Estimation of Tropical Sea Level Anomaly by an Improved Kalman Filter

NGAI HANG CHAN

Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania and Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

JOSEPH B. KADANE

Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania

ROBERT N. MILLER

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

WILFREDO PALMA

Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania

(Manuscript received 27 February 1995, in final form 22 December 1995)

ABSTRACT

Kalman filter theory and autoregressive time series are used to map sea level height anomalies in the tropical Pacific. Our Kalman filters are implemented with a linear state space model consisting of evolution equations for the amplitudes of baroclinic Kelvin and Rossby waves and data from the Pacific tide gauge network. In this study, three versions of the Kalman filter are evaluated through examination of the innovation sequences, that is, the time series of differences between the observations and the model predictions before updating. In a properly tuned Kalman filter, one expects the innovation sequence to be white (uncorrelated, with zero mean). A white innovation sequence can thus be taken as an indication that there is no further information to be extracted from the sequence of observations. This is the basis for the frequent use of whiteness, that is, lack of autocorrelation, in the innovation sequence as a performance diagnostic for the Kalman filter.

Our long-wave model embodies the conceptual basis of current understanding of the large-scale behavior of the tropical ocean. When the Kalman filter was used to assimilate sea level anomaly data, we found the resulting innovation sequence to be temporally correlated, that is, nonwhite and well fitted by an autoregressive process with a lag of one month. A simple modification of the way in which sea level height anomaly is represented in terms of the state vector for comparison to observation results in a slight reduction in the temporal correlation of the innovation sequences and closer fits of the model to the observations, but significant autoregressive structure remains in the innovation sequence. This autoregressive structure represents either a deficiency in the model or some source of inconsistency in the data.

When an explicit first-order autoregressive model of the innovation sequence is incorporated into the filter, the new innovation sequence is white. In an experiment with the modified filter in which some data were held back from the assimilation process, the sequences of residuals at the withheld stations were also white. To our knowledge, this has not been achieved before in an ocean data assimilation scheme with real data. Implications of our results for improved estimates of model error statistics and evaluation of adequacy of models are discussed in detail.

1. Introduction

Simple linear models capture much of the dynamics of the long temporal and large spatial scale behavior of the dynamic topography of the tropical ocean, but data are too sparse to provide models with complete initial and boundary conditions. Analyses based on spatial and temporal interpolation of the data alone without reference to dynamics are also limited by the sparsity of data. Data assimilation is therefore a particularly appropriate tool for this purpose.

Miller and Cane (1989, hereafter MC) used the Kalman filter with a simple dynamical model driven by a wind stress field derived from ship observations to assimilate monthly sea surface height anomaly data from island tide gauge stations. That data assimilation system has since been used by Miller (1990) for an observing system simulation experiment with data from a simulated array of moored instruments and by Fu et al.

Corresponding author address: Prof. Robert N. Miller, College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331.

(1993) to assimilate altimetric data from Geosat. Such a model was also applied to the tropical Indian Ocean by Perigaud and Fu (1990). Results from a similar model were presented by Kawabe (1994).

All data assimilation systems require hypotheses about the residuals of the model proposed. These hypotheses can be tested to decide the adequacy of the model, but there are few examples in the literature in which such tests are performed in detail, and most of these examples are taken from numerical weather prediction, where data are much more plentiful.

It is commonly assumed that the model and observational errors are well described by stationary white noise (i.e., temporally uncorrelated, with zero mean and constant variance). There is, however, no particular reason to believe that this is the case, and there is mounting evidence that it is not. Examination of the residuals should contribute to fitting a better model to the physical phenomenon. In this paper, we examine the observed errors of the MC model and present two modified Kalman filter models. The result is a formulation of the Kalman filter that is optimal according to the criterion of zero autocorrelation of the innovation sequence (e.g., see Daley 1992b; Dee et al. 1985).

An intuitive justification for using the whiteness of the residuals as a yardstick for optimality is that white noise contains no systematic information; if the residuals are white, then most of the signals in the data have been extracted. We shall see that this criterion alone does not guarantee that the filter has all the properties one might desire, such as reliable basinwide estimates of the error statistics. Further improvement of our assimilation scheme by tuning parameters remains possible, but it is likely that our modified Kalman filters are close to the best one can do with these data and this simple assimilation framework. Better results would require more data and/or a much more intricate filtering scheme, in particular, one which would accommodate a more detailed noise model.

This paper is structured as follows: The data are presented in section 2. The model proposed by MC is reviewed and its performance is analyzed in section 3. Section 4 contains descriptions of the suggested approach and discussion of the results of two data assimilation experiments. Final remarks are presented in section 5.

2. The data

The data are exactly those used in MC. We describe them briefly here. Two types of datasets are involved in this study: wind forcing data and sea level height observations. The wind data consist of monthly pseudostress maps provided by The Florida State University. These maps were derived by subjective analysis from merchant ship observations (see Stricherz et al. 1992) and processed according to Zebiak (1989). The resulting maps provide complete data coverage of the tropical Pacific, albeit with unknown error characteristics. Only the zonal component of the wind stress is considered, as in MC. Dimensional arguments support this assumption, which results in considerable simplification of the error model. The most obvious consequence of neglecting the meridional windstress is loss of accuracy near the boundaries.

The sea level data consist of observations of the sea level heights for a group of tide gauge stations in the tropical Pacific. Station data from Rabaul, Nauru, Jarvis, Christmas, Santa Cruz, Callao, Kapingamarangi, Tarawa, Canton, and Fanning were selected in the original MC work because they are near the equator where the model is expected to do its best, and they provide long overlapping time series. For Rabaul, Nauru, Christmas, and Tarawa the observations were taken from 1974 to 1983; for Jarvis from 1977 to 1984; for Santa Cruz from 1978 to 1983; for Callao from 1942 to 1984; for Kapingamarangi from 1978 to 1983; and for Fanning from 1972 to 1983. Thus, the overlapping period is 1978–83. The raw data for the MC model consist of monthly means with tides removed (cf. Wyrtki et al. 1988).

3. Modeling the data

a. The physical model

The physical model used in MC is based on the linearized primitive equations on the equatorial β plane, further simplified by decomposition into vertical modes and a long-wave approximation. The solution to these equations can be expressed in terms of equatorial Kelvin waves and nondispersive Rossby waves, whose amplitudes evolve according to simple advection equations. The model is implemented from 125°E longitude to 80°W in 5° intervals. At each grid point there are two vertical modes and, for each vertical mode, there are five meridional Rossby mode amplitudes and one Kelvin wave amplitude. Thus, there are 12 values at each of the 32 grid points in the model for a total of 384 values each month. We use typical Pacific values for the Kelvin wave speeds and the length scales of the two vertical modes: 2.91 m s^{-1} and 357 km for the first vertical mode and 1.78 m s^{-1} and 279 km for the second, cf. Cane (1984). Details are given in appendix A. The evolution of the state vector $\mathbf{W}_t$, whose components are the 384 wave amplitudes at time t, can be written schematically as

$$\begin{cases} \mathbf{W}_{t+1} = \mathbb{L}\mathbf{W}_t + \boldsymbol{\tau}_t + \boldsymbol{\epsilon}_t, \\ \mathbf{Y}_t = \mathbb{H}\mathbf{W}_t + \mathbf{v}_t, \end{cases}$$

where $\mathbf{Y}_t$ is the observation vector, that is, the monthly mean sea level height anomaly measured at a discrete set of points, and $\mathbb{H}$ is the observation matrix (10×384 in this case), which maps a vector of wave amplitudes into a vector whose components are the sea level height anomalies at the ten selected island stations. Therefore,

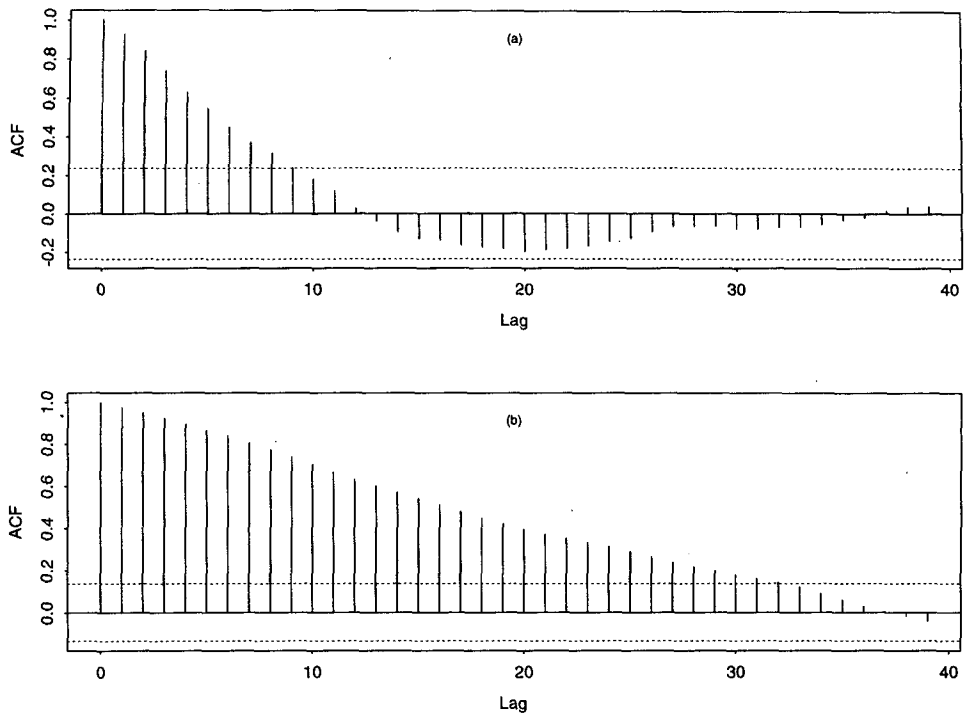

FIG. 1. ACF of the sea level at Rabaul. Abscissa is time lag, in time intervals of one month for plot (a) and 10 days for plot (b). Dashed lines denote the limit of the 95% confidence level for the ACF being different from zero. (a) Monthly observations. (b) ten-day repeated observations. The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}) / \Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

H defines the functional relation between the state vector $\mathbf{W}_t$, whose components are wave amplitudes, and $\mathbf{Y}_t$, whose components are the observed quantities, in this case sea level height anomalies. We refer to the explicit relation between the state vector and the vector of observed quantities as the *measurement model.* Here τ_t is the zonal component of the wind stress vector (384 $\times$ 1) transformed in the meridional direction to forcing of the wave modes, ϵ_t is the state noise vector (384 $\times$ 1) with covariance matrix **Q** (384 $\times$ 384), and $\mathbf{v}_t$ is the observation noise vector (10 $\times$ 1) with covariance matrix **R** (10 $\times$ 10). Both ϵ_t and $\mathbf{v}_t$ are assumed to be independent white noise sequences. The transition matrix **L** (384 $\times$ 384) represents the discretized form of the advection operator, which determines the evolution of the wave amplitudes. Boundary conditions specify that the zonal velocity must vanish everywhere on the eastern boundary, and the meridionally integrated mass flux must vanish on the western boundary. Here **Q** and **R** are taken to be known and constant in time.

The observation vector $\mathbf{Y}_t = [y_t(1), \cdots, y_t(10)]$ is a multivariate time series, $y_t(i)$ being the deviation from the monthly mean sea level at time t for the ith station, cf. MC. The index i represents the stations Rabaul, Nauru, Jarvis, Christmas, Santa Cruz, Callao, Kapingamarangi, Tarawa, Canton, and Fanning. In this study we consider values of the data $\{\mathbf{Y}_t\}$ from 1978 to 1983, one observation per month.

In the MC paper, the Kalman filter model was implemented with time steps of ten days due to considerations of numerical accuracy, but only monthly observations are available for this time period. Each monthly observation was therefore repeated three times within each month. Thus, at the end of the repetition process, there are 214 *observations* $\mathbf{Y}_t$ (t = 1, $\cdots$, 214). Originally, the temporal structure of the data over a month's time did not seem significant. High-frequency motions are filtered in the data and in the model, and there is little change in the quantities of interest over the course of a month. It is, in fact, widely believed that the physical data do not contain timescales shorter than seasonal. We shall see, however, that significant improvement of the data assimilation scheme can be achieved by introducing a measurement model that more closely resembles the measurement process. The repetition scheme not only introduces artificial data into the model but also creates a time series with nonstationary patterns, creating high-order autocorrelations artificially. Figure 1 displays the empirical autocorrelation function $\hat{\gamma}_k$ (ACF) of the sea level data at Rabaul. The ACF is defined as

$$\hat{\gamma}_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2},$$

where $\{y_t\}$ are the observations and $\bar{y}$ is the sample mean. Figure 1a shows the ACF for monthly observations and Fig. 1b presents the ACF for the ten-day repeated observations. Whereas the ACF of the monthly observations decays to zero quickly, the ACF of the ten-day repeated data converges more slowly, indicating possible nonstationary patterns.

In this paper, the repeated observation problem encountered in MC is circumvented by explicitly using the one-month data approach given in section 4a, where the monthly structure of the sea level data is incorporated into the state space system. As seen in section 4a and appendix B, implementation of such a modified algorithm is straightforward and propagation of the covariance error comes directly from the Kalman filter equations.

b. *The Kalman filter*

The Kalman filter methodology can be used to estimate the state vector and predict future observations. In this approach, the state is estimated linearly from the observation $\mathbf{Y}_t$ and the previous estimated state $\hat{\mathbf{W}}_{t|j}$, where the subscript $t|j$ denotes the estimation at time t based on observations up to time j. For the filtering process, $j = t$ and the recursive estimation equations are given by (cf. Gelb 1974):

$$\hat{\mathbf{W}}_{0|0} = \mathbf{0}$$

$$\hat{\mathbf{W}}_{t|t-1} = \mathbf{L}\hat{\mathbf{W}}_{t-1|t-1} + \boldsymbol{\tau}_t$$

$$\mathbf{P}_{t|t-1} = \mathbf{L}\mathbf{P}_{t-1|t-1}\mathbf{L}^{\mathrm{T}} + \mathbf{Q}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}_{t|t-1}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$$

$$\hat{\mathbf{W}}_{t|t} = \hat{\mathbf{W}}_{t|t-1} + \mathbf{K}_t\boldsymbol{\nu}_t$$

$$\boldsymbol{\nu}_t = \mathbf{Y}_t - \mathbf{H}\hat{\mathbf{W}}_{t|t-1},$$

where $\boldsymbol{\nu}_t$ is the observation *residual* sequence (10×1), also known as the *innovation* sequence, and $\mathbf{K}_t$ is the *Kalman gain matrix* (384×10); $\mathbf{P}_{t|t-1}$ is the covariance matrix of the estimation error of $\mathbf{W}_t$ based on observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_{t-1}$:

$$\mathbf{P}_{t|t-1} = E[(\mathbf{W}_t - \hat{\mathbf{W}}_{t|t-1})(\mathbf{W}_t - \hat{\mathbf{W}}_{t|t-1})^{\mathrm{T}}],$$

and $\mathbf{P}_{t|t}$ is the covariance matrix of the estimation error of $\mathbf{W}_t$ based on observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_t$:

$$\mathbf{P}_{t|t} = E[(\mathbf{W}_t - \hat{\mathbf{W}}_{t|t})(\mathbf{W}_t - \hat{\mathbf{W}}_{t|t})^{\mathrm{T}}].$$

Recursive numerical schemes to compute $\mathbf{K}_t$ can be found in Hannan and Deistler (1988), Gelb (1974), Anderson and Moore (1979), or Goodwin and Sin (1984). Description of the Kalman filter in the context of modeling the ocean and the atmosphere can be found in the texts by Bennett (1992) or Daley (1991), or in the review article by Ghil and Malanotte-Rizzoli (1991) and references therein.

The Kalman estimator of $\mathbf{Y}_t$ is given by $\hat{\mathbf{Y}}_t = \mathbf{H}\hat{\mathbf{W}}_{t|t-1}$. By means of a formula of this type, the sea level anomaly can be calculated from the state vector at any location in the grid and *interpolated* to any time within 1978–83.

In order to compute $\hat{\mathbf{W}}_{t|t}$ and $\hat{\mathbf{Y}}_t$, estimates of the observation noise covariance matrix $\mathbf{R} = E(\mathbf{v}_t\mathbf{v}_t^{\mathrm{T}})$ and the system noise covariance $\mathbf{Q} = E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^{\mathrm{T}})$, where $E(\cdot)$ represents expected value, are required. The observation error covariance matrix $\mathbf{R}$ is related to the instrumental error in the tide gauges at the stations and physical effects that are not represented in this state space. This latter source of error includes what some authors have termed *error of representativeness;* see Daley (1993) and references therein. Here, as in MC, we begin by assuming that the observation errors are white in time and uncorrelated from station to station.

In MC, $\mathbf{Q}$ is derived from an assumption about the error covariance of the wind data. Errors from other sources such as computational errors or neglected physics are omitted. This assumption is supported by order of magnitude calculations and by the posterior result that errors in the wind data of reasonable magnitude are sufficient to explain all of the discrepancy between the model output and the tide gauge data. This assumption is further borne out in comparisons between a similar wind-driven model and dynamic height data derived from expendable bathythermograph casts (Miller et al. 1995). The wind errors are assumed to have homogeneous anisotropic Gaussian covariance structure. The zonal and meridional decorrelation scales L_x and L_y, respectively, were determined by examination of the sample covariance matrices of the series of differences between the observed sea level anomalies and the predictions of the model without updating. Details of this parameterization are given in MC. The parameters L_x and L_y, along with an estimate of the variance of the wind stress errors, define $\mathbf{Q}$ and hence the recursion equations for the state error covariance estimation. In section 4 we present the results of an experiment in which the assumptions of whiteness of the system and observation noise sequences are relaxed.

Formally, the estimation also depends on the estimates of the initial state $\hat{\mathbf{W}}_{0|0}$ and the initial prediction covariance for the state, $\mathbf{P}_{0|0} = E[(\hat{\mathbf{W}}_{0|0} - \mathbf{W}_0)(\hat{\mathbf{W}}_{0|0} - \mathbf{W}_0)^{\mathrm{T}}]$, but these dependences are of little consequence in forced-dissipative systems such as this one. Most of the dissipation in this model results from the loss of energy at the western boundary; see, for example, Cane and Sarachik (1977). The evolution of the state vector is determined by a balance of the forcing, dissipation and wave dynamics, and the influence of the initial state decays rapidly with time. The estimated system error covariance $\mathbf{P}_{t|t}$ reaches a stationary state rapidly due to the dynamic balances in the model, cf. MC section 5b. The initial error covariance $\mathbf{P}_{0|0}$ was chosen to be equal to $20\mathbf{Q}$, a rough estimate of the error covariance of a zero guess, hence the signal covariance.

The quality of this estimate is not of great importance. Following a short initial equilibration period, changes in $\mathbf{P}_{t|t}$ result from gaps in the data, and therefore changes in $\mathbf{H}$. This system reaches its stationary state rapidly, and the estimates of the state for most of the period of this study are not very sensitive to the choice of $\mathbf{P}_{0|0}$. According to MC and our experience, the system can be considered stable after a few months.

4. Data assimilation experiments

Three sets of data assimilation experiments were performed, one with the original MC model, the second with the "Average" model, in which the averaging of model outputs over a month was modeled explicitly, and the third, the "AR(1)" model, in which an autoregressive model of the innovation sequence was added to the Average model. AR(1) is the standard terminology for a first-order autoregressive process, that is, one in which the current value can be expressed as the sum of a proportion of its value at the previous time and a random disturbance. More generally, we say that the sequence $\{\mathbf{X}_t\}$ is derived from an AR(p) process if $\mathbf{X}_t = \Sigma_{k=1}^{p} \mathbf{a}_k \mathbf{X}_{t-k} + \epsilon_t$, where $\{\epsilon_t\}$ is a stationary white sequence and the $\mathbf{a}_k$ are constant matrices. Each of the three sets consisted of two experiments, one in which data from six island stations were assimilated and four held back for verification, and one in which data from all ten available stations were assimilated.

The main instrument used for validation of the state space model is the analysis of the innovation process $\boldsymbol{\nu}_t = \mathbf{Y}_t - \hat{\mathbf{Y}}_t$. Some of the features used to select a good model include: few parameters (parsimony), a stationary, white innovation sequence (independent, mean zero, constant variance), and small error variance compared to the variance explained by the model or compared to the total variance in the data. If the model utilizes the data effectively, then the innovations will be a white noise sequence. On the other hand, any other pattern in the innovation process would indicate the possibility of improving the model.

Several standard techniques are available to check whether a sequence is white. The autocorrelation function (ACF) is useful for detecting nonstationary patterns in the data or the presence of seasonal components. The partial autocorrelation function (PACF) is useful for detecting the presence of an autoregressive (AR) structure.

For a given time series $\{Z_1, Z_2, \ldots\}$, the PACF, ρ_{kk}, may be regarded as the correlation between Z_1 and Z_k after adjusting for the intermediate observations $Z_2, \ldots, Z_k$. Specifically, consider regressing Z_{k+1} based on a linear function of $\{Z_k, Z_{k-1}, \ldots, Z_2\}$, say, $\beta_1 Z_k + \ldots + \beta_{k-1} Z_2$ with the β's chosen to minimize the mean square error of prediction. Assuming such β's have been chosen and considering time running backward, the best predictor of Z_1 with minimum mean square error based on the same $\{Z_k, Z_{k-1}, \ldots, Z_2\}$ will then be $\beta_1 Z_2 + \ldots + \beta_{k-1} Z_k$. The PACF at lag k is defined as the correlation between these two prediction errors, that is,

$$\rho_{kk} = \operatorname{corr}(\mathbf{Z}_{k+1} - \beta_1 Z_k - \cdots - \beta_{k-1} Z_2, Z_1 - \beta_1 Z_2 - \cdots - \beta_{k-1} Z_k).$$

For a lucid discussion on ρ_{kk}, see Cryer (1986). Equivalently, if Z_t is an AR(p) process, then the PACF at lag p, ρ_{pp}, will correspond to the last coefficient α_p of the autoregression

$$Z_t = \alpha_1 Z_{t-1} + \cdots + \alpha_p Z_{t-p} + \epsilon_t,$$

where ϵ_t is a white noise sequence, see Chatfield (1994). It can be shown that $\rho_{kk} = 0$ for $k > p$ for an AR(p) model. Consequently, if the data follows an AR(p) process, its sample PACF will not differ significantly from zero for lags $k > p$.

The portmanteau statistic is useful for testing the whiteness of a sequence. It is a chi-square test applied to the sample autocorrelations $\hat{\gamma}_{\nu}(i)$ at lag i of the residuals $\boldsymbol{\nu}_t$:

$$Q_L = N \sum_{i=1}^{L} \hat{\gamma}_{\nu}^2(i).$$

Asymptotically as $N \to \infty$, Q_L follows a chi-square distribution with L degrees of freedom for data from a white noise sequence, where L is an appropriately chosen integer and N is the sample size. Detailed descriptions of these techniques can be found in Brockwell and Davis (1991).

To compare the performance of the proposed models, we have included in sections 4a and 4c maps of the estimated rms residual error. These maps illustrate the propagation of the estimation error in the equatorial region including the ten stations assimilated in this study. We have also included detailed residual analyses of the Kalman filter estimates at Rabaul. Residual analyses carried out at the other stations produce similar results as for Rabaul. Some of these results appear in the tables.

In the work described in MC, the state space model was implemented with observations from six selected stations: Rabaul, Nauru, Jarvis, Christmas, Santa Cruz, and Callao. The remaining stations were utilized to assess the performance of the model in other locations. They also used the repetition scheme discussed in section 2 for the observations within a month.

In our experiments with the original MC model, calculations were performed with updating at six and ten tide gauge stations. Figures 2 and 3 display the performance of the MC model at Rabaul, with assimilation of data from six and ten stations respectively. The predictions follow the data closely in both cases, as expected for a station at which data were assimilated; see Figs. 2a and 3a. From the ACF plots, Figs. 2c and 3c, the residuals appear to be serially correlated. The Par-

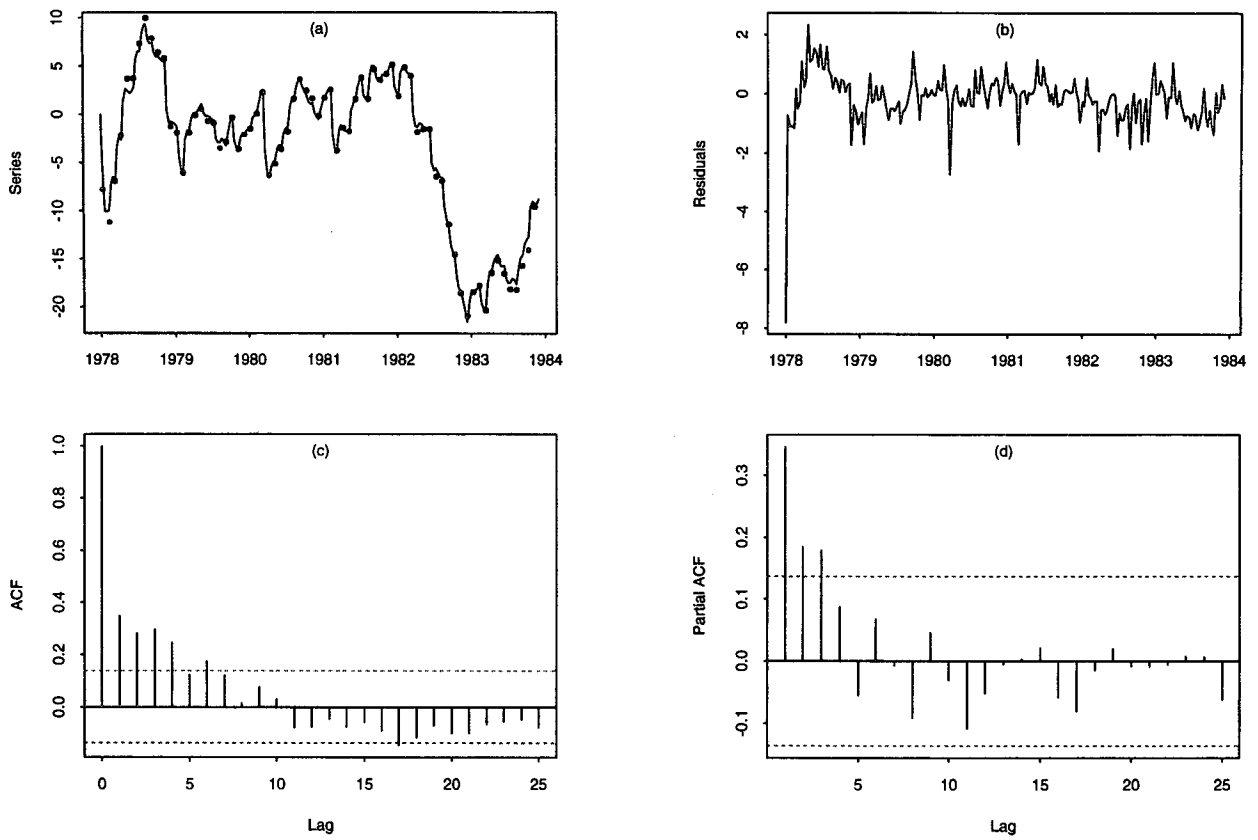

FIG. 2. Results from MC model at Rabaul (six stations): (a) Predictions. Ordinate is sea level height anomaly in cm, abscissa is time in months. Solid line denotes model output, symbols denote observations. (b) Residuals. Axes as in (a). (c) ACF of residuals. Abscissa is time lag, in time intervals of 10 days. Dashed lines denote the limit of the 95% confidence level for the ACF being different from zero. (d) Partial ACF of residuals. Legend as in panel (c). The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})/\Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

tial ACF plots suggest a third-order autoregressive process for the residuals, that is, the residuals are correlated at a time lag of one month. Similar behavior of the ACF and the PACF of the residuals is observed for the remaining nine stations, as shown in Table 2 and Table 3. The 95% confidence bands for the ACF and PACF to be different from zero are ± 0.13.

Results of all three series of experiments are summarized in Table 1. Standard deviations of the residuals $\hat{\sigma}$ are shown, along with the prior estimates $\hat{\sigma}_{\text{est}}$ of these quantities. Examination of the second and third columns shows that in the MC model with data from six stations assimilated, data residuals from Kapingamarangi and Canton, which did not participate in the assimilation process, have standard deviations that are roughly in line with the corresponding prior estimates and are comparable with the observed total standard deviations. At Tarawa, the assimilation is evidently most successful, with small residual standard deviation and a reasonable accurate prior estimate. Of the four stations which did not participate in the assimilation process, the best results are found at Tarawa, as shown in Table 2 of MC. At Fanning, the prior estimate of the residual standard deviation is pessimistic, nearly double the actual value in this realization. It is a measure of the success of this system that the prior estimates of the residual standard deviations are reasonably accurate for the most part. We must conclude that the extrapolation of the model to other locations gives predictable, if not uniformly good, results.

a. The average approach

One conceptual problem with the MC model is that the repetition of the monthly data at each ten-day time interval is not a good representation of the measurement process. It not only introduces artificial data into the model, but also creates a time series with nonstationary patterns. Here our first alternative approach is implemented, in which the observing model is more faithful to the actual observation process. Assuming a ten-day time step for the state evolution system, the monthly observations can be regarded as an average of three ten-day states. Formally, this time-averaged state space system can be written as

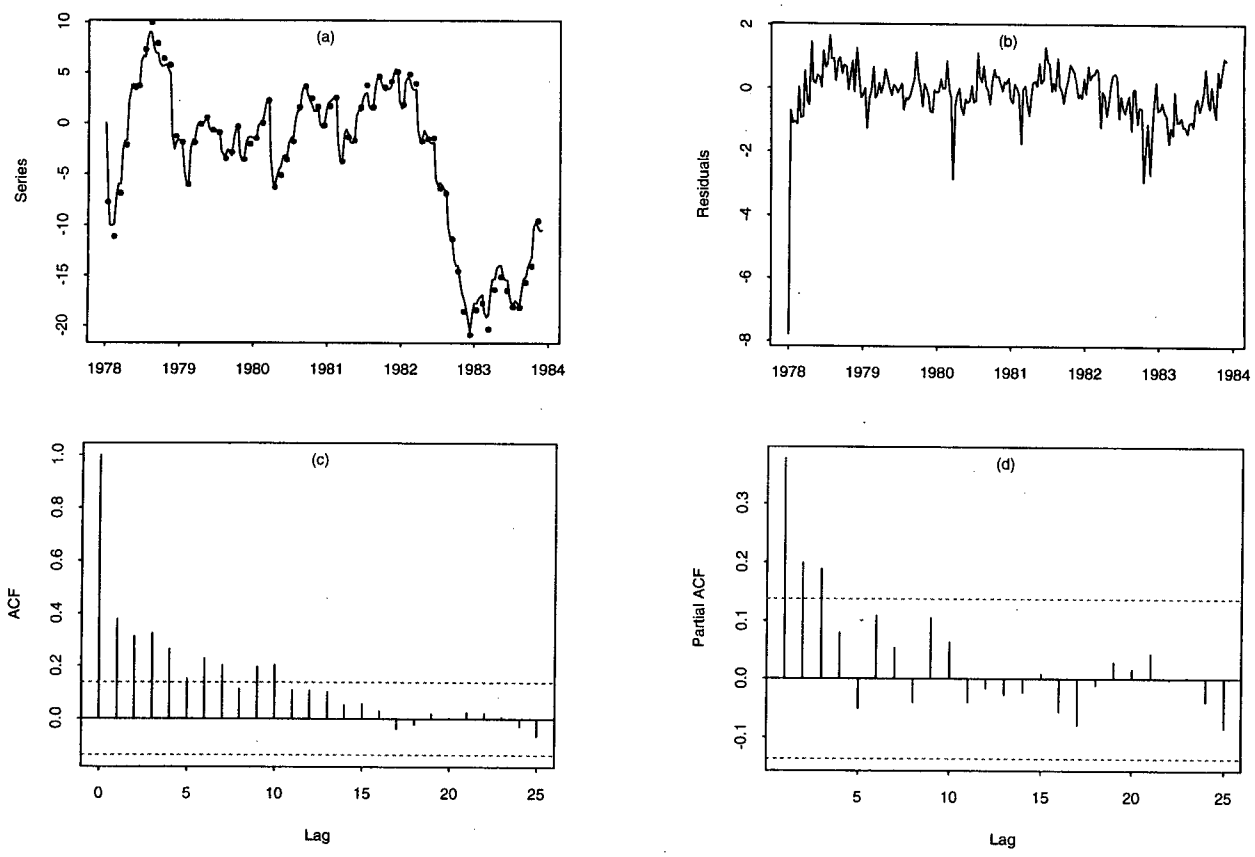

FIG. 3. MC model at Rabaul (ten stations). Legends as in Fig. 2. (a) Predictions, (b) residuals, (c) ACF of residuals, and (d) partial ACF of residuals. The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})/\Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

$$\begin{cases} X_{k+1} = \tilde{\mathbf{L}} X_k + \tilde{\tau}_k + \mathbf{F}\xi_k, \\ Y_k = \tilde{\mathbf{H}} X_k + \mathbf{G}\zeta_k, \end{cases}$$

for $k = 0, \cdots, 71$, where

$$X_k = \begin{pmatrix} W_{3k+1} \\ W_{3k+2} \\ W_{3k+3} \end{pmatrix}, \quad \tilde{\tau}_k = \begin{pmatrix} \tau_{3k+1} \\ \tau_{3k+2} \\ \tau_{3k+3} \end{pmatrix},$$

$$\xi_k = \begin{pmatrix} \epsilon_{3k+3} \\ \epsilon_{3k+4} \\ \epsilon_{3k+5} \end{pmatrix}, \quad \zeta_k = \begin{pmatrix} v_{3k+1} \\ v_{3k+2} \\ v_{3k+3} \end{pmatrix},$$

with system matrices

$$\tilde{\mathbf{L}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}^2 \\ \mathbf{0} & \mathbf{0} & \mathbf{L}^3 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{L} & \mathbf{I} & \mathbf{0} \\ \mathbf{L}^2 & \mathbf{L} & \mathbf{I} \end{bmatrix},$$

$$\tilde{\mathbf{H}} = [\mathbf{H}/3 \quad \mathbf{H}/3 \quad \mathbf{H}/3], \quad \mathbf{G} = [\mathbf{I}/3 \quad \mathbf{I}/3 \quad \mathbf{I}/3]$$

($\mathbf{I}$ denotes the 384×384 identity matrix). The Kalman filter is implemented in the same way as in section 3b. It is not necessary to work with the full $(3 \times 384) \times (3 \times 384)$ matrices in the defining equations. Practical implementation of the algorithm only requires 384×384 matrices, as in the MC model. Details of the average algorithm implementation are given in appendix B.

The Average model was run with six and ten stations data as was the original MC model. Figures 4 and 5 display the results for Rabaul. As before and as expected, the predictions are very close to the observed data. Figures 4a and 5a differ little from the corresponding panels in Figs. 2 and 3. Comparison of Figs. 4b and 5b with the corresponding panels in Figs. 2 and 3 shows that the residuals in the run with the Average model are distinctly smaller than those from the original MC model; note the difference in scale from Figs. 2b and 3b to Figs. 4b and 5b. Examination of Table 1 shows that this is the case at five of the six assimilation stations, the exception being Nauru. Since the Average model represents a change in the measurement model from MC, rather than a change in the observation or system error models, this indicates that the filtering scheme is overfitted. Recall that if the output of the filter were to track the observation exactly, the result would obviously contain the observation error, which in this case has rms amplitude of 3 cm. The combina-

TABLE 1. Statistical summary of residual analyses.

Station	Observed total $\hat{\sigma}_T$	MC 6 $\hat{\sigma}$	MC 6 $\hat{\sigma}_{est}$	MC 10 $\hat{\sigma}$	MC 10 $\hat{\sigma}_{est}$	Average 6 $\hat{\sigma}$	Average 6 $\hat{\sigma}_{est}$	Average 10 $\hat{\sigma}$	Average 10 $\hat{\sigma}_{est}$	AR (1) 6 $\hat{\sigma}$	AR (1) 6 $\hat{\sigma}_{est}$	AR (1) 10 $\hat{\sigma}$	AR (1) 10 $\hat{\sigma}_{est}$
					Stations used								
Rabaul	7.82	0.87	2.4	0.55	2.4	0.12	3.6	0.18	1.8	0.08	3.6	0.15	1.7
Nauru	10.2	1.8	2.4	2.0	2.1	2.1	2.2	2.2	1.1	2.1	2.0	1.5	1.0
Jarvis	7.64	1.2	2.0	1.4	2.0	0.57	2.0	0.74	1.2	0.48	2.0	0.62	1.1
Christmas	9.30	1.5	2.2	1.3	1.9	0.54	2.4	0.59	1.4	0.47	2.1	0.57	1.2
Santa Cruz	10.56	1.9	2.4	1.0	2.4	0.58	2.3	0.57	1.3	0.32	2.2	0.36	1.2
Callao	7.85	0.80	2.6	0.88	2.7	0.32	2.6	0.30	1.4	0.32	2.5	0.30	1.4
					Stations withheld								
Kapinga	6.03	5.5	3.9	1.3	2.3	6.0	4.0	0.72	1.3	4.4	3.7	0.66	1.1
Tarawa	8.0	3.7	3.2	1.4	2.1	3.8	3.8	0.82	1.3	2.9	3.6	0.73	1.2
Canton	7.0	5.0	5.5	0.82	2.4	5.0	4.6	0.25	1.5	3.7	4.4	0.21	1.4
Fanning	7.41	3.3	6.1	0.48	2.3	4.3	5.0	0.13	1.8	3.6	4.8	0.13	1.7

Note that $\hat{\sigma}_T$ is the estimated standard deviation of the sea level observations, $\hat{\sigma}$ is the estimated standard deviation of the residuals, and $\hat{\sigma}_{est}$ represents the Kalman filter estimation of the standard deviation of the residual error. The stations used and withheld are for the six station case.

tion of model and observation should be better than either the model or the observation alone. This is reflected in the fact that the estimated rms errors in the filtered model, $\hat{\sigma}_{est}$ in Table 1, are less than 3 cm.

Table 1 also shows that the standard deviations $\hat{\sigma}$ of the post-assimilation residuals $\mathbf{Y}_t - \mathbf{H}\hat{\mathbf{W}}_{t|t}$ are in general much smaller than we expect them to be; in other words, our analysis is closer to the actual observations than it should be and may contain too much of the observation noise. Since the filter output is a weighted sum of model output and observation, it is natural to suspect that the observations are overweighted. The weight on the model output is determined by our estimate of the forecast error, which should represent the cumulative effects of errors in the model itself, the forcing imposed and the initial conditions, as well as the salutary effect of past data assimilation. From the defining equations of the filter given in section 3b, it is clear that our estimate of the forecast error depends on the evolution matrix $\mathbb{L}$, the sequence of observation matrices $\mathbb{H}$, the observation noise covariance $\mathbb{R}$, and the system noise covariance $\mathbb{Q}$, which contains our prior estimate of the errors in the model and the forcing data. (As noted in that section, there is also a weak dependence on the initial error, but that has little influence in the present case.) A value of $\mathbb{Q}$ with smaller norm will result in a smaller value of $\mathbb{P}_{t|t} = E[(\mathbf{W}_t - \hat{\mathbf{W}}_{t|t})(\mathbf{W}_t - \hat{\mathbf{W}}_{t|t})^{\mathrm{T}}]$ at any given time, which will in

TABLE 2. First and second component of the ACF.

Station	MC 6 $\hat{\gamma}_1$	MC 6 $\hat{\gamma}_2$	MC 10 $\hat{\gamma}_1$	MC 10 $\hat{\gamma}_2$	Average 6 $\hat{\gamma}_1$	Average 6 $\hat{\gamma}_2$	Average 10 $\hat{\gamma}_1$	Average 10 $\hat{\gamma}_2$	AR (1) 6 $\hat{\gamma}_1$	AR (1) 6 $\hat{\gamma}_2$	AR (1) 10 $\hat{\gamma}_1$	AR (1) 10 $\hat{\gamma}_2$
Rabaul	.34	.28	.37	.31	.67	.52	.54	.24	.01	.09	.04	−.17
Nauru	.73	.52	.76	.59	.09	.07	.44	.32	.10	.09	.17	.18
Jarvis	.81	.66	.79	.63	.45	.16	.55	.21	.03	−.02	.12	−.13
Christmas	.73	.61	.69	.54	.44	.24	.34	.12	.00	.17	.01	.10
Santa Cruz	.92	.87	.91	.84	.77	.66	.77	.67	.00	.28	−.07	.22
Callao	.77	.61	.75	.56	.32	.35	.18	.23	−.08	.27	−.03	.23
Kapinga	.84	.68	.70	.49	.57	.34	.49	.32	.09	.18	.06	.23
Tarawa	.84	.68	.69	.48	.55	.40	.58	.43	.04	.22	.04	.19
Canton	.84	.71	.50	.32	.54	.25	.52	.21	.20	−.06	.24	−.09
Fanning	.83	.68	.44	.28	.56	.28	.17	.09	.04	−.14	−.01	.03

Note that $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the sample estimated first- and second-order autocorrelations. The sample ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})/\Sigma_{t=1}^{n}(y_t - \bar{y})^2$.

TABLE 3. First and second component of the PACF.

	MC				Average				AR (1)			
	6		10		6		10		6		10	
Station	$\hat{\rho}_{11}$	$\hat{\rho}_{22}$	$\hat{\rho}_{11}$	$\hat{\rho}_{22}$	$\hat{\rho}_{11}$	$\hat{\rho}_{22}$	$\hat{\rho}_{11}$	$\hat{\rho}_{22}$	$\hat{\rho}_{11}$	$\hat{\rho}_{22}$	$\hat{\rho}_{11}$	$\hat{\rho}_{22}$
Rabaul	.35	.18	.38	.20	.67	.13	.55	−.09	.01	.09	.04	−.17
Nauru	.73	−.04	.76	.01	.09	.06	.44	.17	.10	.08	.17	.15
Jarvis	.81	.04	.79	.01	.45	−.05	.55	−.15	.03	−.02	.12	−.14
Christmas	.73	.17	.69	.12	.44	.06	.34	.01	.00	.17	.01	.10
Santa Cruz	.92	.10	.91	.07	.77	.15	.77	.20	.00	.28	−.07	.21
Callao	.77	.03	.75	.00	.32	.27	.18	.21	−.08	.26	−.03	.23
Kapinga	.84	−.11	.70	.01	.57	.02	.49	.11	.09	.17	.06	.22
Tarawa	.84	−.07	.69	.00	.55	.13	.58	.15	.04	.22	.04	.19
Canton	.84	.00	.50	.09	.54	−.06	.52	−.08	.20	−.10	.24	−.16
Fanning	.83	−.08	.44	.10	.56	−.04	.17	.06	.04	−.14	−.01	.03

Note that $\hat{\rho}_{11}$ and $\hat{\rho}_{22}$ are the sample estimated first- and second-order partial autocorrelations.

turn, result in smaller gain $\mathbf{K}_t$. This will lead to the end result of weakening the correction from $\hat{\mathbf{W}}_{t|t-1}$ to $\hat{\mathbf{W}}_{t|t}$. In this case, more weight would then be given to the model, and less to the observation.

With the exception of Nauru, the standard deviations of the residuals, $\hat{\sigma}$, are consistently much smaller than the data errors or the expected errors in the model-data combination $\hat{\sigma}_{\text{est}}$. A filtering scheme that embodied more confidence in the dynamical model in the form of a system noise covariance matrix $\mathbf{Q}$ with smaller norm would give more weight to the model and less to the observations, and yield a sequence of residuals with

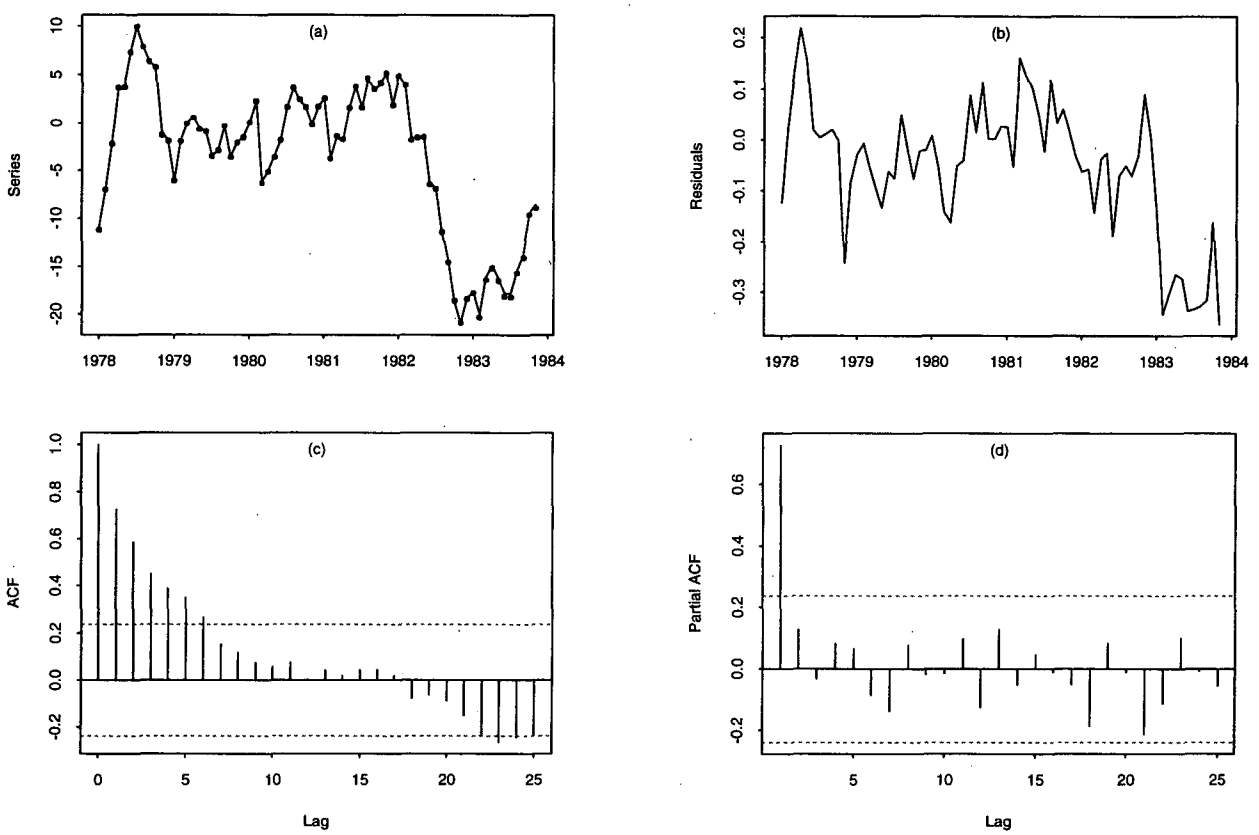

FIG. 4. Average model at Rabaul (six stations): (a) Predictions, (b) residuals, (c) ACF of residuals, and (d) partial ACF of residuals. The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}) / \Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

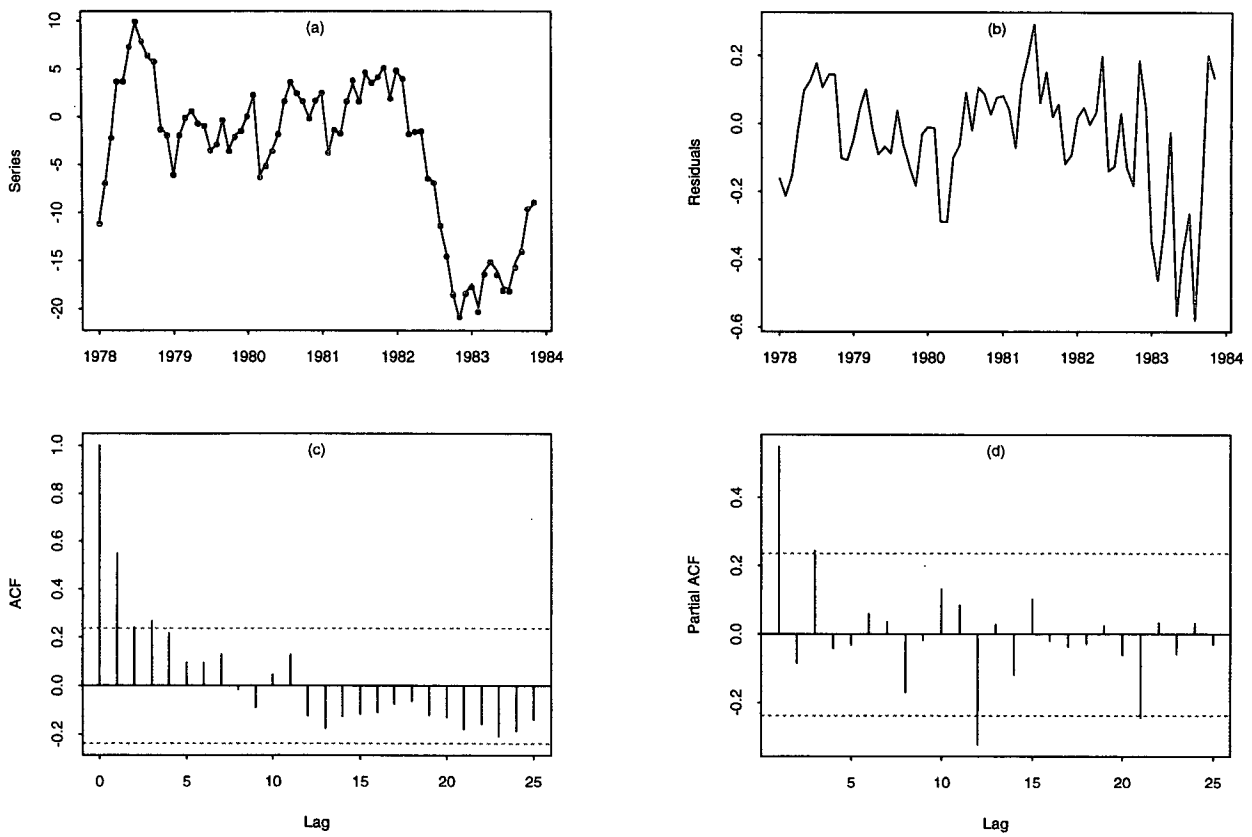

FIG. 5. Average model at Rabaul (ten stations): (a) Predictions, (b) residuals, (c) ACF of residuals, and (d) partial ACF of residuals. The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})/\Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

greater rms amplitude than those found in the present case, but still smaller than the expected observation errors. The resulting overall field might well be more accurate, since it would contain less of the observation errors.

It is also interesting that the residuals seem to decrease steadily toward the end of the 1983. This pattern, which appears more clearly in the residuals from the experiment in which six stations were assimilated, may be the consequence of the removal of average values in computing the wind stress and sea level anomalies, since the averaging period includes the strong El Niño disturbance of 1982–83, cf. MC. It is interesting that it does not appear clearly in the results from the MC model; see Figs. 2b and 3b. Perhaps the effect is masked by the artificial data introduced in MC.

In the experiments with data assimilated from six stations, residuals at the four stations held back for verification have comparable rms amplitudes for the average method and the original MC method with six

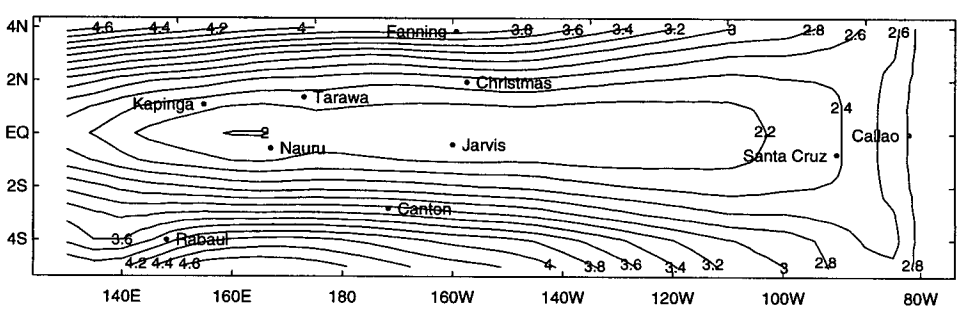

FIG. 6. Average model (six stations): Contour map of expected rms error.

stations. The estimated errors for the average case are a bit closer to the actual rms residual amplitudes except at Kapingamarangi. The greatest improvement in the estimated residual amplitude is at Fanning.

As shown in Table 1, the residuals of the average model have markedly reduced serial correlation at every station but Rabaul, even though they still seem to be serially correlated. The improvement over the MC model is shown by the partial ACF plots shown in Figs. 4d and 5d and in Table 1. The partial ACF plots suggest a first-order autocorrelation model for the residuals in both cases, instead of the third-order model, which was suggested for the MC residuals, cf. Figs. 2d and 3d. This is consistent with the general finding of a one-month autoregressive structure. Tables 2 and 3 show that in most cases the one-month and two-month autocorrelations in the Average model were less than the ten-day and twenty-day autocorrelations in the MC model. The exceptions were at Rabaul, and at Canton, where the one- and two-month lags in both experiments with the Average model were similar to the ten- and twenty-day lags in the MC experiment with assimilation at ten stations. For the Average model, 95% confidence bands for the ACF and PACF of the residuals are ± 0.24.

Maps of the estimated error of the Average model are given in Figs. 6 and 7. Those are contour maps of the expected residual standard deviation, given by $\hat{\mathbf{H}}\hat{\mathbf{P}}\hat{\mathbf{H}}^{\mathrm{T}}$, where $\hat{\mathbf{P}}$ is the stationary residual covariance matrix of the state estimation for the average approach, that is, $\hat{\mathbf{P}} = E[(\mathbf{X}_{t_{71}} - \mathbf{X}_{t_{71}|t_{71}})(\mathbf{X}_{t_{71}} - \mathbf{X}_{t_{71}|t_{71}})^{\mathrm{T}}]$, where t_{71} corresponds to the time of the last observation. The estimated errors are quite a bit smaller than those found in MC Fig. 5b, and are almost certainly overoptimistic, especially in the data void between Santa Cruz and Jarvis.

Using the fact that the partial autocorrelations drop after the first lag, that is, one month in this formulation, an AR(p) model, that is, an autoregressive model with p lags, was fitted to each component of the residuals, for $p = 1, \cdots, 10$. We selected the order of the AR model by minimizing the Akaike Information Criterion (AIC), which is equal to [-2 ln(maximized likelihood) + 2 (number of independent parameters estimated)]; see, for example, chapter 9 of Brockwell and Davis (1991). The result was an AR(1) model for all the stations:

$$\boldsymbol{\nu}_t = \boldsymbol{\Phi}\boldsymbol{\nu}_{t-1} + \boldsymbol{\omega}_t, \qquad (1)$$

where $\boldsymbol{\Phi} = \mathrm{diag}(\phi_i)$. The estimated coefficients and their estimated standard deviations are displayed in Table 4.

TABLE 4. Estimated autoregressive parameters.

	AR(1) Model			
	6		10	
Station	$\hat{\phi}$	$\hat{\sigma}_{\hat{\phi}}$	$\hat{\phi}$	$\hat{\sigma}_{\hat{\phi}}$
Rabaul	.77	.08	.58	.10
Nauru	.31	.12	.57	.10
Jarvis	.60	.11	.69	.10
Christmas	.49	.11	.39	.11
Santa Cruz	.89	.06	.87	.07
Callao	.53	.10	.45	.11
Kapingamarangi	.72	.09	.69	.10
Tarawa	.66	.11	.67	.10
Canton	.76	.09	.81	.08
Fanning	.57	.10	.18	.12

Note that $\hat{\phi}$ is the maximum likelihood estimator of the first-order autoregressive parameter and $\hat{\sigma}_{\hat{\phi}}$ is the estimated standard deviation.

A multivariate AR(1) model to $\boldsymbol{\nu}_t$ was also fitted, but the differences in performance between the scalar model and the multivariate model were negligible. Hence, we use the scalar model. There is no obvious geographical pattern in $\boldsymbol{\Phi}$. For the experiment with ten stations, the Average model produces approximately nonserially correlated residuals at Callao and Fanning. Similarly, the Average model with six stations produces nearly nonserially correlated residuals at Nauru and Callao.

b. *The AR(1) approach*

A greater improvement in the behavior of the residuals is achieved by incorporating (1) into the Kalman filter as follows. According to the diagnostic check of

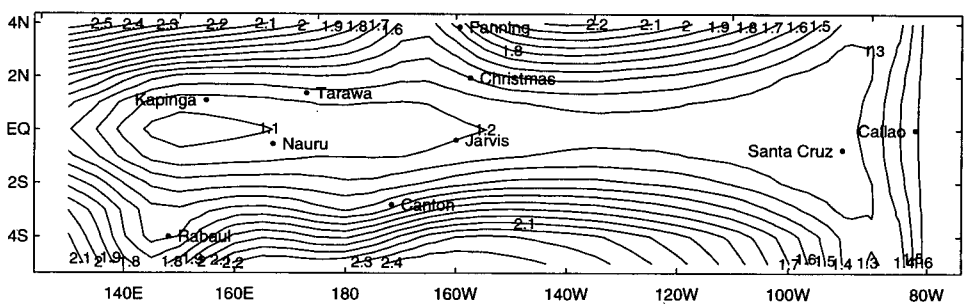

FIG. 7. Average model (ten stations): Contour map of expected rms error.

the average model, the residual sequence $\boldsymbol{\nu}_t$ is serially correlated. Furthermore, it was found that an appropriate model for the residuals sequence is an AR(1) model: $\boldsymbol{\nu}_t = \Phi\boldsymbol{\nu}_{t-1} + \boldsymbol{\omega}_t$, where $\boldsymbol{\omega}_t$ is a white noise sequence with covariance matrix $\mathbf{B}$ and Φ is a diagonal matrix (10×10).

Rather than rebuilding the Kalman filter model from scratch, we incorporated the information provided by the residuals into the model. We chose to do this by modifying the Average model described in section 4 for $t = 0, \cdots, 71$ as follows:

$$\begin{cases} \boldsymbol{X}_{t+1} = \tilde{\mathbf{L}}\boldsymbol{X}_t + \tilde{\tau}_t + \mathbf{F}\boldsymbol{\xi}_t, \\ \boldsymbol{Y}_t = \mathbf{H}\boldsymbol{X}_t + \boldsymbol{\gamma}_t, \\ \boldsymbol{\gamma}_t = \Phi\boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t. \end{cases} \tag{2}$$

The AR(1) structure for $\boldsymbol{\gamma}_t$ is suggested by the first-order autoregressive model for the innovation $\boldsymbol{\nu}_t = \mathbf{G}\boldsymbol{\zeta}_t$. In the modified model, γ_t represents the error in the observation equation, that is, $\boldsymbol{\gamma}_t = \boldsymbol{Y}_t - \mathbf{H}\boldsymbol{X}_t$, the difference between the observation $\boldsymbol{Y}_t$ and its expected value $\mathbf{H}\boldsymbol{X}_t$. This is analogous to $\boldsymbol{\nu}_t = \boldsymbol{Y}_t - \mathbf{H}\hat{\boldsymbol{X}}_{t|t-1}$, which represents the *observed* difference between $\boldsymbol{Y}_t$ and its predicted value $\mathbf{H}\hat{\boldsymbol{X}}_{t|t-1}$, denoted by $\hat{\boldsymbol{X}}_t$ herein. Let M denote the space generated by the history of the observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1}$, up to time $t-1$. A modified estimator of $\boldsymbol{Y}_t$ based on M is $\tilde{\boldsymbol{Y}}_t$ defined as

$$\begin{aligned} \tilde{\boldsymbol{Y}}_t &= E(\boldsymbol{Y}_t | M) \\ &= E(\mathbf{H}\boldsymbol{X}_t + \gamma_t | M) \\ &= \mathbf{H}E(\boldsymbol{X}_t | M) + \Phi E(\gamma_{t-1} + \boldsymbol{\omega}_t | M) \\ &= \mathbf{H}E(\boldsymbol{X}_t | M) + \Phi E(\gamma_{t-1} | M) \\ &= \mathbf{H}\hat{\boldsymbol{X}}_t + \Phi E(\mathbf{Y}_{t-1} - \mathbf{H}\boldsymbol{X}_{t-1} | M) \\ &= \mathbf{H}\hat{\boldsymbol{X}}_t + \Phi\boldsymbol{\nu}_{t-1} \\ &= \hat{\boldsymbol{Y}}_t + \Phi\boldsymbol{\nu}_{t-1}. \end{aligned}$$

Note that $\tilde{\boldsymbol{Y}}_t$ is the predictor produced by the modified model described by Eq. (2). The new prediction error is $\boldsymbol{\omega}_t = \boldsymbol{Y}_t - \tilde{\boldsymbol{Y}}_t = \boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_t - \Phi\boldsymbol{\nu}_{t-1} = \boldsymbol{\nu}_t - \Phi\boldsymbol{\nu}_{t-1}$. Observe that $\text{var}(\boldsymbol{\nu}_t) = \text{var}(\Phi\boldsymbol{\nu}_{t-1} + \boldsymbol{\omega}_t)$. Since $\boldsymbol{\nu}_{t-1}$ and $\boldsymbol{\omega}_t$ are uncorrelated, $\text{var}(\boldsymbol{\nu}_t) = \text{var}(\Phi\boldsymbol{\nu}_{t-1}) + \text{var}(\boldsymbol{\omega}_t)$. Thus, $\text{var}(\boldsymbol{\nu}_t) = \Phi\ \text{var}(\boldsymbol{\nu}_{t-1})\Phi^{\mathrm{T}} + \text{var}(\boldsymbol{\omega}_t)$. Consequently, the variance of the new estimation error $\boldsymbol{\omega}_t$ is less than or equal to the old estimation error $\boldsymbol{\nu}_t$.

c. *Performance of the AR(1) model*

Figures 8 and 9 show the performance of the AR(1) approach at Rabaul, for experiments performed with data assimilated at six and ten stations, respectively. As was the case with the previous methods, the rms amplitude of the residuals was very small; see Figs. 8a and 9a. The residual plots do not present the clear patterns shown in Figs. 4b and 5b. The evident negative trend in the residuals during 1983 might still be present, but not so clearly as in the averaged case; compare Fig. 8b to Fig. 4b. Contrary to the MC and simple average approaches, the ACF's of the residuals in the two AR(1) experiments, shown in Figs. 8c and 9c, decay to zero after the lag at zero time steps, indicating approximately nonserially correlated errors. The Partial ACF, Figs. 8d and 9d, show no significant autoregressive components for the residuals. Thus, the residuals seem to be white noise. As shown in Table 2 and Table 3, the ACF and PACF of the residuals for the AR(1) model in the other nine stations have similar behavior to those of Rabaul. Analogously to the average model, 95% confidence bands for the ACF and PACF are ± 0.24.

From Table 1, the residual standard deviations of the AR(1) approach are smaller than the residual standard deviations of the two other methods at most of the stations. The greatest differences between the AR(1) approach and the average approach are at Santa Cruz in the experiments with six and ten assimilation stations, and at Rabaul and Kapingamarangi in the experiments with six assimilation stations. This latter case is interesting in that the rms error in the MC and average approaches at Kapingamarangi is as large as the signal amplitude. This indicates that the model estimates are useless there; an estimate of zero amplitude would do as well. In the AR(1) case, the rms error is 25% less than the signal amplitude, and is closer to the prior estimate $\hat{\sigma}_{\text{est}}$, which is roughly the same in all three experiments.

It is worth noting that the smaller standard deviation of the residuals and slightly better prior estimates at the four stations omitted from the assimilation process is achieved by introducing an additional vector of parameters Φ to the Kalman filter model. The cost of having a smaller variance is to entertain a slightly less parsimonious model.

Maps of the estimated error of the AR(1) model are given in Figs. 10 and 11. These maps show generally smaller rms amplitudes than those shown for the Average model in Figs. 6 and 7. At first glance these maps appear overoptimistic, and perhaps they are, but Table 1 shows that the posterior estimates of the rms errors at Kapingamarangi, Tarawa, Canton, and Fanning, where no data were assimilated, are reasonably reliable.

5. Discussion

With the modified models, Average and Average with AR(1) error, two main problems of the MC approach appear to be solved. The introduction of artificial data into the model can be avoided by reformulating the MC approach as an average state space model. The application of the average model with AR(1) error produces nearly nonserially correlated residuals. This has been suggested by some authors, for example, Daley 1992b; Dee et al. 1985, as a criterion for optimality, following the work of Kailath (1968). More-

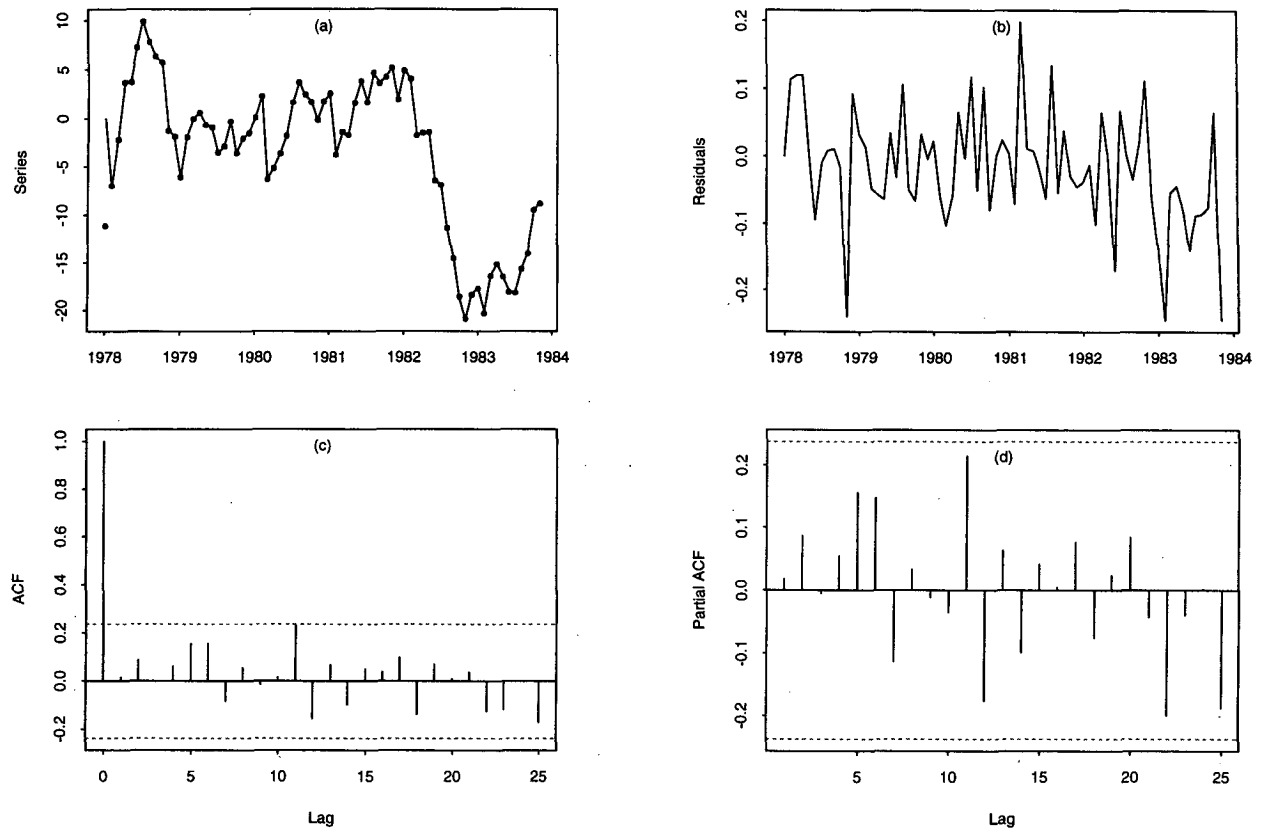

FIG. 8. AR(1) model at Rabaul (six stations): (a) Predictions, (b) residuals, (c) ACF of residuals, and (d) partial ACF of residuals. The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})/\Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

over, our finding of an autoregressive structure in the residuals of the Kalman filter model suggests that its state space is incomplete. As noted by MC, the physical model of the sea level is highly simplified; gross effects such as the spatial variation of the density structure of the water column, and therefore the wave speeds, is neglected, and the dynamics of the meridional variation of the sea level height are approximated by a Kelvin wave and five Rossby waves.

Another probably more important source of problems with the MC model is the assumption of temporal independence of the wind stress error field. The assumption of month to month independence of the wind stress error seems unrealistic and better models for its correlation structure should be investigated. In that sense, it could be worth exploring a first-order autoregressive model for the system noise. This approach can be handled through the Kalman filter theory with *colored noise;* see, for example, chapter 5 of Chui and Chen (1990). Even though the computations would be more complex than those for the independent state noise case, the incorporation of an autoregressive structure for the wind error could eventually improve the performance of the Kalman filter.

Our AR(1) model is a model of the autocorrelation of the innovation sequence. Such autocorrelation could result from serial correlation in either the system noise or the observation noise. Daley (1992a,b) suggested diagnostic procedures for investigating the nature of the noise series based on lagged innovation covariances, but the lagged innovation covariances do not exhibit any significant spatial structure in this system. This approach may well be fruitful for more recent, more dense datasets.

Our finding that the expected analysis error variances are not highly sensitive to serial correlation of the innovation sequences [compare the values of $\hat{\sigma}_{\text{est}}$ in Table 1 for corresponding experiments with the Average and AR(1) models] is consistent with the theory presented by Daley (1992a). In that paper, Daley argued that when the advection speed U exceeds the quotient $\Delta x/\Delta t$, where Δx is the spatial interval between observations and Δt is the assimilation interval, the expected analysis error is not sensitive to serial correlation in the observation errors. The first baroclinic mode Kelvin wave in our model travels nearly 70° in longitude in a month. Analysis errors stemming from errors in the higher meridional mode and higher baroclinic mode

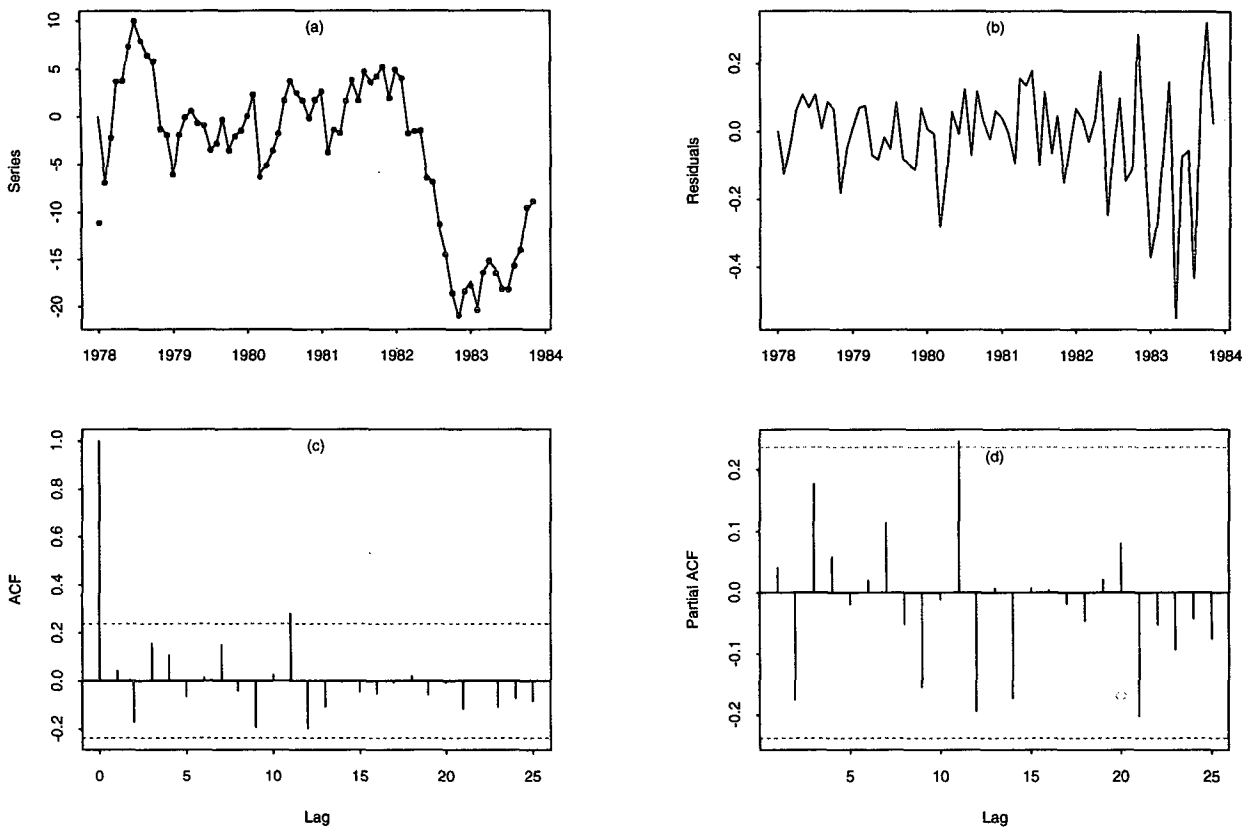

FIG. 9. AR(1) model at Rabaul (ten stations): (a) Predictions, (b) residuals, (c) ACF of residuals, and (d) partial ACF of residuals. The ACF is given by $\hat{\gamma}_k = \Sigma_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}) / \Sigma_{t=1}^{n} (y_t - \bar{y})^2$.

Rossby waves are most affected by serially correlated observation errors. Subsequent studies with greater spatial data density will also be more affected by serially correlated observation errors.

A modification of the state space model to consider daily tide gauge data is another interesting focus for future research. Although the physical model should be reformulated and the noise correlation structure could be more complicated, daily observations of the sea level could eventually provide much more information than monthly data and a better Kalman model can be obtained.

The rms differences between the Kalman filter output and the observations themselves at the assimilation stations, with the exception of Nauru, are much smaller than the prior estimates of these quantities or the rms observation errors. This leads us to believe that the observations are overfitted, that is, the weight given to the observations is too large relative to the model-derived

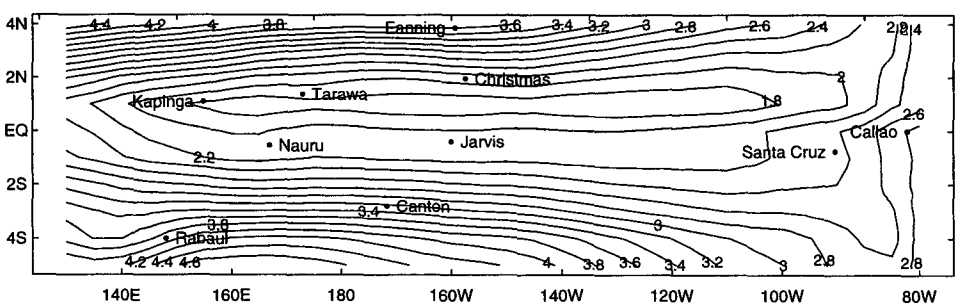

FIG. 10. AR(1) model (six stations): Contour map of expected rms error.

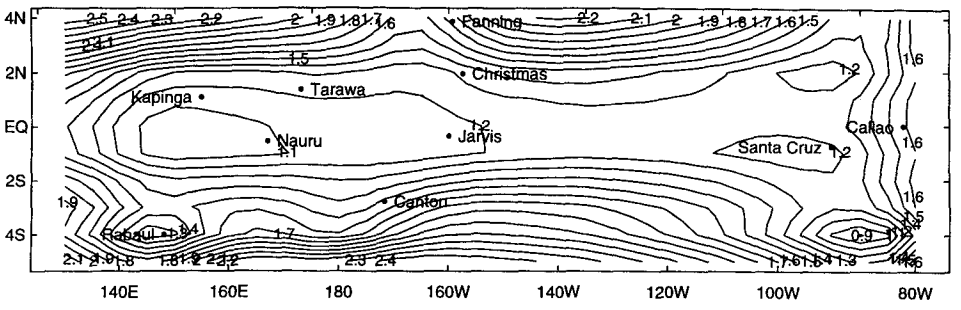

FIG. 11. AR(1) model (ten stations): Contour map of expected rms error.

prior estimate. A better model could probably be derived by reducing **Q** and recalculating the values of $\hat{\phi}$. Such a tuning process would be better performed on a more extensive dataset.

Kelly and Qiu (1995), in their construction of a Kalman filter for assimilation satellite-derived temperature and geostrophic velocities into a mixed layer model of the North Atlantic, examined the spectra of the innovation sequences that resulted from several parameter choices and chose the one that yielded the spectrum that was closest to being white. No tests were performed to determine whether the observed color in the spectrum was significant, but the shape of the spectrum was adequately sensitive to the parameter choices. We know of no other example of a model that has achieved the goal of uncorrelated innovations with actual observed data. Having achieved this goal, we can evaluate the improvement that is available. Evidently, the original MC model was fairly close to optimal for many practical purposes. The additional information extracted from the data does not produce a dramatic change in the actual forecasts or posterior error estimates, even though these latter quantities are somewhat improved. We are not sure whether this improvement is important. It is noteworthy, however, that in the AR(1) experiment with data assimilated at six stations, the residuals at the four stations held back for verification appear to be white. It is likely that Figs. 10 and 11 are overoptimistic, but we have no way of knowing this without additional data.

With the addition of the new models, however, certain things become apparent. The most obvious of these is the appearance of the trend in the innovation sequence at Rabaul in 1983, and its apparent susceptibility to modeling by a first-order autoregressive process. It is this sort of finding, that we hope will lead to improvement in our physical models and enhancement of our understanding of the physical processes, as well as improved mapping of the relevant fields.

Some may object to the approach here on the grounds that our dynamical model does not contain sufficient physical detail and is too coarsely resolved to justify an elaborate statistical analysis such as this one. It is worth noting, however, that this long-wave model embodies a set of ocean physics that has for a decade or more been the basis of our interpretation of the large scale response of the tropical ocean to atmospheric forcing. We believe the present work to contain the essential step toward refinement of this model, whether the deficiencies lie in the data or in the model itself.

As noted in MC, neither nonlinearities nor errors in the wave speed could be expected to contribute as much to the error amplitude as wind errors of the magnitude we expect. But whatever the source of error, it is noteworthy that it is not necessary to look beyond autoregressive models with lags of one month to explain the innovation sequence. Ultimately, we hope that analysis of time series characteristics of residual sequences from data assimilation will lead us toward the most efficient ways to improve the models. In future examination of possible model improvements by addition of physical detail, we might rule out as relatively unimportant refinements which, if neglected, would be expected to lead to residuals with autocorrelation times greater than one month.

Acknowledgments. This research was supported in part by the Office of Naval Research Grant N00014-89-J-1851 and NOAA Office of Global Programs Grant NA-36-GP0123-01. NHC was supported in part by an Earmarked Research Grant from the RGC of Hong Kong. Part of this research was begun as WP's Advanced Data Analysis project for the Ph.D. in statistics at Carnegie Mellon University. The authors would like to thank two referees and the editor for helpful comments. RNM would like to thank Mark Cane and Dudley Chelton for helpful suggestions and critical reading of the manuscript.

APPENDIX A

The Physical Model

The MC physical model, based on linearized equations of motion on an equatorial beta plane, may be described by the following state space equations, cf. Miller and Cane (1989):

$$\begin{cases} \boldsymbol{W}_{t+1} = \mathbf{L}\boldsymbol{W}_t + \boldsymbol{\tau}_t + \boldsymbol{\epsilon}_t \\ \boldsymbol{Y}_t = \mathbf{H}\boldsymbol{W}_t + \mathbf{v}_t. \end{cases}$$

The physical model is derived from the linearized primitive equations on the equatorial β plane. Separation of variables in the vertical and horizontal directions results in a decomposition in vertical modes, the amplitude of which is governed by a set of equations formally identical to the linearized shallow water equations, as described by Cane and Sarachik (1981) and Cane (1984). The solution to the shallow water equations on the equatorial β plane, further simplified by a long-wave approximation, can be expressed as the sum of Kelvin and nondispersive Rossby wave amplitudes, whose evolution is described by simple advection equations, with speeds determined from the separation constants in the modal analysis. The condition that the zonal velocity must vanish at the land boundary determines the eastern boundary condition. This condition cannot be satisfied in the context of the long-wave approximation. The appropriate condition at the western boundary is that the integral of the mass flux into the boundary over the extent of the boundary must vanish.

In the numerical scheme, two vertical modes are included and for each vertical mode, the corresponding series includes one Kelvin mode and five Rossby modes. The advection equations for the wave amplitudes are solved by the method of characteristics. The boundary conditions provide relations between the incident Rossby wave and the reflected Kelvin wave at the western boundary, and the incident Kelvin wave and the reflected Rossby waves at the eastern boundary. Mass is not exactly conserved at the eastern boundary. Some energy is lost at the western boundary. This contributes most of the dissipation in the model.

APPENDIX B

Average State Space Algorithm

This appendix explains with some detail the implementation of the average algorithm. First, all the equations for the standard Kalman filter are provided, as in Gelb (1974). Second, by using the mathematical structure of the average approach, the recursive equations are greatly simplified. In particular, the average algorithm is based on a number of 384 × 384 matrices instead of 1152 × 1152 matrices, as required by the original equations.

The state space equations for the average approach are given by

$$\begin{cases} \mathbb{X}_{k+1} = \tilde{\mathbb{L}}\mathbb{X}_k + \tilde{\tau}_k + \mathbb{F}\xi_k, \\ Y_k = \tilde{\mathbb{H}}\mathbb{X}_k + \mathbb{G}\zeta_k, \end{cases}$$

with the matrices $\tilde{\mathbb{L}}$, $\mathbb{F}$, $\tilde{\mathbb{H}}$, and $\mathbb{G}$ defined as in section 4a. Following Gelb (1974), the Kalman recursions can be written as:

Initial state estimation: $\mathbb{X}_0 = \hat{\mathbb{X}}_{0|0} = 0$.

State noise covariance matrix:

$$\tilde{\mathbb{Q}} = \mathbb{F}[\text{diag}\{\mathbb{Q}, \mathbb{Q}, \mathbb{Q}\}]\mathbb{F}^{\mathrm{T}} \equiv \begin{bmatrix} \mathbb{Q} & \mathbb{Q}_1 & \mathbb{Q}_2 \\ \mathbb{Q}_1^{\mathrm{T}} & \mathbb{Q}_3 & \mathbb{Q}_4 \\ \mathbb{Q}_2^{\mathrm{T}} & \mathbb{Q}_4^{\mathrm{T}} & \mathbb{Q}_5 \end{bmatrix}.$$

State estimation:

$$\hat{\mathbb{X}}_{k|k-1} = \tilde{\mathbb{L}}\hat{\mathbb{X}}_{k-1|k-1}.$$

Given the state decomposition

$$\mathbb{X}_k = \begin{pmatrix} \mathbb{X}_k^{(1)} \\ \mathbb{X}_k^{(2)} \\ \mathbb{X}_k^{(3)} \end{pmatrix},$$

this equation can be written as

$$\hat{\mathbb{X}}_{k|k-1}^{(1)} = \mathbb{L}\hat{\mathbb{X}}_{k-1|k-1}^{(3)}$$

$$\hat{\mathbb{X}}_{k|k-1}^{(2)} = \mathbb{L}^2\hat{\mathbb{X}}_{k-1|k-1}^{(3)}$$

$$\hat{\mathbb{X}}_{k|k-1}^{(3)} = \mathbb{L}^3\hat{\mathbb{X}}_{k-1|k-1}^{(3)}.$$

Error covariance estimation:

$$\tilde{\mathbb{P}}_{k|k-1} = \tilde{\mathbb{L}}\tilde{\mathbb{P}}_{k-1|k-1}\tilde{\mathbb{L}}^{\mathrm{T}} + \tilde{\mathbb{Q}}.$$

If $\tilde{\mathbb{P}}_{k|k-1}$ is decomposed as

$$\tilde{\mathbb{P}}_{k|k-1} = \begin{bmatrix} \mathbb{P}_{k|k-1}^{(1)} & \mathbb{P}_{k|k-1}^{(2)} & \mathbb{P}_{k|k-1}^{(3)} \\ \mathbb{P}_{k|k-1}^{(2)\mathrm{T}} & \mathbb{P}_{k|k-1}^{(4)} & \mathbb{P}_{k|k-1}^{(5)} \\ \mathbb{P}_{k|k-1}^{(3)\mathrm{T}} & \mathbb{P}_{k|k-1}^{(5)\mathrm{T}} & \mathbb{P}_{k|k-1}^{(6)} \end{bmatrix},$$

then

$$\mathbb{P}_{k|k-1}^{(1)} = \mathbb{L}\mathbb{P}_{k-1|k-1}^{(6)}\mathbb{L}^{\mathrm{T}} + \mathbb{Q}$$

$$\mathbb{P}_{k|k-1}^{(2)} = \mathbb{L}\mathbb{P}_{k-1|k-1}^{(6)}\mathbb{L}^{2\mathrm{T}} + \mathbb{Q}_1$$

$$\mathbb{P}_{k|k-1}^{(3)} = \mathbb{L}\mathbb{P}_{k-1|k-1}^{(6)}\mathbb{L}^{3\mathrm{T}} + \mathbb{Q}_2$$

$$\mathbb{P}_{k|k-1}^{(4)} = \mathbb{L}^2\mathbb{P}_{k-1|k-1}^{(6)}\mathbb{L}^{2\mathrm{T}} + \mathbb{Q}_3$$

$$\mathbb{P}_{k|k-1}^{(5)} = \mathbb{L}^2\mathbb{P}_{k-1|k-1}^{(6)}\mathbb{L}^{3\mathrm{T}} + \mathbb{Q}_4$$

$$\mathbb{P}_{k|k-1}^{(6)} = \mathbb{L}^3\mathbb{P}_{k-1|k-1}^{(6)}\mathbb{L}^{3\mathrm{T}} + \mathbb{Q}_5.$$

However, it is only necessary to keep track of the following three matrices:

$$\mathbb{M}_{k|k-1}^{(1)} = \mathbb{P}_{k|k-1}^{(1)} + \mathbb{P}_{k|k-1}^{(2)} + \mathbb{P}_{k|k-1}^{(3)}$$

$$\mathbb{M}_{k|k-1}^{(2)} = \mathbb{P}_{k|k-1}^{(2)\mathrm{T}} + \mathbb{P}_{k|k-1}^{(4)} + \mathbb{P}_{k|k-1}^{(5)}$$

$$\mathbb{M}_{k|k-1}^{(3)} = \mathbb{P}_{k|k-1}^{(3)\mathrm{T}} + \mathbb{P}_{k|k-1}^{(5)\mathrm{T}} + \mathbb{P}_{k|k-1}^{(6)}$$

Error covariance update:

$$\tilde{\mathbb{P}}_{k|k} = \tilde{\mathbb{P}}_{k|k-1} - \tilde{\mathbb{P}}_{k|k-1}\tilde{\mathbb{H}}^{\mathrm{T}}[\tilde{\mathbb{H}}\tilde{\mathbb{P}}_{k|k-1}\tilde{\mathbb{H}}^{\mathrm{T}} + \tilde{\mathbb{R}}]^{-1}\tilde{\mathbb{H}}\tilde{\mathbb{P}}_{k|k-1}$$

If $\tilde{\mathbb{P}}_{k|k}$ is written as

$$\tilde{\mathbb{P}}_{k|k} = \begin{bmatrix} \mathbb{P}_{k|k}^{(1)} & \mathbb{P}_{k|k}^{(2)} & \mathbb{P}_{k|k}^{(3)} \\ \mathbb{P}_{k|k}^{(2)\mathrm{T}} & \mathbb{P}_{k|k}^{(4)} & \mathbb{P}_{k|k}^{(5)} \\ \mathbb{P}_{k|k}^{(3)\mathrm{T}} & \mathbb{P}_{k|k}^{(5)\mathrm{T}} & \mathbb{P}_{k|k}^{(6)} \end{bmatrix},$$

and the following two matrices are defined

$$\mathbf{M}_{k|k-1} = \mathbf{M}^{(1)}_{k|k-1} + \mathbf{M}^{(2)}_{k|k-1} + \mathbf{M}^{(3)}_{k|k-1},$$
$$\mathbf{Z}_{k|k-1} = \mathbf{H}\mathbf{M}_{k|k-1}\mathbf{H} + 3\mathbf{R},$$

then

$$\mathbf{P}^{(1)}_{k|k} = \mathbf{P}^{(1)}_{k|k-1} - \mathbf{M}^{(1)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{(1)\mathrm{T}}_{k|k-1}$$
$$\mathbf{P}^{(2)}_{k|k} = \mathbf{P}^{(2)}_{k|k-1} - \mathbf{M}^{(1)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{(2)\mathrm{T}}_{k|k-1}$$
$$\mathbf{P}^{(3)}_{k|k} = \mathbf{P}^{(3)}_{k|k-1} - \mathbf{M}^{(1)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{(3)\mathrm{T}}_{k|k-1}$$
$$\mathbf{P}^{(4)}_{k|k} = \mathbf{P}^{(4)}_{k|k-1} - \mathbf{M}^{(2)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{(2)\mathrm{T}}_{k|k-1}$$
$$\mathbf{P}^{(5)}_{k|k} = \mathbf{P}^{(5)}_{k|k-1} - \mathbf{M}^{(2)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{(3)\mathrm{T}}_{k|k-1}$$
$$\mathbf{P}^{(6)}_{k|k} = \mathbf{P}^{(6)}_{k|k-1} - \mathbf{M}^{(3)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{(3)\mathrm{T}}_{k|k-1}.$$

Similarly to the previous case, it is only needed to keep track of the following three matrices:

$$\mathbf{M}^{(1)}_{k|k} = \mathbf{P}^{(1)}_{k|k} + \mathbf{P}^{(2)}_{k|k} + \mathbf{P}^{(3)}_{k|k}$$
$$\mathbf{M}^{(2)}_{k|k} = \mathbf{P}^{(2)\mathrm{T}}_{k|k} + \mathbf{P}^{(4)}_{k|k} + \mathbf{P}^{(5)}_{k|k}$$
$$\mathbf{M}^{(3)}_{k|k} = \mathbf{P}^{(3)\mathrm{T}}_{k|k} + \mathbf{P}^{(5)\mathrm{T}}_{k|k} + \mathbf{P}^{(6)}_{k|k}.$$

Kalman gain matrix:

$$\tilde{\mathbf{K}}_k = \tilde{\mathbf{P}}_{k|k}\tilde{\mathbf{H}}^{\mathrm{T}}\tilde{\mathbf{R}}^{-1}.$$

If

$$\tilde{\mathbf{K}}_k = \begin{bmatrix} \mathbf{K}^{(1)}_k \\ \mathbf{K}^{(2)}_k \\ \mathbf{K}^{(3)}_k \end{bmatrix},$$

then

$$\mathbf{K}^{(1)}_k = \mathbf{M}^{(1)}_{k|k}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$$
$$\mathbf{K}^{(2)}_k = \mathbf{M}^{(2)}_{k|k}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$$
$$\mathbf{K}^{(3)}_k = \mathbf{M}^{(3)}_{k|k}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}.$$

State estimation update:

$$\tilde{\mathbf{X}}_{k|k} = \tilde{\mathbf{X}}_{k|k-1} + \tilde{\tau}_k + \tilde{\mathbf{K}}_k[\mathbf{y}_k - \tilde{\mathbf{H}}\tilde{\mathbf{X}}_{k|k-1} - \tilde{\mathbf{H}}\tilde{\tau}_k]$$

Given

$$\tilde{\tau}_k = \begin{bmatrix} \tau^{(1)}_k \\ \tau^{(2)}_k \\ \tau^{(3)}_k \end{bmatrix}$$

the state estimation update can be written as

$$\hat{\mathbf{X}}^{(1)}_{k|k} = \hat{\mathbf{X}}^{(1)}_{k|k-1} + \mathbf{K}^{(1)}_k[y_k - \mathbf{H}(\hat{\mathbf{X}}^{(1)}_{k|k-1} + \hat{\mathbf{X}}^{(2)}_{k|k-1} + \hat{\mathbf{X}}^{(3)}_{k|k-1})/3 - \mathbf{H}\tau^{(1)}_k]$$
$$\hat{\mathbf{X}}^{(2)}_{k|k} = \hat{\mathbf{X}}^{(2)}_{k|k-1} + \mathbf{K}^{(2)}_k[y_k - \mathbf{H}(\hat{\mathbf{X}}^{(1)}_{k|k-1} + \hat{\mathbf{X}}^{(2)}_{k|k-1} + \hat{\mathbf{X}}^{(3)}_{k|k-1})/3 - \mathbf{H}\tau^{(2)}_k]$$
$$\hat{\mathbf{X}}^{(3)}_{k|k} = \hat{\mathbf{X}}^{(3)}_{k|k-1} + \mathbf{K}^{(3)}_k[y_k - \mathbf{H}(\hat{\mathbf{X}}^{(1)}_{k|k-1} + \hat{\mathbf{X}}^{(2)}_{k|k-1} + \hat{\mathbf{X}}^{(3)}_{k|k-1})/3 - \mathbf{H}\tau^{(3)}_k].$$

Simplified recursive scheme:

The simplified scheme applies to the processes involving the error covariance estimation and the error covariance update. These are the most time and memory consuming sections of the average algorithm.

Define the following matrices:

$$\mathbf{N} = \mathbf{L} + \mathbf{L}^{2\mathrm{T}} + \mathbf{L}^{3\mathrm{T}},$$
$$\mathbf{Q}^{(1)} = \mathbf{Q} + \mathbf{Q}_1 + \mathbf{Q}_2,$$
$$\mathbf{Q}^{(2)} = \mathbf{Q}^{\mathrm{T}}_1 + \mathbf{Q}_3 + \mathbf{Q}_4,$$
$$\mathbf{Q}^{(3)} = \mathbf{Q}^{\mathrm{T}}_2 + \mathbf{Q}^{\mathrm{T}}_4 + \mathbf{Q}_5.$$

Then the **Error Covariance Estimation** can be written as

$$\mathbf{P}^{(6)}_{k|k-1} = \mathbf{L}^3\mathbf{P}^{(6)}_{k-1|k-1}\mathbf{L}^{3\mathrm{T}} + \mathbf{Q}_5.$$
$$\mathbf{M}^{(1)}_{k|k-1} = \mathbf{L}\mathbf{P}^{(6)}_{k|k-1}\mathbf{N}^{\mathrm{T}} + \mathbf{Q}^{(1)}$$
$$\mathbf{M}^{(2)}_{k|k-1} = \mathbf{L}^2\mathbf{P}^{(6)}_{k|k-1}\mathbf{N}^{\mathrm{T}} + \mathbf{Q}^{(2)}$$
$$\mathbf{M}^{(3)}_{k|k-1} = \mathbf{L}^3\mathbf{P}^{(6)}_{k|k-1}\mathbf{N}^{\mathrm{T}} + \mathbf{Q}^{(3)},$$

and the error covariance update is given by

$$\mathbf{M}^{(1)}_{k|k} = \mathbf{M}^{(1)}_{k|k-1} - \mathbf{M}^{(1)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{\mathrm{T}}_{k|k-1}$$
$$\mathbf{M}^{(2)}_{k|k} = \mathbf{M}^{(2)}_{k|k-1} - \mathbf{M}^{(2)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{\mathrm{T}}_{k|k-1}$$
$$\mathbf{M}^{(3)}_{k|k} = \mathbf{M}^{(3)}_{k|k-1} - \mathbf{M}^{(3)}_{k|k-1}\mathbf{Z}^{-1}_{k|k-1}\mathbf{M}^{\mathrm{T}}_{k|k-1}.$$

REFERENCES

Anderson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering.* Prentice-Hall, 357 pp.

Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography.* Cambridge University Press, 346 pp.

Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods.* 2d ed. Springer-Verlag, 577 pp.

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

——, and E. S. Sarachik, 1977: Forced baroclinic ocean motions, II, The linear equatorial bounded case. *J. Mar. Res.*, **35,** 395–432.

——, and ——, 1981: The response of a linear baroclinic equatorial ocean to periodic forcing. *J. Mar. Res.*, **39,** 651–693.

Chatfield, C., 1994: *The Analysis of Time Series: An Introduction.* 4th ed. Chapman and Hall, 286 pp.

Chui, C. K., and G. Chen, 1990: *Kalman Filtering with Real-Time Applications.* Springer-Verlag, 195 pp.

Cryer, J., 1986: *Time Series Analysis.* Duxury Press, 286 pp.

Daley, R. A., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

——, 1992a: The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120,** 164–177.

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

——, 1993: Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.*, **11,** 634–647.

Dee, D. P., S. E. Cohn, and M. Ghil, 1985: An efficient algorithm for estimating noise covariance in distributed systems. *IEEE Trans. Autom. Control,* **AC-30,** 1057–1065.

Fu, L.-L., I. Fukumori, and R. N. Miller, 1993: Fitting dynamic models to the Geosat sea level observations in the tropical Pacific Ocean. Part II: A linear wind-driven model. *J. Phys. Oceanogr.*, **23,** 2162–2181.

Gelb, A., Ed., 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.

Goodwin, G. C., and K. S. Sin, 1984: *Adaptive Filtering Prediction and Control.* Prentice-Hall, 540 pp.

Hannan, E. J., and M. Deistler, 1988: *Statistical Theory of Linear Systems.* Wiley, 380 pp.

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.

Kawabe, M., 1994: Mechanisms of interannual variations of equatorial sea level associated with El Niño. *J. Phys. Oceanogr.,* **24,** 979–993.

Kelly, K. A., and B. Qiu, 1995: Heat flux estimates for the western North Atlantic. Part I: Assimilation of satellite data into a mixed layer model. *J. Phys. Oceanogr.,* **25,** 2344–2360.

Miller, R. N., 1990: Tropical data assimilation experiments with simulated data: The impact of the tropical ocean and global atmosphere thermal array for the ocean. *J. Geophys. Res.,* **95,** 11 461–11 482.

——, and M. A. Cane, 1989: Kalman filter analysis of sea level height in the tropical Pacific. *J. Phys. Oceanogr.,* **19,** 775–791.

——, 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.

Périgaud, C., and L.-L. Fu, 1990: Indian Ocean sea level variations optimally estimated from Geosat and shallow-water simulations. *Proc. Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Clermont-Ferrand, France, World Meteorological Organization, 510–514.

Stricherz, J. N., J. J. O'Brien, and D. M. Legler, 1992: Atlas of Florida State University tropical Pacific winds for TOGA. Mesoscale Air-Sea Interaction Group Tech. Rep., The Florida State University, Tallahassee, FL, 261 pp.

Wyrtki, K., K. Constantine, B. J. Kilonsky, G. Mitchum, B. Miyamoto, T. Murphy, S. Nakahara, and P. Caldwell, 1988: The Pacific Island Sea Level Network. JIMAR Contribution No. 88-0137, Data Rep. 002, Joint Institute for Marine and Atmospheric Research, University of Hawaii, Honolulu, HI, 71 pp.

Zebiak, S. E., 1989: Oceanic heat content variability and El Niño cycles. *J. Phys. Oceanogr.,* **19,** 475–486.