A Comparison of Adaptive Kalman Filters for a Tropical Pacific Ocean Model

ISABELLE BLANCHET* AND CLAUDE FRANKIGNOUL

Laboratoire d'Océanographie Dynamique et de Climatologie, Université Pierre et Marie Curie, Paris, France

MARK A. CANE

Lamont-Doherty Earth Observatory, Palisades, New York

(Manuscript received 15 November 1995, in final form 5 June 1996)

ABSTRACT

The Kalman filter is the optimal linear assimilation scheme only if the first- and second-order statistics of the observational and system noise are correctly specified. If not, optimality can be reached in principle by using an adaptive filter that estimates both the state vector and the system error statistics. In this study, the authors compare the ability of three adaptive assimilation schemes at estimating an unbiased, stationary system noise. The adaptive algorithms are implemented in a reduced space linear model for the tropical Pacific. Using a twin experiment approach, the algorithms are compared by assimilating sea level data at fixed locations mimicking the tropical Pacific tide gauges network. It is shown that the description of the system error covariance matrix requires too many parameters for the adaptive problem to be well posed. However, the adaptive procedures are efficient if the number of noise parameters is dramatically reduced and their performance is shown to be closed to optimal, that is, based on the true system noise covariance. The link between those procedures is elucidated, and the question of their applicability and respective computational cost is discussed.

1. Introduction

Since oceanic initial conditions are poorly known and numerical ocean models are imperfect, model predictions need to be corrected by observations. An efficient approach to assimilate oceanic data is based on the Kalman filter (Kalman 1960). When observations become available, the filter computes a linear combination of the model forecast and the data, weighted by their accuracy, to give an improved estimate of the ocean state, which is taken as initial condition for the next forecast. The Kalman filter (hereafter KF) gives the best linear unbiased estimate (BLUE) of the ocean state using all observations available up to the analysis time, provided the statistics of the model errors and the observational errors are known; it also calculates the propagation of the error covariance matrix, so the accuracy of the model predictions is estimated at each assimilation step. The optimality of the KF requires that model and data error statistics be known. In the oceanographic context, the errors in the observations are relatively well determined but not those in the forcing fields, which can be substantial. As the forcing errors are not differentiated from the modeling errors in the KF for the ocean model used here, it is their sum that is referred to here as the system errors. Since the system errors are poorly known at best, the KF is suboptimal. The estimation of the state of the ocean is thus far from optimal, and a bad specification of the system or observational errors can even lead to a divergence of the KF, which means that the state and its estimate diverge, and the residuals grow (Jazwinsky 1969). Nonetheless, the observations can be used to improve the representation of the system noise, if an *adaptive filter* is designed to readjust the noise description based on information obtained as the measurements become available (Gelb 1974).

Adaptive KF can simultaneously estimate the state and the system error statistics, but these algorithms require an even greater amount of computation and storage than the standard KF, so that their application to meteorology and oceanography has been very limited. Dee et al. (1985) designed an adaptive KF that was tested in a simplified meteorological context, updating only a few parameters that were assumed to describe the system noise covariance matrix. Instead of improving the representation of the system noise covariance matrix, Hoang et al. (1995) estimated deep ocean currents in a quasigeostrophic model by directly updating the KF gain, which suffices to make better predictions, but provides little information on the model shortcomings.

* Additional affiliation: Centre Militaire d'Océanographie/Service Hydrographique et Océanographique de la Marine, Brest, France.

Corresponding author address: Dr. Isabelle Blanchet, GRGS/BRESM, 14 avenue Edouard Belin, 31055 Toulouse, France.
E-mail: blanchet@thor.cnes.fr

Because it updates the forecast-error covariance matrix, the KF requires huge computations in oceanographic applications, due to the very large dimension n of the state space, n being the number of variables times the number of grid points [typically $n = O(10^5)$]. Indeed, the system error covariance matrix update requires $2n$ integrations of the numerical model. To reduce the cost, many suboptimal schemes have been developed. Parish and Cohn (1985) calculated the error covariance structure only for points separated by small distances, while Dee (1991) used approximate dynamics to propagate the forecast error covariances, and Fukumori et al. (1992), following Wiener (1949), used the asymptotically constant Kalman gain matrix. An alternative is to reduce the dimensionality of the problem by using a low-resolution model (Miller and Cane 1989) or by computing the forecast error covariances with a coarser-resolution model than the model used to forecast the state itself (LeMoyne and Alvarez 1991). Fukumori and Malanotte-Rizzoli (1995) approximated the state forecast covariance matrix by reducing the dimension of the model (coarse grid) and using the asymptotic steady-state limit. To avoid the drawbacks of coarse resolution, Cane et al. (1997) have adopted a different approach: using a multivariate empirical orthogonal function (EOF) analysis, they reduced the state space for the forecast covariance update to a small set of basis functions, which nonetheless represented all of the significant structures that were predicted by the model. The procedure was shown to lead to a substantial saving without any loss of accuracy compared to a gridpoint KF.

Since Cane et al. (1997) have successfully implemented the KF in reduced space, it becomes computationally possible to use an adaptive filter to also estimate the system errors, which should enhance the filter performance. In the present paper, we compare several adaptive schemes that can be applied to the reduced space model, using a twin experiment approach in the tropical Pacific context. A twin experiment is a simulation of data assimilation in fully known conditions based on synthetic data. A first simulation defines the ''true'' state of the ocean, which is given by the model response to observed winds. The ''observations'' are constructed by adding a noise with known error covariance matrix and zero mean to the true sea level, using realistic location for the measuring stations. The system noise is then added, the noise being stationary and with zero mean (unbiased case), and the observations assimilated in a second simulation using an arbitrary first guess for the system noise covariance matrix. The ability of the adaptive algorithms to retrieve the system noise covariance matrix and the true state of the ocean can then be compared. The maximum likelihood estimator of Maybeck (1982), that of Dee (1995), and the more intuitive approach of Myers and Tapley (1976) are considered. Their similarities and shortcomings are pointed out, and a new procedure is designed that satisfactorily estimates the system error covariances, leading to near-optimal performance.

The paper is organized as follows. In the next section, we present the numerical model and the reduced state space KF approach of Cane et al. (1997). In section 3, we review the KF. In section 4, we discuss the adaptive procedure and present three adaptive algorithms. In section 5, we compare these algorithms using the twin experiment approach. The performance and limitations of the adaptive procedures are discussed, and some clues given to improve them. Conclusions are given in section 6.

2. The tropical ocean model

The numerical model used in our study is a two vertical mode version of the Cane and Patton (1984) algorithm for solving the linear long-wave approximation of the shallow-water equations on an equatorial beta plane. The staggered grid is spaced 2° longitude by 0.5° latitude, and the time step of the numerical scheme is one-quarter of a month. The internal wave speeds for the two vertical modes, 2.86 and 1.85 m s^{-1}, are typical of the mid-Pacific (Cane 1984; Busalacchi and Cane 1985). The model has no thermodynamics nor salinity and is forced by surface wind stress. Here the model will be used in the anomaly mode; that is, it will be forced by wind stress anomalies. The tropical Pacific is represented by a basin extending from 28.75°S to 28.75°N and 124°E to 80°W. The coasts of Australia and Mexico are crudely represented by two rectangles on the southwest and the northeast corners, respectively (Fig. 1). The number of grid points is approximately 9000.

To implement the KF for this ocean model, the covariance matrix of the system errors, that is, the modeling plus forcing errors (see section 3), needs to be specified. This means that one should be able to determine its $n(n + 1)/2$ elements, with n (number of grid points times number of variables) $\sim O(10^5)$. However, the oceanographic time series are very short and it is very unlikely that one could determine so many elements. Thus, Cane et al. (1997) have developed a method based on multivariate EOF (MEOF) analysis to reduce the number of degrees of freedom in the covariance calculation. The procedure, which also makes the KF computationally feasible, is the following. The ocean model was forced by the FSU winds (Goldenberg and O'Brien 1981). The six model variables (for each vertical mode: Kelvin wave amplitude, and Rossby mode displacement and zonal velocity) were saved monthly from January 1967 to September 1991, that is, for 297 months. A first set of EOFs was calculated for each of the variables. After normalization ensuring that the six variables had the same relative weight, a second EOF analysis was performed on the combined sets. For our study, 102 of these MEOFs ordered in terms of decreasing variance were retained, accounting for more than 99% of the entire variance of the model state. This

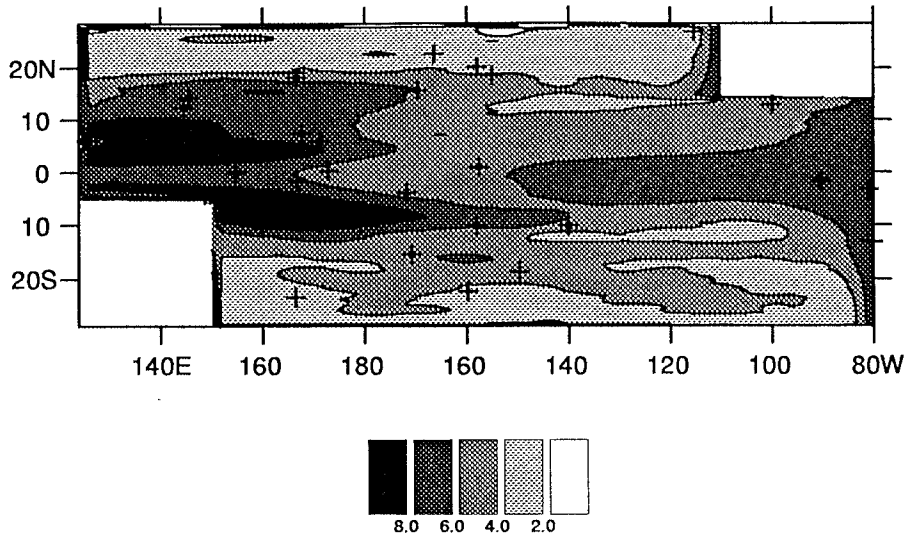

FIG. 1. Standard deviation of the true sea level. Contour interval is 2 cm. The crosses indicate the location of the observing stations.

gives a very efficient basis for the model states occurring during the period 1967–91, but the retained MEOFs may be inadequate to represent all the structures of the true ocean, either because the model is incapable of occupying all regions of phase space where the true ocean can be, or because the wind forcing during this period missed important wind regimes. Means for handling such deficiencies are discussed in Cane et al. (1997).

Cane et al. (1997) compared the results of Miller et al. (1995), who implemented a full grid KF for the Cane and Patton (1984) numerical model with a coarser grid (5° longitude by 2° latitude), with those of a reduced KF experiment with the same coarse grid model. The results were comparable, and the reduced KF performed even better for most of the stations withheld for verification. Cane et al. (1997) argued that the ''higher-resolution'' gridpoint filter tends to overfit the data locally, perhaps causing it to underweight more remote connections.

3. Description of the Kalman filter

The prediction model can be represented by a linear equation:

$$\mathbf{w}^f(k) = \mathbf{L}\mathbf{w}^a(k-1) + \mathbf{F}\mathbf{u}(k-1), \tag{1}$$

where the model prediction $\mathbf{w}^f$ and the analysis $\mathbf{w}^a$ (superscript *f* for forecast, *a* for analysis) are row vectors of dimension *n*, which is normally the number of prognostic variables of the numerical ocean model times its number of grid points, but here is the number of MEOFs. The prediction at time step *k* depends on the model dynamics $\mathbf{L}$, the analyzed state $\mathbf{w}^a$ at the preceding time step, and the forcing term $\mathbf{u}$ of dimension *r* (the wind stress in the present context) projected onto the model variables by the matrix $\mathbf{F}$.

The true state of the ocean $\mathbf{w}^t$ (superscript *t* for true) evolves according to

$$\mathbf{w}^t(k) = \mathbf{L}\mathbf{w}^t(k-1) + \mathbf{F}\mathbf{u}(k-1) + \mathbf{b}^t(k-1), \tag{2}$$

where $\mathbf{b}^t$ accounts for the system errors. Since the wind stress is poorly known, particularly in the Tropics, the system errors take into account both the model errors and the forcing errors. Here, $\mathbf{b}^t$ is a random noise vector that is assumed to be stationary, decorrelated in time with a Gaussian spatial covariance matrix $\mathbf{Q}$:

$$E[\mathbf{b}^t(k)] = \mathbf{q}(k) \tag{3a}$$

$$E\{[\mathbf{b}^t(k)][\mathbf{b}^t(l)]^{\mathrm{T}}\} = \mathbf{Q}(k)\delta_{kl}. \tag{3b}$$

The assumption of serial independence is questionable. Indeed, Dee (1995) showed that the system errors depend on the actual state of the ocean $\mathbf{w}^t$ in an unknown manner. However, the assumption of whiteness is made for simplicity in this first study. Also, model and forcing are assumed to be unbiased, so $\mathbf{q}(k) = 0$. However, our ultimate goal is to design an adaptive filter that can handle the biased, correlated case. If the system noise mean $\mathbf{q}(k)$ is known and differs from zero, then Eq. (1) should be replaced by

$$\mathbf{w}^f(k) = \mathbf{L}\mathbf{w}^a(k-1) + \mathbf{F}\mathbf{u}(k-1) + \mathbf{q}(k). \tag{1b}$$

The observations are related to the model variables by the equation

$$\mathbf{w}^o(k) = \mathbf{H}(k)\mathbf{w}^t(k) + \mathbf{b}^o(k), \tag{4}$$

where $\mathbf{w}^o$ (superscript *o* for observation) is a row vector of dimension *p*, the number of observations, and $\mathbf{H}(k)$ is the observation matrix. This matrix expresses the *p*

observations as a linear combination of the n variables; it can also be used to interpolate the latter to the observing stations. For a fixed observational network, as considered in this paper, $\mathbf{H}$ does not vary in time. The observational noise $\mathbf{b}^o$ is supposed to be Gaussian, stationary, and white in time, with zero mean and spatial covariance matrix $\mathbf{R}$:

$$E[\mathbf{b}^o(k)] = 0 \tag{5a}$$

$$E\{[\mathbf{b}^o(k)][\mathbf{b}^o(l)]^{\mathrm{T}} = \mathbf{R}(k)\delta_{kl}. \tag{5b}$$

The system and observational errors are assumed to be uncorrelated:

$$E[[\mathbf{b}^o(k)][\mathbf{b}^t(1)]^{\mathrm{T}}] = 0. \tag{6}$$

The KF gives an estimate $\mathbf{w}^a(k)$ of the true state of the ocean, $\mathbf{w}^t(k)$, through a linear combination of the observations $\mathbf{w}^o(k)$ and the model predictions $\mathbf{w}^f(k)$:

$$\mathbf{w}^a(k) = \mathbf{w}^f(k) + \mathbf{K}(k)[\mathbf{w}^o(k) - \mathbf{H}(k)\mathbf{w}^f(k)], \tag{7}$$

where $\mathbf{K}$ is a gain matrix, which depends on the relative accuracy of the predictions and the observations. The vector

$$\mathbf{v}(k) = \mathbf{w}^o(k) - \mathbf{H}(k)\mathbf{w}^f(k) \tag{8}$$

represents the additional information provided by the observed data and is called the *observational residual* or *innovation vector*.

The prediction error covariance matrix and the analyzed errors covariance matrix are defined by

$$\mathbf{P}^f(k) = E\{[\mathbf{w}^f(k) - \mathbf{w}^t(k)][\mathbf{w}^f(k) - \mathbf{w}^t(k)]^{\mathrm{T}}\} \tag{9}$$

$$\mathbf{P}^a(k) = E\{[\mathbf{w}^a(k) - \mathbf{w}^t(k)][\mathbf{w}^a(k) - \mathbf{w}^t(k)]^{\mathrm{T}}\}. \tag{10}$$

The Kalman gain $\mathbf{K}$ is determined as the matrix that minimizes the functional J, which represents the mean quadratic error of the analysis and is given by

$$\begin{aligned} J &= E\{[\mathbf{w}^a(k) - \mathbf{w}^t(k)]^{\mathrm{T}}[\mathbf{w}^a(k) - \mathbf{w}^t(k)]\} \\ &= \mathrm{Tr}[\mathbf{P}^a(k)], \end{aligned} \tag{11}$$

which yields

$$\mathbf{K}(k) = \mathbf{P}^f(k)\mathbf{H}(k)^{\mathrm{T}}[\mathbf{H}(k)\mathbf{P}^f(k)\mathbf{H}(k)^{\mathrm{T}} + \mathbf{R}(k)]^{-1}, \tag{12}$$

so that one has

$$\mathbf{P}^f(k) = \mathbf{L}\mathbf{P}^a(k-1)\mathbf{L}^{\mathrm{T}} + \mathbf{Q}(k-1) \tag{13}$$

$$\mathbf{P}^a(k) = \mathbf{P}^f(k) - \mathbf{K}(k)\mathbf{H}(k)\mathbf{P}^f(k). \tag{14}$$

If there is no observation to assimilate at the time step k, then $\mathbf{H}(k)$ is identically zero and so is $\mathbf{K}(k)$, by (12). Note that the matrix $[\mathbf{H}(k)\mathbf{P}^f(k)\mathbf{H}(k)^{\mathrm{T}} + \mathbf{R}(k)]$ can always be inverted since there are no perfect observations, so $\mathbf{R}$ is not singular. The forecast-error covariance matrix $\mathbf{P}^f(k)$ is predicted by (13). This is the most costly step of the KF since it requires $2n$ integrations of the numerical model to obtain the term $\mathbf{L}\mathbf{P}^a\mathbf{L}^{\mathrm{T}}$ [each row of $\mathbf{L}\mathbf{P}^a$ is calculated by using the corresponding row of $\mathbf{P}^a$ as model input and integrating for one time step, then the operation is repeated for each line of $\mathbf{L}\mathbf{P}^a$, which leads to $\mathbf{L}(\mathbf{L}\mathbf{P}^a)^{\mathrm{T}}$].

Equation (14) can be rewritten (Gelb 1974) as

$$\mathbf{P}^a(k)^{-1} = \mathbf{P}^f(k)^{-1} + \mathbf{H}(k)^{\mathrm{T}}\mathbf{R}(k)^{-1}\mathbf{H}(k), \tag{15}$$

making it clear that the accuracy of the analysis is the sum of the accuracy of the forecast, based on the past observations, and that of the current observations. The KF minimizes the variance of the analysis error and therefore the variance of the prediction error not only at each time step but over the entire assimilation interval. The filter extracts all useful information from the innovation vector at each time step, by an application of Bayesian ideas in a dynamical context (Kalman 1960; Ghil and Malanotte-Rizzoli 1991).

The stability of the KF should be examined since stability ensures that any error in the calculation of $\mathbf{w}^a$ (such as computational errors) is bounded. The conditions that guarantee the stability of the filter and the uniqueness of the behavior of $\mathbf{P}^f(k)$ for large k, independently of its initial value $\mathbf{P}(0)$, are complete observability and stochastic uniform complete controllability (Gelb 1974; Fukumori et al. 1992). Observability is the ability to determine the complete state from measurements in the absence of noise. Stochastic controllability is the ability to drive the state from one arbitrary state to another arbitrary state using the stochastic control variables $\mathbf{b}^t$ in Eq. (2). One way of characterizing the observability and controllability of the system is to examine the rank of the observability and controllability matrix. Hautus (1969) provides an alternative method. The system is observable if and only if there is no eigenvector of $\mathbf{L}$ that is in the null-space of $\mathbf{H}$. The system is stochastically controllable if and only if there is no eigenvector of $\mathbf{L}^{\mathrm{T}}$ that is in the null-space of $\mathbf{\Gamma}^{\mathrm{T}}$, where $\mathbf{\Gamma}$ is the matrix mapping the noise forcing $\mathbf{b}^t$ onto the model state.

Following Hautus (1969), we have calculated the eigenvalues and eigenvectors of $\mathbf{L}$ for the reduced space model. All the eigenvalues of $\mathbf{L}$ are less than one, so there are no unstable or neutral modes. No eigenvector of $\mathbf{L}$ is in the null-space of $\mathbf{H}$ (defined in section 5a), so the system is observable. The condition for stochastic controllability is easily verified since, in our case, $\mathbf{\Gamma}$ is the identity matrix. Therefore, we have proved the existence of a stable asymptotic Kalman filter for our experimental context. Note that the stochastic controllability of the system is important, since it implies that the necessarily arbitrary choice of $\mathbf{P}(0)$, the initial value of the forecast covariance matrix, will not influence the performances of the KF after the first few time steps.

4. The adaptive filter

The optimality of the KF depends on the assumption that we know perfectly the statistics of the observational and system errors, which is not the case in the oceanic

context. However, the observations can be used to improve the representation of the system noise as well as the estimation of the ocean state through the adaptive KF (Gelb 1974). In principle, an adaptive filter can estimate both the system and the observational errors. However, adaptive algorithms that try to update both the observational noise and the system noise are not robust, since it is not easy to distinguish between errors in $\mathbf{Q}$ and $\mathbf{R}$ (Groutage et al. 1987; Maybeck 1982). Moreover, using a simple scalar model, Daley (1992) showed that there is no way to distinguish between overspecification of the model error and underspecification of the observation error, and vice versa, since the innovation lag covariances depend only on the ratio of the misspecification factors. Since the observational errors are generally much better known than the system errors, we will only estimate $\mathbf{Q}$ adaptively.

a. The innovation sequence

When the system and observational noise are not serially correlated [(3b) and (3a)], and when the KF is optimal, the innovation sequence $\mathbf{v}(k)$ should be white with zero mean (e.g., Mehra 1970). This means that, if $\mathbf{w}^a$ is an optimal estimate, there is no information left in $\mathbf{v}$ to improve the forecast. The demonstration of this fact is straightforward. Using (4) and (8), one can write

$$\mathbf{v}(k) = \mathbf{H}\mathbf{w}^t(k) + \mathbf{b}^o(k) - \mathbf{H}\mathbf{w}^f(k) \tag{16}$$

$$= \mathbf{H}\mathbf{e}(k) + \mathbf{b}^o(k) \tag{17}$$

with

$$\mathbf{e}(k) = \mathbf{w}^t(k) - \mathbf{w}^f(k), \tag{18}$$

which can be easily related to $\mathbf{e}(\mathrm{k}-1)$ using (1), (2), (4), and (7):

$$\mathbf{e}(k) = \mathbf{L}[\mathbf{I} - \mathbf{K}(k-1)\mathbf{H}]\mathbf{e}(k-1) + \mathbf{b}^t(k-1) - \mathbf{L}\mathbf{K}(k-1)\mathbf{b}^o(k-1). \tag{19}$$

Then, using (3a) and (5a), one has

$$E[\mathbf{e}(k)] = \mathbf{L}[\mathbf{I} - \mathbf{K}(k-1)\mathbf{H}]E[\mathbf{e}(k-1)], \tag{20}$$

so that $E[\mathbf{e}(k)]$ eventually depends only on $E[\mathbf{e}(0)]$, which is assumed to be zero. Since, from (5a) and (17), one has

$$E[\mathbf{v}(k)] = \mathbf{H}E[\mathbf{e}(k)], \tag{21}$$

the innovation vector has zero mean.

The lag j covariance of the innovation sequence, hereafter denoted by $\mathbf{C}_j$, is equal to

$$\mathbf{C}_j = E[\mathbf{v}(k)\mathbf{v}^{\mathrm{T}}(k-j)] = E\{[\mathbf{H}\mathbf{e}(k) + \mathbf{b}^o(k)][\mathbf{H}\mathbf{e}(k-j) + \mathbf{b}^o(k-j)]\}. \tag{22}$$

Since $\mathbf{b}^o(k)$ is independent of $\mathbf{w}^f(k)$ and $\mathbf{w}^t(k)$, using (5b) and (9), the lag zero covariance is given by

$$\mathbf{C}_0 = \mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R}. \tag{23}$$

If $j > 0$, $\mathbf{b}^o(k)$ is independent of $\mathbf{e}(k-j)$ and $\mathbf{b}^o(k-j)$, so one has

$$\mathbf{C}_j = E\{\mathbf{H}\mathbf{e}(k)[\mathbf{H}\mathbf{e}(k-j) + \mathbf{b}^o(k-j)]^{\mathrm{T}}\}. \tag{24}$$

Using (19) recursively, one obtains

$$\mathbf{e}(k) = \left\{\prod_{i=k-j}^{k-1}\mathbf{L}[\mathbf{I} - \mathbf{K}(i)\mathbf{H}]\right\}\mathbf{e}(k-j) + \sum_{i=k-j}^{k-1}\left\{\prod_{m=i+1}^{k-1}\mathbf{L}[\mathbf{I} - \mathbf{K}(m)\mathbf{H}]\right\}\mathbf{b}^t(i) - \sum_{i=k-j}^{k-1}\left\{\prod_{m=i+1}^{k-1}\mathbf{L}[\mathbf{I} - \mathbf{K}(m)\mathbf{H}]\right\}\mathbf{L}\mathbf{K}(i)\mathbf{b}^o(i). \tag{25}$$

So for $j > 0$, (24) can be written as

$$\begin{aligned}\mathbf{C}_j &= \mathbf{H}\left\{\prod_{i=k-j+1}^{k-1}\mathbf{L}[\mathbf{I} - \mathbf{K}(i)\mathbf{H}]\right\} \times \{\mathbf{L}[\mathbf{I} - \mathbf{K}(k-j)\mathbf{H}]\mathbf{P}^f(k-j)\mathbf{H}^{\mathrm{T}} - \mathbf{L}\mathbf{K}(k-j)\mathbf{R}\} \\ &= \mathbf{H}\left\{\prod_{i=k-j+1}^{k-1}\mathbf{L}[\mathbf{I} - \mathbf{K}(i)\mathbf{H}]\right\} \times \mathbf{L}\{\mathbf{P}^f(k-j)\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^f(k-j)\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1} - \mathbf{K}(k-j)\} \\ &\quad \times [\mathbf{H}\mathbf{P}^f(k-j)\mathbf{H}^{\mathrm{T}} + \mathbf{R}].\end{aligned} \tag{26}$$

If $\mathbf{K}(k-j)$ is the optimal Kalman gain defined by (12), it becomes obvious that for $j > 0$ (and similarly for $j < 0$)

$$\mathbf{C}_j = 0 \text{ for } \mathrm{j} \neq 0. \tag{27}$$

This demonstrates that the innovation sequence is white when the filter is optimal, that is, when the noise characteristics are perfectly known.

Jazwinsky (1969) pointed out that the only quantities available for judging filter performance are the magnitude of the residuals (innovation vector) and their expected statistics (zero mean and whiteness). The statistical consistency of the innovation should thus be tested, since the assimilation scheme efficiency cannot be deducted from the size of the residuals alone. The lag j innovation covariance $\mathbf{C}_j$ depends on the difference between the optimal gain $\{\mathbf{P}^f(k-j)\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^f(k-j)\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\}$ and the actual gain $\mathbf{K}(k-j)$ and provides a measure of the optimality of the data assimilation procedure (Daley 1992). If the lagged innovation covariances are not close to zero when the observational and system errors are serially uncorrelated, then the observations are not used efficiently.

The purpose of the adaptive filter is to ''whiten'' the innovation sequence. Using tests for whiteness, the adaptive filter is designed to adjust the system noise parameters on the basis of the sample covariance function $\mathbf{C}_j$ (Mehra 1970; Gelb 1974). Whitening the in-

novation sequence will guarantee the optimality of the filter (Kailath 1968), but will not guarantee that the updated system noise covariance is the true one. In fact, when the number of observations is small, $p \ll n$, as in the present study, $\mathbf{Q}$ cannot be uniquely determined; (26) shows that there are only np equations to estimate $n(n+1)/2$ parameters. One can determine $\mathbf{K}$ but not $\mathbf{Q}$.

The techniques of innovation whitening (Mehra 1970) consist in relating system noise parameters to the sample estimates of the lag-covariance function $\hat{\mathbf{C}}_j = N^{-1}\sum_{i=j+1}^{j+N}\mathbf{v}(i)\mathbf{v}(i-j)^{\mathrm{T}}$, using the KF equations (12), (13), and (26), and then solving them to obtain parameter estimates. However, these methods are more readily applicable to time-invariant systems, that is, fixed observation networks, linear dynamics, and stationary noises. To be more general, we will concentrate on adaptive algorithms, which estimate the noise parameters only from the lag zero covariance $\hat{\mathbf{C}}_0(k)$ at each time step k. The lagged innovation covariances will be used only as a diagnosis for the adaptive schemes performance.

b. Myers and Tapley's estimator

Myers and Tapley (1976) (MT hereafter) have designed an adaptive filter based on ''empirical estimators'' that can estimate both the system errors covariance matrix and the bias of a model. They consider that the system noise was slowly varying but remained stationary over N time steps, but here we make the reasonable assumption that the system noise is fully stationary so that $\mathbf{Q}$ and $\mathbf{q}$ are constant in time.

The true state of the ocean $\mathbf{w}^t$ is not known, so a first estimate, $\mathbf{q}_s(i)$, of the system noise at time step i, $\mathbf{b}^t(i)$, can be derived from (2) when replacing $\mathbf{w}^t$ by $\mathbf{w}^a$:

$$\begin{aligned}\mathbf{q}_s(i) &= \mathbf{w}^a(i) - [\mathbf{L}\mathbf{w}^a(i-1) + \mathbf{F}\mathbf{u}(i-1)]\\ &= \mathbf{w}^a(i) - \mathbf{w}^f(i) + \hat{\mathbf{q}}(i-1). \qquad (28)\end{aligned}$$

If the $\mathbf{q}_s(i)$ are assumed to be representative of the $\mathbf{b}^t(i)$, they may be considered to be independent and identically distributed. An unbiased estimator for $\mathbf{q}$ at time step k is the sample mean over the last N steps,

$$\hat{\mathbf{q}}(k) = \frac{1}{N}\sum_{i=k-N+1}^{k}\mathbf{q}_s(i). \qquad (29)$$

An unbiased estimator for $\mathbf{Q}$ is obtained by first constructing a similar estimator for the covariance. Then, by removing its expected value, MT find the unbiased estimate

$$\hat{\mathbf{Q}}(k) = \frac{1}{N-1}\sum_{i=k-N+1}^{k}\left\{[\mathbf{q}_s(i)-\hat{\mathbf{q}}(k)][\mathbf{q}_s(i)-\hat{\mathbf{q}}(k)]^{\mathrm{T}} - \left(\frac{N-1}{N}\right)[\mathbf{L}\mathbf{P}^a(i-1)\mathbf{L}^{\mathrm{T}} - \mathbf{P}^a(i)]\right\}. \qquad (30)$$

At each time step $k \geq N$, the adaptive filter requires the storage of the last N noise samples, and the estimation of (29) and (30). The state propagation equations (1) and (13) are replaced by

$$\mathbf{w}^f(k) = \mathbf{L}\mathbf{w}^a(k-1) + \mathbf{F}\mathbf{u}(k-1) + \hat{\mathbf{q}}(k-1) \qquad (31)$$

$$\mathbf{P}^f(k) = \mathbf{L}\mathbf{P}^a(k-1)\mathbf{L}^{\mathrm{T}} + \hat{\mathbf{Q}}(k-1), \qquad (32)$$

which shows that if the system noise increases, as does the prediction error covariance matrix $\mathbf{P}^f$. As a consequence, the gain $\mathbf{K}$ (12) increases and more confidence is given to the observations. The initial value $\hat{\mathbf{Q}}(0)$ must be specified.

Myers and Tapley pointed out that the estimator (30) may become negative definite in numerical applications, in particular when the length of the sample N is small. To avoid this problem, MT reset all the negative diagonal elements to their absolute value, while Groutage et al. (1987) preferred to keep the last semidefinite positive estimate of $\mathbf{Q}$. For small N, we found out that negative values occur too often to use Groutage's approach, as it does not allow the estimator to vary freely. Hence, we use a new approach to guarantee the semidefinite positiveness of $\hat{\mathbf{Q}}$: we calculate the eigenvalues of $\mathbf{Q}$ at each time step and reset the negative eigenvalues to zero.

c. Maybeck's estimator

Another approach to determining the system error covariance matrix $\mathbf{Q}$ was suggested by Maybeck (1982), who used a maximum-likelihood estimator. As MT, Maybeck considered that the noise was *essentially stationary over N sample periods*, but he considered only the case of an unbiased system noise ($\mathbf{q} = 0$). To obtain an explicit maximum-likelihood estimator and to make it feasible for on-line computations, Maybeck had to make several simplifications (see Maybeck 1982 for details), yielding for known $\mathbf{R}$ the expression

$$\hat{\mathbf{Q}}_{\mathrm{May}}(k) = \frac{1}{N}\sum_{j=k-N+1}^{k}\{\mathbf{K}(j)\mathbf{v}(j)\mathbf{v}(j)^{\mathrm{T}}\mathbf{K}(j)^{\mathrm{T}} - [\mathbf{L}\mathbf{P}^a(j-1)\mathbf{L}^{\mathrm{T}} - \mathbf{P}^a(j)]\}, \qquad (33)$$

where $\mathbf{v}$ is the innovation vector (8). As before, $\mathbf{Q}_{\mathrm{May}}(0)$ must be specified.

The empirical estimator (30) of MT has not previously been compared to Maybeck's maximum-likelihood estimator (33), but it is easy to show that they are identical under Maybeck's assumption that the bias is known and equal to zero. In this case, the unbiased estimator (30) of MT becomes

$$\hat{\mathbf{Q}}_{\mathrm{MT}}(k) = \frac{1}{N}\sum_{i=k-N+1}^{k}\{[\mathbf{q}_s(i)][\mathbf{q}_s(i)]^{\mathrm{T}} - [\mathbf{L}\mathbf{P}^a(i-1)\mathbf{L}^{\mathrm{T}} - \mathbf{P}^a(i)]\}. \qquad (34)$$

Now, using (1), (7), and (8) one can write (28) in the form

$$\mathbf{q}_s(i) = \mathbf{w}^a(i) - \mathbf{w}^f(i) = \mathbf{K}(i)[\mathbf{w}^o(i) - \mathbf{H}(i)\mathbf{w}^f(i)]$$
$$= \mathbf{K}(i)\mathbf{v}(i). \quad (35)$$

Replacing $\mathbf{q}_s(i)$ in (34) recovers Maybeck's (33).

d. A maximum-likelihood estimator

To reduce the costs of estimating the $n(n + 1)/2$ elements of the system noise covariance matrix, Bélanger (1974) suggested that the latter be parameterized first, so that only a few parameters are needed. Considering a stationary case, Bélanger (1974) assumed that both $\mathbf{R}$ and $\mathbf{Q}$ were linear functions of prescribed matrices, and he estimated them simultaneously. The Bélanger's algorithm was reformulated by Dee et al. (1985), who made it computationally more efficient by formulating the noise covariance estimation as a secondary filter. This algorithm was successfully applied to a simple shallow-fluid model of a form similar to that used in operational numerical weather prediction. However, its convergence was very slow when data were processed too soon and the assumed initial errors in the parameters too small.

Dee (1995) presented a simpler adaptive scheme based on maximum-likelihood estimation. The elements of the innovation vector $\mathbf{v}$ were supposed to be jointly distributed with

$$E(\mathbf{v}) = 0 \quad (36)$$

$$E(\mathbf{v}\mathbf{v}^T) = \mathbf{C}_0(\boldsymbol{\alpha}), \quad (37)$$

where $\boldsymbol{\alpha}$ is the M-dimensional vector of the unknown parameters describing the covariance matrix, $M << p$ (p being the number of observations). Dee (1995) suggested estimating $\boldsymbol{\alpha}$ on the basis of a single sample of $\mathbf{v}$ by maximizing its conditional probability density function $p(\mathbf{v} \mid \boldsymbol{\alpha})$, which was assumed to be Gaussian:

$$p(\mathbf{v} \mid \boldsymbol{\alpha}) = (2\pi)^{-p/2}[\det\mathbf{C}_0(\boldsymbol{\alpha})]^{-1/2} \times \exp\left[-\frac{1}{2}(\mathbf{v}^T\mathbf{C}_0^{-1}(\boldsymbol{\alpha})\mathbf{v})\right]. \quad (38)$$

The maximum-likelihood estimate of $\boldsymbol{\alpha}$ is then given by

$$\boldsymbol{\alpha}_{\text{ml}} = \underset{\boldsymbol{\alpha}}{\text{argmax}}\, p(v \mid \boldsymbol{\alpha}) = \underset{\boldsymbol{\alpha}}{\text{argmax}}\, \text{f}(\boldsymbol{\alpha}) \quad (39)$$

with

$$\text{f}(\boldsymbol{\alpha}) = \ln[\det\mathbf{C}_0(\boldsymbol{\alpha})] + \mathbf{v}^T\mathbf{C}_0^{-1}(\boldsymbol{\alpha})\mathbf{v}. \quad (40)$$

Here, $\mathbf{C}_0(\boldsymbol{\alpha})$ can be related to the system noise covariance matrix, using (23) and (13):

$$\mathbf{C}_0(k, \boldsymbol{\alpha}) = \mathbf{H}\mathbf{L}\mathbf{P}^f(k)\mathbf{L}^T\mathbf{H}^T + \mathbf{R}$$
$$= \mathbf{H}\mathbf{L}\mathbf{P}^a(k-1)\mathbf{L}^T\mathbf{H}^T + \mathbf{H}\mathbf{Q}(k-1, \boldsymbol{\alpha})\mathbf{H}^T + \mathbf{R}. \quad (41)$$

Since the estimation of $\boldsymbol{\alpha}$ is done on the basis of a single sample of $\mathbf{v}$, $\mathbf{P}^a$, which depends only on $\mathbf{Q}(i)$ with $i < k$, is independent of $\boldsymbol{\alpha}$, so that only $\mathbf{Q}(\boldsymbol{\alpha})$ is to be determined.

Dee (1995) pointed out that single-sample covariance estimation is only reasonable if the number of observations is more than two orders of magnitude larger than the number of parameters to be estimated. He also showed that the simultaneous estimation of several parameters led to a large variance of the single-sample estimates. Since the number of observations is rather small in the oceanographic context, it does not seem to be reasonable to use the algorithm in this form.

Indeed, since the system noise is stationary, we can modify the algorithm and directly use the entire sequence of the innovation, rather than only one sample at a time. The new estimator is written $p(\mathbf{V}(\text{kstep}) \mid \boldsymbol{\alpha})$, where $\mathbf{V}(i)$ denotes the innovation sequence during the first i steps, and kstep is the total number of time steps of the assimilation run. Using repeated applications of Bayes's rule, we can write

$$p[\mathbf{V}(\text{kstep}) \mid \boldsymbol{\alpha}] = \left\{\prod_{i=2}^{\text{kstep}} p[\mathbf{v}(i) \mid \mathbf{V}(i-1), \boldsymbol{\alpha}]\right\} p[\mathbf{v}(1) \mid \boldsymbol{\alpha}]. \quad (42)$$

Each probability density in the right-hand side of (42) is given as in (38) by

$$p[\mathbf{v}(i) \mid \mathbf{V}(i-1), \boldsymbol{\alpha}] = (2\pi)^{-p/2}[\det\mathbf{C}_0(\mathbf{i}, \boldsymbol{\alpha})]^{-1/2} \times \exp\left\{-\frac{1}{2}[\mathbf{v}(i)^T\mathbf{C}_0^{-1}(i, \boldsymbol{\alpha})\mathbf{v}(i)]\right\} \quad (43)$$

so that the new functional to minimize is given by

$$f(\boldsymbol{\alpha}) = \sum_{i=1}^{\text{kstep}} \{\ln[\det\mathbf{C}_0(i, \boldsymbol{\alpha})] + \mathbf{v}(i)^T\mathbf{C}_0^{-1}(i, \boldsymbol{\alpha})\mathbf{v}(i)\}, \quad (44)$$

where $\mathbf{C}_0(i, \boldsymbol{\alpha})$ is related to $\mathbf{Q}(\boldsymbol{\alpha})$ by (41), with $\mathbf{Q}$ being time independent. This functional is simply the sum over all time steps of the functional (40), and $\boldsymbol{\alpha}$ is estimated from kstep times p observations instead of only p. A reasonable number of parameters can thus be estimated.

For the single-sample estimator (40), Dee (1995) showed that the gradient of f can be easily calculated if the dependence of $\mathbf{Q}$ on $\boldsymbol{\alpha}$ is of a simple form. However, this is not true for the functional (44) since $\mathbf{P}^a(k)$ depends on $\boldsymbol{\alpha}$ in a complicated way through (12), (13), and (14). To solve the minimization problem, we use a quasi-Newton method and calculated the Hessian matrix numerically.

The maximum-likelihood estimator defined by (42) turns out to be directly related to that of Maybeck (1982, section 4c), a fact obscured by the simplifications introduced by Maybeck to attain on-line applicability. In the appendix, we show that the two estimators essentially differ by the number of time steps that are considered and by the method for solving the minimization problem.

5. Testing the adaptive algorithms

a. A twin experiment

The assimilation experiments, including the forecast calculations (1), are done entirely in the reduced space. The small dimension of the reduced space ($n = 102$) allows us to run many experiments. To test the various adaptive algorithms, we use a twin experiment approach. The true state of the tropical Pacific ocean (Fig. 1) is simulated by the outputs of the model projected onto the reduced space and forced by the unperturbed wind stress anomalies over the 216-month period December 1974 to December 1992,

$$\mathbf{w}^t(k) = \mathbf{L}\mathbf{w}^t(k-1) + \mathbf{u}_{\mathrm{FSU}}(k-1). \qquad (45)$$

The matrix $\mathbf{L}$ thus represents the projection of the ocean model dynamics onto the reduced space. The wind stresses are derived from the FSU pseudostress analysis based on ship observations (Goldenberg and O'Brien 1981), smoothed and detrended in the manner described by Cane et al. (1986). The forcing $\mathbf{u}$ in (45) corresponds to the monthly response of the ocean model to the prescribed wind stress, starting from a model state equal to zero, and projected onto the reduced space; hence $\mathbf{F}$ in (1) is therefore the identity matrix. The time step of the dynamical model in the physical space is one-quarter month, but as observations are assimilated only every month, it is taken to be one month in the reduced space.

The sea level observations are simulated using the observational network in Fig. 1, which reproduces 34 tide gauge stations at various Pacific islands (Wyrtki et al. 1988). They are derived from the true sea level by adding a zero mean, spatially decorrelated white noise with diagonal covariance matrix $\mathbf{R}$. Following Miller and Cane (1989), we consider that the tide gauge data are accurate within 3 cm, so that $\mathbf{R} = 9\ \mathrm{cm}^2 \times \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of order 34. The observation matrix $\mathbf{H}$ is independent of time and given by the projection of the observation matrix in the physical space.

The model and forcing errors (i.e., the system errors) are assumed for simplicity to be dominated by wind stress errors and simulated by perturbing the FSU wind stress by a noise $\mathbf{u}'$, following Miller and Cane (1989) and Cane et al. (1997). Although the FSU wind stress errors are highly inhomogeneous, because of the distribution of the ship tracks and the variations of the wind field, we take them as spatially homogeneous. In full space, the wind stress errors $e(x, y, t)$ are assumed to be Gaussian, serially decorrelated, with a spatial structure given by

$$E[e(x, y, k)e(x', y', \mathrm{l})] = W\delta(\mathrm{k} - \mathrm{l})\exp\left[-\frac{(x-x')^2}{L_x^2} - \frac{(y-y')^2}{L_y^2}\right], \qquad (46)$$

where $W = (\rho_a/\rho_w C_D)^2 \times (266\ \mathrm{m}^2\ \mathrm{s}^{-2})^2$, $L_x = 10°$, and $L_y = 4°$. The amplitude W was determined as in Miller and Cane (1989).

The system error covariance matrix in the reduced space is determined using a ''Monte Carlo run.'' The model is forced by a random zonal wind stress generated from (46) for 2001 months. The model state is reset to zero at the midpoint of each month, and the model fields are projected after one month onto the MEOFS. The model noise $\mathbf{b}^t$ is obtained as the projection onto the reduced space of the model response to the monthly forcing stress. The covariance matrix in the reduced space is a sample estimate of $\mathbf{Q}$ calculated from the 1965 months that remain after the 36 months of spinup are discarded. In the present paper, this matrix will be referred to as the true system error covariance matrix $\mathbf{Q}_{\mathrm{true}}$. The system noise corresponds to an rms sea level error of 2.3 cm, a conservative estimate that is reasonable here, where the fields have been smoothed by the EOFs truncation. Figure 2 (upper left) shows that the system errors project most strongly on the first few EOFs and that there are large error covariances.

An unfiltered run (hereafter UR) obtained by forcing the reduced model with the perturbed FSU wind stress defines the ''false'' state of the tropical Pacific ocean,

$$\mathbf{w}^f(k) = \mathbf{L}\mathbf{w}^f(k-1) + [\mathbf{u}_{\mathrm{FSU}}(k-1) + \mathbf{u}'(k-1)]. \qquad (47)$$

To be consistent with the notations of section 3, we note

$$\mathbf{u}(k) = \mathbf{u}_{\mathrm{FSU}}(k) + \mathbf{u}'(k), \qquad (48)$$

so that $\mathbf{b}^t(k) = -\mathbf{u}'(k)$. The unfiltered run represents the traditional forecast run when no data are assimilated, using an imperfect model.

We then define an a priori covariance matrix that is supposed to reflect our imperfect knowledge of the system errors. As commonly done (e.g., Gourdeau et al. 1992), we choose the simplest covariance matrix, noted $\mathbf{Q}_{\mathrm{a\ priori}}$: the unity matrix scaled by an amplitude factor. Thus, no account is taken of the particular space we are working in (the EOFs space). In keeping with most modelers' optimism, we strongly underestimate the system errors: the trace of $\mathbf{Q}_{\mathrm{a\ priori}}$ amounts to 2‰ of that of the true one. The KF run with $\mathbf{Q}_{\mathrm{a\ priori}}$ (hereafter PKF) is stable and is a first reference run allowing us to evaluate the improvement brought by the adaptive procedure. This filter is suboptimal and the innovation sequence is red at all the observing stations. Their strong serial correlation is seen in Fig. 3 (dashed line).

To compare the different runs in identical settings, we have calculated the correlation and the rms difference between the sea level observations and predictions at the observations points. We have also calculated the correlation and the rms differences with the true state in the EOF space. Each calculation is done over the 216 time steps, for both the forecast and the analyzed state. The results of UR are given in Table 1 (top line). PKF with $\mathbf{Q}_{\mathrm{a\ priori}}$ does not improve the predictions skill, which remains close to that of UR (Table 1, second line). Indeed, the a priori system errors are small so the Kalman

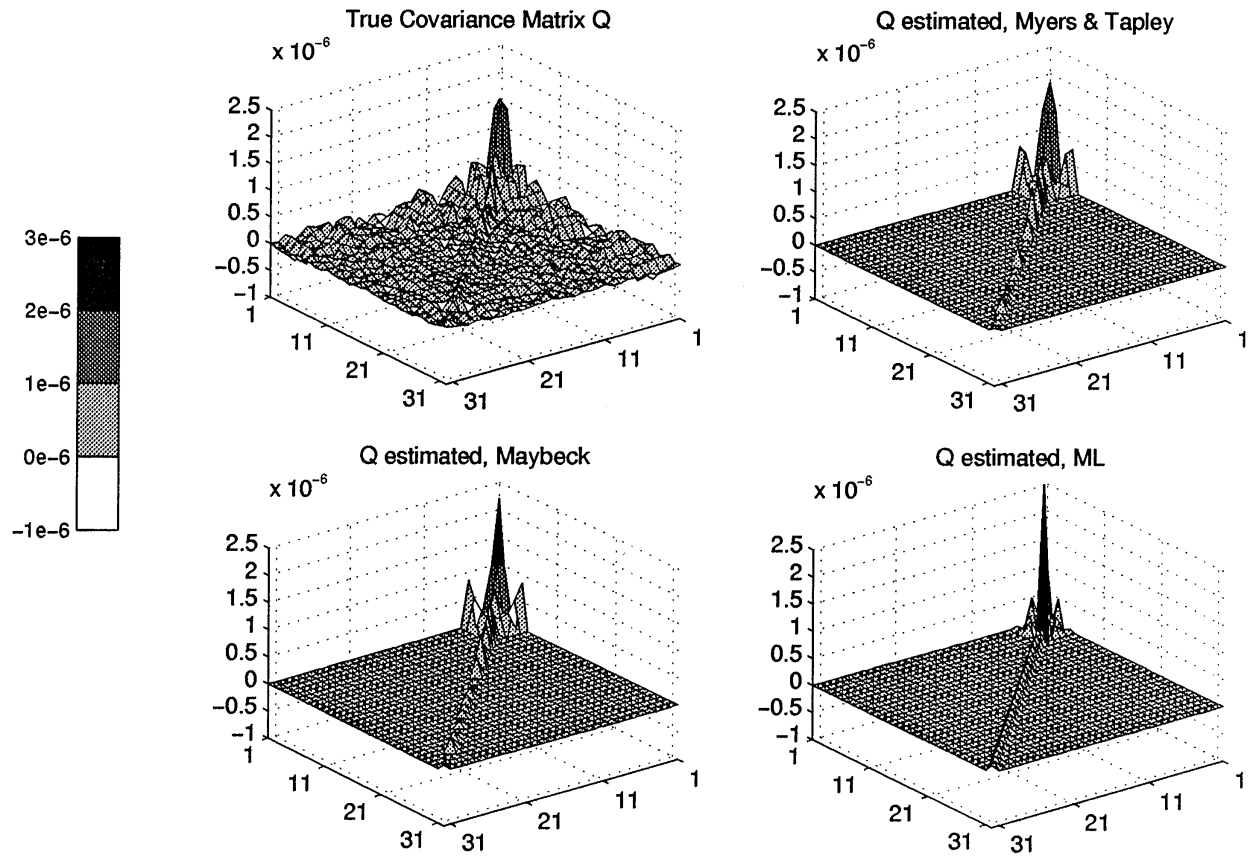

FIG. 2. Mapping onto the first 31 EOFs of the true system noise covariance matrix (upper left) and of the covariance matrix estimated using the adaptive procedures of MT (112 parameters, upper right) and Maybeck (112 parameters, lower left) and the ML estimator (16 parameters, lower right). The *x* and *y* axes represent, respectively, the line and row numbers of the elements of the covariance matrix, and the *z* axis represents their amplitude.

gain is small, and therefore the observations have little influence on the estimation [Eq. (7)].

A second KF reference run is made with the true system error covariance matrix $\mathbf{Q}_{\text{true}}$ (hereafter TKF). This new run gives, by definition, the optimal results for a KF, and the innovation sequence is white (Fig. 3, dotted line). Thus, TKF defines for the forecasts the lower limits of the rms differences and the upper limits of the correlations (Table 1, bottom line). The efficiency of the adaptive algorithms will be given by how closely the rms differences and correlations approach these optimal values.

To test the adaptive algorithms of MT and Maybeck, two runs will be considered below:

1) an adaptive KF run (hereafter AKF), which starts with $\mathbf{Q}_{\text{a priori}}$ as a first guess; an updated matrix is then calculated at each time step; and
2) a straightforward KF run based on a mean covariance matrix estimate $\mathbf{Q}_{\text{est}}$ derived from the AKF run as described below (hereafter UKF, where U stands for updated). Since the true model noise is stationary, this final run should best characterize the performance of the adaptive algorithm.

b. Myers and Tapley's algorithm

1) STABILITY OF THE ALGORITHM

To test the robustness of MT's algorithm, a first adaptive AKF run was performed with sample length $N = 10$. At each time step we calculated the Frœbenius norm and the trace of the estimated $\mathbf{Q}$ matrix, and we compared them with their true values (these two measures discriminate between errors in the diagonal and off-diagonal terms). The run showed a suspicious increase in these two quantities, so we made a longer run obtained by simply repeating the FSU winds sequence five times, which clearly demonstrated that the adaptive procedure diverges (Fig. 4, top). The experiment was repeated with $N = 5$ and $N = 20$, with similar results (the smaller the value of N, the faster the divergence). It turns out that the first five diagonal terms of the matrix are stable, while the higher diagonal terms diverge. It is thus in directions of little variability that the algorithm becomes unstable. This instability arises from the ill-conditioning of the estimation problem, since the number of parameters to be estimated in the covariance matrix and the bias is 5355

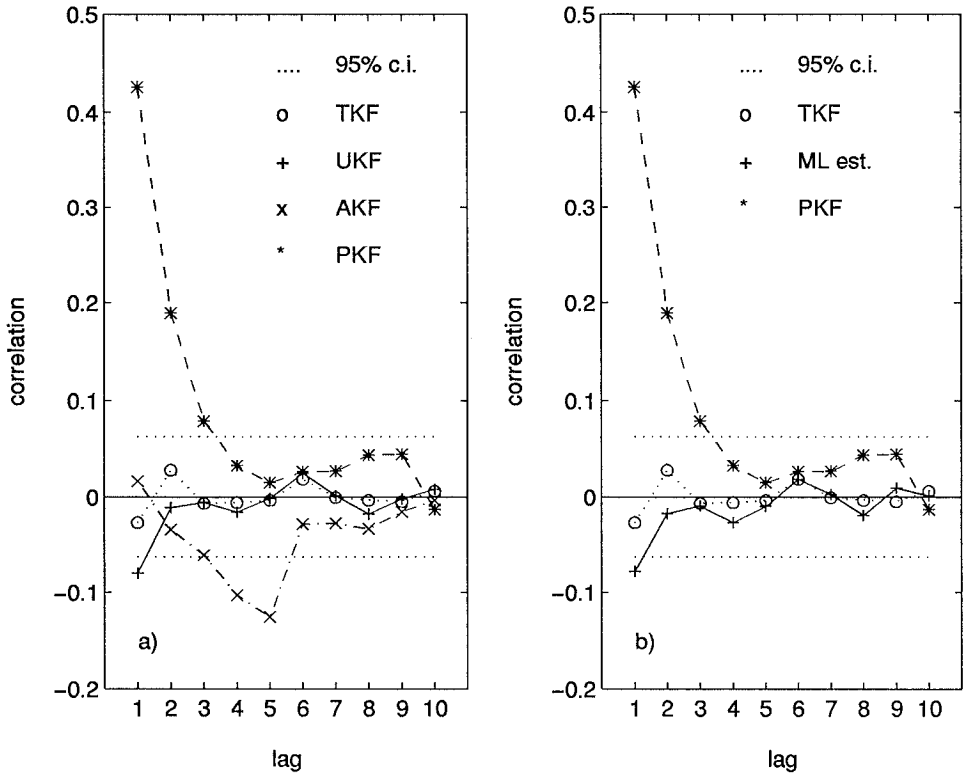

FIG. 3. Mean over the 34 observing stations of the lagged autocorrelation of the innovation sequence for the two reference runs PKF (∗) and TKF (○) for (a) the AKF (×) and UKF (+) runs of the MT algorithm and (b) the ML estimator (+); lag 1–10. The dotted horizontal lines give the 95% confidence interval for zero correlation at the 34 stations. The correlations have been calculated using the last 156 time steps when the filters are in steady state.

[(102 × 103/2) + 102], while there are only 216 × 34 = 7344 observations. The number of parameters must thus be dramatically reduced.

2) PARAMETERS REDUCTION AND DEFINITION OF THE ESTIMATED MATRIX

Assuming that the atmospheric flow in the midlatitudes evolves on two distinct timescales, Dee et al. (1985) parameterized separately the slow and fast parts of the system noise, thus reducing the number of parameters. However, it is better here to take advantage of the particular nature of the EOF coordinates to identify the crucial terms that need to be updated since the EOFs are ordered in terms of decreasing variance. This points to an inherent advantage of the reduced space KF for adaptive error estimation. Thus, rather than estimating the entire system error covariance matrix, we choose to update only its diagonal terms and the first five covariances, keeping the other elements to zero. The number of parameters for the

TABLE 1. Performance of the adaptive algorithms of MT and of Maybeck when estimating 112 parameters. AKF is the adaptive KF run. UKF is a KF run with the covariance matrix estimated from AKF. Root-mean-square differences and correlations are given for the observations at the 34 stations (left) and for the true state of the ocean (right) for both the forecasts and the analysis. For reference, values are also given for UR, the unfiltered run; PKF, the KF run with $\mathbf{Q}_{\text{a priori}}$; and TKF, the KF run with $\mathbf{Q}_{\text{true}}$.

		Observations				True state			
		Correlation		rms (cm)		Correlation		rms (× 10^{-4})	
		Forecast	Analyses	Forecast	Analysis	Forecast	Analyses	Forecast	Analysis
Unfiltered		0.68		6.21		0.66		7.48	
PKF		0.69	0.70	6.09	6.00	0.66	0.66	7.37	7.34
Myers and Tapley	AKF	0.73	0.87	5.53	3.70	0.68	0.71	7.45	6.92
	UKF	0.77	0.95	4.95	2.31	0.76	0.80	5.84	5.25
Maybeck	AKF	0.75	0.89	5.21	3.38	0.74	0.77	6.25	5.80
	UKF	0.77	0.95	4.95	2.25	0.76	0.80	5.84	5.25
TKF		0.79	0.94	4.72	2.37	0.81	0.86	5.05	4.05

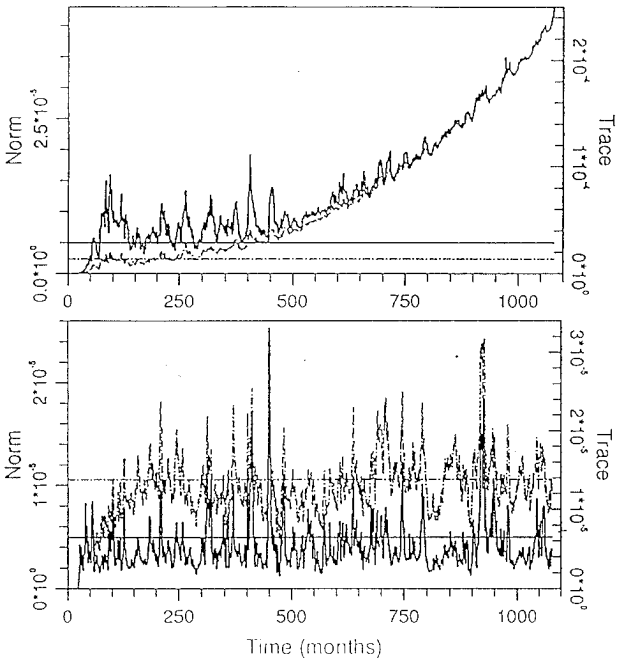

FIG. 4. Time evolution over 90 years of the Frœbenius norm (continuous line) and trace (dash–dotted line) of the estimator $\hat{\mathbf{Q}}$ of the system noise covariance matrix, using the MT adaptive scheme and estimating all covariances (top) or only the covariances between the first five EOFs and the diagonal terms (bottom). The horizontal lines indicate the true values.

covariance matrix drops from 5253 to 112. In fact, the results turn out to be insensitive to the update of the covariances, and estimating only the diagonal of the matrix (102 parameters) gives comparable performance.

The simplified covariance matrix is updated with (30). As seen in Fig. 4 (bottom), the robustness of the procedure is greatly improved: after a rapid initial increase, norm and trace fluctuate around their true value. This variability arises from the limited length N of the sample used in (29) and (30). Consider, for example, the trace, which is the sum of squares of Gaussian variables with zero mean but different variance. Following Stuart and Ord (1991), one can show that the trace is approximately distributed as g times a χ^2 variate with ν degrees of freedom, where g and ν are determined by identifying the moments of order 1 and 2. Based on the trace of $\mathbf{Q}_{\text{true}}$, the 95% confidence interval for the trace of the estimated matrix is found to be in good agreement with the fluctuations in Fig. 4.

To decrease the sampling variability, we define an estimated matrix $\mathbf{Q}_{\text{est}}$ as the sample mean of $\hat{\mathbf{Q}}$ over the last 50 assimilation steps. Its structure is compared to that of the true one in Fig. 2, which shows that the variances are a bit too large. The system error covariances between the first five EOFs are also slightly overestimated, presumably because we have neglected all the other covariances. To make the comparison easier between the estimated and the true matrices, Fig. 5 shows the root-mean-square system noise variance projected onto the sea level. This quantity is actually the square root of the diagonal term of the matrix $\mathbf{H}_{\text{grid}}\mathbf{Q}\mathbf{H}_{\text{grid}}^{\text{T}}$, where $\mathbf{H}_{\text{grid}}$ is the matrix that maps the state vector onto the gridded map of sea level anomalies. Globally the system errors given by $\mathbf{Q}_{\text{est}}$ are overestimated (more than 1 cm) in the extraequatorial regions (north of 10°N and south of 10°S), giving too much weight to the observations (Fig. 6). On the contrary, between 5°S and 5°N the errors are underestimated in the central and eastern part of the basin. This is consistent with the dynamics of the model that performs best near the equator, and it reflects equatorial wave propagation.

3) PERFORMANCE OF THE ALGORITHM

The adaptive run (AKF) and the run with $\mathbf{Q}_{\text{est}}$ (UKF) are compared in Table 1 and Fig. 7, which shows the performance in terms of the index

$$\text{perf} = \frac{\text{rms(filtered run)} - \text{rms(UR)}}{\text{rms(TKF)} - \text{rms(UR)}}, \qquad (49)$$

so that 0 corresponds to the unfiltered run and 1 to the KF run with the true covariance matrix.

First, one should notice that the on-line adaptive procedure AKF does not clearly improve the predictions of the model: although the agreement with the observations is slightly better in AKF, the rms error with the true state increases. The poor performance of AKF is linked to its use of a poorly estimated bias [section 5b(5)]. On the other hand, the bias is ignored in UKF, where the noise is assumed to have zero mean and UKF performs nearly as well as TKF.

The UKF results are very close to those of TKF in the observation space, since the observations have been used to update the covariance matrix $\mathbf{Q}$. In fact, the rms difference between observations and analysis is even lower for UKF than for TKF, consistent with the slight overestimation of $\mathbf{Q}_{\text{est}}$, which leads to bigger forecast errors [Eq. (13)], a bigger Kalman gain [Eq. (12)], hence an analyzed field that is closer to the observations. In the model space, the results have been considerably improved by using $\mathbf{Q}_{\text{est}}$, but they fall short of optimal.

The geographical distribution of the rms differences between forecast and true sea level anomalies (Fig. 8) confirms that the adaptive procedure with the simplified covariance matrix significantly improves the predictions of the KF, which has become close to optimal. The predictions are improved not only at the observing stations, but over the entire basin through the propagation of information by the model dynamics.

4) "WHITENESS" OF THE INNOVATION SEQUENCE

Since the innovation sequence is white when the filter is optimal, a good adaptive algorithm should whiten the innovation sequence (section 4a). Following Jazwinsky (1969), who pointed out that looking only at the size of the residuals is not sufficient to judge the performance

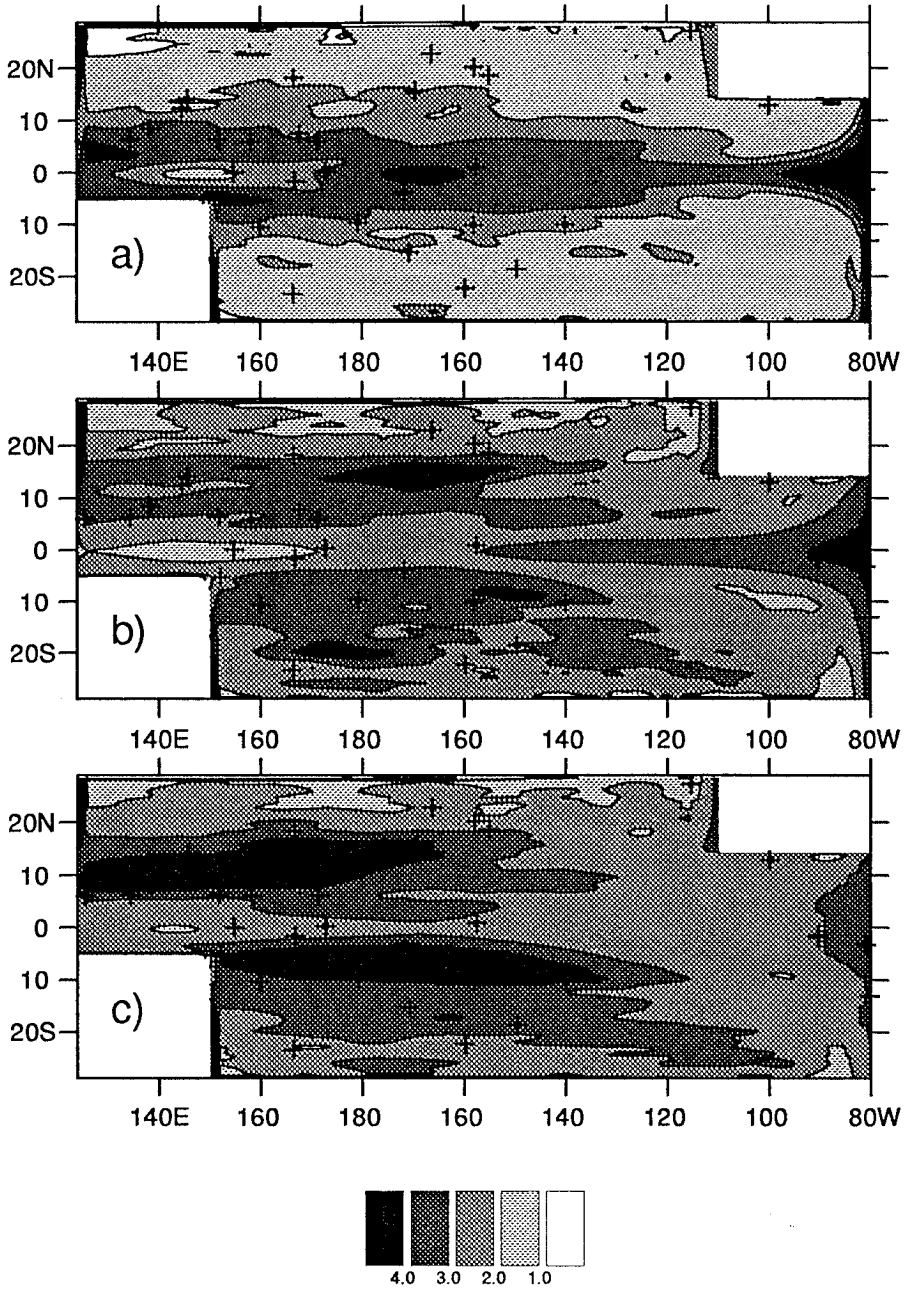

FIG. 5. Root-mean-square of the system noise projected onto sea level for (a) the true system noise, (b) the noise estimated by MT, and (c) the noise estimated by ML. Contour interval is 1 cm.

of the adaptive procedure, we have investigated the statistical characteristics of the innovation sequence of the AKF and UKF runs and compared them with those of the PKF and TKF runs. Figure 3a shows that, when the filter is in a steady state the adaptive procedures are efficient at ''whitening'' the innovation sequences, which are slightly correlated for AKF and nearly uncorrelated for UKF. Pure whiteness is not expected, since the structure of the simplified covariance matrix $\mathbf{Q}_{\text{est}}$ differs from that of $\mathbf{Q}_{\text{true}}$.

5) DEPENDENCE ON N AND $\hat{\mathbf{Q}}(0)$

The results above correspond to a sample length $N = 5$, but we tried several other values. Figure 9 shows that the performance of the UKF runs improved with decreasing N. Indeed, during the adaptive filter initialization, $\mathbf{q}_s(i)$ in (28) is a poor indicator of the noise $\mathbf{b}^t(i)$. If the initial guess for $\mathbf{Q}$ underestimates the noise, then the gain $\mathbf{K}$ will be small, and the analysis $\mathbf{w}^a$ will remain close to the forecast $\mathbf{w}^f$, so the estimated system noise

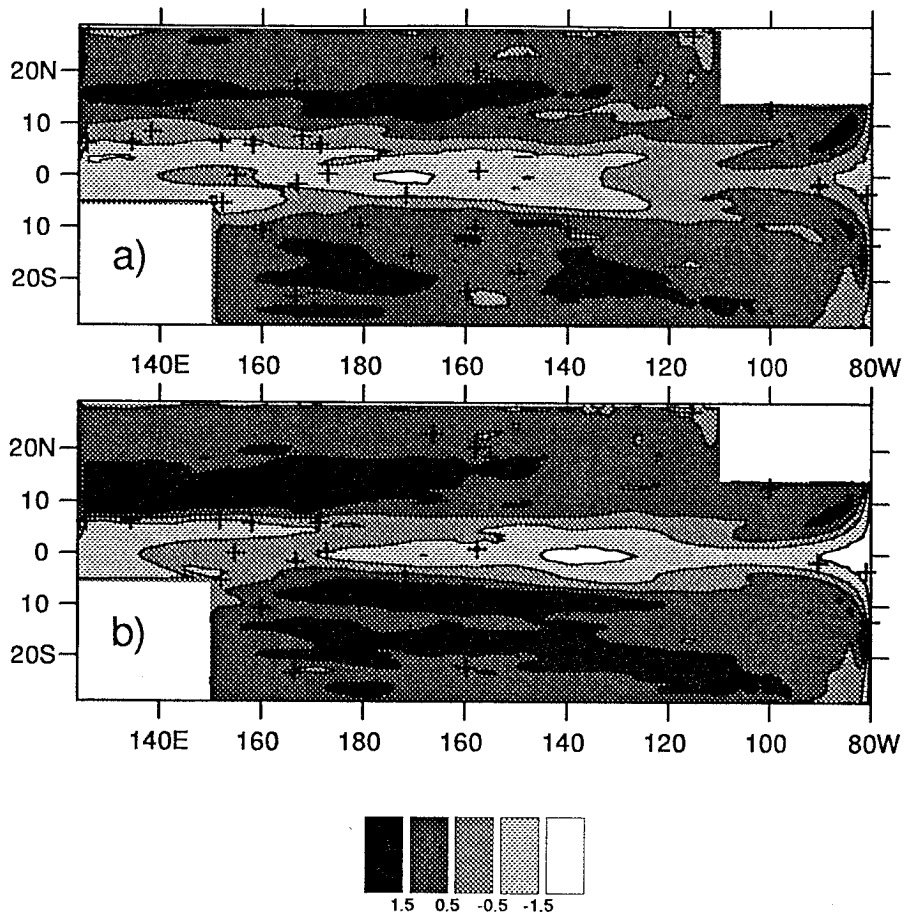

FIG. 6. Differences of the estimated system noise projected onto sea level with its true value for (a) MT scheme (Fig. 5b minus Fig. 5a), and (b) ML scheme (Fig. 5c minus Fig. 5a). Contour interval is 1 cm.

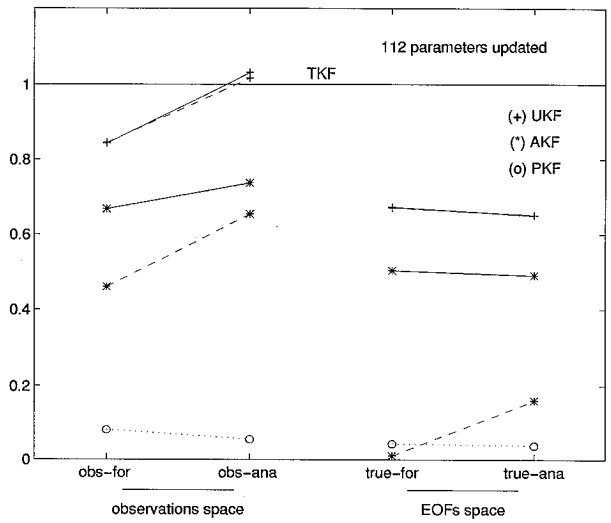

FIG. 7. Performance index of the adaptive algorithms of MT (dashed line) and Maybeck (continuous line) when estimating 112 parameters. Values with the a priori covariance matrix are shown by the dotted line. Values of the true covariance matrix are equal to one.

$\mathbf{q}_s$, which depends on $\mathbf{w}^a - \mathbf{w}^f$, will be too small. However, updating $\mathbf{Q}$ with the iterative procedure will improve $\mathbf{q}_s$ as the run proceeds. For small N, the estimator $\hat{\mathbf{Q}}$ is allowed to vary rapidly, getting closer to the true covariance matrix and leading to better UKF performance.

On the other hand, the AKF runs improve with increasing N. This occurs because the sample mean $\hat{\mathbf{q}}$ used in AKF is a poorer estimate of the mean of the true bias for small N. Indeed, its average over the last 50 steps has an rms value (difference with the true one) decreasing from 3.08×10^{-4} for $N = 2$ to 2.28×10^{-4} for $N = 5$ and 3.57×10^{-5} for $N = 20$. Since (31) makes use of the estimator $\hat{\mathbf{q}}$ at each time step, the performance of the adaptive KF is lower when N decreases, even though $\hat{\mathbf{Q}}$ is better. This suggests that the UKF run is best for small N because we have used our knowledge that the true bias is zero. If the bias was unknown, different conclusions might been reached, unless the bias was estimated by some other mean.

To test the robustness of the adaptive procedure, we made an adaptive run with an overestimated system error covariance matrix $\hat{\mathbf{Q}}$ as initial guess and $N = 5$. The

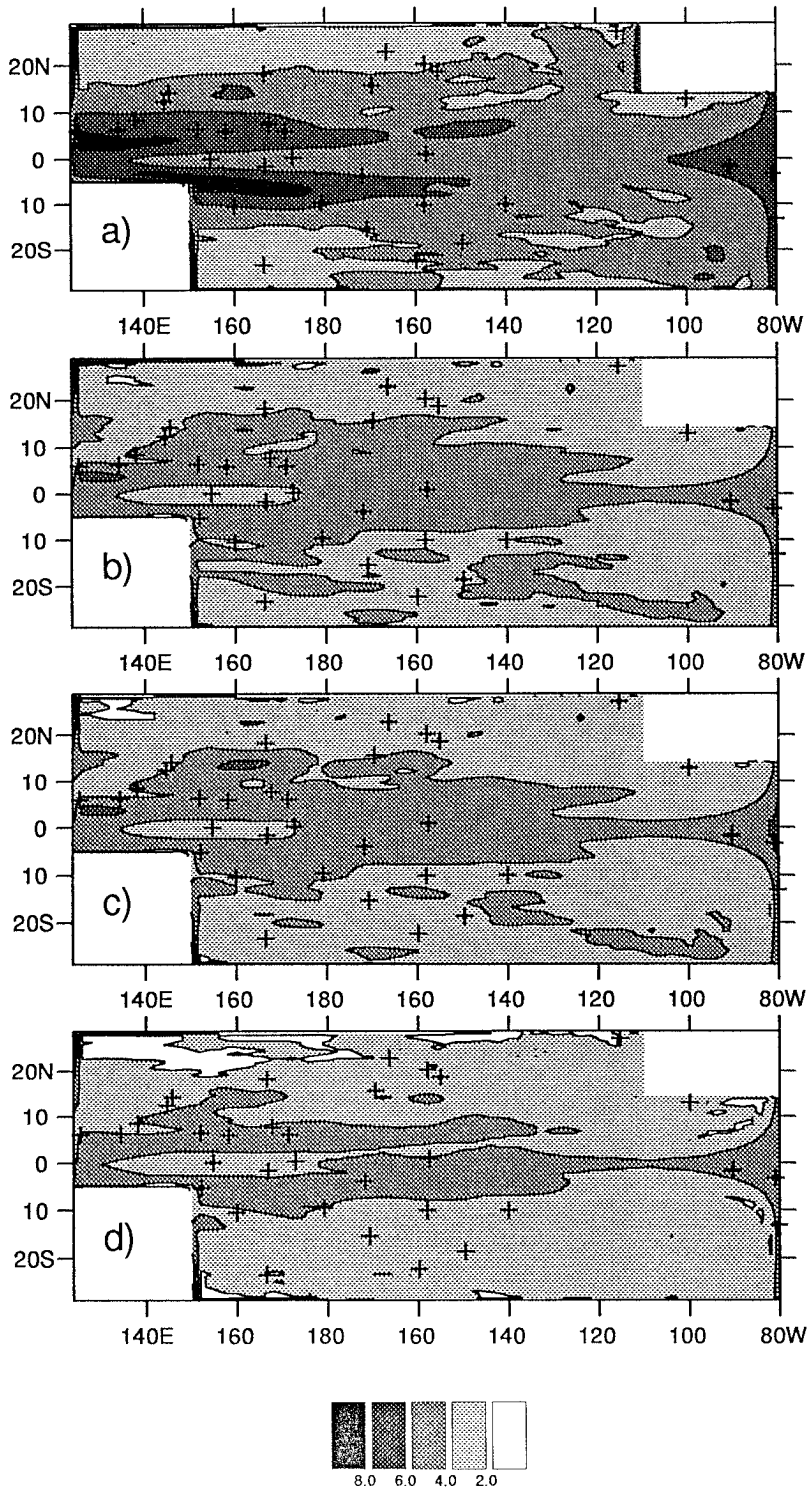

FIG. 8. Root-mean-square difference between true and predicted sea level using as system noise covariance matrix (a) the a priori matrix (b) that estimated by the MT adaptive procedure (c) that estimated by the ML function, and (d) the true covariance matrix. Contour interval is 2 cm.

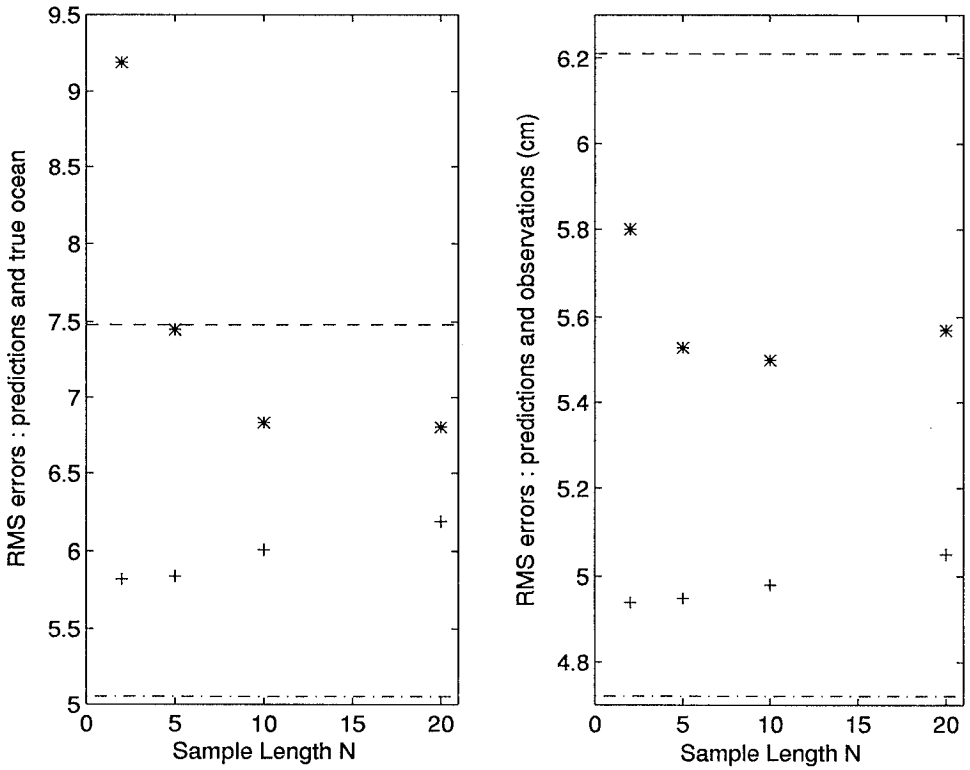

FIG. 9. Root-mean-square difference between the predictions and the true state of the ocean (left) and between the predicted sea level at the observing stations and the observations (right) versus sample length. The results are given for the MT adaptive scheme AKF (∗) and for UKF (+). For reference, corresponding values are given for UR (dashed line) and TKF (dash–dotted line).

norm and trace of the estimator decreased rapidly and, after about 60 time steps, fluctuated as before around their true value (not shown). After about 150 time steps, the variability of $\hat{\mathbf{Q}}$ was indistinguishable from the case of an underestimated initial choice. The performance of the new UKF run was very similar (rms error between observed and predicted sea level at the observing stations 4.96 cm, correlation 0.77, rms error between predicted and true state 5.90×10^{-4}, and correlation 0.76). Hence, $\mathbf{Q}_{\text{est}}$ does not seem to depend on the initial choice if enough data are available for the estimators to adjust.

c. *Maybeck's algorithm*

We first performed an extended adaptive run with the sample length $N = 10$, estimating all the terms of the covariance matrix as in section 5b(1). Figure 10 (top) shows that the adaptive algorithm does not diverge but the estimators are biased and the covariance matrix clearly overestimated. Maybeck (1982) pointed out that the maximum-likelihood estimator is biased for small samples, anticipating this result.

In view of the similarity between MT's and Maybeck's estimators, it is remarkable that the MT algorithm leads to the divergence of the filter while that of Maybeck is stable in the same ill-conditioned setting. The instability of MT arises from the simultaneous identification of the mean, in agreement with Godbole (1974), who showed that the simultaneous estimation of the mean and the covariance matrix of the system errors may not be well behaved. For stationary noise, he demonstrated that the noise covariances can be identified without the knowledge of noise means, but not the contrary. Indeed, using the sample mean of the innovation, an unbiased estimator of the innovation covariance and, through an adaptive scheme, of the system noise covariance can be calculated. On the other hand, estimating the mean $\mathbf{q}$ requires computing the gain matrix $\mathbf{K}$, which depends on $\mathbf{Q}$ by (12) and (13), so a good estimation of $\mathbf{Q}$ is necessary.

In this first run with Maybeck's algorithm, the diagonal terms of $\mathbf{Q}_{\text{est}}$ are only slightly overestimated (10%), but the covariances are not well estimated, presumably because of their too large number. As in section 5b(2), a simplified matrix with 112 parameters, gave a much better behavior, although the trace was still a bit overestimated, unlike MT (Fig. 10, bottom). Figure 2 (lower left) shows that, although the error variance of the first EOF is much too big, the global structure of $\mathbf{Q}_{\text{est}}$ is preserved.

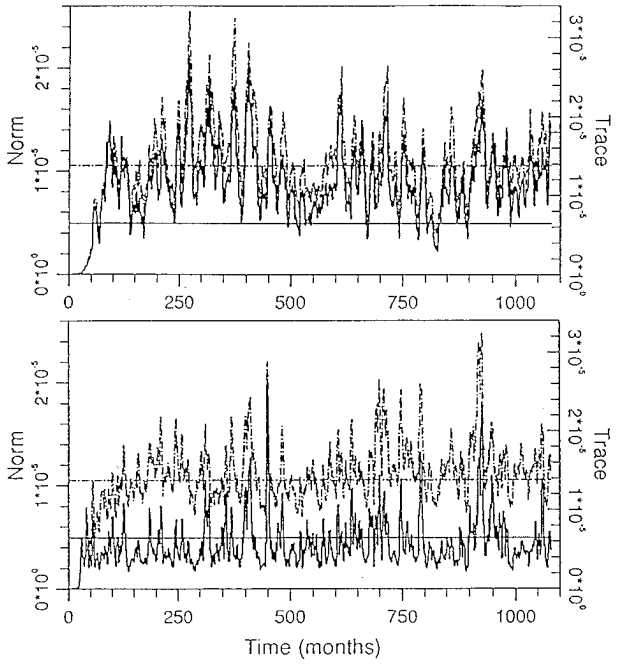

FIG. 10. Time evolution over 90 years of the Frœbenius norm (continuous line) and trace (dash–dotted line) of $\hat{\mathbf{Q}}$, using Maybeck's adaptive scheme and estimating all covariances (top) or the covariances between the first five EOFs and the diagonal terms (bottom). The horizontal lines represent the values of the norm of the true covariance matrix (continuous) and its trace (dash–dotted).

The AKF run is superior to that of the MT algorithm because the forecast is not perturbed by any bias estimation as in MT. No bias is estimated in either algorithm in the UKF runs, and they are nearly identical (Table 1 and Fig. 7). The only difference is that the rms difference between analyzed sea level and observations is smaller in Maybeck's algorithm, consistent with the slight overestimation of $\mathbf{Q}_{\text{est}}$.

d. Maximum-likelihood algorithm

The adaptive run based on the maximum-likelihood estimator derived from Dee's algorithm (hereafter ML) is given by minimizing the functional (44). Since $\mathbf{Q}(\boldsymbol{\alpha})$, which enters (44) via (41), should be a positive semidefinite matrix, we represent it with a Cholesky factorization, $\mathbf{S}$ being the lower triangular Cholesky factor of $\mathbf{Q}$,

$$\mathbf{Q}(\boldsymbol{\alpha}) = \mathbf{S}(\boldsymbol{\alpha})\mathbf{S}^{\mathrm{T}}(\boldsymbol{\alpha}), \tag{50}$$

with

$$\mathbf{S}(\boldsymbol{\alpha}) = \sum_{i=1}^{M} \alpha_i \mathbf{S}_i. \tag{51}$$

To precisely compare the ML algorithm with those of MT and Maybeck, the elements of the covariance matrix that need to be estimated are the covariances of the first five EOFs and the diagonal [see section 3b(2)]. Because of the computational burden, it was difficult to estimate so many parameters, so the diagonal terms from the sixth EOF on were assumed to decay exponentially with an *e*-folding scale of 0.03, leaving only one adjustable amplitude. Hence, the dimension of the unknown parameter in (51) is 16. The first 15 matrices $\mathbf{S}_i$ are zero everywhere except for one element $\{S_i(j, k) = 1; j \le 5, k \le j\}$, while $\mathbf{S}_{16}$ is zero except for the diagonal elements for $j > 5$, given by

$$S_{16}(j, j) = e^{-0.015(j-6)}. \tag{52}$$

The first guess $\mathbf{Q}_{\text{a priori}}$ was chosen to be diagonal, and exactly equal to that used in sections 5a and 5b for the first six elements, with the rest determined by (52). The variance of the a priori matrix was then 0.7‰ of that of the true one. Since the system noise is more underestimated than in Table 1, the PKF run is slightly worse than before (Table 2).

Table 2 compares ML with the adaptive procedures of MT and of Maybeck in similar settings, that is, with the same parameterization (52) of the last 97 terms of the diagonal of $\mathbf{Q}$. The ML estimator gives results very similar to those of the UKF runs of the two other adaptive procedures. Figure 3b illustrates that the innovation lag-correlation closely resembles that of the UKF run with the algorithm of MT. Except at lag 1, the estimated correlation falls within the 95% confidence interval.

The structure of the ML system error covariance matrix is somewhat different from that estimated with the MT and Maybeck algorithms: the first five diagonal

TABLE 2. Performance of the adpative algorithms of MT, Maybeck, and ML when estimating 16 parameters. As in Table 1.

		Observations				True state			
		Correlation		rms (cm)		Correlation		rms ($\times 10^{-4}$)	
		Forecast	Analyses	Forecast	Analysis	Forecast	Analyses	Forecast	Analysis
Unfiltered		0.68		6.21		0.66		7.48	
PKF		0.69	0.69	6.15	6.10	0.66	0.66	7.42	7.41
Myers and Tapley	AKF	0.73	0.86	5.57	3.73	0.71	0.74	6.92	6.36
	UKF	0.77	0.96	4.97	2.09	0.76	0.80	5.81	5.18
Maybeck	AKF	0.75	0.89	5.24	3.32	0.74	0.76	6.21	5.74
	UKF	0.77	0.96	4.97	2.11	0.77	0.80	5.81	5.18
ML		0.77	0.95	4.96	2.28	0.77	0.80	5.81	5.18
TKF		0.79	0.94	4.72	2.37	0.81	0.86	5.05	4.05

terms are slightly overestimated and the covariance elements are lower (Fig. 2, lower right), and the system noise for sea level is slightly overestimated when mapped in gridpoint space (Fig. 5c). Interestingly, the differences between the errors estimated with ML and the true ones have the same spatial structures as those obtained with MT but with bigger amplitudes, which emphasizes the contrast between the regions near and away from the equator (Fig. 6). In both cases, the estimated errors are too low along the equator, giving too much confidence to the model predictions in this region. Despite the differences between the estimated matrices, the geographical distribution of the rms differences between forecast and true sea level anomalies (Fig. 8d) resembles that given by the MT adaptive procedure.

The fact that the results of the three adaptive procedures are very close although the estimated matrices are different confirms that the system error covariance matrix that whitens the innovation sequence is not determined uniquely when the number of observations p is much smaller than the dimension of the state vector n (section 4c).

In summary, as expected from the similarities between the three estimators, the results of the adaptive procedure are very much alike. However, the ML procedure requires much more computation to minimize the nonlinear functional.

6. Conclusions

The goal of the present work was to develop and test a reduced space adaptive KF for a linear model of the tropical Pacific Ocean. Using a twin experiment approach, we tested three different adaptive algorithms. The first two, the empirical estimator of MT (1976) and the maximum-likelihood estimator of Maybeck (1982), were shown to be equivalent if the system noise has zero mean. The two algorithms had to be modified in view of the still large dimension of the state vector; estimating all the elements of the system error covariance matrix led to ill conditioning, since the amount of observations was small. Taking advantage of the EOFs coordinates used to define the reduced space, we were able to efficiently reduce the number of matrix parameters and to adaptively obtain quite realistic values.

Both algorithms show similar performance in forecasting the state of the ocean when using an averaged system noise covariance matrix $\mathbf{Q}_{\text{est}}$ based on the last part of the adaptive runs. As MT's adaptive procedure is very efficient at whitening the innovation sequence, the results of the runs with $\mathbf{Q}_{\text{est}}$ were really close to optimal, meaning that all information in the observations had been extracted. However, the forecasts of the adaptive run itself were far less satisfactory with MT's algorithm because it makes use of on-line estimates of the system noise mean, and the latter cannot be identified without the knowledge of the noise covariance matrix (Godbole 1974).

The third algorithm is a maximum-likelihood estimator inspired by Dee (1995), which has been designed to estimate a few parameters of the system error covariance matrix using the entire sequence of the innovation vector. Its performance is comparable even though the estimated system errors covariance matrix differs notably from those estimated with the algorithms of MT and Maybeck, which are alike and more realistic. The minimization of the nonlinear functional is computationally very costly, however, and the algorithm is not recommended.

In summary, the adaptive algorithm of MT with a simplified covariance matrix is efficient (nearly optimal) in the unbiased case. Moreover, it is easily applicable to a biased system noise. However, the noise mean should be estimated by some other procedure since the simultaneous estimation of the covariance matrix and the mean is not well behaved. Indeed, this was confirmed by the poor performance of the adaptive algorithm (AKF) compared to that of the KF run with the estimated matrix. In forthcoming papers, the algorithm will be applied to a real dataset and the problem of a biased system noise addressed.

Acknowledgments. We would like to thank Alexey Kaplan for his help and fruitful discussions and the reviewers for suggestions that led to an improvement of the presentation. Here, IB was supported by CMO/SHOM. This research was supported in France by the PNEDC. It was done in part at the Lamont–Doherty Earth Observatory and supported by ONR Contract 5-21736 and NOAA G-POP Grant NA56GP0221 with Columbia University, and at the Massachusetts Institute of Technology, whose hospitality to one of us (CF) is gratefully acknowledged.

APPENDIX

Comparison of Maybeck's Estimator and the ML Estimator

The maximum-likelihood function from which Maybeck's estimator (33) was derived is given by $p[\mathbf{w}(k), \mathbf{W}_N^o(k) \mid \boldsymbol{\alpha}]$, where $\mathbf{W}_N^o(k)$ denotes the observation history over the last N steps, and $\mathbf{w}(i)$ the state vector. It simultaneously gives an estimator of the state vector $\mathbf{w}_{\text{ml}}(k)$ and the noise parameters $\boldsymbol{\alpha}_{\text{ml}}$. Successive application of the Bayes's rule gives

$$p[\mathbf{w}(k), \mathbf{W}_N^o(k) \mid \boldsymbol{\alpha}] = p[\mathbf{w}(k) \mid \mathbf{W}^o(k), \boldsymbol{\alpha}] \left\{ \prod_{i=k-N+1}^{k} p[\mathbf{w}^o(i) \mid \mathbf{W}^o(i-1), \boldsymbol{\alpha}] \right\}, \quad \text{(A1)}$$

where $\mathbf{W}^o(i)$ is the measurement history from step 1 to step i, with

$$p[\mathbf{w}(k) \mid \mathbf{W}^o(k), \boldsymbol{\alpha}] = (2\pi)^{-n/2}[\det\mathbf{P}^a(k)]^{-1/2} \times \exp\left\{-\frac{1}{2}[\mathbf{w}(k) - \mathbf{w}^a(k)]^{\mathrm{T}}\mathbf{P}^a(i)[\mathbf{w}(k) - \mathbf{w}^a(k)]\right\} \quad \text{(A2)}$$

and

$$p[\mathbf{w}^o(i) \mid \mathbf{W}^o(i-1), \boldsymbol{\alpha}] = (2\pi)^{-p/2}[\det\mathbf{C}_0(i)]^{-1/2} \times \exp\left\{-\frac{1}{2}[\mathbf{w}^o(i) - \mathbf{H}\mathbf{w}^f(i)]^{\mathrm{T}}\mathbf{C}_0^{-1}(i) \times [\mathbf{w}^o(i) - \mathbf{H}\mathbf{w}^f(i)]\right\}. \quad \text{(A3)}$$

Using (8), it is easy to show that (43) and (A3) are identical. The functional f_{May} to be minimized can be defined as

$$f_{\mathrm{May}}(\mathbf{w}, \boldsymbol{\alpha}) = \ln[\det\mathbf{P}^a(k)] + [\mathbf{w}(k) - \mathbf{w}^a(k)]^{\mathrm{T}}\mathbf{P}^a(k)^{-1} \times [\mathbf{w}(k) - \mathbf{w}^a(k)] + \sum_{i=k-N+1}^{k}\{\ln[\det\mathbf{C}_0(i)] + \mathbf{v}(i)^{\mathrm{T}}\mathbf{C}_0(i)^{-1}\mathbf{v}(i)\}. \quad \text{(A4)}$$

The maximum-likelihood estimators $\mathbf{w}_{\mathrm{ml}}(k)$ and $\boldsymbol{\alpha}_{\mathrm{ml}}$ should satisfy the two equations

$$\frac{\partial f_{\mathrm{May}}}{\partial \mathbf{w}}(\mathbf{w}_{\mathrm{ml}}, \boldsymbol{\alpha}_{\mathrm{ml}}) = [\mathbf{w}_{\mathrm{ml}}(k) - \mathbf{w}^a(k)]^{\mathrm{T}}\mathbf{P}^a(k)^{-1} = 0 \quad \text{(A5)}$$

and

$$\frac{\partial f_{\mathrm{May}}}{\partial \alpha_j}(\mathbf{w}_{\mathrm{ml}}, \boldsymbol{\alpha}_{\mathrm{ml}}) = \mathrm{Tr}\left\langle\{\mathbf{P}^a(k)^{-1} - \mathbf{P}^a(k)^{-1}[\mathbf{w}_{\mathrm{ml}}(k) - \mathbf{w}^a(k)] \times [\mathbf{w}_{\mathrm{ml}}(k) - \mathbf{w}^a(k)]^{\mathrm{T}}\mathbf{P}^a(k)^{-1}\}\frac{\partial\mathbf{P}^a(k)}{\partial\alpha_j}\right\rangle - 2\frac{\partial\mathbf{w}^a(k)^{\mathrm{T}}}{\partial\alpha_j}\mathbf{P}^a(k)^{-1}[\mathbf{w}_{\mathrm{ml}}(k) - \mathbf{w}^a(k)] - 2\sum_{i=k-N+1}^{k}\frac{\partial\mathbf{w}^f(i)}{\partial\alpha_j}\mathbf{H}(i)^{\mathrm{T}}\mathbf{C}_0^{-1}(i)\mathbf{v}(i) + \sum_{i=k-N+1}^{k}\mathrm{Tr}\left\{[\mathbf{C}_0^{-1}(i) - \mathbf{C}_0^{-1}(i)\mathbf{v}(i) \times \mathbf{v}(i)^{\mathrm{T}}\mathbf{C}_0^{-1}(i)]\frac{\partial\mathbf{C}_0(i)}{\partial\alpha_j}\right\} = 0. \quad \text{(A6)}$$

The solution of (A5) is given by

$$\mathbf{w}_{\mathrm{ml}}(k) = \mathbf{w}^a(k), \quad \text{(A7)}$$

so that the maximum-likelihood estimator of the model state is given by the KF equations. Equation (A6) can thus be simplified to

$$\frac{\partial f_{\mathrm{May}}}{\partial \alpha_j}(\mathbf{w}_{\mathrm{ml}}, \boldsymbol{\alpha}_{\mathrm{ml}}) = \mathrm{Tr}\left[\mathbf{P}^a(k)^{-1}\frac{\partial\mathbf{P}^a(k)}{\partial\alpha_j}\right] - 2\sum_{i=k-N+1}^{k}\frac{\partial\mathbf{w}^f(i)}{\partial\alpha_j}\mathbf{H}(i)^{\mathrm{T}}\mathbf{C}_0^{-1}(i)\mathbf{v}(i) + \sum_{i=k-N+1}^{k}\mathrm{Tr}\left\{[\mathbf{C}_0^{-1}(i) - \mathbf{C}_0^{-1}(i)\mathbf{v}(i)\mathbf{v}(i)^{\mathrm{T}}\mathbf{C}_0^{-1}(i)] \times \frac{\partial\mathbf{C}_0(i)}{\partial\alpha_j}\right\} = 0, \quad \text{(A8)}$$

and its solution $\boldsymbol{\alpha}_{\mathrm{ml}}$ minimizes the equivalent functional

$$f_{\mathrm{May}}(\boldsymbol{\alpha}) = \ln[\det\mathbf{P}^a(k)] + \sum_{i=k-N+1}^{k}[\ln[\det\mathbf{C}_0(i)] + \mathbf{v}(i)^{\mathrm{T}}\mathbf{C}_0(i)^{-1}\mathbf{v}(i)]. \quad \text{(A9)}$$

For k and N = kstep, the maximum-likelihood functions (A9) and (44) only differ by the term $\ln[\det\mathbf{P}^a(k)]$ in f_{May}. In fact, this term was neglected by Maybeck (1982) to reach the solution (33), so the two functionals are otherwise equivalent. However, Maybeck's solution introduces further simplifications and is calculated over N samples whereas we minimize directly the functional (44) over the whole time history.

REFERENCES

Bélanger, P. R., 1974: Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatika*, **10,** 267–275.

Busalacchi, A. J., and M. A. Cane, 1985: Hindcasts of sea level variations during the 1982–83 El Niño. *J. Phys. Oceanogr.*, **15,** 213–221.

Cane, M. A., 1984: Modeling sea level during El Niño. *J. Phys. Oceanogr.*, **14,** 1864–1874.

——, and R. J. Patton, 1984: A numerical model for low-frequency equatorial dynamics. *J. Phys. Oceanogr.*, **14,** 1853–1863.

——, S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321,** 827–832.

——, A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, 1997: Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman Filter. *J. Geophys. Res.*, in press.

Daley, R., 1992: The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.*, **120,** 178–196.

Dee, D. P., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.*, **117,** 365–384.

——, 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.
——, S. E. Cohn, A. Dalcher, and M. Ghil, 1985: An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Trans. Autom. Control,* **30,** 1057–1065.
Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model. *J. Geophys. Res.,* **100,** 6777–6793.
——, J. Benveniste, C. Wunsch, and D. B. Haidvogel, 1993: Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother. *J. Phys. Oceanogr.,* **23,** 1803–1820.
Gelb, A., 1974: *Applied Optimal Estimation.* The MIT Press, 374 pp.
Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics,* Vol. 33, Academic Press, 141–266.
Godbole, S. S., 1974: Kalman filtering with no a priori information about noise—White noise case: Identification of covariances. *IEEE Trans. Autom. Control,* **19,** 561–563.
Goldenberg, S., and J. J. O'Brien, 1981: Time and space variability of tropical wind stress. *Mon. Wea. Rev.,* **109,** 1190–1205.
Gourdeau, L., S. Arnault, Y. Meynard, and J. Merle, 1992: Geosat sea-level assimilation in a tropical Atlantic model using Kalman filter. *Oceanologica Acta,* **15,** 567–574.
Groutage, F. D., R. G. Jacquot, and R. L. Kirlin, 1987: Techniques for adaptive state estimation through the utilization of robust smoothing. *Control Dyn. Syst.,* **23,** 273–308.
Hautus, M. L. J., 1969: Controllability and observability conditions of linear autonomous systems. *Proc. K. Ned. Akad. Wet. Ser. A: Math. Sci.,* 443–448.
Hoang, H. S., P. De Mey, O. Talagrand, and R. Baraille, 1995: Assimilation of altimeter data in a quasi-geostrophic ocean model by simple nonlinear adaptive filter. *Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, World Meteor. Org., 521–526.
Jazwinsky, A. H., 1969: Adaptive filtering. *Automatika,* **5,** 475–485.
Kailath, T., 1968: An innovations control approach to least square estimation. Part I: Linear filtering in additive white noise. *IEEE Trans. Autom. Control,* **13,** 646–655.
Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.,* **82D,** 35–45.
LeMoyne, H., and J. C. Alvarez, 1991: Analysis of dynamic data assimilation for atmospheric phenomena. Effect of the model order. *Atmosfera,* **4,** 145–164.
Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.,* **111,** 46–59.
Maybeck, P. S., 1982: *Stochastic Models, Estimation, and Control.* Vol. 2. Academic Press, 423 pp.
Mehra, R. K., 1970: On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Autom. Control,* **15,** 175–184.
Miller, R. N., and M. A. Cane, 1989: A Kalman filter analysis of sea level height in the tropical Pacific. *J. Phys. Oceanogr.,* **19,** 773–790.
——, A. J. Busalacchi, and E. C. Hackert, 1995: Sea surface topography fields of the tropical Pacific from data assimilation. *J. Geophys. Res.,* **100,** 13 389–13 425.
Myers, K. A., and B. D. Tapley, 1976; Adaptive sequential estimation with unknown noise statistics. *IEEE Trans. Autom. Control,* **21,** 520–523.
Parish, D. F., and S. E. Cohn, 1985: A Kalman filter for a two-dimensional shallow-water model: Formulation and preliminary experiments. Office Note 304, 64 pp. [Available from National Centers for Environmental Prediction, Washington, DC 20233.]
Stuart, A., and J. K. Ord, 1991: *Kendall's Advanced Theory of Statistics.* Edward Arnold, 1323 pp.
Wiener, N., 1949: *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications.* Wiley, 163 pp.
Wyrtki, K., K. Constantine, J. J. Kilonsky, G. Mitchum, B. Miyamoto, T. Murphy, S. Nakahara, and P. Caldwell, 1988: The Pacific Island Sea Level Network. JIMAR Contribution 88-0137, Data Rep. 002, Joint Institute for Marine and Atmospheric Research, University of Hawaii at Manoa, Honolulu, Hawaii, 71 pp.