4.1a}$$

while $\underset{\sim}{x}_k$ is the atmospheric state, governed by

$$\underset{\sim}{x}_{k+1} = \Psi \underset{\sim}{x}_k + \underset{\sim}{\xi}_k \quad , \qquad k = 0,1,2,... \tag{4.1b}$$

In this linear model, $\hat{\underset{\sim}{x}}_k$ and $\underset{\sim}{x}_k$ will never become actually uncorrelated when they start from the same initial state, $\hat{\underset{\sim}{x}}_0 = \underset{\sim}{x}_0$, and hence have the same mean. However, the variance of their difference will grow with time due to the system noise $\underset{\sim}{\xi}_k$.

We would like to choose Q, i.e., D_1 and D_2 in (3.12a), so as to have

$$\text{trace } P_N = E(\hat{\underset{\sim}{x}}_N - \underset{\sim}{x}_N)^T(\hat{\underset{\sim}{x}}_N - \underset{\sim}{x}_N) = 2\hat{\underset{\sim}{x}}_N^T\hat{\underset{\sim}{x}}_N \tag{4.2}$$

at time N, which would correspond to $\hat{\underset{\sim}{x}}_N$ and $\underset{\sim}{x}_N$ being uncorrelated. We prescribed N, in rough agreement with predictability estimates, to be N = 10 days. If the model (4.1a) were conservative, (4.2) would be equivalent to

$$\text{trace } P_N = 2\hat{\underset{\sim}{x}}_0^T\hat{\underset{\sim}{x}}_0 \ .$$

In fact, we set $D_1 = \gamma D$, $D_2 = 0.25\gamma D$, similarly to Eq. (3.12c), with γ chosen so as to satisfy

$$\text{trace } P_N = 2\alpha\hat{\underset{\sim}{x}}_0^T\hat{\underset{\sim}{x}}_0 \ , \tag{4.3}$$

and $\alpha = 0.3$. This was easily achieved by trial and error, since P_k evolves simply according to (2.11b, 2.11d'). We found $\gamma = 0.028$ when $\Delta t = 30$ min.

The reason $\alpha = 0.3$ was chosen instead of the $\alpha = 1$ suggested by predictability theory, cf. Eq. (4.2), is that trace P_k in the linear, observation-free model (4.1) continues to grow after time N. It does actually level off also, as a result of dissipation. Its leveling-off time, however, is not related to the predictability limit N and is in fact much larger. Our attempt to account for loss of predictability in our experiments results, therefore, in a choice of Q which, even with $\alpha = 0.3$, is considerably larger than the Q which would appear in a nonlinear model.

We expect actually that the estimation-theoretical framework, applied to experiments with a nonlinear model, will lead to new insights into the nature of atmospheric predictability. One possible approach is the adaptive determination of Q, in which Q is actually determined from the observations themselves (Chin, 1979; Ohap and Stubberud, 1976).

4.2.2 Numerical results. Having prescribed Q, we turn now to the actual experiments with a noisy system. Fig. 6 shows the expected rms error for a run with the K-filter. As in Fig. 2, (a) shows "land", (b) the "ocean", and (c) the entire region.

At the first synoptic time, the total expected rms error (T) over land drops below the observational error level, which in our nondimensional units is 0.088. It grows much more sharply between synoptic times than in Fig. 2a, due to the presence of system noise, which is added to the error advected from the ocean. However, the estimation error just <u>after</u> each synoptic time is smaller than just after the previous synoptic time; the same is true, moreover, of the error just <u>before</u> successive synoptic times.

The <u>monotone decrease</u> of the components of (trace $P_k(-))^{1/2}$, as well as those of (trace $P_k(+))^{1/2}$, from one synoptic period to the next is even more striking in Figs. 6b,c. We notice, however, that in contradistinction to Figs. 2b,c, neither the total error (T) over the ocean, nor that over the entire region ever drop below the observational error level. The expected rms errors now increase between synoptic times instead of decreasing: the effect of the system noise ξ is stronger than the effect of dissipation in Ψ.

What does happen is that the expected rms errors very quickly settle into an asymptotically periodic pattern with the synoptic interval of 12 h as the period. The convergence occurs within 1-2 days over land, and within 4-5 days over the ocean. In particular, the values of trace $P_k(\pm)$ at synoptic times tend to a constant. This leads us to suspect that in fact the matrices $P_k(\pm)$ themselves, and hence the filter K_k, tend to a constant.

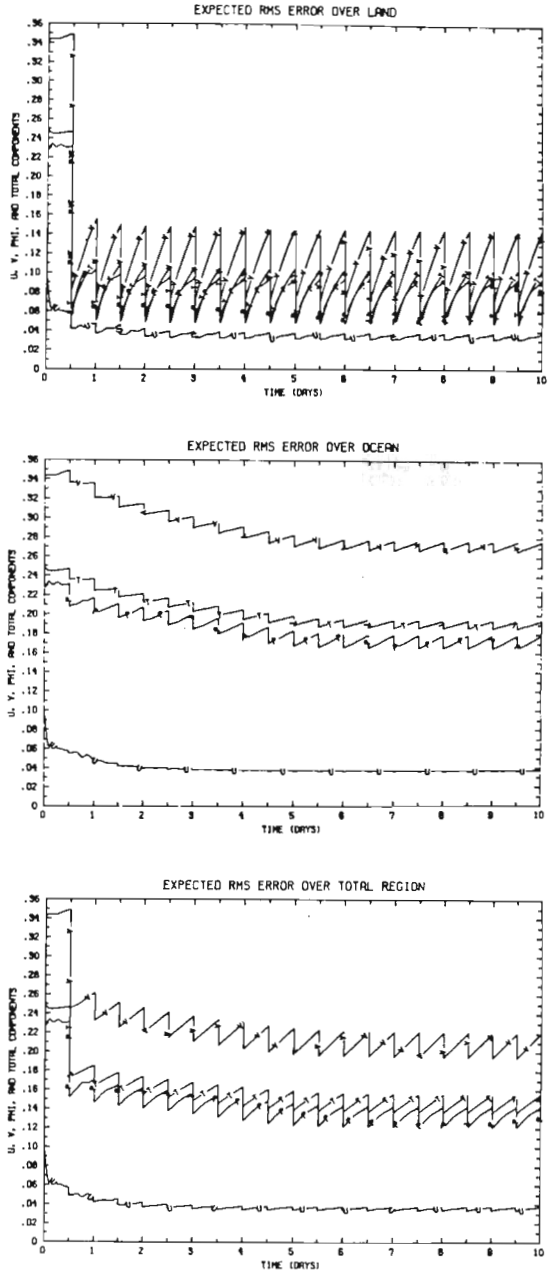


Fig. 6 This figure and the following ones show the properties of the estimated algorithm (2.11) in the presence of system noise, $Q \neq 0$. This figure gives the Erms estimation error, and is homologous to Fig. 2. Notice the sharper increase of error over land between synoptic times, and the convergence of each curve to a periodic, nonzero function.

4.2.3 Asymptotic filter: properties. Examining the behavior of all entries of the gain matrix K as a function of time confirmed our conjecture. Moreover, the asymptotic matrix K_∞, to which K_k tends, is banded. Recall that P is an n × n matrix, with n = 3M = 48, H is a p × n matrix, with p = n/2 = 24, and R is p × p:

$$P = \left[\begin{array}{c|c} P_\ell & P_{\ell-o} \\ \hline P_{o-\ell} & P_o \end{array}\right] , \quad H = (I \mid 0) . \tag{4.4a}$$

Here P_ℓ is the submatrix of the auto-covariances for estimation errors over land, $P_\ell^T = P_\ell$, P_o the autocovariance of errors over the ocean, $P_o^T = P_o$, while $P_{\ell-o}$ is the cross-covariance of errors over land and over the ocean, with $P_{o-\ell} = P_{\ell-o}^T$. By (2.12b) and (4.4a),

$$K = \left[\begin{array}{c} P_\ell R^{-1} \\ \hline P_{o-\ell} R^{-1} \end{array}\right] , \tag{4.4b}$$

which is n × p.

All entries away from the diagonal of the upper block in K_∞ become rapidly smaller with distance from the diagonal. Periodicity in x leads to the appearance of a few larger elements in the corners of both blocks, the upper and the lower.

To visualize better the behavior of K_k in time and to study the structure of K_∞, we considered contour plots of the elements of K (not shown), and selected cross-sections of these plots. The cross-sections correspond to the influence functions of observations at the selected location. In other words, they show the weight given to an observation at such a location when updating a point situated a certain distance away from it.

The chosen locations, or "upper-air stations", were SF, SL (for Saint Louis) and NY. There is no influence function for mid-ocean points, like HA, since no observations are made there. Cross-sections were plotted at every synoptic time, i.e., every 24 time steps. It was clear that convergence occurred within 4-5 days, as it did for trace P_k over the entire region.

Fig. 7 shows the influence functions for the selected locations at the end of day 10. Fig. 7a marked (u-u) gives the influence of a u observation at the selected stations on u updates at any grid point in the L-domain. Fig. 7b, marked (u-v), gives the weight of a v observation at a station in a u update at every grid point, and so on; Fig. 7i gives the influence of v on ϕ.

All the weighting coefficients involving u are rather small (Figs. 7a,b,c,e,h). This is due to our choice of the

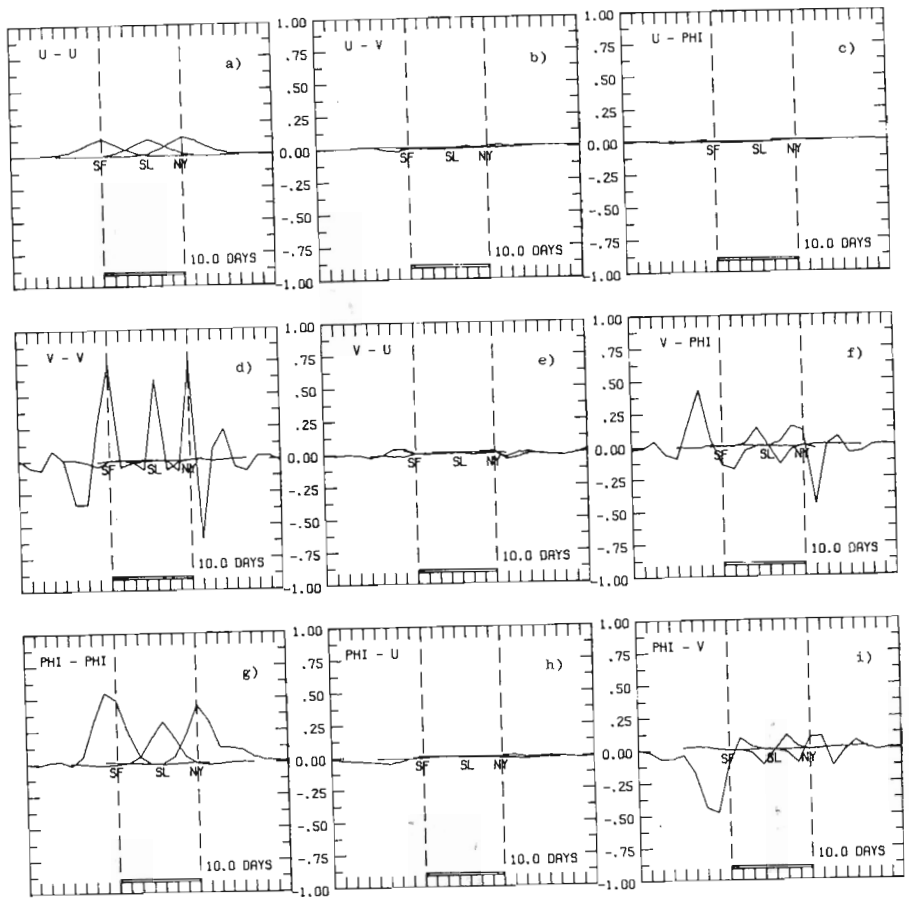


Fig. 7 Influence functions of selected stations at which observations are made. The stations selected are SL (for Saint-Louis, an interior continental location), SF and NY (see Fig. 4). These functions are simply cross-sections of the K-B algorithm's gain matrix K at day 10. They correspond to the weight given an observation at station SF, say, when making a correction to the forecast field at any one of the M grid points in the L-domain. The nine panels, 7a-7i, give the influence of a) $\underline{u}$ observations on $\underline{u}$ corrections, of b) $\underline{v}$ on $\underline{u}$, c) ϕ on $\underline{u}$, d) $\underline{v}$ on $\underline{v}$, e) $\underline{u}$ on $\underline{v}$, f) ϕ on $\underline{v}$, g) ϕ on ϕ, h) $\underline{u}$ on ϕ, i) $\underline{v}$ on ϕ. The particularities of the curves are discussed in the text.

system noise covariance matrix Q: its $\underline{u}$-components were chosen small, cf. (3.12a) and the 4-to-1 ratio of D_1 to D_2. This choice entails relatively good predictions of $\underline{u}$, which have to be corrected only to a small extent by the observations. The (u-u) coefficients (Fig. 7a) are the largest of the coefficients involving $\underline{u}$; they still do not exceed 0.125. The (u-u) influence functions are approximately equal for SF, SL and NY, positive and symmetric in the E-W direction. They are the only ones to have the latter properties.

The influence function for (ϕ-ϕ) centered at SL is the smallest one shown in Fig. 7g. It is positive over land, becoming nearly zero at SF and NY and slightly negative out into the ocean. The relative small size and symmetry of this function is due to its station, SL, being located in the middle of a data-dense region: neighboring stations receive almost equal weights and advection plays but a small role.

The peaks of the (ϕ-ϕ) influence functions centered at NY and at SF are considerably higher than the SL peak. This is due to the absence of observations on the ocean side of these stations. In fact, the peak for the SF influence function is slightly higher than the NY peak. Moreover, the former is located one grid point West of SF, rather than at SF itself, while the NY peak is at NY. Both data density and advection thus play a role.

It makes sense for the point upstream of SF to give even more weight to SF information than SF itself: SF is closer

to inland points and their information is also weighted heavily. Due to the advection of error, the forecast error at synoptic time for this ocean point is considerably larger than that for the point downstream from New York, although they are equidistant from land. Hence the larger weight given to adjacent land observations for the Pacific point than for the Atlantic point.

As in Fig. 7g, the (v-v), (v-ϕ) and (ϕ-v) influence functions (Figs. 7d,f,i) all show strong inhomogeneity-differences between the SF, SL and NY functions, as well as anisotropy-differences in the East and the West direction. The SL function for (v-ϕ) and (ϕ-v) is very nearly antisymmetric. This antisymmetry reminds us of the same feature being exhibited by the (v-ϕ) and (ϕ-v) correlations in Schlatter (1975, Fig. 3). The latter were based on assumptions of geostrophy and verified against the network of U. S. radiosonde stations.

Notice from (4.4) that $P_\infty(+)$ and K_∞ have similar symmetry properties, since we took R to be diagonal. The diagonal elements of R^{-1} simply multiply the columns of P_ℓ and $P_{o-\ell}$, yielding the influence functions in Fig. 7. Hence it is legitimate to compare the symmetry properties of the asymptotic influence functions in the case at hand with those of the steady-state covariance matrices in some current objective analysis schemes.

Our (v-ϕ) and (ϕ-v) influence functions for SF and NY, however, are far from being antisymmetric: they do not even equal zero at SF or NY, respectively. We conclude that it is reasonable to use the geostrophy assumption for wind-height correlations (cf. also Bergman, 1979) in data-dense regions, where the estimation error covariance, i.e., the covariance of forecast-minus-observed fields, is nearly homogeneous and isotropic. Close to the borderline of data-dense and data-poor regions, this assumption will seriously distort the optimal weighting coefficients.

In fact, the influence functions determined by the filtering procedure at the first synoptic time (Fig. 8) are either perfectly symmetric ((u-u), (v-v) and (ϕ-ϕ)) at SL, or perfectly antisymmetric for that station (all six wind-wind and wind-height cross-sections). Furthermore, in all nine panels of Fig. 8, the influence function of NY is either the mirror image ((u-u), (v-v) and (ϕ-ϕ)) of the one of SF, or the inverted mirror image thereof (all other cross-sections).

The comparison of Fig. 8 with Fig. 7 allows us to distinguish between the effect of inhomogeneous data density and the effect of advection on the optimal K-B filter. Fig. 8 shows the effect of data distribution only, since at the first synoptic time no information has been advected yet from previous data insertions. Fig. 7 shows the combination of the two effects.

Different data densities result in different influence functions according to station location (Fig. 8): stations

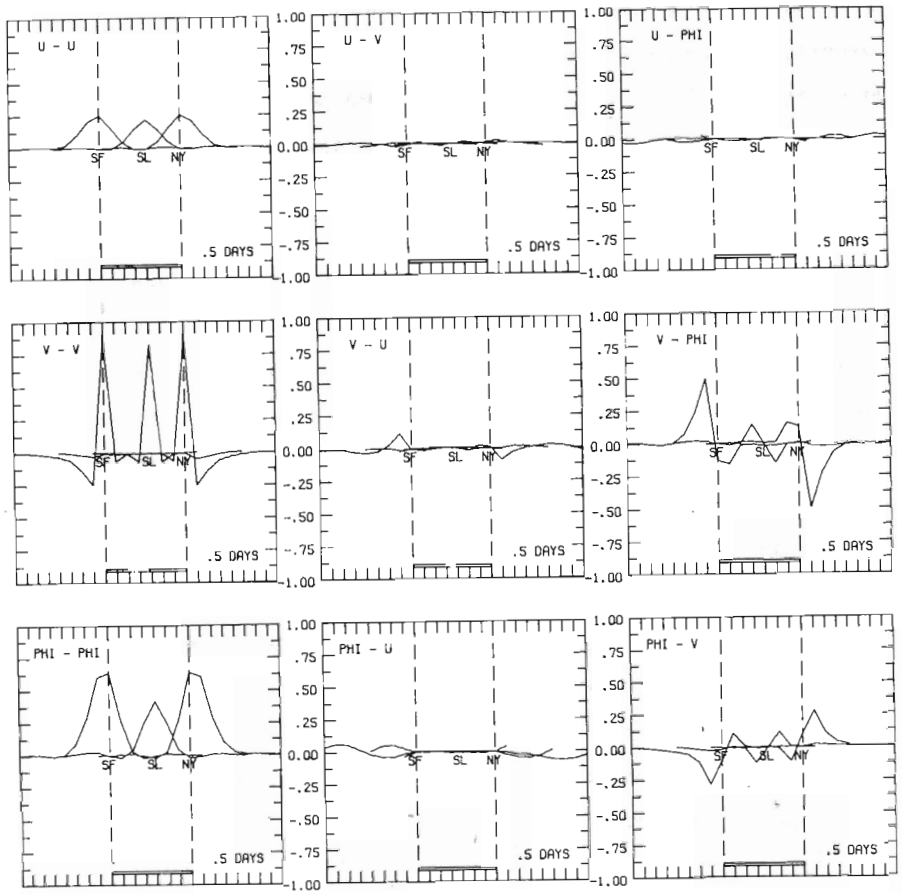


Fig. 8 Same as Fig. 7, but for K used at the first synoptic time. The influence functions here are much more symmetric than in Fig. 7.

located in sharp gradients of observation availability, such as SF and NY, have more influence than inland stations (SL); their influence out to sea is also greater than their influence inland. It is advection, however, which leads to the difference between the influence functions of stations on the West Coast (SF) and East Coast of continents (NY = Tokyo). The latter difference was discussed in connection with Fig. 7g, and can also be found in Figs. 7d,f,i.

4.2.4 Asymptotic filter: results

We used the gain matrix K at day 10 from this run as a time-invariant gain matrix for another run which was otherwise identical to the previous one. This matrix is a very good approximation to the exact asymptotic filter K_∞. Estimation errors after 1-2 days were practically indistinguishable from those obtained when using the time-varying K-B filter, K_k. There is therefore no need, in our constant-coefficient system, to compute the filter K at every synoptic time: the approximate computation of the asymptotic filter K_∞ once and for all is sufficient for any practical purposes. Furthermore, K_∞ is independent of P_o. As indicated in Sec. 4.3, it depends only on Ψ, Q and R.

The asymptotic filter is sometimes called the Wiener filter (W-filter). Wiener (1949) in fact solved the estimation problem for stationary time series, using all past information. It was the contribution of Kalman (1960) to devise a practical sequential filter for stochastic

processes governed by differential equations and using only past information over a finite time interval.

The individual rms errors in estimation for the K-filter run with system noise, as well as for the W-filter run (not shown), are even noisier than those in Fig. 3. We plotted the ratios of individual m-s errors to the expected m-s error (given in Fig. 6). A comparison of these ratios with Figs. 3b,d showed that the system noise between observations causes fluctuations of higher frequency than did the observational noise alone in the perfect model. The same remarks as in Sec. 4.1 about the actual spread of estimation errors around their mean apply.

4.2.5 Filtering of fast waves. Fig. 9 gives the time histories of the estimated solution at HA, SF and NY. System noise clearly excites the fast, inertia-gravity waves in the solution even more than observational noise alone did in Fig. 4. The time histories for a run with the Π-filter, instead of the K-filter, are shown in Fig. 10. Clearly the fast waves have been removed, and the evolution of the estimated solution is perfectly smooth.

Expected rms errors (not shown) with the Π-filter were, for the v- and φ-components, almost indistinguishable from those with the K-filter. Expected rms errors for the u-component, however, were significantly greater in the case of the Π-filter: after 10 days, they were about twice the corresponding errors with the K-filter. This is still well below the level of observational noise, though.

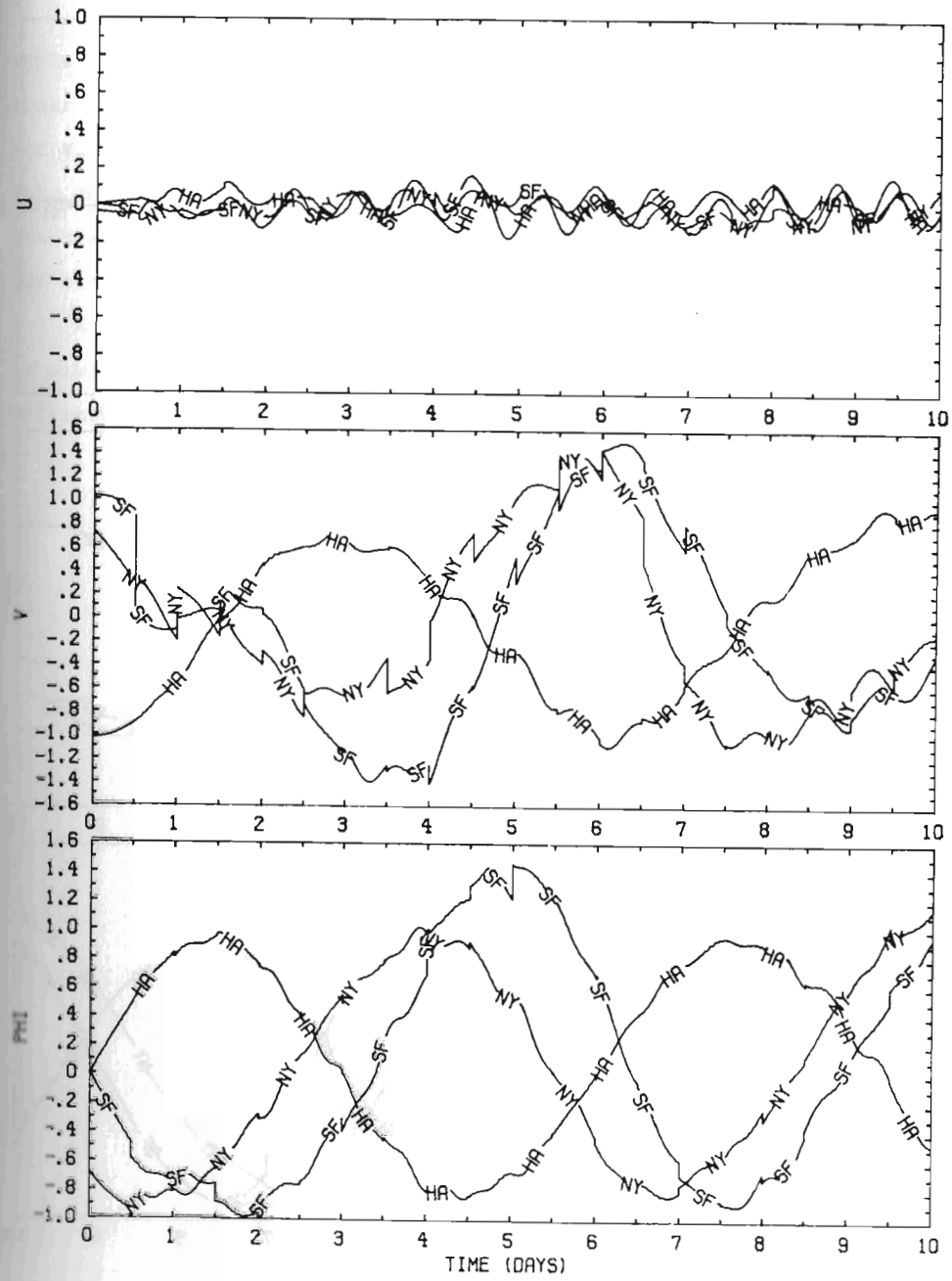


Fig. 9 Same as Fig. 4, but for Q ≠ 0. The fast oscillations are larger than in the noise free case.

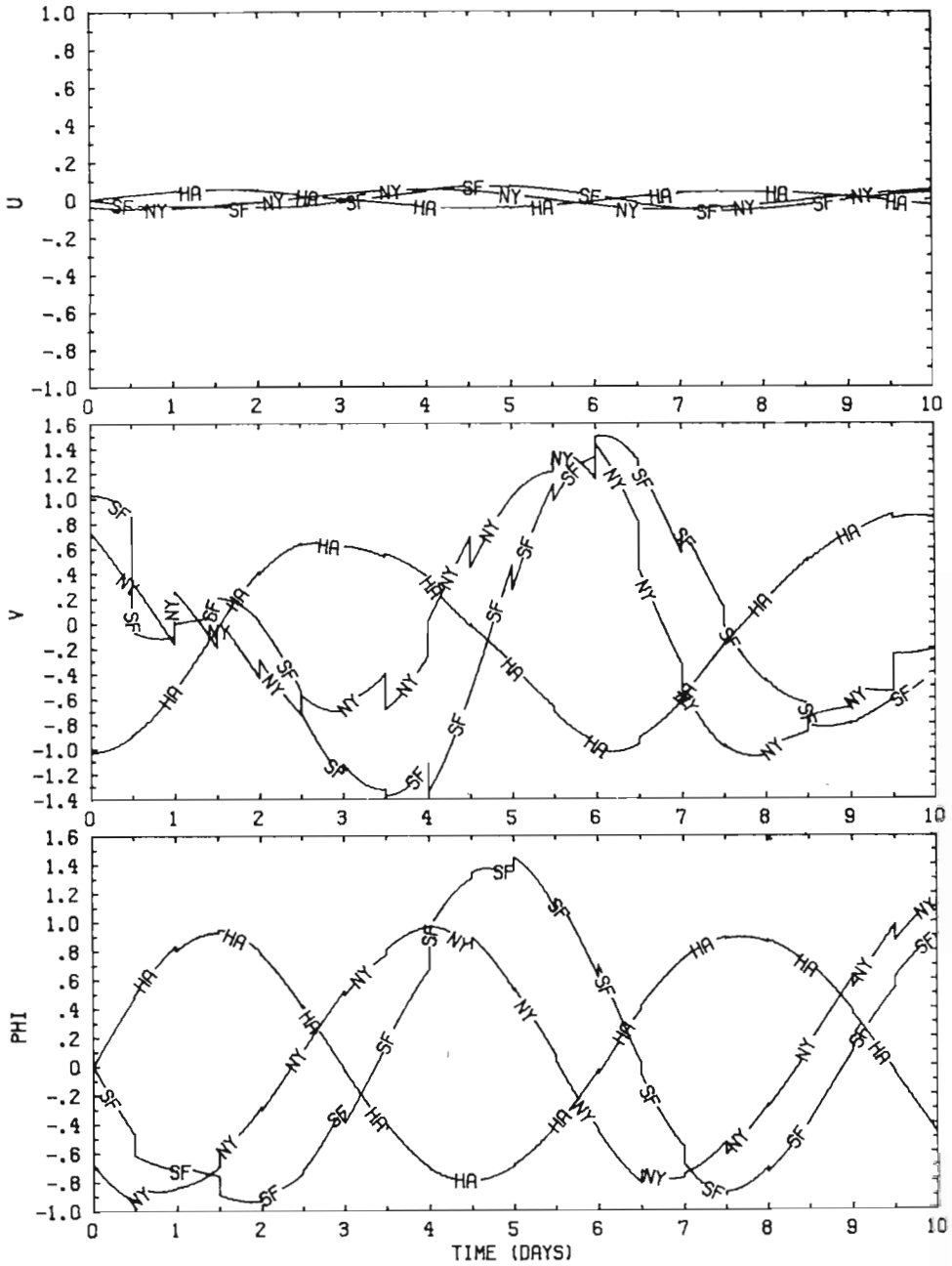


Fig.10 Same as Fig. 9, but for a run using the Π-filter, rather than the K-filter. The fast oscillations have been eliminated.

The larger u-component errors for the Π-filter are explained by the fact that the Π-filter allows almost no observational correction to be performed on the u-components: our slow wave subspace R has very small u-components (cf. Eq. (3.11)). The estimation errors in u produced by the system noise $\underset{\sim}{\zeta}$ cannot be counteracted therefore by the observations.

A run with the asymptotic form of the Π-filter, which is simply ΠK_∞, gave practically the same results as the Π-filter itself.

4.3 Theoretical Analysis of the Scalar Case

In order to help explain some of the qualitative features of the numerical results in Sec. 4.1 and 4.2, we perform an asymptotic analysis of the filtering equations for a scalar state x. The matrices of interest will now actually be scalars, and we assume $\Psi \neq 0$, $H = 1$, $Q \geq 0$, $R > 0$, and $P_0 > 0$. The positivity assumptions on Q, R, and P_0 are due to the fact that they are variances. We assume furthermore that the observation is performed only every r time steps; in the numerical experiments reported herein we would have $r = 24$, as $\Delta t = 30$ min. and observations are taken at the standard 12 hour synoptic intervals.

The filtering algorithm (2.11) yields in this case

$$P_k(-) = \Psi^2 P_{k-1}(+) + Q \ , \tag{4.5a}$$

$$P_k(+) = \begin{cases} P_k(-)R/(P_k(-)+R) \ , & \text{when } k=jr,\ j=1,2,3,\ldots, \\ P_k(-) \ , & \text{otherwise,} \end{cases} \tag{4.5b}$$

$$K_{jr} = P_{jr}(+)/R. \tag{4.5c}$$

From Eq. (4.5b), one immediately finds that

$$P_{jr}(+) \leq \min \{P_{jr}(-), R\}, \tag{4.5d}$$

an analogue of Eq. (2.3b). In particular, the estimation error variance drops below the observational noise level at each observation time, although it may grow in between

observation times.

Defining S_j,

$$S_j = P_{jr}(+) \ , \quad j = 0,1,2,\ldots,$$

to be the estimation error variance after the j^{th} observation, and the quantities

$$A = \Psi^{2r} \ , \quad B = \sum_{p=0}^{r-1} \Psi^{2p} \ ,$$

Eqs. (4.5a,b) can be converted into a nonlinear difference equation for S_j:

$$S_j = g(S_{j-1}) \ , \tag{4.6a}$$

where

$$g(s) = \frac{(As + BQ)R}{As + BQ + R} \ . \tag{4.6b}$$

To determine the asymptotic behavior of Eq. (4.6), and hence of the filter K, we now distinguish between the two cases $Q = 0$ and $Q > 0$. In the case of a perfect model, $Q = 0$, the solution of (4.6) may be written explicitly as

$$S_j = \frac{A^j(A-1)S_0R}{A(A^j-1)S_0 + (A-1)R} \ , \quad \text{if } |\Psi| \neq 1, \tag{4.7a}$$

or

$$S_j = \frac{S_0R}{jS_0 + R} \ , \quad \text{if } |\Psi| = 1 \ . \tag{4.7b}$$

As $j \to +\infty$, therefore,

$$S_j \to 0 \ , \quad \text{if } |\Psi| \leq 1 \ , \tag{4.7c}$$

$$S_j \to (1 - \tfrac{1}{A})R \ , \qquad \text{if } |\Psi| > 1. \tag{4.7d}$$

Eq. (4.7c) states that for stable dynamics, i.e., $|\Psi| \leq 1$, the estimation error variance, and hence the filter $K_{jr} = S_j/R$, always tend to zero. System (3.1) itself is conservative, while our difference scheme (3.6) is dissipative. Hence all eigenvalues of the difference scheme matrix Ψ have modulus less than or equal to unity. We are thus in the stable case (4.7c) and it is only to be expected that our estimation error covariance matrices and filter K_k approach zero in the absence of system noise. This is in accordance with our discussion of Fig. 2 in Sec. 4.1.

We wish to determine now the asymptotic nature of S_j in the case $Q > 0$, i.e., in case system noise is present. The quadratic equation

$$s = g(s) \tag{4.8}$$

has a positive discriminant, hence it has real roots. Its free term is negative, hence the roots are of opposite sign. Let S_+ denote the unique positive root of (4.8).

Notice that

$$dg/ds = AR^2/(As+BQ+R)^2 > 0 \ .$$

Therefore $g(s)$ is a monotone increasing function of s, with $g(0) > 0$, while its derivative is monotone decreasing

and tending to zero as $s \to +\infty$. It follows that the root S_+ of (4.8) is approached monotonically by the solutions of the recursion (4.6a) (Isaacson and Keller, 1966, Ch. 3.1, Fig. 2a),

$$S_j \to S_+ \quad \text{as} \quad j \to +\infty \ .$$

If S_0 is greater than (less than) S_+, then S_j will decrease (increase) monotonically to S_+. Since $K_{jr} = S_j/R$, we have also

$$K_{jr} \to S_+/R \quad \text{as } j \to +\infty,$$

the convergence being monotone as well. This is in accordance with the monotone decrease of trace $P_{jr}(+)$ in Fig. 6; in fact $P_{jr}(-)$ decreases also monotonically. Furthermore, S_+ is independent of $S_0 = P_0$, and hence the asymptotic filter $K_\infty = S_+/R$ is independent of P_0 as well.

To find an approximate value for S_+, we now assume that $|\Psi| < 1$ and $r \gg 1$, so that $A = \Psi^{2r} \ll 1$. Then the quadratic term in Eq. (4.8) is negligible and we have, approximately,

$$S_+ = \frac{QR}{Q + (1-\Psi^2)R} \ . \tag{4.9a}$$

It follows in general, by analogy with (2.3b) and (4.5d), that

$$S_+ \leq \min \{R, \ Q/(1-\Psi^2)\} \ ,$$

at least approximately. In particular, when the observational

error variance R is small enough, so that

$$(1-\Psi^2)R << Q \ , \qquad (4.9b)$$

then S_+ is roughly equal to R, and the size of Q has little influence. If, on the other hand, the observational error variance is large enough so that

$$Q << (1-\Psi^2)R \ , \qquad (4.9c)$$

then S_+ is roughly proportional to Q, and the size of R has little influence.

The two extreme cases (4.9b,c) explain much of the qualitative nature of the results in Sec. 4.2. Indeed, a matrix version of Eq. (4.9b) is satisfied over land, and we see that the expected m-s errors at observation times are approximately equal to the observational error variances. In fact, they are slightly smaller than R ; this is in accordance with (4.5d), as well as (4.9a,b). It also agrees with operational experience, as stated in Sec. 3.4.

Over the ocean, we can write $R = \infty$, so that a matrix version of (4.9c) is satisfied. Experiments with different magnitudes of Q (not shown), have confirmed that the size of Q is indeed the determining factor in the size of P over the ocean. This also agrees with operational experience: analysis error in data-poor regions is essentially equal to the error in forecasts from one analysis time (synoptic or subsynoptic) to the next.

The asymptotic properties of P and K discussed above have counterparts in the full, vector-matrix case, under the assumptions of complete observability and complete controllability. These assumptions concern properties of the matrices Ψ and H; they are satisfied for our model. The interested reader is referred to Bucy and Joseph (1968, Ch. 5) and to Jazwinski (1970, Sec. 7.6) for a full discussion of the general results.

5. CONCLUDING REMARKS

We have shown the ways in which the concepts and formalism of sequential estimation theory are relevant to the 4-D assimilation of meteorological data, by applying them to a simple model. The stochastic-dynamic model used for the illustration of the theory was governed by the linear shallow-water equations, including the β-effect of latitudinal changes in planetary vorticity. The dynamics of this model are similar to those of operational NWP models in that they admit as solutions slow, quasi-geostrophic Rossby waves, as well as fast inertia-gravity waves.

We have modified the standard Kalman-Bucy (K-B) filter in order to obtain optimal estimates of the slow, meteorologically significant waves, while eliminating entirely the fast, undesirable waves. In this way, our modified K-B filter achieves simultaneously the optimal 4-D assimilation of data and the initialization of model states for the purpose of noise-free forecasts.

It was shown that the optimal filter for the linear problem converges rapidly to an asymptotic matrix, the Wiener filter. Furthermore, the asymptotic filter (W-filter) performs nearly as well as the exact, time-varying filter (K-B filter). The Wiener filter depends on system dynamics Ψ, observational pattern H, system noise covariance Q and

observational noise covariance R; it does not depend on the initial errors, P_0. The rapid convergence and good performance of the W-filter for the linear problem hold hope for the filtering of nonlinear problems with similar dynamic and stochastic properties.

Nonlinear estimation theory is not completely understood mathematically, at least not in a practically applicable form. The filter which is most widely used in engineering applications is the extended K-B filter (EKF: Bucy and Joseph, 1968, Ch. 8; Gelb, 1974, Ch. 6; Jazwinski, 1970, Ch. 6 and Ch. 9). The principle of the EKF is simple: linearize the problem around an estimated state, and apply the corresponding linear filter over a time interval, T say, over which the solution of the nonlinear problem is not expected to change much. In large-scale NWP this time could equal 6 h to 24 h. After time T, relinearize around the new state and proceed. Clearly, T is limited by the dynamics of the system in the problem. When choosing T, a trade-off between accuracy and expediency has to be made.

The EKF gives good results when the true characteristic correlation time τ of the perturbations, which we model as white noise, i.e., $\tau = 0$, is actually short compared to T, $\tau \ll T$. This is certainly the case in NWP. Moreover, it appears from our experience with linear problems that:

a) the asymptotic filter for each one of the successive linearizations will work sufficiently well, and there is no need to compute time-varying filters over a time interval T;

b) the dependence of the W-filter on (linear) system dynamics Ψ is rather weak.

We conclude that the W-filter for the succeeding T-interval will be easily computable from the W-filter valid over the preceding T-interval by a perturbation procedure.

It is more realistic in NWP to let the observation matrix H be a function of time, $H = H(t)$, rather than a constant. Different observations are made at the synoptic times, 0000 GMT and 1200 GMT, the subsynoptic times, 0600 GMT and 1800 GMT, and in between. The distribution of observations, however, is not too far from being time-periodic. Let $H^*(t)$ be a periodic matrix function of time, with period 24 h, and assume that the actual observation matrix, $H(t)$, differs from $H^*(t)$ at every t only by a matrix of small rank, i.e., only by the presence, absence or location of relatively few observations.

The asymptotic filter corresponding to $(\Psi(t), H^*(t))$, $K^*(t)$ say, should also be periodic, with a period of 24 h, rather than constant. We expect some easily computed modification of $K^*(t)$ to be a good approximation to the optimal filter for $(\Psi(t), H(t))$. We plan to study, therefore, time-periodic observation patterns and their modifications.

Sequential estimation accounts explicitly for the fact that the system whose state we wish to estimate is governed by certain dynamics. It is this aspect of the theory which distinguishes it from the so-called "optimal interpolation" currently used in operational NWP. The latter essentially assumes that the system obeys trivial dynamics, $\Psi = I$.

We saw already that taking into account the dynamics allowed us to unify the assimilation and intialization aspects of preparing initial data for numerical weather forecasts. Furthermore, it allows us to account in a systematic way for the advection of information from data-rich to data-poor areas, and of "negative information", i.e., large errors, from data-poor to data-rich regions.

In particular, the theory shows that weighting coefficients for observations should be skewed in the direction of the prevailing winds, with larger weights upstream; the amount of skewness should depend on average wind intensity, i.e., on season. Also weights used on the Western edge of the continents should be different from those used on the Eastern edge. The present results on the anisotropy and inhomogeneity of estimation error structure in the zonal direction should also be supplemented by the results on inhomogeneity in the latitudinal direction of Ghil et al. (1980). It is this aspect of the theory which we expect to have the largest impact in terms of improving operational procedures.

One further aspect of the theory merits attention: the covariance matrices Q and R do not need to be prescribed a priori. They can be determined in the estimation process itself (e.g., Chin, 1979; Ohap and Stubberud, 1976) by using an adaptive filter. The determination of system noise would have important consequences for predictability theory, as well as for stochastic modifications of numerical schemes (Faller and Schemm, 1977). The determination of observational noise, eliminating the well known problem of "ground truth",

would also help greatly in improving operational objective analysis and data assimilation.

Acknowledgement.

The potential usefulness of estimation theory in meteorological applications was first pointed out to one of us (M.G.) by Edward S. Sarachik. It is a pleasure to acknowledge many useful discussions on linear and nonlinear filtering with George Papanicolaou. Practical aspects of the filters' eventual implementation in an operational environment were discussed with Wayman Baker, Mark Cane, Milton Halem and Eugenia Kalnay-Rivas. The final version of the manuscript benefited from comments by G. Cats, R. Daley, A. Hollingsworth, A. Lorenc and O. Talagrand, made during and after the Seminar. The manuscript could not have been prepared without the whole-hearted support of Constance Engle.

Our work on this problem has been supported by NASA Grants NSG-5034 and NSG-5130, administered by the Laboratory for Atmospheric Sciences, Goddard Space Flight Center. Computations were carried out on the CDC 6600 of the Courant Mathematics and Computing Laboratory, New York University, under Contract DE-AC02-76ER03077 with the U. S. Department of Energy.

References

Bengtsson, L., 1975: 4-Dimensional Assimilation of Meteorological Observations, GARP Publications Series, No. 15, World Meteorological Organization — International Council of Scientific Unions, CH-1211 Geneva 20, Switzerland, 76 pp.

Bergman, K. H., 1979: Multivariate analysis of temperatures and winds using optimum interpolation, Mon. Wea. Rev. 107, 1423-1444.

Browning, G., A. Kasahara and H.-O. Kreiss, 1979: Initialization of the primitive equations by the bounded derivative method, NCAR Ms. 0501-79-4, National Center for Atmospheric Research, Boulder, Colo. 80307, 44 pp.

Bube, K. P., and M. Ghil, 1980: Assimilation of asynoptic data and the initialization problem, this volume.

Bucy, R. S., and P. D. Joseph, 1968: Filtering for Stochastic Processes with Applications to Guidance, Wiley-Interscience, New York, 195 pp.

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.

Chin, L., 1979: Advances in adaptive filtering, in Control and Dynamic Systems, Vol. 15, C. T. Leondes (ed.), Academic Press, 278-356.

Curtain, R. F., and A. J. Pritchard, 1978: Infinite Dimensional Linear Systems Theory, Lecture Notes in Control and Information Sciences, Vol. 8, Springer-Verlag, New York, 297 pp.

Davis, M.H.A., 1977: Linear Estimation and Stochastic Control, Halsted Press, John Wiley and Sons, New York, 224 pp.

Faller, A. J., and C. E. Schemm, 1977: Statistical corrections to numerical prediction equations II, Mon. Wea. Rev. 105, 37-56.

Fleming, R. J., T. M. Kaneshige and W. E. McGovern, 1979a: The global weather experiment I. The observational phase through the first special observing period, Bull. Amer. Met. Soc. 60, 649-659.

_____ , _____ ,_____, and T. E. Bryan, 1979b: The global weather experiment II. The second special observing period, Bull. Amer. Met. Soc. 60, 1316-1322.

Gelb, A. (ed.), 1974: Applied Optimal Estimation, The M.I.T. Press, Cambridge, Mass., 374 pp.

Ghil, M., 1980: The compatible balancing approach to initialization, and four-dimensional data assimilation, Tellus 32, 198-206.

_____, and R. Mosebach, 1978: Asynoptic variational method for satellite data assimilation, in Halem et al. (1978), pp. 3.32-3.49.

_____, R. C. Balgovind and E. Kalnay-Rivas, 1980: A stochastic-dynamic model for global atmospheric mass field statistics, NASA Tech. Memo. 82009, NASA Goddard Space Flight Center, Greenbelt, Md. 20771, 50 pp.

_____, M. Halem and R. Atlas, 1979: Time-continuous assimilation of remote-sounding data and its effect on weather forecasting, Mon. Wea. Rev. 107, 140-171.

Halem, M., M. Ghil, R. Atlas, J. Susskind and W. J. Quirk, 1978: The GISS sounding temperature impact test. NASA Tech. Memo. 78063, NASA Goddard Space Flight Center, Greenbelt, Md. 20771, 421 pp. [NTIS N7831667].

Isaacson, E., and H. B. Keller, 1966: Analysis of Numerical Methods, Wiley, New York, 541 pp.

Jazwinski, A. H., 1970: Stochastic Processes and Filtering Theory, Academic Press, New York, 376 pp.

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

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

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems, Trans. ASME, Ser. D, J. Basic Eng., 82, 35-45.

_____, and R. S. Bucy, 1961: New results in linear filtering and prediction theory, Trans. ASME, Ser. D.: J. Basic Eng., 83, 95-108.

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

_____, 1980: Nonlinear normal mode initialization and quasi-geostrophic theory, J. Atmos. Sci., 37, 958-968.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion, Tellus, 21, 289-307.

McPherson, R. D., K. H. Bergman, R. E. Kistler, G. E. Rasch and D. S. Gordon, 1979: The NMC operational global data assimilation system, Mon. Wea. Rev., 107, 1445-1461.

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

Ohap, R. F., and A. R. Stubberud, 1976: Adaptive minimum variance estimation in discrete-time linear systems, in Control and Dynamic Systems, Vol. 12, C. T. Leondes (ed.), Academic Press, 583-624.

Palmén, E., and C. W. Newton, 1969: Atmospheric Circulation Systems, Academic Press, New York, 603 pp.

Parzen, E., 1960: Modern Probability Theory and Its Applications, Wiley, New York, 464 pp.

Pedlosky, J., 1979: Geophysical Fluid Dynamics, Springer-Verlag, New York, 624 pp.

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

_____, 1970: Algorithms for sequential and random observations, Meteorol. Mono., 11, 100-109.

_____, 1973a: Transient suppression in optimal sequential analysis, J. Appl. Meteor., 12, 437-440.

_____, 1973b: A comparison of the performance of quasi-optimal and conventional objective analysis schemes, J. Appl. Meteor, 12, 1093-1101.

_____, 1973c: Static and dynamic constraints on the estimation of space-time covariance and wavenumber-frequency spectral fields, J. Atmos. Sci., 30, 1252-1266.

_____, 1976: Linear sequential coding of random space-time fields, Info. Sci., 10, 217-241.

Phillips, N. A., 1971: Ability of the Tadjbakhsh method to assimilate temperature data in a meteorological system, J. Atmos. Sci., 28, 1325-1328.

_____, 1976: The impact of synoptic observing and analysis systems on flow pattern forecasts, Bull. Amer. Met. Soc., 57, 1225-1240.

Richtmyer, R. D., and K. W. Morton, 1967: Difference Methods for Initial-Value Problems, 2nd. ed., Wiley-Interscience, New York, 405 pp.

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

Schlatter, T. W., 1975: Some experiments with a multivariate stiatistical objective analysis scheme, Mon. Wea. Rev., 103, 246-257.

_____, G. W. Branstator and L. G. Thiel, 1976: Testing a global multivariate statistical objective analysis scheme with observed data, Mon. Wea. Rev., 104, 765-783.

_____, 1977: Reply (to Comments by H. J. Thiebaux on "Testing a global ..."), Mon. Wea. Rev., 105, 1465-1468.

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

Wiener, N., 1949: Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, Wiley, New York, 163 pp.

Appendix A. List of Major Symbols

$c_\ell^{(m)}$ for solutions of the continuous system (3.1), the three phase speeds, m = 1,2,3, corresponding to wave number ℓ, as defined by Eq. (3.3b). The same notation is used for the phase speeds of solutions of the discrete system (3.6), which agree with the phase speeds of solutions of the continuous system to order $(\Delta t)^2$.

E expectation, or ensemble averaging, operator

f Coriolis parameter, $f = 10^{-4}s^{-1} \cong 2\Omega \sin 45°$

G fast wave subspace

g gravitational acceleration constant of the Earth

H observation matrix; defines the observed linear combinations of state variables

j superscript indicating the spatial grid point, $x^j = j\Delta x$

K Kalman gain matrix, defined by (2.11c)

k subscript indicating the time, $t_k = k\,\Delta t$

L length of the computational domain, about half the circumference of the Earth at 45°N

ℓ wave number; $\ell L/2\pi$ is an integer

M number of grid points in the L-domain

n number of state variables, $n = 3M$; the dimension of $\underset{\sim}{x}$

$P_k(-)$ estimation error covariance matrix just prior to observations at time k

$P_k(+)$ estimation error covariance matrix just after observations at time k

p number of observations at a synoptic time; the dimension of $\underset{\sim}{z}$

Q covariance matrix of the system noise

R slow wave subspace

R covariance matrix of the observational noise

t time

Δt temporal increment in the discrete system (3.6)

U mean zonal wind component in the linearization of (3.2)

u perturbation zonal wind component

v perturbation meridional wind component

v_{max} initial amplitude of v

$\underset{\sim}{W}$ a solution $(u,v,\phi)^T$ of the continuous system (3.1)

$\underset{\sim}{W}_k^j$ a solution (u_k^j, v_k^j, ϕ_k^j) of the discrete system (3.6)

x distance along the spatial L-domain

Δx spatial increment in the discrete system (3.6)

$\underset{\sim}{x}$ "true" atmospheric state given by the stochastic model (2.5)

$\hat{\underset{\sim}{x}}_k(-)$ estimated atmospheric state just prior to observations at time k

$\hat{\underset{\sim}{x}}_k(+)$ estimated atmospheric state just after observations at time k

$\underset{\sim}{z}$ vector of observations, given by (2.6)

$\underset{\sim}{\zeta}$ observational noise, or random part of the observation model (2.6)

$\underset{\sim}{\xi}$ system noise, or random part of the atmospheric model (2.5)

Π orthogonal projection operator onto the slow wave subspace R

Φ mean geopotential at 45°N in the linearization of (3.2), i.e., g times the mean equivalent atmospheric height at 45°N

ϕ perturbation geopotential

ϕ_0 initial amplitude of ϕ

Ψ matrix defining the atmospheric dynamics, given for our model by Eqs. (3.6)

Ω angular rate of rotation of the Earth