Dynamics of Atmospheres and Oceans, 13 (1989) 171–218
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

METEOROLOGICAL DATA ASSIMILATION FOR OCEANOGRAPHERS. PART I: DESCRIPTION AND THEORETICAL FRAMEWORK *

M. GHIL

Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024-1565 (U.S.A.)

(Received March 1, 1988; revised August 10, 1988; accepted September 9, 1988)

ABSTRACT

Ghil, M., 1989. Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. Dyn. Atmos. Oceans, 13: 171–218.

The main theme of this article is the central role that dynamics plays in estimating the state of the atmosphere and of the ocean from incomplete and noisy data. The evolution of meteorological data usage in estimation and prediction is reviewed, from manual synoptic analysis through objective analysis to four-dimensional assimilation into a prediction model. Four-dimensional data assimilation tries to balance properly the roles of dynamical and observational information. Sequential estimation is presented as the proper framework for understanding this balance, and the Kalman filter as the optimal procedure for data assimilation.

The optimal filter computes forecast error covariances of a given atmospheric or oceanic model exactly, and hence data assimilation should be closely connected with predictability studies. A simple barotropic model of geophysical flow is used to illustrate the basic concepts. This model's results with simulated data are compared with the accumulated experience from full-scale meteorological models, assimilating real data.

The so-called initialization problem of eliminating fast, inertia-gravity waves in meteorological data assimilation is described within the context of sequential estimation. Improvements to operational schemes of data assimilation in current meteorological use are suggested, with a view to ocean applications.

1. INTRODUCTION AND MOTIVATION

The current and expected explosion in remotely sensed oceanographic data is ushering in a new age of physical oceanography. In this new age,

* Part II is in preparation.

0377-0265/89/$03.50

sophisticated theories about ocean circulation, developed up till now from very little field information, will be tested against the incoming data; new theories will be developed to explain the varied phenomena transpiring from the more plentiful data; and increased understanding of ocean circulation will be translated into prediction.

The data revolution knocking at the door will bring closer the daily practice of physical oceanography to that of dynamic meteorology. Even so, the number of oceanographic data to become available in the mid-1990s is expected to be smaller than that currently available in meteorology by a factor of ten, roughly speaking, after allowing for the difference in spatial and temporal scales. Moreover, oceanic data are and will be even less uniformly distributed than atmospheric data, with a preponderence of data at the surface and in the dynamically interesting parts of the world ocean, such as western boundary currents. A detailed comparison between the two sister disciplines with respect to data availability is given in the Appendix.

This comparison, well known in its essentials to the reader, indicates that current and expected data in oceanography fall far short of complete, uniform and accurate coverage of the mass and velocity fields throughout the world ocean's width and depth. To compensate for these shortcomings in the data actually observed, it is necessary to call upon the accumulated experience of past data, and the theoretical knowledge distilled from this experience, a knowledge best incorporated into numerical models of oceanic flow.

As in meteorology, flow models can be used to assimilate the data, creating a dynamically consistent, complete and accurate 'movie' of the oceans in motion. One key problem is how to determine variables not directly observed, such as the velocity components, from the observed variables, such as surface height or wind stress. The other key problem is how to use information in one part of the ocean, at the surface or in a western boundary current, in order to infer the state of the other parts, at depth and throughout a subtropical gyre, say.

The answers to these two problems lie, as we shall see, in the dynamical coupling between variables, for the one, and the advection of information with the flow, for the other. This is the central role that dynamics plays in estimating the state of the ocean from incomplete data. The other side of the coin is that numerical models are not and never will be perfectly accurate representations of the ocean's large-scale motions. Both models and data have errors; hence the need to properly balance dynamical and observational information.

Meteorological data usage can thus provide some guidance to oceanographers, first in what to do with their new data, and then in how to do it better than the meteorologists, who are still much richer in data. Hence, the

purpose of this article is twofold: (i) to provide an introduction to the current operational practice of data assimilation in numerical weather prediction (NWP), and how this practice was arrived at; and (ii) to outline a theoretical framework for the evaluation of data assimilation methods, and for the development of better methods. The two topics will not be treated in complete separation from each other, but rather as intertwined threads.

In section 2, the history of data usage in meteorology is outlined, and a number of methods for combining data with models are briefly mentioned. In section 3, the mathematical framework of estimation theory is presented, with an emphasis on sequential estimation as conceptually closest to the spirit of meteorological data assimilation. More general aspects of estimation, and connections with control theory, are postponed for Part II.

In section 4, sequential estimation is illustrated for a one-dimensional (1-D) shallow-water model. In section 5, the optimal filter given by sequential estimation theory, the Kalman filter, is modified to account for the two types of waves present in primitive-equation (PE) models, fast and slow, and to eliminate the fast, undesirable modes. This paper concludes with brief comments on practical aspects of operational data assimilation, not covered by the theoretical framework presented here.

Part II will be published soon. It starts with a quick review of the theoretical results so far. In its section 2, so-called 'optimal interpolation' (OI), the method currently in widest use in NWP, is shown to be a particular suboptimal filter within the broader framework of sequential estimation, and ways to improve it are proposed. In section 3, computational considerations are raised, and an application to a two-dimensional (2-D) shallow-water model is reviewed, with operation counts.

The question of adaptive determination of observational error is addressed in section 4, and the determination of model error is shown to be related to the predictability question. Sequential estimation for non-linear flow equations is discussed in section 5. Conclusions for the upcoming era of oceanographic data assimilation follow in section 6 of Part II.

2. DATA ANALYSIS AND DATA ASSIMILATION IN METEOROLOGY

Plentiful data on a routine basis are expensive to acquire, in both meteorology and oceanography, and for the price of these data society is entitled to useful as well as interesting information. Useful information is not limited to the climatological or current state of the atmosphere or the ocean, but extends to a prediction of their future state. A qualitative understanding of the geofluid is thus necessary but not sufficient for our purposes, and a quantitative estimate of its state in the past and present, as well as quantitative prediction of future states, is required. The estimate of

the present state is a prerequisite for future prediction, and the accuracy of past prediction is essential for an accurate estimate of the present.

How does the estimation of the present proceed in meteorology? A good starting point for an answer is Wiener's (1956) article on prediction and dynamics. At the time of his writing, meteorology, like econometrics, could still be considered a semi-exact science, as opposed to the allegedly exact science of celestial mechanics (Horton et al., 1983). Dynamical processes in the atmosphere were still poorly known, while observations were sparse in space and time as well as inaccurate.

Relying theoretically on the hope for the system's ergodicity and stationarity, and therewith on certain related theorems of Birkhoff and von Neumann, Wiener argued that the best approach to atmospheric estimation and prediction was statistical. In practice, this meant ignoring any quantitative dynamical knowledge of system behavior, requiring instead a complete knowledge of the system's past history and using the Wiener–Hopf filter to process this infinite but inaccurate information into yielding an estimate of present and future (Wiener, 1949).

During roughly the same period, synoptic meteorologists were actually producing charts of atmospheric fields at present and future times guided by tacit principles similar to those explicitly formulated by Wiener. The main tool was smooth interpolation and extrapolation of observations in space and time. Still, rudimentary but quantitative dynamical knowledge was incorporated into these estimates of atmospheric states, to wit, the geostrophic relation between winds and heights, and the advection of large-scale features by the prevailing winds. Some analogies with the state of upper ocean dynamics just a short while ago are already emerging from this description.

The first step into the present period of estimation in meteorology was objective analysis, which replaced manual, graphic interpolation of observations by automated, mathematical methods, such as two-dimensional (2-D) polynomial interpolation (Panofsky, 1949). Not surprisingly, this step was largely motivated by the use of then rapidly improving knowledge of atmospheric dynamics to produce numerical weather forecasts (Charney et al., 1950). The word 'analysis' is derived in this context from the traditional practice of identifying the result of the synoptician's analysis of the data, a result plotted manually as a chart or contour map, with the analysis process itself.

The main ideas underlying objective analysis were statistical (Eliassen, 1954; Gandin, 1963; Phillips, 1976). Observations are considered to sample a random field, with a given spatial covariance structure which is predetermined and stationary in time. This generalizes, in fact, the ideas of Wiener (1956) from a finite-dimensional system governed by ordinary dif-

ferential equations (ODEs) to an infinite-dimensional one governed by the partial differential equations (PDEs) of geophysical fluid dynamics (GFD). In practice, these statistical ideas appeared too complicated and computationally expensive at the time to be adopted, as they stood, into the fledgling NWP process. Instead, various short cuts, such as the successive-correction method (SCM) were implemented in the operational routine of weather bureaus (Cressman, 1959).

Two related developments led to the next step, in which the connection between statistical interpolation, in the broadest sense, on the one hand, and dynamics, on the other, became apparent and started to be used systematically. One development was the increasingly accurate nature of numerical weather forecasts; the other was the advent of time-continuous, space-borne observing systems. Together, they produced the concept of four-dimensional (4-D) space–time continuous data assimilation in which a model forecast of atmospheric fields is sequentially updated with incoming observations (Charney et al., 1969; Smagorinsky et al., 1970; Rutherford, 1972). Here the model carries forward in time the knowledge of a finite number of past observations, subject to the appropriate dynamics, to be blended with the latest observations.

The design of the Global Atmospheric Research Program (GARP) in the late 1960s (Charney et al., 1966) raised for the first time in GFD the question of trade-offs between observed and unobserved variables, on the one hand, and between sampling density in space and time, on the other. As we shall see in the subsequent sections, and in Part II, such questions can be best formulated as observability questions about the governing equations with respect to a proposed observing system. These are precisely the questions that also need to be addressed by the presently contemplated global programs in oceanography.

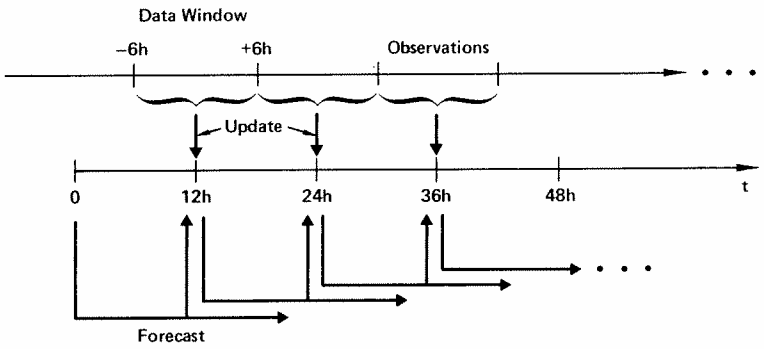


Fig. 1. Operational cycle of a weather service which combines the forecasting and data assimilation processes.

Combining the 4-D assimilation of the new satellite, aircraft and drifting buoy data with the usual objective analysis of the earlier, conventional data from radiosondes, ships and land stations (see Fig. A1) led to an interesting realization. In fact, NWP operations had, of necessity, combined dynamics with observations all along in determining the state of the atmosphere at all times, and in particular at those times from which forecasts had to be issued. This is illustrated in Fig. 1.

Any weather bureau carries out two processes in parallel: one is the numerical forecast, from a particular moment in time, or epoch, which we shall call initial time; the other is the 4-D assimilation of incoming data in order to estimate as well as possible the state of the atmosphere at the next epoch from which a forecast has to be issued. The process illustrated in the figure is intermittent updating, in which all data within a certain interval, or window, are used together at the same epoch to update the state of the system as forecast by the NWP model (Bengtsson, 1975).

Forecasts are typically started at so-called synoptic times, 00 GMT (Greenwich Mean Time) and 1200 GMT, in which case a 12-h assimilation cycle with ± 6-h windows is used. The subsynoptic times 0600 GMT and 1800 GMT also intervene when using a 6-h cycle with ± 3-h windows. At analysis or update times the numerical forecast is first verified against the new data, then it is combined or blended with them, i.e., the data are assimilated into the model, and finally a new forecast is issued from the newly estimated state of the atmosphere.

The new estimate is clearly based on the past observations, as carried forward in time by the model, and on the current observations. Each forecast has to extend at least up to the next update time, but certain forecasts, issued at selected epochs, e.g., every 24 h, can extend beyond it, for 24 h, 48 h, 72 h or even 10 days.

The intermittent updating process described so far was entirely appropriate as long as most data where taken, by international agreement, at the same time, in order to provide a 'synopsis' of the global weather, hence synoptic times and synoptic maps. With the advent of satellite data, time-continuous data assimilation, i.e., in practice at every model time step, became possible (Ghil et al., 1979).

In any case, considerable interest developed throughout the 1970s in objective analysis and data assimilation methods, in preparation for the First GARP Global Experiment (FGGE), later relabeled the Global Weather Experiment (GWE). Table I lists the methods in use by 10 advanced weather services at the end of the decade. These methods will not be reviewed extensively here, because of lack of space. The SCM, polynomial interpolation and statistical methods have already been mentioned, with references. Among statistical methods, univariate interpolation refers to linear regres-

TABLE I

Characteristics of data assimilation schemes in operational use at the end of the 1970s (after Gustavsson, 1981)

Organization or country	Operational analysis methods	Analysis area	Analysis/ forecast cycle
Australia	Successive correction method (SCM)	SH [c]	12 h
	Variational blending techniques	Regional	6 h
Canada	Multivariate 3-D statistical interpolation	NH [c] Regional	6 h (3 h for the surface)
France	SCM; wind-field and mass-field balance through first guess	NH	6 h
	Multivariate 3-D statistical interpolation	Regional	
Federal Republic of Germany	SCM; upper-air analyses are built up, level by level, from the surface	NH	12 h (6 h for the surface)
	Variational height/wind adjustment		Climatology only as preliminary fields
Japan	SCM	NH	12 h
	Height-field analyses are corrected by wind analyses	Regional	
Sweden	Univariate 3-D statistical interpolation	NH	12 h
	Variational height/wind adjustment	Regional	3 h
United Kingdom	Hemispheric orthogonal polynomial method	Global	6 h
	Univariate statistical interpolation (repeated insertion of data)		
U.S.A.	Spectral 3-D analysis	Global	6 h
	Multivariate 3-D statistical interpolation	Global	
U.S.S.R.	2-D [b] statistical interpolation	NH	12 h
ECMWF [a]	Multivariate 3-D statistical interpolation	Global	6 h

[a] European Centre for Medium Range Weather Forecasts.
[b] 2-D are in a horizontal plane.
[c] Southern Hemisphere (SH) and Northern Hemisphere (NH).

sion for each meteorological variable separately. Multivariate interpolation attempts to take into account the dynamical, non-linear coupling between field variables at least in a linearized, statistical sense.

The only methods in Table I not explicitly mentioned so far are variational (Sasaki, 1958). It is clear, however, that noisy, inaccurate data should not be fitted by exact interpolation, but rather by a procedure which has to achieve two goals simultaneously: (i) to extract the valuable information contained in the data, and (ii) to filter out the spurious information, i.e., the noise. Thus the analyzed field should be close to the data, but not too close.

The statistical approach to this problem is linear regression, which in the presence of certain constraints is sometimes called ridge regression. The variational approach is to minimize the distance, e.g., in a quadratic norm, between the analyzed field and the data, subject to constraints that yield a smoother result. The connection between these two approaches in a stationary, ergodic context is intuitively obvious, and is reflected in the fact that root-mean-square (r.m.s.) minimization is used in popular parlance for both approaches.

The exact correspondence between variational weights and covariance matrices follows from a duality result of Kimeldorf and Wahba (1970). The polynomial interpolation referred to earlier (Panofsky, 1949) and in Table I is in fact a variational method with analyzed fields prescribed to be of 2-D polynomial form, with the total number of coefficients much smaller than the number of data points; this ensures smoothness of the analyzed field without imposing any dynamic constraints.

The analysis method in widest operational use today in NWP is a particular form of statistical interpolation, commonly referred to as optimal interpolation (OI: Lorenc (1981) and McPherson et al. (1979)). OI is described within the broader context of estimation theory in section 2 of Part II. A particular implementation of variational ideas, using the equations of motion as a strong constraint, is also being considered at present by some weather services (Courtier and Talagrand, 1987). This implementation by the adjoint method is discussed in Part II in the context of an important duality result in estimation and control theory.

General reviews of meteorological analysis and assimilation methods are Bengtsson (1975), Bengtsson et al. (1981), Bourke et al. (1985), Hollingsworth (1987), Thiébaux and Pedder (1987) and Williamson (1982). Brief unifying treatments are given by Lorenc (1986), Phillips (1982) and Wahba (1982). The present article provides its own unifying point of view, that of sequential estimation. Estimation and control theory deals with the solutions of randomly perturbed systems of differential equations, subject to various distributions of data and to different types of forcing. It permits therefore a proper understanding of the combination between imperfectly known dy-

namics and inaccurate observations, as well as an evaluation of various algorithms for weighting incoming data against a forecast based on past data (Ghil, 1986).

Rather than present this theory in all its generality, I use a simple model of geophysical flow to bring out the salient aspects, as they apply to data assimilation for the atmosphere and oceans. These points are then buttressed by comparison with accumulated experience from full-scale NWP models and real meteorological data. In the course of the presentation, relations to other points of view, such as Bayesian estimation (Lorenc, 1986), OI with variational constraints (Phillips, 1982), regularized spline interpolation (Wahba, 1982), and variational minimization by the adjoint method (Talagrand and Courtier, 1987) should become reasonably clear.

This article is loosely based on Ghil et al. (1981, 1982, 1983) and on Cohn (1982). Of these primary references, only Ghil et al. (1981) has appeared in the open literature. Its results are presented here for tutorial purposes, and are greatly abridged. The additional results of Cohn (1982) and of Ghil et al. (1982, 1983) are given greater emphasis, as are the comparisons with NWP models and real data (Ghil et al., 1979; Halem et al., 1982; Balgovind et al., 1983; and others). Part II is based on recent reports and unpublished work leading up to the present. The two papers also address specific issues of data assimilation not covered by the preceding references, with a special emphasis on future applications to physical oceanography.

3. ESTIMATION THEORY AND DATA ASSIMILATION

3.1. Multiple measurements, estimation and minimization

The previous two sections have given a feeling for the need to combine statistics and dynamics in using incomplete and inaccurate information to estimate the state of the geofluid. To develop an ability to do so, however, requires a few steps, the last of which are non-trivial.

To begin, take two independent measurements y and z of a quantity x, such as the temperature of the mixed layer at a point; it is natural to seek an estimate of x, call it $\hat{x}$, in the linear form

$$\hat{x} = \alpha_1 y + \alpha_2 z \tag{1}$$

We assume that the two measuring instruments are unbiased

$$Ey = Ez = Ex \tag{2a}$$

E is the expectation operator, i.e., it represents the 'mean' or 'average' of an infinite number of measurements. Requiring the estimate itself to be unbi-

ased, $E\hat{x} = Ex$, immediately implies that $\alpha_1 + \alpha_2 = 1$ and hence eqn. (1) can be rewritten as

$$\hat{x} = y + \alpha_2(z - y) \tag{1'}$$

We also assume that the measurement errors are uncorrelated

$$E(y - x)(z - x) = 0 \tag{2b}$$

and that their variances σ_1^2 and σ_2^2 are known

$$\sigma_1^2 := E(y - x)^2 \tag{2c}$$

$$\sigma_2^2 := E(z - x)^2 \tag{2d}$$

the special equality sign ' := ' is used for defining identities.

The optimal linear unbiased estimate of x is given by choosing α_2 and $\alpha_1 = 1 - \alpha_2$ so as to minimize the variance

$$\sigma^2 := E(\hat{x} - x)^2 \tag{3}$$

of the estimation error. The required minimum is achieved by choosing

$$\alpha_1 = \hat{\sigma}^2/\sigma_1^2 \tag{4a}$$

$$\alpha_2 = \hat{\sigma}^2/\sigma_2^2 \tag{4b}$$

here $\hat{\sigma}^2$ is just the variance of the optimal estimate given by

$$\hat{\sigma}^{-2} = \sigma_1^{-2} + \sigma_2^{-2} \tag{4c}$$

The weights α_1 and α_2 thus reflect the relative uncertainties in y and z, respectively, and $\hat{\sigma}^2$ is smaller than both σ_1^2 and σ_2^2. In fact, it is convenient to call accuracy $A := \sigma^{-2}$ the inverse of the variance of a random variable. With this terminology, eqn. (4c) states that the accuracy of a linear, unbiased, optimal estimate equals the sum of the accuracies of unbiased, mutually uncorrelated measurements.

Formally, the variational approach to estimating $\hat{x}$ requires one to minimize

$$J := \beta_1(x - y)^2 + \beta_2(x - z)^2 \tag{5}$$

for arbitrary weights β_1 and β_2. The result will be the same, eqn. (1) with eqn. (4), provided

$$\beta_1 = \alpha_1^2 \tag{6a}$$

$$\beta_2 = \alpha_2^2 \tag{6b}$$

Problem (5) appears to be simpler, since no statistical assumptions (eqn. (2a–d)) need to be made. But eqn. (6) shows that some access to information like eqn. (2c and d) is required.

As discussed by Lorenc (1986), Wahba (1982), and their references, this information can also be retrieved variationally, given a distribution of observations in space or time, which, by ergodicity, is equivalent to a distribution in probability space. The variational problem has to be reformulated for these purposes as minimizing

$$J' := (x-y)^2 + (x-z)^2 + \lambda x^2 \tag{7}$$

and the regularization or smoothing parameter λ has to be determined from the data. This can be done by a resampling scheme (Efron, 1982), such as the bootstrap, (generalized) cross-validation, or the jack-knife. The result should still be the same, to within sampling error.

At this point, it becomes clear that preference for the statistical approach (eqns. (1) and (4)) or the variational approach (eqn. (7)) hinges on computational considerations. The relative efficiency of numerical algorithms derived from either approach cannot be determined from our little two-measurement example, and so we move on to the next step in this section's exposition of estimation theory.

3.2. Sequential estimation and optimal data assimilation

Estimation theory deals with the solutions of randomly perturbed systems of differential equations, ODEs or PDEs, as determined from noisy data distributed arbitrarily in space and time. This is an extremely active area of research in engineering and in mathematics, as witnessed by the large number of periodicals and books (e.g., Bucy and Joseph, 1987; Jazwinski, 1970; Gelb, 1974) dedicated to it.

It is entirely sufficient for our present purposes to consider the ODE, or lumped-parameter case, in discrete time, since any numerical model of the atmosphere or ocean has to be presented in such a finite form to modern computational devices (Ghil et al., 1981). Furthermore, in the next few sections we shall deal with linear flow equations. Non-linear models are taken up in section 5 of Part II, where it is shown that the quadratic, advective non-linearities of GFD are well handled by the successive linearizations associated with the so-called extended Kalman filter (EKF).

With these caveats, and the insight gained from the previous subsection, a linear unbiased data assimilation scheme for the geofluid can be written as

$$\mathbf{w}_k^{\mathrm{f}} = \boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^{\mathrm{a}} \tag{8a}$$

$$\mathbf{w}_k^{\mathrm{a}} = \mathbf{w}_k^{\mathrm{f}} + \mathbf{K}_k\left(\mathbf{w}_k^{\mathrm{o}} - \mathbf{H}_k\mathbf{w}_k^{\mathrm{f}}\right) \tag{8b}$$

The state vector $\mathbf{w}$ represents all model variables, such as temperature and velocity components, at a set of grid points or in the form of spectral

coefficients; the forecast model (eqn. (8a)) is advanced in discrete time steps Δt, $\mathbf{w}_k = \mathbf{w}(t_k)$, $t_k = k\Delta t$; superscript f stands for the forecast, o for observations and a for the analysis; Ψ is the system matrix, describing its (at first linear) dynamics; and $\mathbf{H}_k$ is the observation matrix.

The observation vector $\mathbf{w}_k^o$ has dimension $p_k \ll N$, where N is the dimension of $\mathbf{w}_k^f$ and $\mathbf{w}_k^a$. The matrix $\mathbf{H}$ represents the fact that only certain variables or combinations thereof are observed, at a set of points much smaller than the total number of grid points (see Figs. A1 and A3). Thus remote soundings of radiance by polar-orbiting satellites combine atmospheric temperatures, or tomographic soundings of acoustic travel times combine oceanic densities. Matrix $\mathbf{H}$ also represents the interpolation of grid values to data location for a grid-point model and (inverse) spectral transforms to physical space for a spectral model. This being said, I shall concentrate hereafter for simplicity on grid-point, finite-difference models.

The vector $\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f$ contains the new information provided by the data. It is called innovation vector in the engineering literature and observational residual in the meteorological literature.

Equation (8b) has the form of eqn. (1′), with $y = \mathbf{w}_k^f$, $z = \mathbf{w}_k^o$ and $\alpha_2 = \mathbf{K}_k$. The conceptual difference between eqns. (8) and (1) is that $\mathbf{w}_k^f$ represents past observations, and the practical difference is that $p_k \neq N$, i.e., $\mathbf{H}_k$ is not square, and it may have a different size at each time step.

In fact, all operational data assimilation schemes have the form of eqn. (8b), whether the model (eqn. (8a)) is linear or non-linear. Existing assimilation schemes, such as successive corrections or OI, differ from each other by the weight matrix $\mathbf{K}_k$, and we wish to find the optimal $\mathbf{K}_k$, in a precise sense to be defined forthwith; in the engineering literature, $\mathbf{K}_k$ is often called the gain matrix.

To be precise, we need a well-defined set of assumptions, and here they are. First the true evolution of the geofluid, $\mathbf{w}_k^t$, is governed by

$$\mathbf{w}_k^t = \Psi_{k-1}\mathbf{w}_{k-1}^t + \mathbf{b}_{k-1}^t \tag{9a}$$

where $\mathbf{b}_k^t$ is a (Gaussian) white-noise sequence, i.e.

$$E\,\mathbf{b}_k^t = 0 \tag{9b}$$

$$E\,\mathbf{b}_k^t\left(\mathbf{b}_l^t\right)^{\mathrm{T}} = \mathbf{Q}_k\delta_{kl} \tag{9c}$$

δ_{kl} is the Kronecker delta, and superscript T indicates the transpose (of a column vector, in this case); $\mathbf{b}_k^t$ is called the system noise.

No difficulty arises by adding a deterministic forcing $\bar{\mathbf{b}}_k$ to the governing eqn. (8a), so that eqn. (9b) becomes

$$E\,\mathbf{b}_k^t = \bar{\mathbf{b}}_k \neq 0 \tag{9b′}$$

Forcing is, in fact, more important for oceanic flows than for atmospheric ones. But in a linear problem one can always separate the particular, forced solution from the homogeneous one. It is the lack of complete initial data for the latter that we wish to compensate for by observations distributed in time. Deterministic forcing will be reintroduced when considering the duality between deterministic control and stochastic estimation in Part II.

Current NWP models are in fact close to perfect, in the sense that their error is almost white, in space and time. Balgovind et al. (1983) have shown that the potential vorticity error of the second-order accurate NWP model then in use at the NASA Goddard Laboratory for Atmospheres (GLA), verified at 24 h and 36 h against a special set of satellite and conventional data comparable to that in Fig. A1, is essentially random, stationary in time, and nearly white in space. Their results are consistent, at least, with an error equation forced by temporally-white noise (Balgovind et al., 1983, their eqns. (2.1–2.5)), which leads to an accumulated system noise covariance growing linearly in time.

In meteorological OI it is commonly assumed that the spatial correlations of mass-field forecast errors are constant in time, while the variances are constant in space and grow linearly in time (McPherson et al., 1979; Lorenc, 1981). This linear growth of variances is typical of Brownian motion with so-called independent increments, i.e., driven by temporally-white noise. This is again consistent with the error of state-of-the-art NWP models being close to white. Numerical models of the ocean are not quite at this stage, but systematic errors can and will be eliminated by physical insight, numerical trial and error, and by applying systematically estimation and control theory (Part II).

The second assumption used in optimizing the weight matrix $\mathbf{K}_k$ concerns the error model for the observations

$$\mathbf{w}_k^o = \mathbf{H}_k \mathbf{w}_k^t + \mathbf{b}_k^o \tag{10a}$$

where $\mathbf{b}_k^o$ is the observational noise. One assumes that $\mathbf{b}_k^o$ is also a (Gaussian) white-noise sequence

$$E\mathbf{b}_k^o = 0 \tag{10b}$$

$$E\mathbf{b}_k^o (\mathbf{b}_l^o)^T = \mathbf{R}_k \delta_{kl} \tag{10c}$$

For convenience of the presentation, it is assumed furthermore that system noise and observational noise are uncorrelated with each other

$$E\mathbf{b}_k^f (\mathbf{b}_k^o)^T = 0 \tag{11}$$

The assumption (eqn. (10c)) of observational errors being uncorrelated in time with each other, as well as that of their being uncorrelated with forecast

errors (eqn. (11)), needs further examination. Temperature observations derived from satellite radiances can have errors correlated from one time step to the next, violating assumption (10c). If forecast temperatures are used as 'first guesses' in the inversion of the non-linear radiative transfer equation (Susskind et al., 1984), then assumption (11) might also be violated. Similar considerations will arise in oceanographic prediction systems deriving density and velocity observations from different versions of acoustic tomography (Cornuelle et al., 1985), for instance.

It is fortunate, therefore, that neither of these two assumptions is crucial to the subsequent results. Correlations between system noise $\mathbf{b}_k^{\mathrm{t}}$ and observational noise $\mathbf{b}_k^{\mathrm{o}}$ can be handled by the formulation of an equivalent problem of the same vector size, for which assumption (11) is satisfied (Gelb, 1974, p. 124, eqns. (4.3.15–4.3.18)). Correlated measurement errors can be handled by careful state-vector augmentation or, better even, by measurement-differencing techniques (Gelb, 1974, section 4.5, especially p. 136 and reference given there); the latter lead to a problem of the same size, and correspond essentially to subtracting from each new observation the part correlated with the previous observation.

Assumptions (9–11) are thus valid either for the model equations derived directly from physical principles, or for the equations modified slightly as discussed above. These assumptions permit us to derive the evolution in time of the error covariance matrices

$$\mathbf{P}_k^{\mathrm{f,a}} := E\left(\mathbf{w}_k^{\mathrm{f,a}} - \mathbf{w}_k^{\mathrm{t}}\right)\left(\mathbf{w}_k^{\mathrm{f,a}} - \mathbf{w}_k^{\mathrm{t}}\right)^{\mathrm{T}} \tag{12}$$

of the forecast $\mathbf{w}_k^{\mathrm{f}}$ and the analysis $\mathbf{w}_k^{\mathrm{a}}$, respectively. This evolution follows from eqns. (8), (9a) and (10a), using eqns. (9b and c), (10b and c) and (11), and it is governed by

$$\mathbf{P}_k^{\mathrm{f}} = \boldsymbol{\Psi}_{k-1}\mathbf{P}_{k-1}^{\mathrm{a}}\boldsymbol{\Psi}_{k-1}^{\mathrm{T}} + \mathbf{Q}_{k-1} \tag{13a}$$

$$\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^{\mathrm{f}}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}} \tag{13b}$$

Hence, by advancing $\mathbf{P}_k^{\mathrm{f,a}}$ along with $\mathbf{w}_k^{\mathrm{f,a}}$, one can know how well the true state $\mathbf{w}_k^{\mathrm{t}}$ is estimated, for any weight matrix $\mathbf{K}_k$, i.e., for any data assimilation scheme, operational or contemplated. Knowledge of the estimation error for arbitrary $\mathbf{K}$ permits one to determine the optimal $\mathbf{K}_k$, which minimizes this error.

There are two problems that arise at this point. First and foremost, the computational complexity of advancing in time the error covariance matrices. While eqn. (8a and b) represent $0(N)$ computations per time step, eqn. (13a and b) represent, at face value, $0(N^2)$ computations. This is quite tolerable for typical engineering applications with $N \leqslant 1000$ say, but prohibitively expensive for atmospheric and oceanic prediction or simulation models with

$N \geqslant 10^5$. In section 3 of Part II it is shown that, by exploiting special features of the dynamics matrix Ψ and the covariance matrix $\mathbf{P}$ which arise in the latter applications, the operation count can be reduced to $0(N)$, i.e., it can be made comparable to that for currently operational, less sophisticated data assimilation methods.

Second, the noise covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ are assumed to be known in the subsequent derivation of the optimal $\mathbf{K}_k$. This is not so in practice, and finding the actual magnitude of system errors and observational errors is an important function of the data assimilation process. A computationally efficient, adaptive filter to do this is described in section 4 of Part II.

The optimal weight matrix $\mathbf{K}_k$ at each time step is obtained by minimizing the expected mean-square (m.s.) estimation error

$$J \equiv \operatorname{tr} \mathbf{P}_k^{\mathrm{a}} := E\left(\mathbf{w}_k^{\mathrm{a}} - \mathbf{w}_k^{\mathrm{t}}\right)^{\mathrm{T}}\left(\mathbf{w}_k^{\mathrm{a}} - \mathbf{w}_k^{\mathrm{t}}\right) \tag{14}$$

This is done by using eqn. (13b) for the matrix $\mathbf{P}_k^{\mathrm{a}}$ and setting the derivative of J with respect to $\mathbf{K}_k$ equal to zero. The crucial ingredient is the identity for the derivative of the trace of a product of matrices

$$\frac{\partial}{\partial \mathbf{A}} \operatorname{tr}(\mathbf{A}\mathbf{B}\mathbf{A}^{\mathrm{T}}) = 2\mathbf{A}\mathbf{B}$$

where $\mathbf{B}$ is symmetric; this applies to $\mathbf{P}_k^{\mathrm{a}}$ in (13b), since both $\mathbf{P}_k^{\mathrm{f}}$ and $\mathbf{R}_k$ are symmetric by definition.

A unique, absolute minimum is attained for

$$\mathbf{K}_k = \mathbf{K}_k^* := \mathbf{P}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}_k^{\mathrm{f}}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1} \tag{15}$$

The linear unbiased data assimilation scheme given by eqn. (8a and b), with the optimal gain matrix $\mathbf{K}_k^*$ in eqn. (15), is called the Kalman filter (Kalman, 1960). Its continuous-time counterpart is often called the Kalman–Bucy filter (Kalman and Bucy, 1961).

To complete the analogy between the Kalman filter and the two-measurement example of subsection 3.1, it is useful to rewrite eqns. (13b) and (15) as

$$\left(\mathbf{P}_k^{\mathrm{a}}\right)^{-1} = \left(\mathbf{P}_k^{\mathrm{f}}\right)^{-1} + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k^{\mathrm{T}} \tag{16a}$$

$$\mathbf{K}_k^* = \mathbf{P}_k^{\mathrm{a}}\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1} \tag{16b}$$

It then becomes clear that eqn. (16a and b) is the counterpart of eqn. (4a–c), i.e., the weight given to the current observations is inversely proportional to their variance, and the accuracy of the analysis is the sum of the accuracies of the forecast, based on the past observations, and of the current observations.

The formula (eqn. (13b)) for $\mathbf{P}_k^a$ can be simplified when $\mathbf{K}_k = \mathbf{K}_k^*$, and the entire filter, with this simplification, is rewritten here for easy reference

$$\mathbf{w}_k^f = \Psi_{k-1}\mathbf{w}_{k-1}^a \tag{17a}$$

$$\mathbf{P}_k^f = \Psi_{k-1}\mathbf{P}_{k-1}^a\Psi_{k-1}^T + \mathbf{Q}_{k-1} \tag{17b}$$

$$\mathbf{K}_k^* = \mathbf{P}_k^f\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \tag{17c}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k^*\mathbf{H}_k)\mathbf{P}_k^f \tag{17d}$$

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k^*\left(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f\right) \tag{17e}$$

At times when no observations are available $\mathbf{H}_k = 0$ and, by eqn. (17c), $\mathbf{K}_k^* = 0$ as well. In this case, $\mathbf{w}_k^a = \mathbf{w}_k^f$ and $\mathbf{P}_k^a = \mathbf{P}_k^f$.

It turns out that the Kalman filter (K-filter hereafter) is much more general than the derivation above would indicate. First of all, the minimized variance $J_\mathbf{A}$ can contain an arbitrary symmetric, non-negative semi-definite weight matrix $\mathbf{A}^T = \mathbf{A} \geqslant 0$, to wit

$$J_\mathbf{A} := E\left(\mathbf{w}_k^a - \mathbf{w}_k^t\right)^T\mathbf{A}\left(\mathbf{w}_k^a - \mathbf{w}_k^t\right) \tag{18}$$

The result is still $\mathbf{K}_k^*$ of eqn. (15), independently of $\mathbf{A}$ (see Ghil et al. (1981) for a two-line derivation of this generalization). It follows that the same optimally estimated solution will be obtained whether one wishes, on physical grounds, to minimize the expected kinetic energy or the expected enstrophy of the velocity field errors.

Second, the K-filter minimizes the estimation error variance not only at every time step, but over the entire interval over which data are provided. This fact, and connections to deterministic variational methods, via control theory, are discussed in Part II. Here it suffices to note that the filter (eqn. 17)) is sequential, or recursive, i.e., current observations are discarded as soon as they are processed, or assimilated. This is due simply to the filter's extracting all useful information from the innovation vector, or observational residual, at each time step, by an application of Bayesian ideas in a dynamical context (Kalman, 1960; Lorenc, 1986; Part II).

The sequential nature of the K-filter (eqn. (17)) makes it conceptually easy to grasp, and it has great practical advantages, as we shall see in the next sections. It is probably the major reason for the astounding success of the K-filter, and of its various computational modifications (e.g., Bierman, 1977; Budgell, 1986), in engineering applications.

This being said, it is time to see the filter in action on a GFD problem.

4. SEQUENTIAL ESTIMATION FOR A SIMPLE BAROTROPIC MODEL

4.1. Governing equations

To illustrate the performance and properties of the Kalman filter, and of modified and suboptimal versions thereof, let us consider a simple model problem. The model problem is governed by a linearized, spatially 1-D version of the shallow-water equations

$$u_t + Uu_x + \phi_x - fv = 0 \tag{19a}$$

$$v_t + Uv_x + fu = 0 \tag{19b}$$

$$\phi_t + U\phi_x + \Phi u_x - fUv = 0 \tag{19c}$$

The features that make this system worthy of interest, in spite of its great simplicity, are the presence of advection, of the Coriolis acceleration and β-effect, and of two physically distinct types of waves, slow Rossby waves and fast inertia-gravity waves. Non-stationary Rossby waves arise in this constant-f model from the equivalent β-effect due to the $-fUv$ term in the continuity eqn. (19c). The equivalent β is given by $\beta_* \cong f^2U/\Phi$ (Phillips, 1971).

As usual, the coordinate x points eastward, u and v are perturbation velocities, eastward and northward, while ϕ is the perturbation geopotential. The parameters are chosen with meteorological, mid-latitude applications in mind. Thus the mean zonal current is taken to be $U = 20$ m s^{-1}, the mean geopotential is $\Phi = 3 \times 10^4$ m^2 s^{-2}, and the Coriolis parameter is $f = 10^{-4}$ s^{-1}. The resulting equivalent β_* is 6.7×10^{-12} m^{-1} s^{-1}, so that $\beta_* \cong \beta/2$ with β the usual value at 45° latitude.

The components of the state vector $\mathbf{w}_k$ are the values of (u, v, ϕ) on a space–time grid $(j\Delta x, k\Delta t)$ over which eqn. (19) is discretized by a finite-difference approximation (Ghil et al., 1981). The approximation in question is the Richtmyer two-step version of the Lax–Wendroff scheme, which is second-order accurate in both space and time. The number of points used, $1 \leqslant j \leqslant M$, is $M = 16$, so that $N = 3M = 48$. A spatially 2-D version of system (19), with $N = 3 \times 60 \times 61 = 10980$, is discussed in section 3 of Part II.

The time step, chosen close to the Courant–Friedrichs–Lewy stability limit, is $\Delta t = 30$ min. In this simple case, the dynamics matrix Ψ_k is constant in time, $\Psi_k \equiv \Psi$. But the reason for using $U \neq 0$ and the equivalent β-term in the first place is the desire to build towards a satisfactory solution of the data assimilation problem for non-linear models. The EKF and its adaptation to GFD problems requires successive linearizations about realistic flows (Ghil et al., 1981, 1982; Budgell, 1986; Lacarra and Talagrand, 1988), i.e.,

Ψ_k will change, albeit slowly, in time. It is shown in section 5 of Part II that the estimation can still proceed quite successfully in this more general and realistic case.

4.2. Dynamics and observing patterns

Different observing patterns are presented here, to reflect both conventional, land-based, and current, remotely sensed data sets (Fig. A1). The simplified geometry of the conventional observing system is shown in Fig. 2.

There are two equal land masses, which we think of as Eurasia and North America, and two equal oceans, the Pacific and the Atlantic. The total length of the physical domain is $2L$, with each ocean and each continent $L/2$ long. Only two-periodic solutions are considered, so that the 16 grid points occupy the computational domain of length $L = 14000$ km. Individual grid points that appear in the discussion of the results in this and subsequent sections are identified in Fig. 2 and its caption.

The solution of eqn. (19) to be estimated consists of a single Rossby wave of wave length $L/2$, i.e., of wavenumber 4, and of amplitude $\phi_0 = 2.5 \times 10^3$ m^2 s^{-2}. This amplitude corresponds to a typical ridge-to-trough difference of 500 m in the height of the 500 mbar pressure surface (Ghil et al., 1981). It follows that $\phi_0/\Phi = 1/12$, which is consistent with the linearization about a constant mean state for moderate evolution times. The travel time of this wave over a distance L, given to first order by L/U, is roughly 12 days.

The idealized conventional observing system we consider first consists of complete observations of winds and heights over land, and no observations over the ocean. A similar situation in oceanography corresponds to the availability of expendable bathythermograph (XBT) observations along shipping lanes (see the Appendix) and, in future field experiments, of buoy clusters and acoustic-tomography arrays (Munk and Wunsch, 1979), with no observations in other ocean areas. The observations in the present case are assumed to be available synoptically, every 12 h.

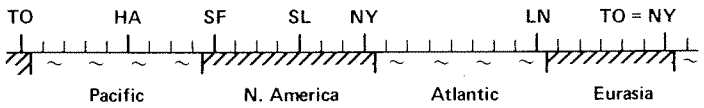


Fig. 2. Distribution of land masses and oceans for the model system (19). Individual grid points of interest are identified as follows: the mid-ocean point (Hawaii, HA); the mid-continent point (Saint-Louis, SL); the continental point nearest to the west coast (San Francisco, SF); nearest to the east coast (New York, NY; identical by sectorial periodicity to Tokyo, TO); and the ocean point just off the west coast (London, LN).

Thus $\mathbf{H}_k$ is periodic, with a period of 12 h, or $24\Delta t$. At synoptic times $\mathbf{H}_k = \mathbf{H}_1 = (\mathbf{I}\ \mathbf{0})$, where $\mathbf{I}$ and $\mathbf{0}$ are 24×24 identity and null matrices, respectively. At all other times, $\mathbf{H}_k \equiv \mathbf{H}_0 = (\mathbf{0}\ \mathbf{0})$.

4.3. Results and interpretation

4.3.1. Conventional observing system

The initial error covariance, $\mathbf{P}_0$, and the observational noise covariance $\mathbf{R}$ are chosen in accordance with current NWP practice (McPherson et al., 1979; Ghil et al., 1981; Lorenc, 1981). As we shall see, the value of $\mathbf{P}_0$ is not important in a linear problem, although it will certainly be so in non-linear problems with multiple attractor sets (Ghil et al., 1985; Ghil and Childress, 1987, Chapter 6). Covariance $\mathbf{R}$ is chosen to be diagonal, with values of r.m.s. geopotential error of 200 $m^2\ s^{-2}$ and r.m.s. wind error of 2 $m\ s^{-1}$. Relative errors in both wind and geopotential are therewith of about 10%. The system noise covariance $\mathbf{Q}$ is taken so as to yield a characteristic decorrelation time between two realizations of eqn. (9a) close to 2 weeks, the typical predictability time of atmospheric disturbances (Lorenz, 1985).

The results are shown in Fig. 3. The curves are marked U, V, P and E for the expected r.m.s. error in the estimation of u, v, ϕ and the total energy, $E = u^2 + v^2 + \phi^2/\Phi$, respectively. These results do not depend on the particular choice of initial state (u, v, ϕ) at $t = 0$.

Over the data-dense region ('land'), the error immediately decreases below the level of observational error, at the first observation time. It increases again in between successive observation times, due to the system noise, as well as to the advection of error from over the ocean (Fig. 3a).

Over the data-sparse ocean (Fig. 3b), the error decreases much less at observation times. This decrease is strictly due to the filter's spreading the information just received over land to the adjacent ocean points. The increase of error in between observation times is much milder than over land: the gradual advection of information from over land partially compensates for the local system noise. The net result in the long run is a decrease of estimation error over the entire region (Fig. 3c).

Particularly remarkable is the fact that, after a few days, the entire prediction–observation system settles into an asymptotically periodic behavior. Estimation errors after observation time and in between repeat almost exactly every half-day. In upper ocean prediction, for an assimilation cycle to reach this asymptotic regime, the time required would be a few weeks or months, rather than a few days. If the ratio of system noise $\mathbf{Q}$ to observational noise $\mathbf{R}$ is similar to that in the case illustrated, error levels in the asymptotic regime would be considerably smaller than initially, in both data-dense and data-sparse regions. For the ocean, of course, reasonable

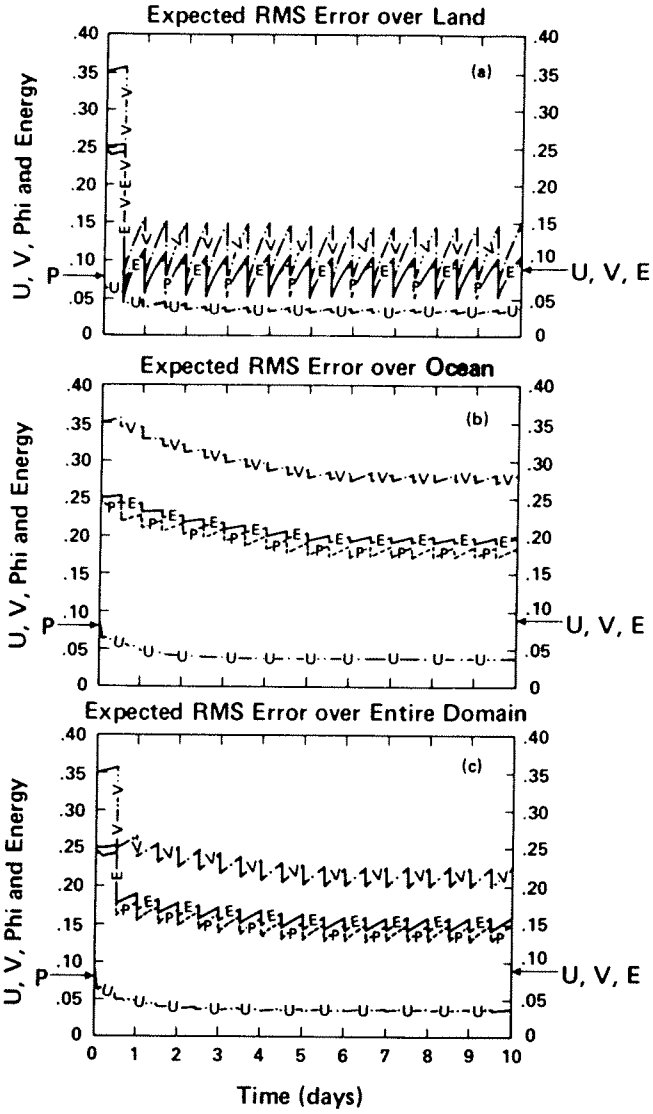


Fig. 3. Evolution in time of the estimation error for a conventional observing system. The curves labeled U, V, and P are obtained by summing those diagonal elements of $\mathbf{P}_k$ which correspond to u, v and ϕ, respectively. The ordinate is normalized so that 1.0 corresponds to an expected r.m.s. error of ϕ_0 for the P curve (dotted), of $v_0 = 4\phi_0/f$ for the U and V curves (dash-dotted) and of $2v_0^2 + \phi_0^2/\Phi$ for the E curve (solid). The observational error level is indicated in each panel; it equals 0.08 for the P curve and 0.09 for the U, V and E curves. The panels show the expected r.m.s. error over (a) land, (b) ocean, and (c) the entire domain (after Ghil et al., 1983).

estimates of $\mathbf{Q}$ and $\mathbf{R}$ are yet to be determined: $\mathbf{Q}$ by predictability studies, and $\mathbf{R}$ by the assimilation of the new and expected remote-sensing data.

To understand the connection between the asymptotic error level and the relative noise levels in dynamics and observations, it suffices to consider a simple, scalar case. For a single equation with constant coefficients, and observations every r time steps ($r = 24$ above), the covariances and weights of the K-filter (eqn. (17)) reduce to

$$\mathbf{P}_k^{\mathrm{f}} = \Psi^2 \mathbf{P}_{k-1}^{\mathrm{a}} + \mathbf{Q} \tag{20a}$$

$$\mathbf{P}_k^{\mathrm{a}} = \mathbf{P}_k^{\mathrm{f}} \mathbf{R} / \left(\mathbf{P}_k^{\mathrm{f}} + \mathbf{R}\right) \qquad (\text{for } k = jr) \tag{20b}$$

$$= \mathbf{P}_k^{\mathrm{f}} \qquad (\text{otherwise}) \tag{20b'}$$

$$\mathbf{K}_{jr} = \mathbf{P}_{jr}^{\mathrm{a}} / \mathbf{R} \tag{20c}$$

From eqn. (20b) it immediately follows that

$$\mathbf{P}_k^{\mathrm{a}} \leqslant \left\{\mathbf{P}_k^{\mathrm{f}}, \mathbf{R}\right\} \tag{21}$$

In particular, the estimation error variance drops to or below the observational noise level at each observation time, although it may grow in between model updates.

To study the asymptotic behavior of eqn. (20), let

$$\mathbf{S}_j := \mathbf{P}_{jr}^{\mathrm{f}}, \quad j = 0, 1, 2, \ldots \tag{22a}$$

$$\mathbf{A} := \Psi^{2r} \tag{22b}$$

$$\mathbf{B} := \sum_{p=0}^{r-1} \Psi^{2p} \tag{22c}$$

(Ghil et al., 1981). Then eqn. (20a and b) yields

$$\mathbf{S}_j = g\left(\mathbf{S}_{j-1}\right) \tag{23a}$$

for

$$g(s) := \frac{(\mathbf{A}s + \mathbf{BQ})\mathbf{R}}{\mathbf{A}s + \mathbf{BQ} + \mathbf{R}} \tag{23b}$$

In the case of a truly perfect model, $\mathbf{Q} = 0$, the non-linear difference eqn. (23) has the explicit solution

$$\mathbf{S}_j = \frac{\mathbf{S}_0 \mathbf{R}}{j\mathbf{S}_0 + \mathbf{R}} \qquad (\text{if } |\Psi| = 1) \tag{24a}$$

$$\mathbf{S}_j = \frac{\mathbf{A}^j (\mathbf{A} - 1)\mathbf{S}_0 \mathbf{R}}{\mathbf{A}(\mathbf{A}^j - 1)\mathbf{S}_0 + (\mathbf{A} - 1)\mathbf{R}} \qquad (\text{if } |\Psi| \neq 1) \tag{24b}$$

It follows that, as $j \to +\infty$

$$\mathbf{S}_j \to 0 \quad \text{for } |\Psi| \leqslant 1 \tag{25a}$$

$$\mathbf{S}_j \to \left(1 - \frac{1}{\mathbf{A}}\right)\mathbf{R} \quad \text{for } |\Psi| > 1 \tag{25b}$$

The former is the case for our conservative PDE (eqn. (19)) discretized by a dissipative numerical scheme. The latter would be the case for an unstable system, either due to baroclinic instability or to an (intentional) numerical one (Miller, 1986).

For a neutrally stable, perfect model, eqn. (24a) exhibits exactly the same dependence of r.m.s. error $\mathbf{S}_j^{1/2}$ on the number of observations j, $\mathbf{S}_j^{1/2} \sim 1/(j)^{1/2}$ as for estimating the mean of a stationary time series from independent observations. The inverse-square-root behavior of estimation error is confirmed by model results with $\mathbf{Q}_k \equiv 0$ (Ghil et al., 1981, fig. 2; not shown here), and indicates that, for perfectly known model dynamics $\Psi_k \neq \mathbf{I}$, the K-filter reduces the estimation process to the well-known one for $\Psi_k \equiv \mathbf{I}$. This is a remarkable result, since the number of model variables is twice the number of observed ones, and the unobserved variables over the 'ocean' could not have been estimated at all had simple persistence, $\Psi_k \equiv \mathbf{I}$, been assumed.

Of course, models are not perfect, so that we need to consider, as in Fig. 3, the case of system noise $\mathbf{Q} > 0$. The quadratic equation

$$s = g(s) \tag{26}$$

has real roots of opposite sign. Let $\mathbf{S}_\infty$ denote its unique positive root. $\mathbf{S}_\infty$ is approached monotonically by solutions of the recursion (23a), from above or below, according to the value of $\mathbf{S}_0$. It follows that the convergence of the weights

$$\mathbf{K}_{jr} \to \mathbf{K}_\infty \equiv \mathbf{S}_\infty/\mathbf{R} \tag{27}$$

is also monotone as $j \to +\infty$. Neither $\mathbf{S}_\infty$ nor $\mathbf{K}_\infty$ depend on $\mathbf{S}_0$.

The result for the scalar case is in agreement with the monotone decrease of $\mathrm{tr}\mathbf{P}_{jr}^{\mathrm{a}}$ in Fig. 3a–c. In fact, both $\mathrm{tr}\mathbf{P}_k^{\mathrm{f}}$ and $\mathrm{tr}\mathbf{P}_k^{\mathrm{a}}$ decrease monotonically for all $k = jr + p$, from one cycle j to the next $(j+1)$, given p fixed.

To obtain an approximate value for $\mathbf{S}_\infty$, let us assume that $|\Psi| < 1$ and $r \gg 1$, so that $\mathbf{A} = \Psi^{2r} \ll 1$. Then the quadratic term in eqn. (26) is negligible and, to good approximation

$$\mathbf{S}_\infty^{-1} = \mathbf{R}^{-1} + (1 - \Psi^2)\mathbf{Q}^{-1} \tag{28}$$

Thus the asymptotic analysis accuracy is slightly less than the sum of the observational accuracy and the model accuracy.

In the case of good observations, $\mathbf{R} \ll \mathbf{Q}$ we have $\mathbf{S}_\infty \cong \mathbf{R}$, in agreement with the results for v, ϕ and E in Fig. 3a; the variable u is only weakly coupled to v and ϕ in eqn. (19) (see Cohn (1982) and Ghil et al. (1981) for details). For poor observations or none, $\mathbf{R} \gg \mathbf{Q}$ and $\mathbf{S}_\infty \cong \mathbf{Q}/(1-\Psi^2)$, as seen in Fig. 3b and confirmed by experiments with different values of $\mathbf{Q}$ (not shown). But always (cf. Fig. 3c)

$$\mathbf{S}_\infty \leqslant \min\{\mathbf{R}, \mathbf{Q}/(1-\Psi^2)\} \tag{29}$$

The asymptotic filter $\mathbf{K}_\infty$ obtained by substituting $\mathbf{P}_k^a = \mathbf{S}_\infty$ into eqn. (16b) is the original Wiener–Hopf filter (see section 2). Kalman's (1960) work gave essentially a much more efficient way to compute an excellent approximation to it. Using the gain matrix $\mathbf{K}_T^*$, where $T = 10$ days, yields in fact estimation results for system (19) which are indistinguishable from those in Fig. 3 after 1 to 2 days (Ghil et al., 1981). This point will come up again in sections 3 and 5 of Part II.

4.3.2. Satellite observing systems

As indicated in section 2, the modern era of data assimilation in meteorology started with the advent of satellite data, and the attempt to use them in operational NWP (Atkins and Jones, 1975; Bengtsson, 1975; Ghil et al., 1979). I shall proceed therefore with the study of observational patterns simulating polar-orbiting and geostationary satellites in system (19).

To further motivate this theoretical study, let us consider first the effects of advection of information in a realistic, fourth-order accurate NWP model, with and without remote-sensing data (Halem et al., 1982). Figure 4 reports results obtained at GLA using data from the first Special Observing Period (SOP-1) of the GWE (formerly FGGE), and a 6-h analysis/forecast cycle (see section 2, especially Fig. 1).

Figure 4a shows the distribution of conventional upper air stations at one synoptic time during SOP-1 (compare with Fig. A1g): they are dense over the continents of the Northern Hemisphere, with less concentrated patches over Australia, South America and some oceanic areas. Figure 4b–d show the results of assimilation experiments for the first two weeks of SOP-1, 5–21 January 1979. These results are presented as the r.m.s. difference between pairs of fields, where the mean is over all grid points and over all synoptic and subsynoptic times, i.e., every 6 h during the given period.

In Fig. 4b we see the difference between the 6-h forecast, or 'first guess' $\mathbf{w}^f$, and the objective analysis $\mathbf{w}^a$ in the NOSAT assimilation cycle. The NOSAT cycle is based essentially on the data in Fig. 4a (see also Fig. A1d and g), and uses no satellite data; hence its name. The elongated error maxima ($\Delta_{\text{r.m.s.}}\phi_{300} \geqslant 60$ m, hatched) along the western part of the Americas and from China to Australia are quite striking: the model's estimate of the

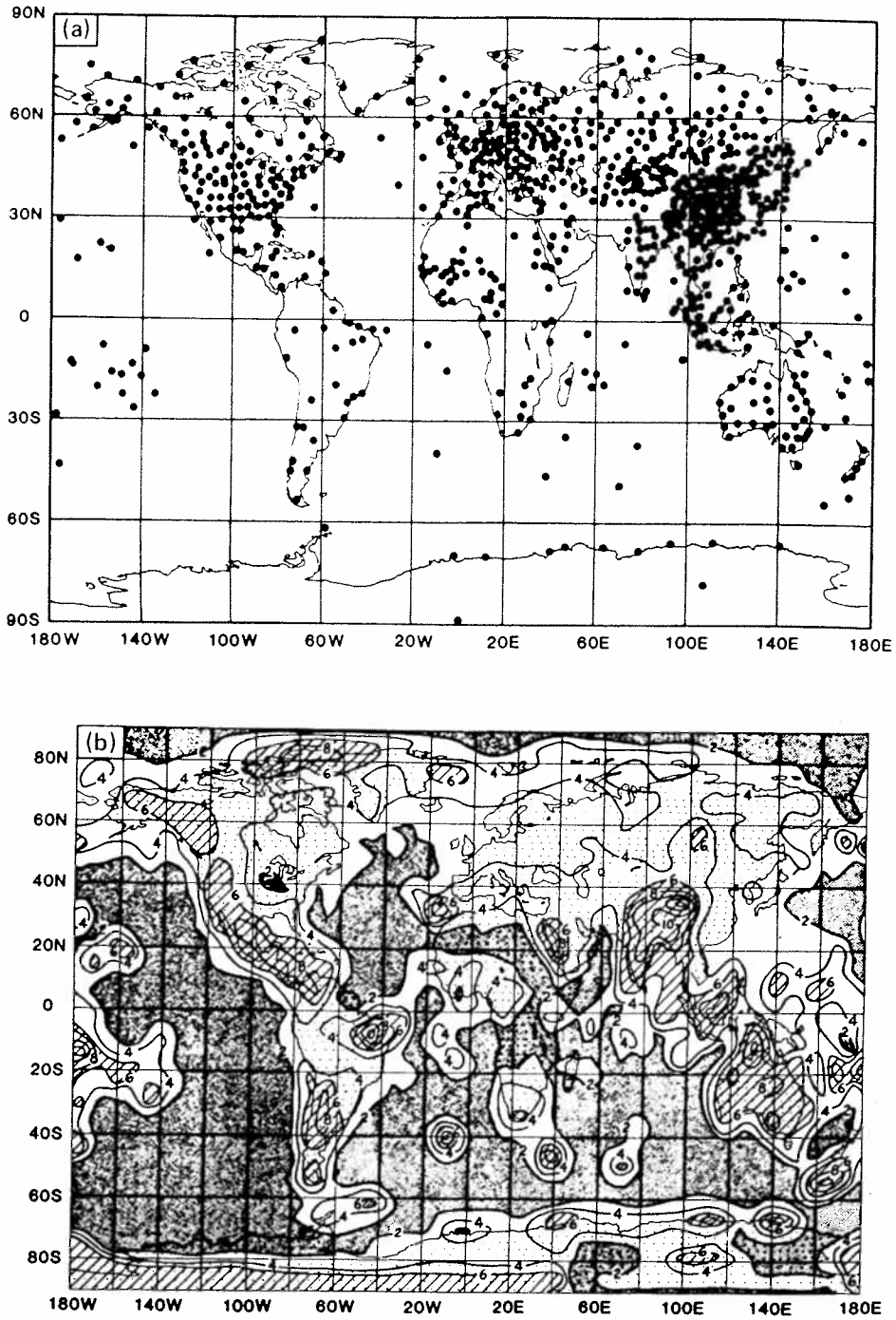
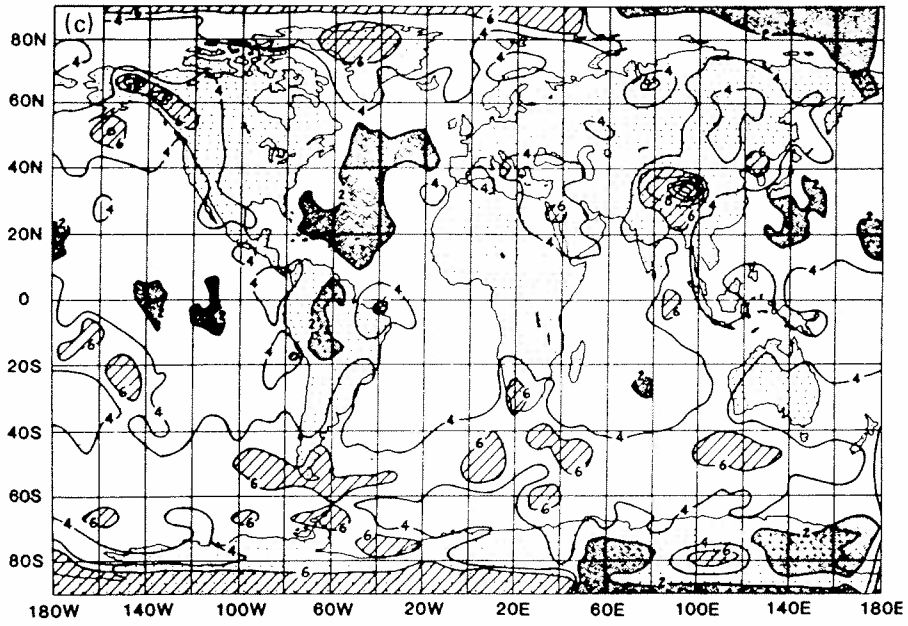


Fig. 4. Advection of information in the data assimilation cycle of a nearly operational NWP model. (a) Rawinsonde stations reporting at 00 Greenwich Mean Time (GMT) 9 January 1979. (b–d) Root-mean-square differences between pairs of 300 mbar height fields ($\Delta_{r.m.s.}\phi_{300}$): (b) 6-h forecast and analysis of the NOSAT cycle; (c) 6-h forecast and analysis of the FGGE (GWE) cycle; and (d) NOSAT and FGGE analysis. Contour interval is 20 m; areas with $\Delta_{r.m.s.}\phi_{300} \leqslant 20$ m are shaded; those with $\Delta_{r.m.s.}\phi_{300} \geqslant 60$ m are hatched (after Halem et al., 1982).



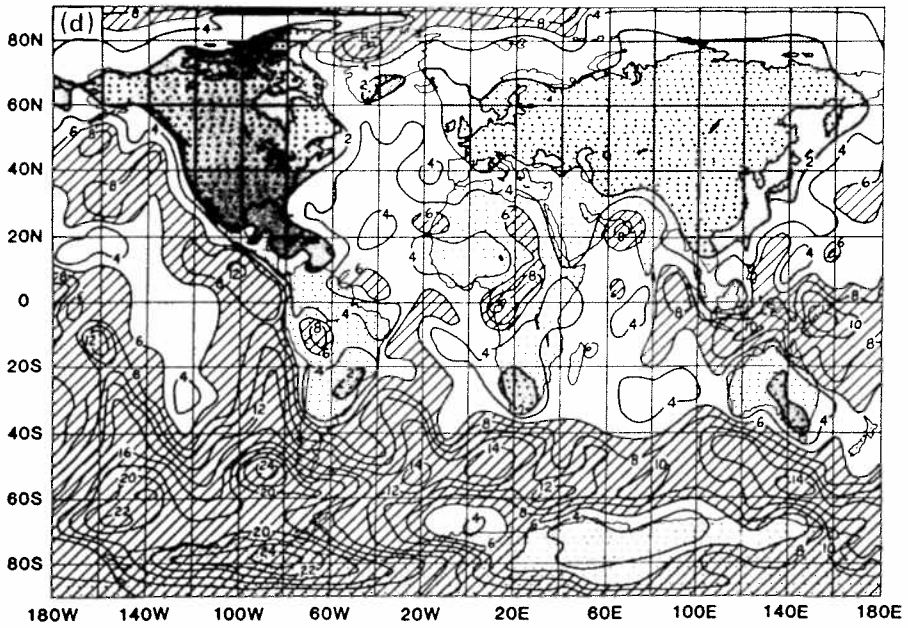


Fig. 4 (continued).

atmospheric state over the Pacific and the Indian Ocean has large errors. These errors are propagated by the flow over the data-dense areas (cf. Fig. 4a), where they clash with the analysis every 6 h (see Fig. 3a). This discrepancy does not develop over Europe, since the Atlantic is too narrow for the accurately estimated state over North America to deteriorate sufficiently.

The elongated error maxima, owing to poor estimation over the oceans, disappear in the difference (Fig. 4c) between the GWE (FGGE) first guess and analysis, based on all the SOP-1 data (see Fig. A1a–g). The satellite data available over the oceans (Fig. A1b and e) help reduce the development of large errors in the assimilation cycle's asymptotic steady state (see days 5–10 of Fig. 3b).

The difference between the objective analyses of the NOSAT and the FGGE cycles (Fig. 4d) is quite small over the extensive data-dense areas of the Northern Hemisphere, and over the limited patches of shared observations in the Southern Hemisphere. It is very large in those areas where only remotely sensed observations anchor the FGGE assimilation cycle, and none are present in the NOSAT cycle. Since the methodology of the assimilation did not provide a built-in error esimate (eqn. (13b)), it is hard to tell how much of the difference in Fig. 4d over the southern oceans is systematic error in the NOSAT cycle or random error in the GWE cycle.

These results with a quasi-operational NWP model and GWE observing systems permit us to draw three conclusions: (1) that the advection of information illustrated in the simple, idealized model (eqn. (19)) occurs in much more complex, realistic data assimilation cycles (Fig. 4b and c); (2) that asynoptic, time-continuous coverage by satellite data can improve the estimate of the geofluid's state (Fig. 4c), in spite of the inherent limitations in the nature and accuracy of the remotely sensed observations (see Table II in section 6); and (3) that a reasonable estimate of the analysis error will greatly increase the utility of the estimate itself (Fig. 4d).

Let us turn therewith to the promised study of satellite observing patterns in the simplified system of sections 4.1 and 4.2. This study will highlight the interaction between observed and unobserved variables, as opposed to the interaction between data-dense and data-sparse areas. The results of sequential estimation in eqn. (19) using data from polar-orbiting or from geostationary satellites are shown in Fig. 5. For purposes of comparison with Fig. 3, the r.m.s. geopotential error and wind error at each point where they are observed is the same as before, although actual error levels for remotely sensed data are typically larger than for conventional data.

In the numerical experiment shown as Fig. 5a, the geopotential ϕ is measured at all grid points every 12 h, but the velocity components u and v are not measured at all. Thus $\mathbf{H}_k$ is again a periodic projection matrix, of

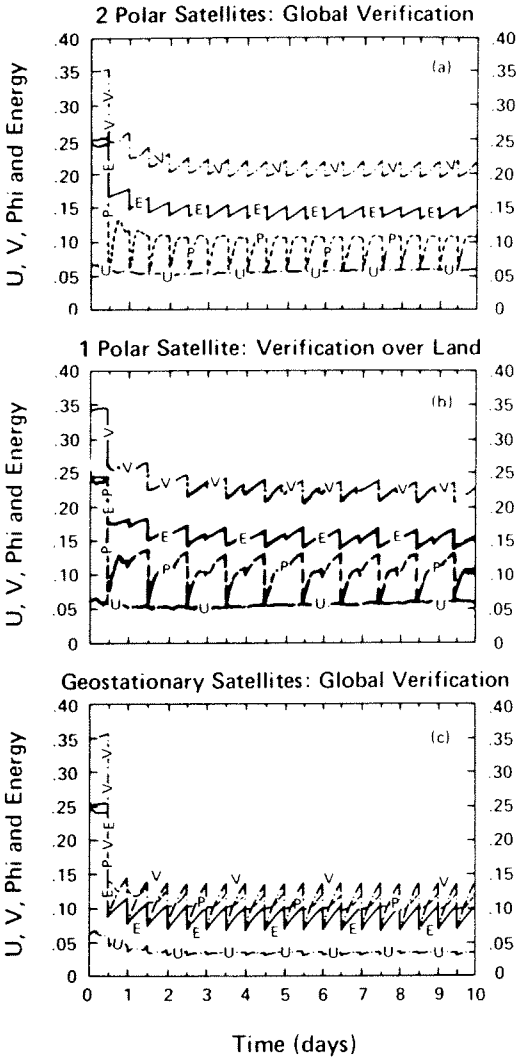


Fig. 5. Evolution in time of the estimation error for satellite observing systems. The labeling of the curves and normalization of the ordinate is the same as in Fig. 3. (a) Geopotential observations over the entire domain, every 12 h; (b) geopotential measured every 12 h, alternatively over 'land' and 'ocean'; (c) velocity observations over the entire domain, every 12 h (after Ghil et al., 1983).

dimension 16×48, identically zero except at synoptic times, when $\mathbf{H}_k = \mathbf{H}_2 = (\mathbf{I}\ \mathbf{0})$, with $\mathbf{I}$ being 16×16 and $\mathbf{0}$ being 16×32. This choice is made to mimick currently operational, intermittent data assimilation cycles (McPherson et al., 1979; Lorenc, 1981), rather than a more desirable, time-continuous cycle (Ghil et al., 1979).

In the atmospheric case, this observing pattern corresponds to the presence of two polar-orbiting satellites, giving almost complete mass-field coverage every 12 h (Ghil et al., 1983). In the oceanic case, partial correspondence is with altimetric measurements attaining similar areal coverage over a longer time interval, of the order of 2 weeks (Kindle, 1986), for sea-surface heights only.

When global measurements of a field are available there is no distinction between the behavior of estimation error at various points, so I show only the expected r.m.s. error over the total region. The error in all variables tends quickly to the same periodic behavior as in Fig. 3. The error in ϕ is noticeably smaller, and that in v is only slightly larger than in the lowermost panel of Fig. 3. The total r.m.s. error is almost the same.

The middle panel (Fig. 5b) shows the results of an experiment in which ϕ is measured every 12 h over half the domain only, first over 'land', then over the 'ocean', as would be the case for a single polar orbiting satellite of the Tiros series. Again u and v are not measured at all. Here the asymptotic periodicity is also quickly reached. A phase difference of 12 h in error behavior over one-half of the region (shown) and the other half (not shown) appears. The individual amplitudes of the error component oscillations are larger than in the panel above. The amplitudes of the r.m.s. error over the total region (not shown) are about the same as in the upper panel for v and the energy, and smaller for ϕ. This experiment shows that information which is partial, both in geographic coverage and in the nature of the variables measured, still suffices to obtain state estimates of the geofluid. Given a nearly optimal data assimilation scheme, the error levels in this case need not greatly exceed those obtained with more bountiful measurement networks.

The last panel of Fig. 5 shows the results of a numerical experiment in which the velocity components u and v are measured at all grid points every 12 h, while ϕ is measured at only one, arbitrary, grid point. Such a measurement is necessary to determine the geopotential field completely, since $\phi(x, t) \equiv \text{const.} \neq 0$, $u(x, t) \equiv v(x, t) \equiv 0$, is a solution of eqn. (19). In the atmospheric case, cloud tracking from geostationary satellites provides velocity measurements, while in the oceanic case, scatterometer data can be used to determine surface winds and wind stress.

The same periodic behavior as in the last panel of Fig. 3 and the first one of Fig. 5 obtains. Now the error in v is considerably less and that in ϕ only slightly larger than before. The total error (E) is about 50% smaller on the average than in Fig. 5a.

4.3.3. Discussion

On the face of it, the smaller estimation errors obtained from total

velocity measurements could be thought to result from there being twice as many velocity measurements as geopotential ones. This, however, is not the case, since errors when v alone is measured (not shown) are only very slightly worse. A more complete explanation of this interesting result, which seems to contradict classical geostrophic adjustment theory, is provided by Daley (1980) within the framework of initialization and slow manifold theory (Leith, 1980; Lorenz, 1980). Connections to this theory will be made in the next section, and in section 4 of Part II.

Variational analysis with a geostrophic constraint (Phillips, 1983) suggests that velocity measurements contribute more or less to the accuracy of the estimated total state than geopotential measurements, according to whether $(fL/2)^2\mathbf{R}_v/\mathbf{R}_\phi$ is much less or much larger than one. Phillips's result is obtained assuming complete coverage with both types of observations. In the numerical experiments reported here, the normalized ratio between mean-square velocity errors $\mathbf{R}_v$ and geopotential errors $\mathbf{R}_\phi$ is close to one and coverage is by either geopotential observations or velocity observations.

Thus our results agree with Daley's (1980), derived by conceptual intersections between the slow and data manifold (Leith, 1980) for linear and non-linear 2-D shallow-water equations, but with no regard to error structure, and disagree with Phillips's (1983) results, obtained by using prescribed error structures in a simple geostrophic-balance model. Hence the need for both simple and realistic model studies, with more or less sophisticated statistics, in planning observing systems for the atmosphere, as well as for the oceans.

The asymptotically periodic nature of the r.m.s. error curves in Figs. 3 and 5 shows that, in fact, the error covariance matrices $\mathbf{P}_k^{\mathrm{f,a}}$ themselves are periodic in the ongoing prediction–observation cycle. The gain matrix $\mathbf{K}_k^*$ used at update times, eqn. (17c), is therefore either periodic (Fig. 5b) or constant (all other cases) from one update time to the next. This is due to the use of time-independent matrices $\mathbf{\Psi}_k \equiv \mathbf{\Psi}$, $\mathbf{Q}_k \equiv \mathbf{Q}$ and $\mathbf{R}_k \equiv \mathbf{R}$, and of constant or periodic observation patterns $\{\mathbf{H}_k\}$.

In all the simple cases treated above, the computational effort can be reduced considerably: a constant or periodic gain matrix can be used in a continuous assimilation cycle, and costly computations are not required at every update time. In a more realistic setting, the gain matrix would have to be recomputed only when the transition matrix $\mathbf{\Psi}_k$ or the noise covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ change substantially. Such changes in the model correspond to long-term changes of the mean circulation pattern, of the flow's variability and of observational systems.

In this section we have learned that (nearly) optimal filters enhance the flow of information from one atmospheric or oceanic field to another, as well as from one geographic region to another. We have also seen that it is

probably preferable to measure velocity components, rather than mass-field variables, for comparable error levels and cost. Finally, computational simplifications have been suggested by the filter's asymptotically constant or periodic behavior.

5. INITIALIZATION AND THE MODIFIED KALMAN FILTER

5.1. Fast waves and initialization

Many aspects of synoptic-scale atmospheric and (mesoscale) oceanic motions are well approximated by relatively slow Rossby waves. These are the only type of waves described by the (linearized) quasi-geostrophic (QG) equations. In NWP, however, primitive-equation (PE) models have replaced QG models at all major operational centers. (Linearized) PE models also describe relatively fast inertia-gravity waves, which carry a much smaller, but non-vanishing amount of the total energy of the flow.

In terms of describing and predicting the slow, meteorologically and oceanographically significant flow features, such as mid-latitude storms or meanders, eddies and rings, the faster waves would seem at first to be more of a nuisance than a help. Hence, the inspired use of the QG approximation in early NWP (Charney et al., 1950), and its continued use in theoretical studies of long-term behavior (Pedlosky, 1987; Ghil and Childress, 1987). Hence also the attempt to justify rigorously the QG approximation by the existence of a slow manifold in the PE system (Leith, 1980; Lorenz, 1980).

Unfortunately, it turns out that the slow manifold does not exist in a rigorous mathematical sense (Vautard and Legras, 1986), and that inertia-gravity waves are an inseparable part of the total behavior of the synoptic scales (Errico, 1982; Lacarra and Talagrand, 1988). In oceanography, global or basin-wide PE models are necessary in order to account correctly for the interaction between the thermohaline and the wind-driven circulation (Bryan and Sarmiento, 1985), and they are generally accepted for the description and prediction of tropical phenomena (Gill, 1982).

In the process of data assimilation, NWP experience has shown that the discrepancy between current data, with their random errors, and model first guess, with its errors, can excite a spuriously large amount of inertia-gravity waves in a PE model. These fast waves are damped out over 12–24 h, and have been shown not to affect 24–48 h forecasts substantially (e.g., Balgovind et al., 1983). However, in an assimilation scheme without proper, built-in error estimation, they can lead to a rejection of data at the next subsynoptic update time, being too different from the first guess (see discussion in the next section; also Daley (1981) for additional undesirable features of the fast waves).

Therefore, a long-standing approach in NWP has been to eliminate entirely, or reduce as much as possible, the amount of inertia-gravity waves at initial forecast time. The minimization of the fast-wave energy at initial time goes by the name initialization in NWP. In other disciplines, initialization often means just the assignment of initial values, whatever their properties otherwise, to a forecast field. The word is used in its narrow, technical NWP meaning throughout this article.

5.2. Initialization and projection

The optimal compromise between statistical minimization of the errors in the initial state, on the one hand, and dynamical minimization of the fast components in this state, on the other, is a topic of considerable current interest in NWP, as witnessed by an entire volume of contributions dedicated to it (Williamson, 1982; see also Ghil, 1980). The relevance to oceanographic data assimilation is discussed in section 2 of Part II, with additional results in hand, and in Ghil and Malanotte-Rizzoli (1989).

A reasonable recipe for this compromise can be given in the simple model (19). In this model, the Rossby waves form a linear subspace, denoted by $\mathcal{R}$ in Fig. 6, and the inertia-gravity waves form a complementary subspace, denoted by $\mathcal{G}$ in the figure.

In the standard formulation of slow manifold theory (Daley, 1980, 1981; Leith, 1980) the two linear subspaces $\mathcal{R}$ and $\mathcal{G}$ are presented as orthogonal

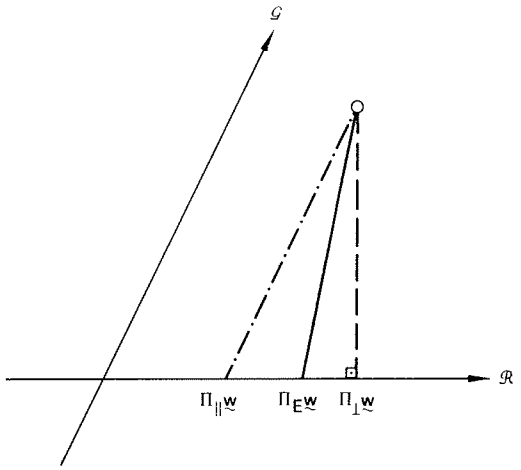


Fig. 6. Schematic representation of the slow subspace $\mathcal{R}$ of Rossby waves and the fast subspace $\mathcal{G}$ of inertia-gravity waves. Three projections onto $\mathcal{R}$ are shown: the parallel projection $\Pi_{\|}$ (dash-dotted), the perpendicular projection $\Pi_{\perp}$ (dashed), and the **E**-perpendicular projection $\Pi_{\mathbf{E}}$ (solid: after Cohn, 1982).

to each other. This is only the case if the linear system under study is self-adjoint, in particular if the full governing equations are linearized about a state of rest. In practice, GFD flows have large shear and linearization about a particular solid-body rotation is not a good approximation for the purposes of data assimilation, as we shall see in section 5 of Part II. Equation (19) has been obtained on purpose by linearization about non-zero mean flow, and hence the associated linear operator is not self-adjoint.

As a consequence, projection onto the slow subspace $\mathscr{R}$ of the state $\mathbf{w}^f$ or $\mathbf{w}^a$ can be carried out in more than one way. The parallel projection $\Pi_{\|}$ eliminates the fast modes of $\mathbf{w}$ without changing the slow ones. The perpendicular projection operator $\Pi_\perp$ minimizes the distance between $\mathbf{w}$ and its projection onto $\mathscr{R}$, $\Pi_\perp \mathbf{w}$, in the usual, Euclidean metric of the phase space. The oblique, or $\mathbf{A}$-perpendicular projection $\Pi_{\mathbf{A}}$ minimizes this distance in a modified metric, with non-negative semi-definite weight matrix $\mathbf{A} \geqslant 0$ (cf. eqn. (18)).

Details about the linear subspaces $\mathscr{R}$ and $\mathscr{G}$ in the continuous system (19), as well as in the actual discrete system used in the numerical examples, can be found in Cohn (1982). The different projections are written down explicitly there as matrix operators for the discrete system. The projection used in the following numerical example is the one most appealing physically, namely the minimum-energy projection, or $\mathbf{E}$-perpendicular projection, which minimizes the expected energy of the analysis error, $\Pi_{\mathbf{E}}$. In this special case, the weight matrix $\mathbf{A}$ in eqn. (18) will be denoted by $\mathbf{E}$; it is positive definite, diagonal, and the diagonal entries are, at each grid point, unity for the velocity components u and v and $1/\Phi$ for the geopotential ϕ.

5.3. The dynamically modified Kalman filter

With these dynamical facts in mind, we can address the issue of the compromise between minimum errors and minimum fast waves, by modifying the standard K-filter $\mathbf{K}_k^*$. The modified filter has to minimize the error functional (18), subject to the constraint that

$$\mathbf{w}_k^a \in \mathscr{R} \tag{30}$$

at all update times k. It is assumed that $\mathbf{w}_0^a \in \mathscr{R}$, i.e., that initialization has been performed at the outset.

The solution of this constrained minimization problem (Cohn, 1982; Ghil et al., 1982) is to take for the gain matrix

$$\mathbf{K}_k = \mathbf{K}_k^{\Pi} := \Pi_{\mathbf{A}} \mathbf{K}_k^* \tag{31}$$

where $\Pi_{\mathbf{A}}$ is the $\mathbf{A}$-orthogonal projection matrix onto $\mathscr{R}$, defined by

$$\text{Range } \Pi = \mathscr{R} \tag{32a}$$

$$\Pi^2 = \Pi \tag{32b}$$

$$(\mathbf{A}\Pi)^{\mathrm{T}} = \mathbf{A}\Pi \tag{32c}$$

The (dynamically) modified K-filter, or ΠK-filter, is the data assimilation scheme (eqn. (8)) based on the choice of gain matrix $\mathbf{K}_k^{\Pi}$. Notice that $\Pi_{\mathbf{A}}$, and therefore $\mathbf{K}_k^{\Pi}$, actually depend on the weighting matrix $\mathbf{A}$: as opposed to the standard K-filter (eqn. (17)), one must now choose the error functional to be minimized.

For any given choice of $\mathbf{A}$, the ΠK-filter also has the property that it minimizes the functional

$$\bar{J}_{\mathbf{A}} := E(\mathbf{w}_k^a - \bar{\mathbf{w}}_k^a)^{\mathrm{T}}\mathbf{A}(\mathbf{w}_k^a - \bar{\mathbf{w}}_k^a) \tag{33}$$

subject to the constraint (30), where $\bar{\mathbf{w}}_k^a$ denotes the analyzed field that would be produced by using the standard K-filter at time k. In fact, we have

$$\mathbf{w}_k^a = \Pi\bar{\mathbf{w}}_k^a \tag{34}$$

Thus, the ΠK-filter combines the standard K-filter with variational normal mode initialization (Daley, 1981; Tribbia, 1982), i.e., with variational projection onto $\mathscr{R}$; $\bar{\mathbf{w}}_k^a$ is an objective analysis, $\mathbf{w}_k^a$ is the initialized field, and the elements of $\mathbf{A}$ are the variational weights. The ΠK-filter, though, minimizes not only the $\mathbf{A}$-distance of eqn. (33) between the final, initialized field $\mathbf{w}_k^a$ and the 'analyzed' field $\bar{\mathbf{w}}_k^a$, but also the $\mathbf{A}$-distance of eqn. (18) between $\mathbf{w}_k^a$ and the true field $\mathbf{w}_k^t$, which is a measure of the actual analysis error.

A particular realization of the estimation process for the standard K-filter is shown in Fig. 7. Figures 3 and 5 showed results that do not depend on the particular stochastic realization of initial error, system noise or observational noise, but follow from the deterministic filtering algorithm (17). In fact, for a truly linear system the sequences $\{\mathbf{P}_k, \mathbf{K}_k\}$ depend only on $\{\Psi_k, \mathbf{H}_k, \mathbf{P}_0, \mathbf{Q}_k, \mathbf{R}_k\}$ and not on the state $\{\mathbf{w}_k\}$ itself. Thus $\mathbf{K}_k$ can be calculated once and for all.

Since we are interested eventually in the non-linear systems of GFD (section 5 of Part II), this approach was not taken here, and we encounter for the first time the evolution of the state itself. The behavior of the actual, rather than expected, r.m.s. estimation error $(\mathbf{w}_k^{f,a} - \mathbf{w}_k^t)$ is shown in Ghil et al. (1981, fig. 3). It is not exactly monotone decreasing and exhibits fluctuations, but it follows in broad outline the expected behavior shown in Fig. 3 here.

Notice that the change in solution at observation times is typically larger at SF, the westernmost observation point, than at NY, the easternmost point: the forecast at SF has larger errors than the forecast at NY (cf. Fig. 4b), since SF is downwind from the data-sparse ocean, while NY is downwind from data-dense land. The solution jumps at the mid-ocean point HA are quite small, since the weighting coefficients of the K-filter's gain matrix decrease rapidly with distance from the observations (see figs. 7 and 8 of Ghil et al., 1981; not shown here).

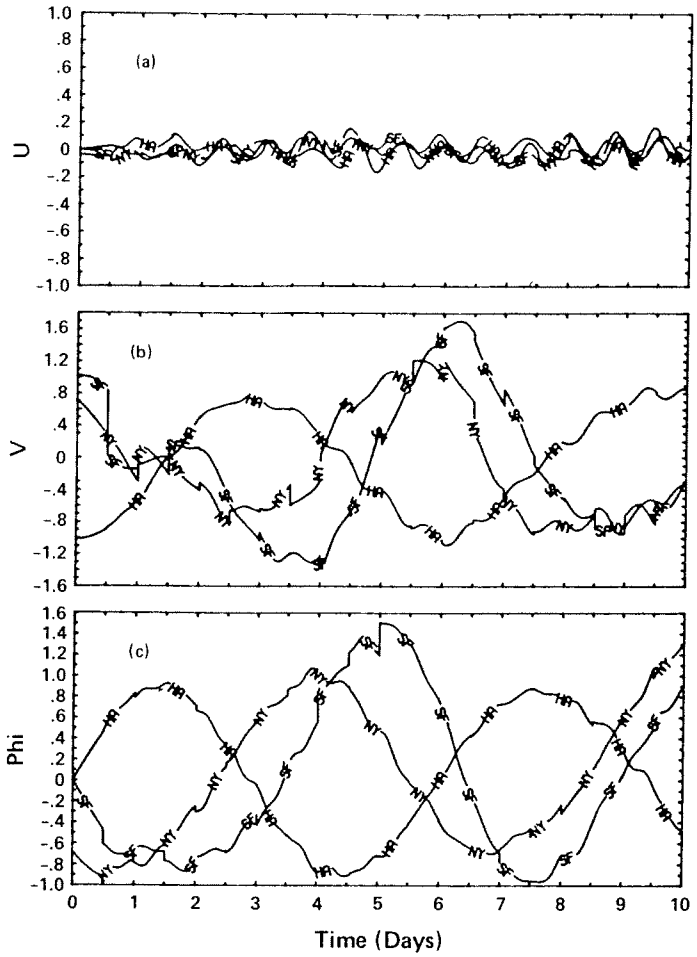


Fig. 7. Evolution of the state estimate $\mathbf{w}^{f,a}$ at three locations, using the standard K-filter and the conventional observing network. See Fig. 2 and its caption for identification of points. (a) Zonal velocity u; (b) meridional velocity v; (c) geopotential ϕ (after Ghil et al., 1981).

The overall picture in Fig. 7 is that of slowly evolving, large-amplitude Rossby waves, with a period of roughly 6 days. Upon these are superimposed smaller amplitude, rapidly evolving inertia-gravity waves, excited by the system noise and the observational noise. The effect of the initializing IIK-filter on this situation is shown in Fig. 8.

The evolution of the Rossby waves is clearly the same as in the preceding figure, while the fast waves have been completely eliminated. In particular, fast waves are no longer excited at update time, even when the analysis $\mathbf{w}_k^a$ differs markedly from the first guess $\mathbf{w}_k^f$. Changing the type of projection to

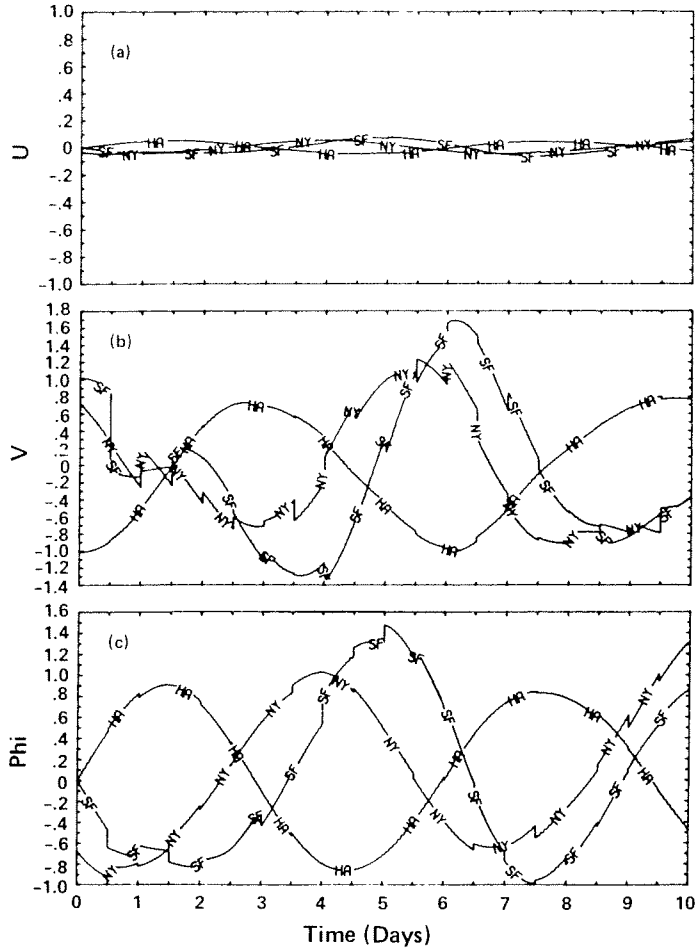


Fig. 8. Evolution of the state estimate at the same mid-ocean point and continental-margin points as in Fig. 7. The minimum-energy projection Π_E is used to modify dynamically the K-filter. Panels (a), (b) and (c) show u, v and ϕ, respectively (after Cohn, 1982).

$\Pi_{\|}$ or $\Pi_{\perp}$ does not seem to make too much of a difference in the estimate (see fig. 10 in Ghil et al., 1981, for $\Pi_{\perp}$; $\Pi_{\|}$ not shown).

At what cost to the estimation error are the fast waves eliminated? It is obvious that constrained optimization (eqns. (18) and (30)) can only yield a minimum larger than or equal to the result of unconstrained optimization (18). In Fig. 9 we see the expected r.m.s. errors for the K-filter and ΠK-filter, side by side.

The excess estimation error of the ΠK-filter over the K-filter, for all the components of the energy, as well as for the total, increases with time in the

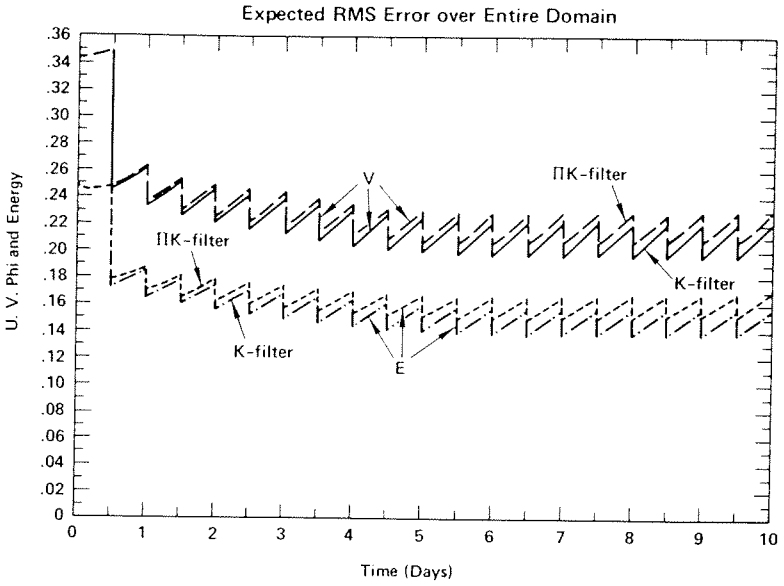


Fig. 9. Evolution of the estimation error for a conventional network, with and without initialization of the K-filter. Only the expected r.m.s. error for v (solid for the standard K-filter, dashed for the ΠK-filter) and for the total energy E (dashed-dotted for K, dotted for ΠK) is shown (after Ghil et al., 1981).

assimilation cycle, but is still quite small in the asymptotic regime at day 10. So the loss of accuracy in estimation is not too great. But what is the gain?

As pointed out in section 5.1, inertia-gravity waves are an inseparable part of the geofluid's behavior. They are essential in tropical phenomena, and in fact their suppression in operational NWP practice by non-linear normal-mode initialization (Daley, 1981) has led to serious estimation errors in tropical analyses (Kanamitsu, 1981). It turns out that the correct amount of fast-wave energy can be determined from the observations by using optimal or nearly-optimal filters (section 4 of Part II). But large errors in the fast waves are very harmful to the correct estimation of the energetic, slow waves in an assimilation scheme which is far from truly optimal, such as OI. Thus initialization, albeit easy, is neither necessary, nor particularly useful when a nearly optimal data assimilation scheme is implemented, but it is very helpful as an improvement to the highly suboptimal assimilation schemes in current operational use (see section 2 of Part II).

6. PRACTICAL CONSIDERATIONS

Section 2 gave a glimpse of the rich history of data assimilation in meteorology, concluding with a few more-detailed references. A critical

review of sequential estimation ideas applied to NWP, from Jones (1965) to Petersen (1976), was given by Ghil et al. (1981, pp. 178–180), and will not be repeated here. An omission in that review, which was brought to my attention later, and should be corrected at this point, are the articles of Epstein (1969) and of Pitcher (1977), who in turn did not mention the seminal work of Kalman.

Epstein (1969) carried out Monte-Carlo experiments with Lorenz's (1960) minimal hydrodynamic system, using a random observing pattern. Pitcher (1977) used an equivalent-barotropic model and 500 mbar geopotential heights from about 450 radiosonde stations in the Northern Hemisphere for 3 days in December 1969. These authors worked with deterministic flow equations, randomizing only over the initial data. Their main conclusion was the need to take into account random perturbations of the equations of motion, i.e., system noise $\mathbf{Q}_k \neq 0$, in order to fit the data more closely and avoid large discrepancies between true forecast error and apparent forecast error. This is a result well worth remembering in the use of variational methods, which assume perfect model dynamics.

To give a better feeling for the gap between the theoretical ideas presented here and the realities of an operational data assimilation cycle, let us consider the various characteristics of meteorological data in current use. These are listed in Table II, while the corresponding horizontal distributions are shown in Fig. A1a–g of the Appendix.

Each observing system measures a different set of meteorological parameters, with differing horizontal and vertical resolution, continuously or intermittently in time. The random errors indicated are crude estimates of standard deviations, the corresponding variances being each the sum of an instrumental and a sampling error variance. Instrumental error is easily determined in the laboratory, but sampling error depends on the motions of the fluid, which contain energy at all scales (Balgovind et al., 1983; Thiébaux and Pedder, 1987, Chapter 6). As we shall see in section 4 of Part II, determining the observational noise from the data themselves is one of the most interesting and useful things a nearly optimal filter can do.

This noise, as well as the system noise, changes as the dominant flow pattern changes. The easiest way to think about large-scale flows as a stochastically perturbed process, rather than a purely deterministic one, is to think about the variance associated with a certain mean. The stochastic perturbations are commonly referred to as subgrid-scale processes. These change as the flow resolved by the simulation or prediction model changes.

One salient reason to determine observational noise adaptively is to improve quality control of data. The capital letter acronyms in Table II and Fig. A1 are code names for types of observations transmitted on the Global Telecommunication System (GTS), as a part of the World Weather Watch

TABLE II

Characteristics of data from operational meteorological systems (after Gustavsson, 1981)

Observing system	Type of observation	Field variables [a]	Vertical resolution (mbar)	Horizontal resolution (km)	Time resolution (h)	Random errors	Other errors; remarks
Surface stations (SYNOP, SHIP)	Synoptic	P_s	–	–	3	±0.1 mbar	Quality of data from commercial ships is variable.
		T_s	–	–	3	±1°	
		$\mathbf{V}_s$	–	–	3	±1 m/s^{-1}	
		r_s	–	–	3	±5%	
Upper-air soundings (TEMP, PILOT)	Synoptic	T	≅ 50	–	6–12	±1°	Larger errors of ±2° in the stratosphere. Larger wind errors occur, e.g., for small radar elevation angles. Errors are vertically correlated.
		$\mathbf{V}$	≅ 50	–	6–12	±1 m s^{-1}	
		r	≅ 50	–	6–12	±5%	
Aircraft reports (AIREP, ASDAR)	Asynoptic	T	–	500–1000 (AIREP) 150 (ASDAR)	Time-continuous	±3 m s^{-1}	Frequent large errors occur in AIREP reports. Quality of ASDAR reports is excellent.
		$\mathbf{V}$	–				
Satellite vertical temperature soundings (TOVS) (polar-orbiting)	Asynoptic	T	≅ 200	500 (GTS-data)	Time-continuous (6)	±1.5–±2°	Large errors in cloudy areas when only microwave measurements are utilized. Errors are horizontally and vertically correlated.
Satellite sea-surface temperatures (polar-orbiting)	Asynoptic	T_s	–	2–10	Time-continuous	±1–2°	Systematic errors seem to occur, especially in the tropics (according to ECMWF).
Satellite cloud wind vector (geostationary)	Asynoptic	$\mathbf{V}_{low}$ (≅ 850 mbar)	–	500	6	±6 m s^{-1}	Large systematic errors occur in assignment of heights to the wind vectors.
	Asynoptic	$\mathbf{V}_{high}$ (≅ 200 mbar)	–	500	6	±12 m s^{-1}	

[a] $\mathbf{V}$ is horizontal velocity and r is relative humidity. Temperature errors are in degrees K.

(WWW). Thus TEMPS are temperature, humidity and wind measurements from radiosondes, SYNOPS are surface observations from land and SHIPS are, well, from ships.

Obviously each observation has a natural error, which can be modeled as random with or without bias, caused by instrumental inaccuracies or by sampling. These natural errors are smoothed out, more or less optimally, by statistical or variational methods. But some observations are just plain wrong, i.e., errors occur in the (human) encoding process or in transmission. Quality control attempts to eliminate these artificial observations, as well as so-called outliers, with abnormally large (natural) deviations, without reducing the number of valid, and hence useful, observations.

Each observation can be checked against a background field, model forecast or climatology, or against nearby observations of the same type, if any. The latter kind of check is colloquially referred to as a 'buddy-check'. Buddy checks can involve continuity in space or in time, depending on the observing system in question.

For orientation purposes, a simple, statistical buddy check is described here, following Gustavsson (1981). Let $w^{o}_{i,k}$ be an observed value at station i and time k, and $\overline{w}_{i,k}$ be the interpolated value from the available nearest neighbors, e.g., two in 1-D, three in 2-D and four in 3-D, if linear interpolation is used. Assuming random, uncorrelated errors, let σ_0^2 be the variance of the (natural) observational errors and σ_x^2 be the corresponding estimated mean-square error of interpolation. Then

$$\Sigma_{i,k} := E\left(w^{o}_{i,k} - \overline{w}_{i,k}\right)^2 = \sigma_0^2 + \sigma_x^2 \tag{35}$$

The buddy check is based on the algorithm

$$\Sigma_{i,k} \leqslant K\sigma \Rightarrow \text{accept } w^{o}_{i,k} \tag{36a}$$

$$\Sigma_{i,k} > K\sigma \Rightarrow \text{reject } w^{o}_{i,k} \tag{36b}$$

here $\sigma^2 = \sigma_0^2 + \sigma_x^2$, and the choice of K is the tricky one. If $w^{o}_{i,k} - \overline{w}_{i,k}$ is normally distributed, the risk r_{ik} of rejecting a correct value is $r_{ik} \cong 0.32$ for $K = 1$, $r_{ik} = 0.04$ for $K = 2$, and $r_{ik} \cong 3 \times 10^{-4}$ for $K = 3$. A compromise has to be reached between the risk of accepting an erroneous value and that of eliminating a correct one.

Of course outliers are often the physically most interesting ones, and numerous failed forecasts can be traced back to rejection of a (nearly) correct observation of an exceptional event. Hence the experience of skilled forecasters is brought to bear on the cases of greatest doubt, permitting manual intervention into an otherwise automated process of data transmission, reception, decoding and assimilation.

An alternative to manual intervention is the use of a non-parametric method for robust estimation (Efron, 1982). Non-parametric means that no

a priori assumptions about the distribution of variables, such as normality, are made. These methods are robust to the presence of outliers, but are computationally rather expensive. A trade-off in delay and cost between automatic robust estimation and manual intervention thus depends on eminently practical considerations.

The way is long from basic principles, however sound, to an actual data assimilation system. A few more steps along this way are taken in Part II.

Quasi-operational data assimilation and forecasting in oceanography has been initiated for the Gulf Stream system by the Harvard Open Ocean Modeling Group (Robinson and Leslie, 1985; Robinson et al., 1987). Other interesting approaches to the estimation problem in oceanography, including statistical methods, direct insertion, variational methods and control-theoretical methods, are due to Bretherton et al. (1976), Malanotte-Rizzoli and Holland (1985), Provost and Salmon (1986) and Wunsch (1988). A bridge between the theory of sequential estimation and of variational methods, on the one hand, and the practice of data assimilation in oceanography, on the other, is provided by Miller (1987). His primer can be used with considerable advantage by the interested readers of this article as a complementary testing ground for their understanding. A more elaborate comparison between data assimilation methods in meteorology and in oceanography will appear in Ghil and Malanotte-Rizzoli (1988).

ACKNOWLEDGEMENTS

Great is my debt to numerous colleagues, past and present: I have learned with R. Atlas, D. Boggs, R. Balgovind, K. Bube, A. Dalcher, M. Halem, E. Isaacson, E. Kalnay, J. Tavantzis and R. Todling; but more than all with my former students and friends Steve Cohn and Dick Dee. Over the years, Norman A. Phillips has made the most penetrating comments on this work; R. Daley, A. Hollingsworth and A. Lorenc have kept up a barrage of criticism, part serious, part light-hearted. It is a pleasure to acknowledge the constant encouragement of Allan R. Robinson in organizing these thoughts and bringing them before the oceanographic community. Other oceanographers, including M. Abbott, B. Cornuelle, D. Halpern, P. Malanotte-Rizzoli, C.N.K. Mooers, R.N. Miller, J.D. Thompson and C. Wunsch, have stimulated me by discussion of these ideas or invitations to present them. The manuscript was measurably improved by the comments of three anonymous referees; two of them were unusually helpful and thorough. This work has benefited from continuing support under NASA grant NAG-5713, as well as from more recent support under ONR grant N 00014-87K-0331 and under a grant from the Institute of Naval Oceanography. B. Gola typed the manuscript; D. Kingsmill, M. Ralph and J. Spehar prepared the figures for publication.

APPENDIX

Data availability in meteorology and oceanography

To measure the distance with respect to availability of data between physical oceanography and dynamical meteorology let us compare Fig. A1

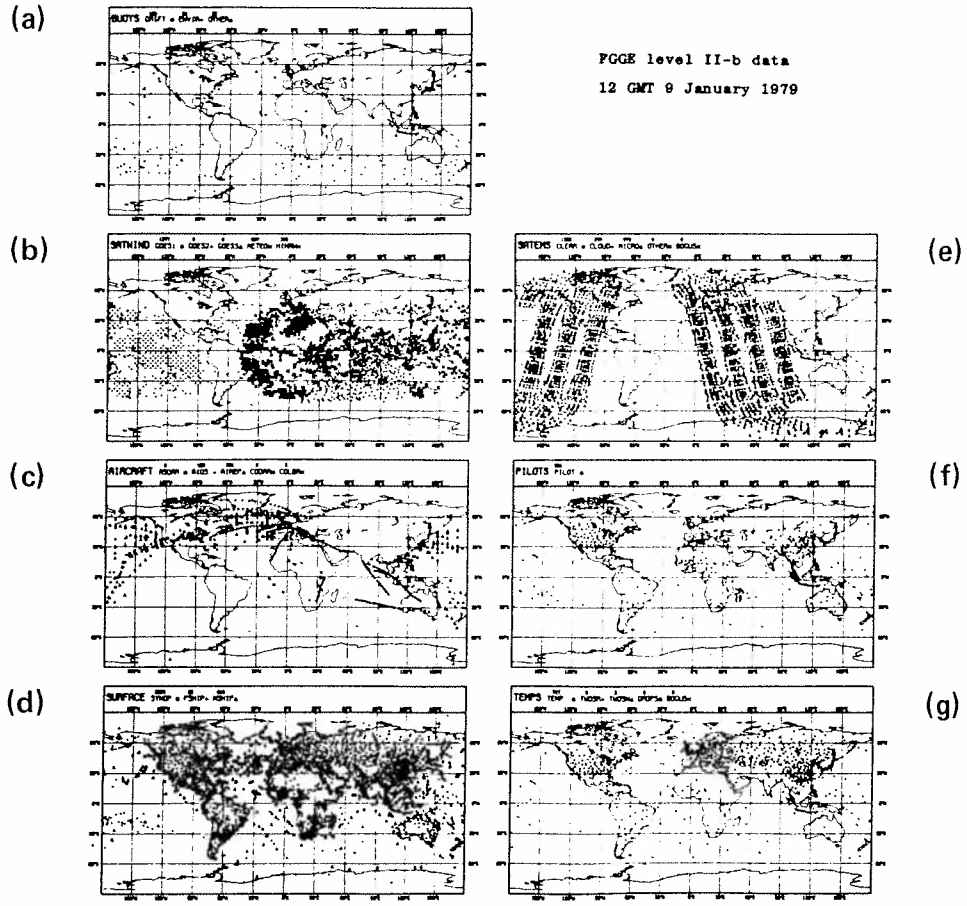


Fig. A1. Meteorological observations available during one 12-h period centered at 1200 GMT 9 January 1979. Each panel gives one type of observations, with data type, instrument source and exact number of soundings for that source at the top left; numbers in parentheses here are typical of measurements available for a 12-h period: (a) Drifting buoys, surface pressure $\mathbf{p}_s$ (270); (b) cloud-drift wind vectors $\mathbf{V}$ (two velocity components) from geostationary satellites, at one of two levels (2250 vectors); (c) $\mathbf{V}$ (two scalars) from aircraft and constant-level balloons (1100); (d) surface temperature T_s, wind $\mathbf{V}_s$ and pressure $\mathbf{p}_s$ (four) from land stations and ships (3450); (e) temperature T from polar-orbiting satellites (2050×5 levels); (f) $\mathbf{V}$ (two) from pilot balloons (660×10); (g) T, $\mathbf{V}$ and humidity q (four scalars) from radio- and dropsondes (750×10) (from Bengtsson et al., 1981).

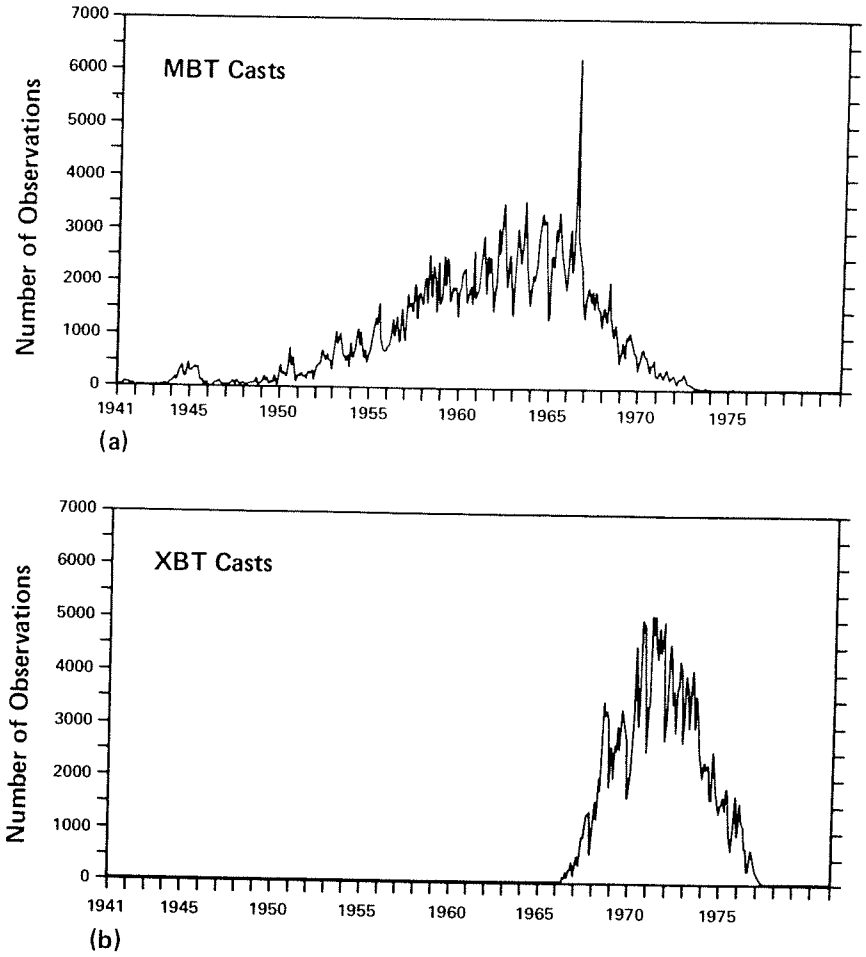


Fig. A2. Number of bathythermograph casts per month, 1941–1977. (a) MBTs; (b) XBTs (after Levitus, 1982).

with Figs. A2 and A3. Figure A1 represents the data typically available at present over one synoptic period, i.e., over 12 h, for the global atmosphere. Figures A2 and A3 represent the distribution in space and time of all oceanographic data up to 1978, archived by the National Oceanographic Data Center (NODC), Washington, DC.

In Fig. A1, the total number of scalar measurements of the atmospheric mass and velocity fields over 12 h is of the order of 10^5 (Ghil, 1986). This number is essentially adequate for a description of large-scale atmospheric fields, by using the methods of data assimilation into weather prediction models that are currently operational in major weather bureaux. The test of

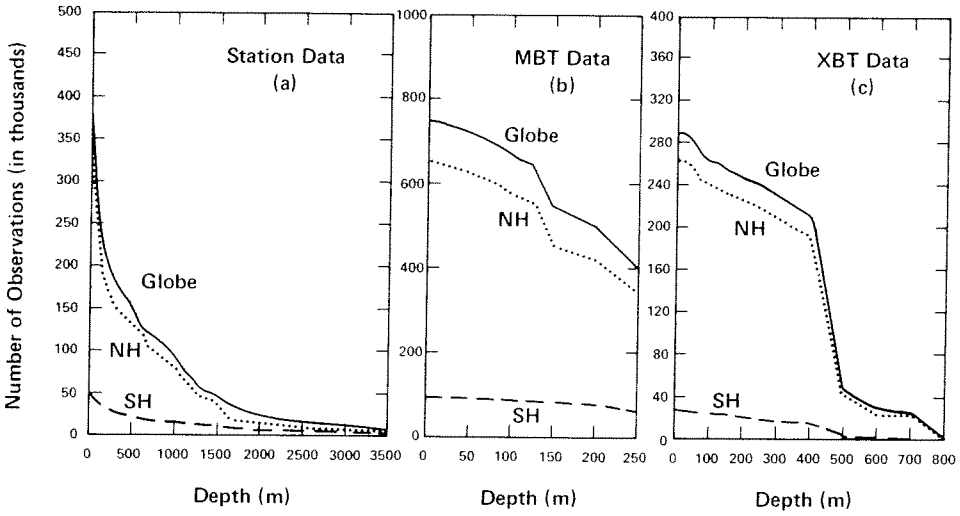


Fig. A3. Distribution of temperature observations at standard levels in NODC archives. (a) Station data (SD); (b) MBTs; (c) XBTs. Each panel shows the distribution with depth over the globe (solid), Northern Hemisphere (dotted) and Southern Hemisphere (dashed). Notice that both the ordinate and the abscissa in each panel have different scales (after Levitus, 1982).

adequacy here is relatively accurate prediction for a few days, or a few synoptic periods.

The total number of archived oceanographic mass-field measurements over a period of 80 years or so is of the order of 10^7: (a) temperature T and salinity S from Nansen casts at about 500 000 hydrographic stations; (b) T from about 785 000 mechanical bathythermograph (MBT) and about 300 000 expendable bathythermograph (XBT) soundings, each with its own vertical distribution of individual measurements (Levitus, 1982). The situation for the oceans' velocity field is rather worse than for the mass field.

On the face of it, taking the number of atmospheric observations as the yardstick, there are 10^2 times more oceanic observations for a period of 10^5 times longer, i.e., 10^3 times fewer observations. This first estimate has to be corrected by allowing for the different time and space scales of the basic phenomena to be observed, and predicted, in the atmosphere and in the ocean. Let us take these to be mid-latitude synoptic eddies, often called mesoscale eddies in oceanography.

In the ocean the Rossby radius of deformation, which is the characteristic length scale, is about 10^2 km, versus 10^3 km in the atmosphere, thus requiring an observational density 10^2 times higher. This is only partially compensated by the longer characteristic time in the oceans, requiring a

frequency of observation 10 times lower than in the atmosphere. Hence the corrected estimate is of 10^4 times fewer observations in the ocean.

Not only have oceanographers been used to so few observations, but these are even more unevenly distributed in space and time than in meteorology. Figure A2 shows the distribution in time of MBT and XBT casts. The XBT instrument is more accurate and convenient than the MBT, which it has essentially replaced. Unfortunately, the number of XBT casts has actually decreased, and there is also a lag in their entering the NODC files.

The distribution of observations in space, horizontally (not shown) and with depth (Fig. A3), is also very uneven. Most data are in the Northern Hemisphere (NH: dotted line in Fig. A3), and there is further concentration of data in western boundary currents and along shipping lanes. The amount of data below the permanent thermocline is a tiny fraction of the total, and decrease of information with depth is quite rapid in the upper ocean as well.

In contrast to this situation, valid until just a few years ago, there are already about 40 000 satellite sea-surface temperature measurements daily. In addition, in the early 1990s, about 50 000 sea-surface height measurements and 180 000 surface wind vectors will be available daily (Halpern, 1987). Thus the daily number of measurements in oceanography will become comparable to that currently available in meteorology. Even so, two problems remain: first, this is still a factor of 10 smaller, due to the difference in characteristic scales; and second, the additional data mentioned are all surface data.

It is hoped that the number of vertical soundings will increase somewhat, due to acoustic-tomography arrays and other advanced systems. But it is unlikely that this increase will be anywhere as dramatic as that for surface data.

REFERENCES

Atkins, M.J. and Jones, M.V., 1975. An experiment to determine the value of satellite infrared spectrometer (SIRS) data in numerical forecasting. Meteorol. Mag., 104: 125–142.

Balgovind, R., Dalcher, A., Ghil, M. and Kalnay, E., 1983. A stochastic-dynamic model for the spatial structure of forecast error statistics. Mon. Weather Rev., 111: 701–722.

Bengtsson, L., 1975. Four-Dimensional Data Assimilation of Meteorological Observations. GARP Publ. Ser., No. 15, WMO/ICSU, Geneva, 76 pp.

Bengtsson, L., Ghil, M. and Källén, E. (Editors), 1981. Dynamic Meteorology: Data Assimilation Methods. Springer-Verlag, New York, 330 pp.

Bierman, G.J., 1977. Factorization Methods for Discrete Sequential Estimation. Academic Press, Orlando, FL, 241 pp.

Bourke, W., Seaman, R. and Puri, K., 1985. Data assimilation. Adv. Geophys., 28B: 124–149.

Bretherton, F.P., Davis, R.E. and Fandry, C., 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. Deep-Sea Res., 23: 559–582.

Bryan, K. and Sarmiento, J.L., 1985. Modeling ocean circulation. Adv. Geophys., 28A: 433–459.

Bucy, R.S. and Joseph, P.D., 1987. Filtering for Stochastic Processes with Applications to Guidance. 2nd ed., Chelsea, New York, 217 pp.

Budgell, W.P., 1986. Nonlinear data assimilation for shallow water equations in branched channels. J. Geophys. Res., 91: 10633–10644.

Charney, J.G., Fjørtoft, R. and von Neumann, J., 1950. Numerical integration of the barotropic vorticity equation. Tellus, 2: 237–257.

Charney, J.G., Fleagle, R., Lally, V., Riehl, H. and Wark, D., 1966. The feasibility of a global observation and analysis experiment. Bull. Am. Meteorol. Soc., 47: 200–220.

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

Cohn, S.E., 1982. Methods of Sequential Estimation for Determining Initial Data in Numerical Weather Prediction. Ph.D. thesis, Courant Institute of Mathematical Sciences, New York University, 183 pp.

Cornuelle, B., Wunsch, C., Behringer, D., Birdsall, T., Brown, M., Heinmiller, R., Knox, R., Metzger, K., Munk, W., Spiesberger, J., Spindell, R., Webb, D. and Worcester, P., 1985. Tomographic maps of the ocean mesoscale. Part 1: Pure acoustics. J. Phys. Oceanogr., 15: 133–152.

Courtier, P. and Talagrand, O., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. Q. J. R. Meteorol. Soc., 113: 1329–1347.

Cressman, G., 1959. An operational objective analysis system. Mon. Weather Rev., 87: 367–374.

Daley, R., 1980. On the optimal specification of the initial state for deterministic forecasting. Mon. Weather Rev., 108: 1719–1735.

Daley, R., 1981. Normal mode initialization. In: L. Bengtsson, M. Ghil and E. Källén (Editors), Dynamic Meteorology: Data Assimilation Methods. Springer-Verlag, New York, pp. 77–109.

Efron, B., 1982. The Jacknife, the Bootstrap and other Resampling Plans. Society of Industrial and Applied Mathematics, Philadelphia, 92 pp.

Eliassen, A., 1954. Provisional Report on Calculation of Spatial Covariance and Autocorrelation of the Pressure Field. Rept. No. 5, Institute of Weather and Climate Res., Academy of Science Oslo, 11 pp. (reprinted in Bengtsson et al. (1981) pp. 319–330).

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

Errico, R.M., 1982. The strong effects of non-quasigeostrophic dynamic processes on atmospheric energy spectra. J. Atmos. Sci., 39: 961–968.

Gandin, L.S., 1963. Objective Analysis of Meteorological Fields. Gidrometeorol. Izd., Leningrad (in Russian); English translation by Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp. (available from NTIS, as N66-18047).

Gelb, A. (Editor), 1974. Applied Optimal Estimation. M.I.T. Press, Cambridge, MA, 374 pp.

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

Ghil, M., 1986. Sequential estimation and satellite data assimilation in meteorology and oceanography. In: Y.K. Sasaki, T. Gat-Chen, L. White, M.M. Zaman, C. Ziegler, L.P. Chang and D.J. Rusk (Editors), Variational Methods in Geosciences. Elsevier, Amsterdam pp. 91–100.

Ghil, M. and Childress, S., 1987. Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics. Springer-Verlag, New York, 485 pp.

Ghil, M. and Malanotte-Rizzoli, P., 1989. Data assimilation in meteorology and oceanography. In preparation.

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

Ghil, M., Cohn, S., Tavantzis, J., Bube, K. and Isaacson, E., 1981. Applications of estimation theory to numerical weather prediction. In: L. Bengtsson, M. Ghil and E. Källén (Editors), Dynamic Meteorology: Data Assimilation Methods. Springer-Verlag, New York, pp. 139–224.

Ghil, M., Cohn, S.E. and Dalcher, A., 1982. Sequential estimation, data assimilation, and initialization. In: D. Williamson (Editor), The Interaction between Objective Analysis and Initialization. Publ. Meteorol. 127 (Proc. 14th Stanstead Seminar), McGill University, Montreal, pp. 83–97.

Ghil, M., Cohn, S.E. and Dalcher, A., 1983. Applications of sequential estimation to data assimilation. In: Large-Scale Oceanographic Experiments in the World Climate Research Programme, WCRP Publ. Series, No. 1, Vol. II, WMO/ICSU, Geneva, pp. 341–356.

Ghil, M., Benzi, R. and Parisi, G. (Editors), 1985. Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics. North Holland, Amsterdam, 449 pp.

Gill, A.E., 1982. Atmosphere–Ocean Dynamics. Academic Press, New York, 662 pp.

Gustavsson, N., 1981. A review of methods for objective analysis. In: L. Bengtsson, M. Ghil and E. Källén (Editors), Dynamic Meteorology: Data Assimilation Methods. Springer-Verlag, New York, pp. 17–76.

Halem, M., Kalnay, E., Baker, W.E. and Atlas, R., 1982. An assessment of the FGGE satellite observing system during SOP-1. Bull. Am. Meteorol. Soc., 63: 407–426.

Halpern, D., 1987. Data assimilation and ocean general circulation models. EOS (Trans. Am. Geophys. Union), 68: 731–733.

Hollingsworth, A., 1987. Objective analysis for numerical weather prediction. In: T. Matsuno et al. (Editors), Short- and Medium-Range Weather Prediction, pp. 11–59.

Horton, C.W., Reichl, L.E. and Szebehely, V.G. (Editors), 1983. Long-Time Prediction in Dynamics. Wiley-Interscience, New York, 496 pp.

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

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

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

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

Kanamitsu, M., 1981. Some climatological and energy budget calculations using the FGGE III-b analyses during January 1979. In: L. Bengtsson, M. Ghil and E. Källén (Editors), Dynamic Meteorology: Data Assimilation Methods. Springer-Verlag, New York, pp. 263–318.

Kimmeldorf, G. and Wahba, G., 1970. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Statist., 41: 495–502.

Kindle, J.C., 1986. Sampling strategies and model assimilation of altimetric data for ocean monitoring and prediction. J. Geophys. Res., 91: 2418–2432.

Lacarra, J.F. and Talagrand, O., 1988. Short-range evolution of small perturbations in a barotropic model. Tellus, 40A: 81–95.

Leith, C.E., 1980. Nonlinear normal mode initialization and quasi-geostrophic theory. J. Atmos. Sci., 37: 958–968.

Levitus, S., 1982. Climatological Atlas of the World Ocean. NOAA Professional Paper 13, National Oceanic and Atmospheric Administration, Rockville, MD, 173 pp. + 17 microfiches.

Lorenc, A., 1981. A global three-dimensional multivariate statistical interpolation scheme. Mon. Weather Rev., 109: 701–721.

Lorenc, A., 1986. Analysis methods for numerical weather prediction. Q. J. R. Meteorol. Soc., 112: 1177–1194.

Lorenz, E.N., 1960. Maximum simplification of the dynamic equations. Tellus, 12: 243–254.

Lorenz, E.N., 1980. Attractor sets and quasi-geostrophic equilibrium. J. Atmos. Sci., 37: 1685–1699.

Lorenz, E.N., 1985. The growth of errors in prediction. In: M. Ghil, R. Benzi and G. Parisi (Editors), Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics. North-Holland, Amsterdam, pp. 243–265.

Malanotte-Rizzoli, P. and Holland, W.R., 1985. Data constraints applied to models of the general ocean circulation. Part I: the steady case. J. Phys. Oceanogr., 16: 1665–1682.

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

Miller, R.N., 1986. Toward the application of the Kalman filter to regional open ocean modeling. J. Phys. Oceanogr., 16: 72–86.

Miller, R.N., 1987. Theory and Practice of Data Assimilation for Oceanography. Rep. Meteorol. Oceanogr., No. 26, Harvard University, Cambridge, MA, 59 pp.

Munk, W. and Wunsch, C., 1982. Observing the ocean in the 1990's. Philos. Trans. R. Soc. London, Ser. A, 307: 439–464.

Panofsky, H., 1949. Objective weather map analysis. J. Meteorol., 6: 386–392.

Pedlosky, J., 1987. Geophysical Fluid Dynamics, 2nd edn. Springer-Verlag, New York, 710 pp.

Petersen, D.P., 1976. Linear sequential coding of random space-time fields. Info. Sci. (NY), 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.

Phillips, N.A., 1976. The impact of synoptic observing and analysis systems on flow pattern forecasts. Bull. Am. Meteorol. Soc., 57: 1225–1250.

Phillips, N.A., 1982. On the completeness of multi-variate optimum interpolation for large-scale meteorological analysis. Mon. Weather Rev., 110: 1329–1334.

Phillips, N.A., 1983. An accuracy goal for a comprehensive satellite wind measuring system. Mon. Weather Rev., 111: 237–239.

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

Provost, C. and Salmon, R., 1986. A variational method for inverting hydrographic data. J. Mar. Res., 44: 1–34.

Robinson, A.R. and Leslie, W.G., 1985. Estimation and prediction of oceanic eddy fields. Prog. Oceanogr., 14: 485–510.

Robinson, A.R., Spall, M.A., Leslie, W.G., Walstad, L.J. and McGillicuddy, D.J., 1987. Gulfcasting: Dynamical Forecast Experiments for Gulf Stream Rings and Meanders, Nov. 1985–June 1986. Rep. Meteorol. Oceanogr., No. 22, Harvard University, Cambridge, MA, 31 pp. + 72 figs.

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

Sasaki, Y., 1958. An objective analysis based on the variational method. J. Meteorol. Soc. Jpn., 36: 77–88.

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

Susskind, J., Rosenfield, J., Reuter, D. and Chahine, M.T., 1984. Remote sensing of weather and climate parameters from HIRS2/MSU on TIROS-N. J. Geophys. Res., 89C: 4677–4697.

Talagrand, O. and Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory. Q. J. R. Meteorol. Soc., 113: 1311–1328.

Thiébaux, H.J. and Pedder, M.A., 1987. Spatial Objective Analysis. Academic Press, London, 299 pp.

Tribbia, J.J., 1982. On variational normal mode initialization. Mon. Weather Rev., 110: 455–470.

Vautard, R. and Legras, B., 1986. Invariant manifolds, quasi-geostrophy and initialization. J. Atmos. Sci., 43: 565–584.

Wahba, G., 1982. Variational methods in simultaneous optimum interpolation and initialization. In: D. Williamson (Editor), The Interaction between Objective Analysis and Initialization. Publ. Meteorol. 127 (Proc. 14th Stanstead Seminar), McGill University, Montreal, pp. 178–185.

Williamson, D. (Editor), 1982. The Interaction between Objective Analysis and Initialization. Publ. Meteorol. 127 (Proc. 14th Stanstead Seminar), McGill University, Montreal, 193 pp.

Wiener, N., 1949. Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications. M.I.T. Press, Cambridge, MA, 163 pp.

Wiener, N., 1956. Nonlinear prediction and dynamics. In: Proc. 3rd Berkeley Symposium Math. Stat. Probability, Vol. 3, University of California Press, pp. 247–252.

Wunsch, C., 1988. Transient tracers as a problem in control theory. J. Geophys. Res., 93C: 8099–8110.