Fitting Dynamic Models to the Geosat Sea Level Observations in the Tropical Pacific Ocean. Part I: A Free Wave Model

LEE-LUENG FU, JORGE VAZQUEZ AND CLAIRE PERIGAUD

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

(Manuscript received 16 July 1990, in final form 17 December 1990)

ABSTRACT

Free, equatorially trapped sinusoidal wave solutions to a linear model on an equatorial beta plane are used to fit the Geosat altimetric sea level observations in the Tropical Pacific Ocean. The Kalman filter technique is used to estimate the wave amplitude and phase from the data. The estimation is performed at each time step by combining the model forecast with the observation in an optimal fashion utilizing the respective error covariances. The model error covariance is determined such that the performance of the model forecast is optimized. It is found that the dominant observed features can be described qualitatively by basin-scale Kelvin waves and the first meridional mode Rossby waves. Quantitatively, however, only 23% of the signal variance can be accounted for by this simple model.

1. Introduction

In a recent paper Gaspar and Wunsch (1989) demonstrated the use of Kalman filtering and optimal smoothing techniques for fitting a dynamic ocean model to altimetric observations. They applied a model of linear barotropic Rossby waves to the Geosat altimeter sea level observations in the western North Atlantic Ocean. The amplitude and phase of these waves were estimated from the altimetric observations. The observational errors were rigorously treated in the approach such that one could distinguish between noise and signal in the estimation process. The uncertainty in the estimates of the wave parameters were also explicitly stated. Even though the model was too simple to account for the dynamics of the region they studied, a small fraction (6%–15%) of the variance was found to be consistent with the model.

In this study we apply the same technique to fitting a simple dynamic model to the Geosat sea level observations in the equatorial Pacific Ocean. Evidence for equatorial Kelvin waves and Rossby waves in the Geosat data has been documented (e.g., Cheney and Miller 1988; Delcroix et al. 1990; White et al. 1990). It is therefore interesting to examine the effectiveness of a simple model of equatorially trapped waves in describing the observation. The model solution is represented by the summation of a number of free, linear, equatorially trapped waves with selected zonal wavenumbers. This is a highly simplified model for the equatorial ocean variability, without taking account of the nonlinear mechanisms, the wind forcing, or the boundary effects. However, the amplitude and phase of each wave component, which are to be estimated from the observations, are allowed to vary with time to account for the missing physics. A quantitative assessment of the performance of such a simple model serves as the basis for the application of more sophisticated models.

2. Geosat-observed sea level anomalies

The first 30 repeat cycles of the Geosat sea level data were used for the study, spanning a period from November 1986 to March 1988. The ground tracks of the observations in the tropical Pacific Ocean are shown in Fig. 1. Sea level observations are made along these tracks every 17.05 days. The data used in the study are the geographically gridded version produced by the Oceanography Group of the Jet Propulsion Laboratory (Zlotnicki et al. 1990), based on the data distributed by NOAA. The raw altimetric sea level data were corrected for ocean and solid earth tides, wet and dry tropospheric range delays, ionospheric range delays, and atmospheric loading as described in Cheney et al. (1987). The corrected data were first edited by a series of procedures including iterative polynomial fitting to remove blunder points. Then the data were interpolated along track to a fixed geographic grid at approximately 7 km intervals.

At each grid point a mean sea level was calculated from all the available observations. With this mean removed from each sea level observation, the resulting sea level anomalies form the database for the study. If the number of observations at a given grid is less than

Corresponding author address: Dr. Lee Lueng Fu, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109.

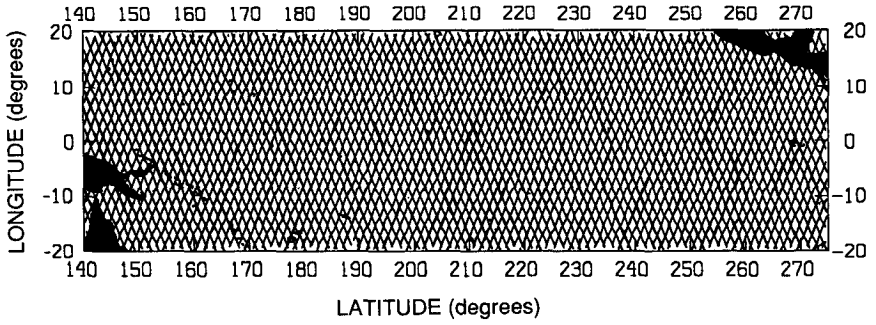

FIG. 1. Geosat ground track coverage in the tropical Pacific Ocean.

21 (70% of the total), then all the data at that grid were discarded to avoid undersampled mean sea level and erroneous sea level anomalies.

The dominant errors in the sea level anomalies are due to the uncertainty in the radial orbit height of the satellite. This orbit error can be characterized by a slowly varying sinusoid with a wavelength of 40 000 km (the circumference of the earth). To reduce this error a first-degree polynomial fit was estimated and removed from each sea level anomaly profile extending from 20°S to 20°N. This procedure has a very small effect on the signals associated with equatorially trapped waves because they are primarily confined within 10 degrees from the equator and their meridional structures are either symmetric or antisymmetric with respect to the equator. A nine-point running averaging process was performed on the corrected sea level anomalies to reduce the measurement noise. The resulting data were subsampled every ninth point to reduce the data volume for subsequent calculations.

For the purpose of producing maps, the sea level anomalies were interpolated to a regular space–time grid using the successive correction method (Bratseth 1986; Tripoli and Krishnamurti 1975). The grid interval is 1° in latitude, 1° in longitude, and 10 days in time. Displayed in Fig. 2 are contour plots of the sea level anomalies as functions of time and longitude at three latitudes: the equator, 4°N and 4°S. At the equator, the positive sea level anomalies that propagated eastward across the Pacific basin during the first 100 days of the observation were described by Cheney and Miller (1988) as a pulse of the first baroclinic-mode equatorial Kelvin waves generated by an anomalous westerly wind event occurring in the western Pacific. These Kelvin waves triggered the 1986–87 El Niño–Southern Oscillation (ENSO) event. Following this positive sea level anomaly is a series of less prominent Kelvin wavelike features. Delcroix et al. (1991) attempted to explain most of these features in terms of observed wind events in the western Pacific.

At 4°N and 4°S, the sea level anomalies show a clear pattern of westward propagation after about day 150. The similarity between these two latitudes suggests that the dominant features are symmetric with respect to the equator. Delcroix et al. (1991) demonstrated that the meridional structure of these westward propagating features were indeed consistent with that of the first meridional mode/first baroclinic mode equatorial Rossby wave, which has a wave function symmetric about the equator with two maxima located at a distance of one radius of deformation from the equator (approximately at 4°N and 4°S; see Fig. 3). The phase speed of the propagation as estimated by them, 1.0 $\pm$ 0.4 m s^{-1}, was also consistent with the Rossby wave theory.

3. The model and Kalman filtering

The model used in this study is based on the linearized equations of motion on an equatorial beta plane (e.g., Gill 1982). The effects of wind forcing and the ocean boundaries are not explicitly considered. Instead, we used free, equatorially trapped sinusoidal wave solutions to fit the observations and allow the amplitude and phase of the waves to vary with time to account for those effects. We used the four kinds of equatorially trapped waves that have the simplest meridional structures: the Kelvin wave, the mixed gravity–Rossby wave, the first and the second meridional-mode Rossby waves. These wave modes include the two symmetrical ones analyzed by Cheney and Miller (1988) and by Delcroix et al. (1991) and two antisymmetrical ones of similar meridional scales. Only the first baroclinic vertical mode was included in the solution because higher vertical modes are more difficult to resolve from the altimetric observations. Displayed in Fig. 3 are the wave functions of these waves. The meridional scales for the wave functions are determined by the Rossby radius of deformation defined as

$$\left(\frac{c}{2\beta}\right)^{1/2},$$

where $\beta = 2\Omega/R$ with R being the Earth's radius and Ω being the Earth's rotation rate, and c is the phase speed of the Kelvin wave. With a typical value of 2.6

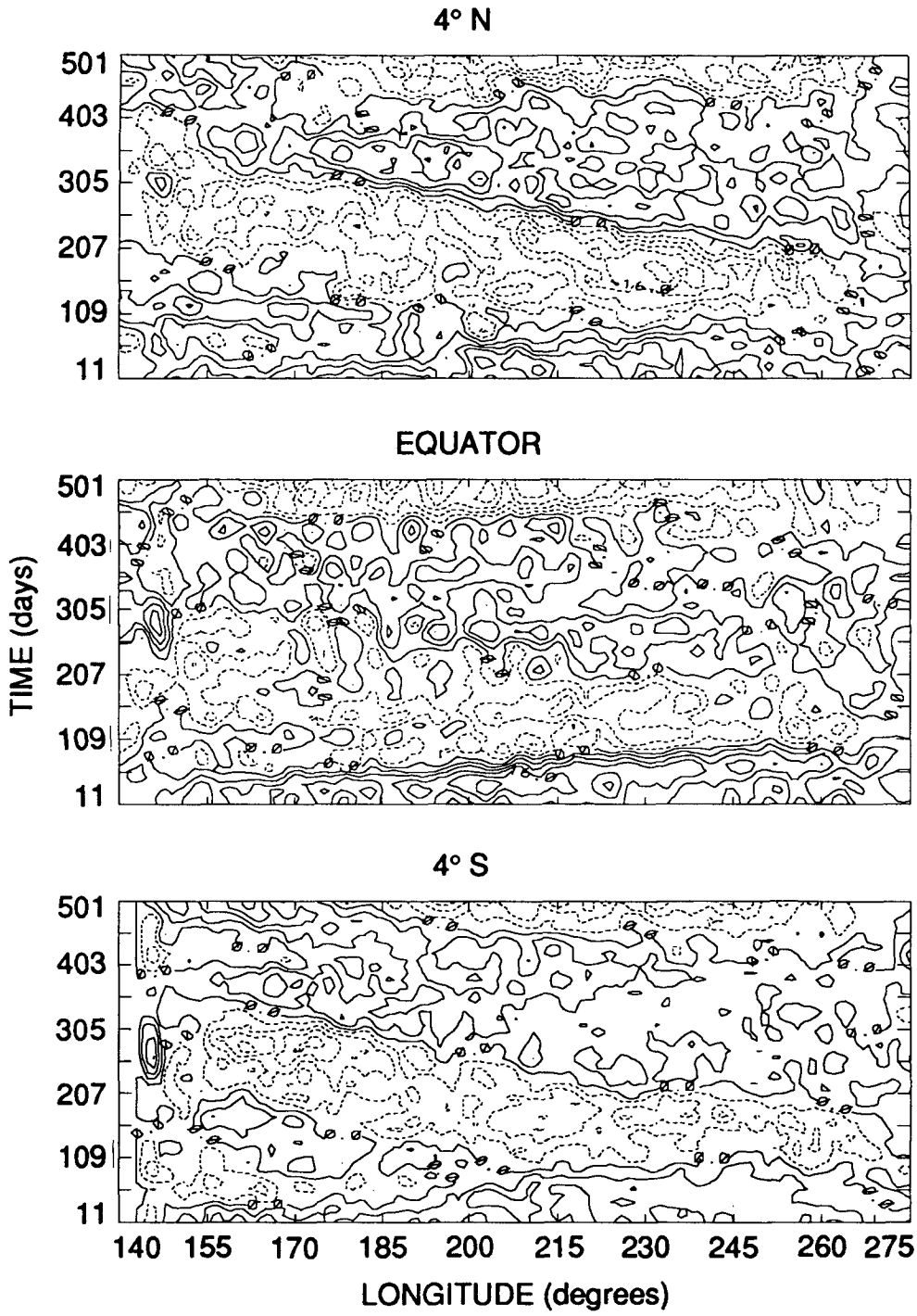

FIG. 2. Contours of the Geosat-observed sea level anomalies at 4°N (upper), the equator (middle), and 4°S (lower). Dashed lines are for negative contours. The contour interval is 4 cm.

m s^{-1} for c in the tropical Pacific (e.g., Wunsch and Gill 1976; Delcroix et al. 1991), the Rossby radius of deformation is 477 km.

The model solution for the sea level, denoted by η, can be expressed as follows:

$$\eta(x, y, t) = \sum_{m=1}^{M} \alpha_m F_m(y) \sin(k_m x - \omega_m t + \theta_m) \quad (1)$$

where $M = 40$; the first ten components are Kelvin waves, the second ten mixed gravity–Rossby waves,

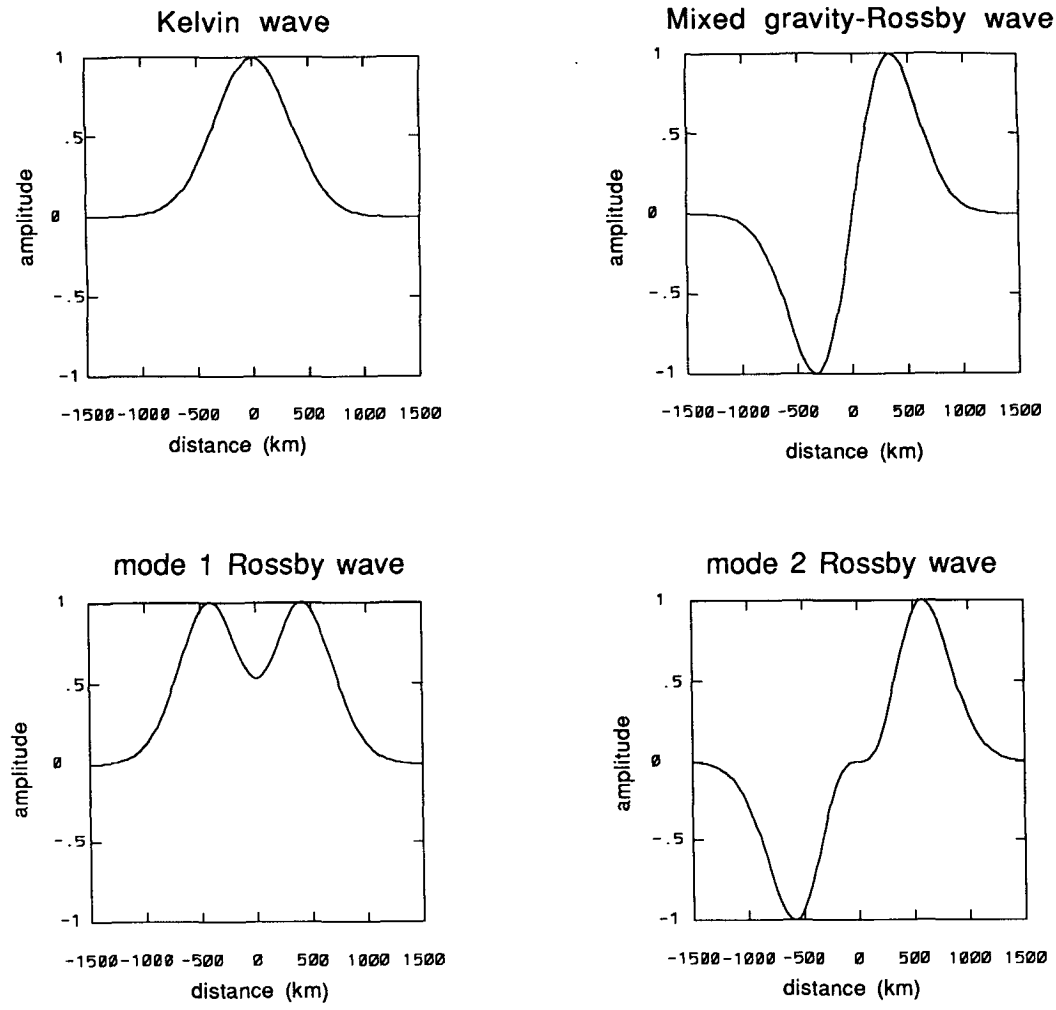

FIG. 3. The wave functions of the equatorially trapped wave modes used in the model.

the third ten first meridional-mode Rossby waves, and the fourth ten the second meridional Rossby waves; $F_m(y)$ is the wave function, α_m the wave amplitude, θ_m the wave phase, k_m the wavenumber, and ω_m the wave frequency. For each kind of wave the ten wavenumbers are chosen to be $2n\pi/L$ with $n = 1, \cdots, 10$ and $L = 15\,000$ km, the size of the data domain. They represent the ten lowest wavenumber Fourier harmonics spanning the domain. The frequency for each wave component is given by the corresponding dispersion relation. Following are the wave functions and dispersion relations for each of the waves (Gill 1982, pp. 436–439).

Kelvin wave:

$$F(y) = \exp\left(\frac{-\beta y^2}{2c}\right)$$

$$\omega = kc$$

Mixed gravity–Rossby wave:

$$F(y) = \left(\frac{\omega}{g}\right) y \exp\left(\frac{-\beta y^2}{2c}\right)$$

$$\frac{\omega}{c} - k - \frac{\beta}{\omega} = 0$$

Rossby wave:

$$F(y) = \frac{c}{g}\left(\frac{\beta c}{2}\right)^{1/2}\left\{(ck-\omega)^{-1}D_{n+1}\left[\left(\frac{2\beta}{c}\right)^{1/2} y\right] + (ck+\omega)^{-1}D_{n-1}\left[\left(\frac{2\beta}{c}\right)^{1/2} y\right]\right\}$$

$$\omega = -\beta k \Big/ \left(k^2 + \frac{(2n+1)\beta}{c}\right)$$

where n is the meridional mode number, and D_{n+1} denotes the parabolic cylinder function of order $n + 1$. The value of c is chosen to be 2.6 m s^{-1}.

Following the formulation of Gaspar and Wunsch

(1989), (1) can be written in the following form with the use of basic trigonometric relations:

$$\eta(x, y, t) = \sum_{m=1}^{M} F_m(y)[q_{2m-1}(t)\cos(k_m x) + q_{2m}(t)\sin(k_m x)] \quad (2)$$

$$q_{2m-1}(t) = \alpha_m \sin(\theta_m - \omega_m t)$$
$$q_{2m}(t) = \alpha_m \cos(\theta_m - \omega_m t). \quad (3)$$

The wave amplitude and phase can be determined from the new variables (q_{2m-1}, q_{2m}). The evolution of (q_{2m-1}, q_{2m}) is given by the following equation:

$$\mathbf{q}(j) = \mathbf{A}(j-1)\mathbf{q}(j-1) + \mathbf{e}(j-1) \quad (4)$$

where j indicates time step, $\mathbf{q}$ is the state vector comprised of all the (q_{2m-1}, q_{2m}), $\mathbf{A}$ is the state transition matrix that transforms $\mathbf{q}$ with time according to (3). The term $\mathbf{e}$ represents the errors of the model due to the missing physics such as nonlinear mechanisms, wind forcing, dissipation, etc. The notation is such that the bold face san serif letters represent matrices and the bold face lower case letters represent vectors.

The altimetric observations of sea level along a given pass, denoted by h, can be written

$$h = \eta + r \quad (5)$$

where r represents the measurement errors including any sea level variations that are not described by the equatorial wave dynamics such as tides. In matrix form (5) can be expressed as

$$\mathbf{h}(j) = \mathbf{C}(j)\mathbf{q}(j) + \mathbf{r}(j) \quad (6)$$

where $\mathbf{h}$ and $\mathbf{r}$ are vectors comprised of the observations and errors at time step j, respectively; $\mathbf{C}$ is the following matrix that relates the state vector to the observations:

$$\begin{matrix} F_1(y_1)\cos k_1 x_1 & F_1(y_1)\sin k_1 x_1 & F_M(y_1)\cos k_M x_1 & F_M(y_1)\sin k_M x_1 \\ \vdots & & \vdots & \\ F_1(y_i)\cos k_1 x_i & F_1(y_i)\sin k_1 x_i \ldots & F_M(y_i)\cos k_M x_i & F_M(y_i)\sin k_M x_i \\ \vdots & & \vdots & \\ F_1(y_n)\cos k_1 x_n & F_1(y_n)\sin k_1 x_n & F_M(y_n)\cos k_M x_n & F_M(y_n)\sin k_M x_n \end{matrix}$$

where x_i, y_i $(i = 1, \cdots, n)$ are the grid locations along the altimeter pass.

Given the dynamic model equation (4) and the observation equation (6), an optimal, sequential estimation of the state vector $\mathbf{q}$ at each time step can be achieved by the technique of Kalman filtering (e.g., Sorenson 1985; Ghil et al. 1981; Gaspar and Wunsch 1989). At each time step, the Kalman filter combines the model forecast (based on the best estimate at the preceding step) with observation to form an optimal estimate of the state vector, which is then used by the model to make a forecast for the next time step. The process then goes on. It begins with an initial guess for $\mathbf{q}$ and its error covariance, denoted by $\mathbf{q}_a(0)$ and $\mathbf{P}_a(0)$, respectively. At a given time step j, the model forecast for the state vector, $\mathbf{q}_f(j)$, is given by the dynamic model,

$$\mathbf{q}_f(j) = \mathbf{A}(j-1)\mathbf{q}_a(j-1) \quad (7)$$

where $\mathbf{q}_a(j-1)$ is the optimal estimate made at time step $j-1$. The error covariance of $\mathbf{q}_f$, denoted by $\mathbf{P}_f$, is given by

$$\mathbf{P}_f(j) = \mathbf{A}(j-1)\mathbf{P}_a(j-1)\mathbf{A}^{\mathrm{T}}(j-1) + \mathbf{\Gamma}(j-1) \quad (8)$$

where $\mathbf{A}^{\mathrm{T}}$ is the transpose of $\mathbf{A}$, and $\mathbf{\Gamma}$ is the model error covariance matrix with elements $\Gamma_{ik} = \langle e_i e_k \rangle$, where $\langle\ \rangle$ represents the expected value. The model forecast is then combined with the observation made at time step j to form a new optimal estimate of $\mathbf{q}$ as follows (a data update operation)

$$\mathbf{q}_a(j) = \mathbf{q}_f(j) + \mathbf{G}(j)[\mathbf{h}(j) - \mathbf{C}(j)\mathbf{q}_f(j)] \quad (9)$$

where $\mathbf{G}(j)$ is the Kalman–Bucy gain matrix defined by

$$\mathbf{G}(j) = \mathbf{P}_f(j)\mathbf{C}^{\mathrm{T}}(j)[\mathbf{C}(j)\mathbf{P}_f(j)\mathbf{C}^{\mathrm{T}}(j) + \mathbf{R}(j)]^{-1} \quad (10)$$

where $\mathbf{R}$ is the observation error covariance matrix with elements $R_{ik} = \langle r_i r_k \rangle$. The matrix $\mathbf{G}$ determines the contribution of the new observation in modifying the model forecast, based on the error estimates of the model and the observation. The error covariance matrix for $\mathbf{q}_a(j)$ is given by

$$\mathbf{P}_a(j) = [\mathbf{I} - \mathbf{G}(j)\mathbf{C}(j)]\mathbf{P}_f(j) \quad (11)$$

where $\mathbf{I}$ is the identity matrix. The new estimates $\mathbf{q}_a(j)$ and $\mathbf{P}_a(j)$ are used by (7) and (8) to make forecasts for the next time step $j+1$.

The result of the Kalman filtering process is critically dependent on the $\mathbf{G}$ matrix that, in turn, is determined by the error covariance of the model forecast and the observation. The estimation of the observation error covariance is discussed in the next section.

4. The observation errors

The processed along-track sea level anomalies discussed in section 2 are used as the observation for the Kalman filtering process. The model is updated by the observation at the times when an altimeter pass traverses the region of study. All the observations made within one pass are treated as being made simultaneously and are used to update the model forecast as synoptic observations. Only the data between 10°N and 10°S are used in the calculation. It takes less than six minutes for the satellite to cover 20 degrees of latitude, so the assumption of simultaneity is justified. The observation error covariance matrix **R** can then be constructed from the one-dimensional along-track autocovariance functions of the observation errors.

There are many error sources in the altimetric measurement system (e.g., Tapley et al. 1982; Fu et al. 1988). Only the following four are considered in this study: orbit, ocean tide, the wet tropospheric range delay (due to water vapor), and the altimeter measurement noise. In the tropics, the first three are the dominant errors. Because there are no known correlations among these errors, the total error autocovariance function can be expressed as the sum of the individual autocovariance functions.

As noted earlier the orbit error can be approximated by a sinusoid with a wavelength of 40 000 km. For the Geosat Exact Repeat Mission data the amplitude for the sinusoid is about 220 cm (Zlotnicki et al. 1990). A large simulated database for the orbit error was generated by repeated sampling of such a sinusoid with a random phase over a distance of 4 800 km, the along-track distance from 20°S to 20°N. Recall that the orbit error has already been reduced by removing a first-order polynomial fit from the altimeter data. Such a polynomial fit was also removed from the simulated orbit error, leaving only the part that is relevant for the residual orbit error. Using this simulated dataset, we obtained an estimate for the autocovariance function of the residual orbit error as shown in Fig. 4a. The variance at zero lag, 9.1 cm^2, represents the energy of the residual orbit error. This value is consistent with the result of Tai (1989) who has evalated the performance of a wide range of orbit correction methods.

The wet tropospheric range delay in the Geosat data was corrected by using the water vapor product of the atmospheric model of the Navy's Fleet Numerical Oceanography Center (FNOC). The residual error due to tropospheric water vapor after the correction was studied by Fu and Zlotnicki (1989) using the differences between the FNOC products with coincident water vapor observations made by a microwave radiometer. In that case the microwave radiometer was the Scanning Multichannel Microwave Radiometer (SMMR) aboard Seasat. The autocovariance function, derived from the wavenumber spectrum estimated by Fu and Zlotnicki (1989), of this error is displayed in Fig. 4a. The variance of the error was estimated to be 11.7 cm^2 (the value at the zero lag). The zero crossing occurs at a lag of 400 km, reflecting the mesoscale variabilities in the tropospheric water vapor content. The negative values at lags greater than 400 km are due to large-scale variabilities.

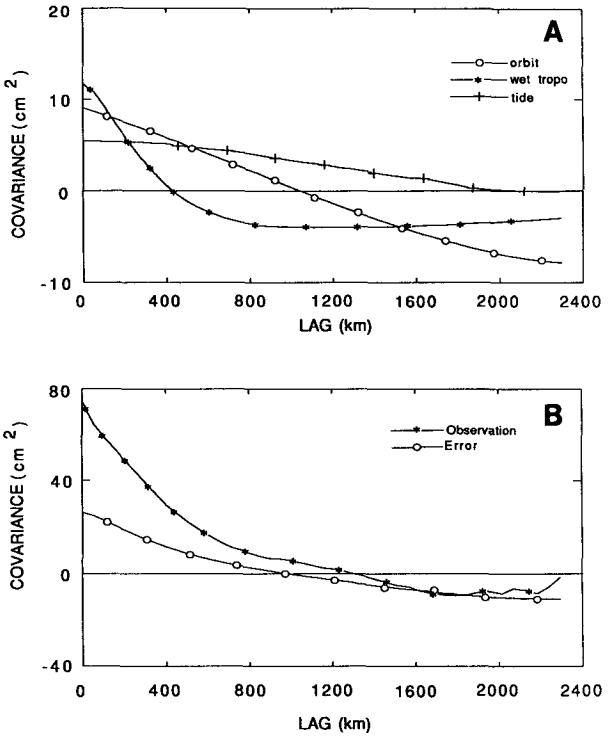

FIG. 4. The covariance function of (a) the various leading altimetry errors, and (b) the Geosat-observed sea level anomalies and the sum of the leading altimetry errors. Note the difference in the vertical scale between the two panels.

The ocean tides in the Geosat data were corrected by using the model of Schwiderski (1980). The global root-mean-square accuracy of this model was estimated to be 10 cm. To estimate the autocovariance function for the residual tidal errors, we assume that the spatial characteristics of the errors are the same as the tide itself. The tidal errors were simulated by sampling the tide model product along the Geosat tracks with the rms variability scaled down to 10 cm. A first-degree polynomial fit to the data was removed to account for the orbit error reduction procedure applied to the altimeter data. The autocovariance function of the simulated residual tidal error is also shown in Fig. 4a. The variance of the error is only 5.5 cm^2, indicating that the tidal error has been effectively reduced by the removal of the first-degree polynomial fit.

The rms noise level of the Geosat altimeter sea-level observation is about 3.5 cm (Fu and Zlotnicki 1989). The nine-point running averaging applied to the data reduces the noise level to 1.2 cm. The autocovariance function for the noise is thus a delta function with a

value of 1.4 cm^2 at the origin and zero elsewhere. By summing the autocovariance functions of the individual errors, we obtained the total error covariance function that is shown in Fig. 4b.

Also shown in Fig. 4b is the autocovariance function of the altimeter sea level observation. The difference between the two autocovariance functions is that of the signal we are trying to interpret with the equatorial wave model. The difference at the zero lag, about 47 cm^2, is an estimate of the variance of the signal, indicating an rms signal level of 7 cm. The decorrelation scale for the signal is about 1600 km.

A number of altimetric measurement errors have not been included in the foregoing error analysis. These include the dry tropospheric range delay, the atmospheric loading effects, the ionospheric range delay, and the sea-state bias. The dry tropospheric range delay was corrected by using the FNOC sea level pressure data. Near the equator, the rms sea-level pressure variability is only 2–3 millibars, yielding an error of only about 0.5 cm in path delay if no correction was made at all. The error induced by the atmospheric loading effects was corrected by resorting to the conventional inverted barometer correction using the FNOC sea level pressure data. Although the validity of this correction has been in doubt, the residual error after the correction is probably of the same magnitude as if no correction had been made, i.e., 2–3 cm. The ionospheric range delay was corrected by using a GPS climatology model. Because the Geosat data used in the study were collected during a period of minimum solar activity, the ionospheric correction rarely exceeds 5 cm (Cheney et al. 1988). The sea-state bias in the Geosat data has been estimated to be 2% of the significant wave height (Cheney et al. 1988). The average significant wave height in the equatorial area is on the order of 2 m, resulting in a sea-state bias of 4 cm.

All of the four missing error sources have dominant spatial scales greater than the width of the equatorial waveguide. The procedure of the removal of a first-degree polynomial fit to reduce orbit error has also greatly reduced the effects of these errors. Their omission from the error analysis is not believed to have significant effects on the results.

5. The model error

The model error is more difficult to estimate than the observation error; the latter is dictated by the characteristics of the measurement, while the former by the missing physics which are difficult to quantify. Without attempting to analyze the effects of the missing physics on the behavior of the wave solutions, we made the same assumption as did Gaspar and Wunsch (1989): The errors in all the wave components are of the same magnitude and uncorrelated with one another, so that the $\mathbf{\Gamma}$ matrix is of the following form:

$$\mathbf{\Gamma} = \sigma_f \mathbf{I}$$

where the model error variance σ_f (known as the process noise variance in the Kalman filtering literature) is yet to be determined.

The criterion used to determine σ_f is the model performance. We seek the value for σ_f that maximizes the model's capability to forecast the observation. If we choose too small a value for σ_f, then we put too little weight on the data in the data update process represented by (9). It then takes a long time for the filtering process to adjust the amplitude and phase of each wave component to fit the data. Consequently, the model has little capability to forecast the observation at each time step. If we choose too large a value for σ_f, then we put too much weight on the data in the data update process. The amplitude and phase of the wave components are adjusted too quickly so as to fit the data. The model then propagates more errors than signals and begins to lose its forecast capability. The optimal value for σ_f is found to be 0.003 cm^2 (see section 6 for details). Gaspar and Wunsch (1989) found that the optimal model error variance for their calculation was 0.01 cm^2.

6. Results

a. The solution with 40 waves

Having determined both observation and model covariances, we applied Eqs. (7)–(11) to the altimeter data, pass by pass, to estimate sequentially the amplitude and phase of the wave components that comprise the state vector. The initial estimate for the wave amplitude was taken to be zero for all of the wave components, i.e., $\mathbf{q}_a(0) = 0$. The error covariance of this initial estimate was assumed to be of the form, $\mathbf{P}_a(0) = p_0\mathbf{I}$, with $p_0 = 50$ cm^2. The value for p_0 was determined based on the estimated signal variance in the altimeter data (see section 4).

It was found that among the 40 wave components, only 4 of them had significant amplitudes: the two longest Kelvin wave components and the two longest first-mode Rossby wave components. The amplitudes and phases of these waves are shown in Fig. 5. It is noted that the model solution is dominated by the longest Kelvin wave and the first-mode Rossby wave (Figs. 5a and 5c), of which both had amplitudes greater than the noise level most of the time. These two wave components are clearly the response of the model to the main features in the data as shown in Fig. 2. The large Kelvin wave amplitudes at day 50–100 correspond to the 1986–87 ENSO event; the subsequent shorter pulses are related to the other less prominent Kelvin wavelike features in the data at the equator. The Rossby wave response at day 150–300 corresponds to the observed feature of westward propagation in the data at 4°N and 4°S.

The variability of the phase is an indicator of the validity of the wave model. A slowly varying phase

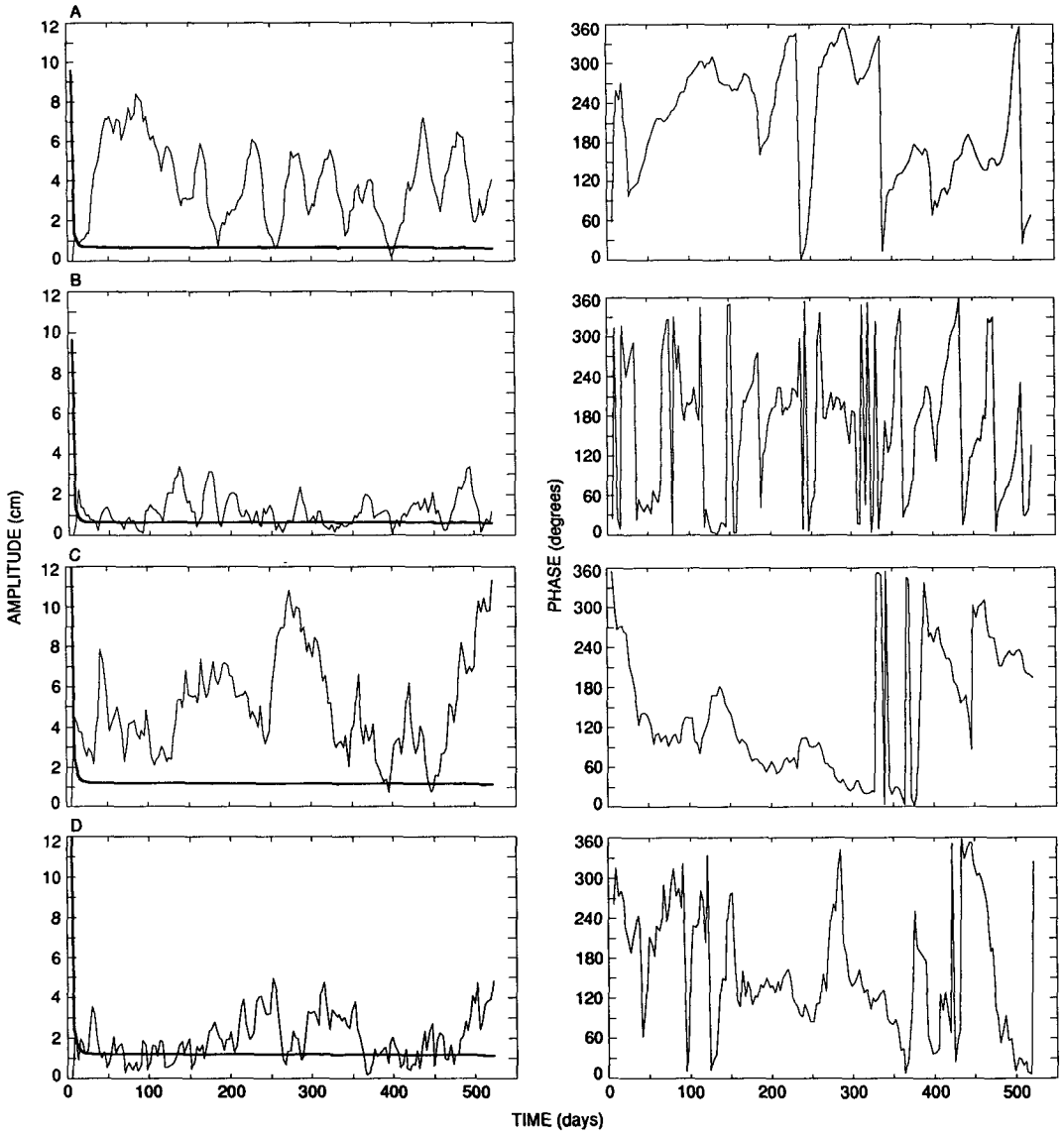

FIG. 5. (a) The thin lines represent the amplitude (left panel) and phase (right panel) for the Kelvin wave with a wavelength of 15 000 km. The thick line shows the estimation error for the wave amplitude. (b) Same as (a) except for the Kelvin wave with a wavelength of 7500 km. (c) Same as (a) except for the Rossby wave with a wavelength of 15 000 km. (d) Same as (a) except for the Rossby wave with a wavelength of 7500 km.

indicates that the observation can be described by waves locally, whereas a rapidly varying phase indicates that the wave model is probably inadequate for the data. The variability of phase generally increases with decreasing wave amplitude. It is noted that the two shorter wave components (Figs. 5b and 5d) have smaller amplitudes and noisier phases. However, the phase tends to be more stable when the amplitude is above the noise level.

The error of the state estimate (the root-mean-square of the diagonal elements of $\mathbf{P}_a$) is dependent on the model forecast error and the data error. At the beginning of the process, this error is dominated by the large error of the initial state; the model forecast thus has very little weight in competing with the data update in the estimation process. When the data are progressively assimilated into the process, the error variance decreases roughly as $1/N$, where N is the number of independent observations. This explains the rapid initial drop of the error estimate shown in Fig. 5. As the error for the state estimate goes down, the model forecast begins to gain more weight in the estimation process. Eventually the error reduction due to the data update is balanced by the model forecast error and the state

error estimate becomes constant in time as shown in Fig. 5 (see section 7 for more discussion on the error characteristics).

To analyze the performance of this model solution, we followed the procedure of Gaspar and Wunsch (1989). At a given time step, we define the following parameters:

$$E_t = \mathbf{h}^T\mathbf{h}$$

$$\epsilon_f = (\mathbf{h} - \mathbf{C}\mathbf{q}_f)^T(\mathbf{h} - \mathbf{C}\mathbf{q}_f)$$

$$\epsilon_a = (\mathbf{h} - \mathbf{C}\mathbf{q}_a)^T(\mathbf{h} - \mathbf{C}\mathbf{q}_a)$$

where E_t is the variance of the data, ϵ_f is the portion of E_t unexplained by the model forecast, ϵ_a is the portion of E_t unexplained by the optimal state estimate made by the model data combination through the Kalman filter. The parameters E_a and E_f, defined as $E_a = E_t - \epsilon_a$ and $E_f = E_t - \epsilon_f$, then represent the variances explained by the optimal state estimate and by the model forecast alone, respectively. The variance explained by the optimal state estimate can be split into two parts: E_f and E_d (i.e., $E_a = E_f + E_d$), where E_d is the variance explained by the data update. As shown in Fig. 6, at the initial stage the model does not have any skill because of the large initial state error. The model forecast is thus uncorrelated with the observation, leading to negative values for E_f. Even after more data have been assimilated into the process, the model performance (as measured by E_f) is generally poor, with E_f being negative quite often. When E_f is positive, it accounts for only a small portion of the signal variance, which is estimated to be 47 cm^2 (see section 4). Although the variance explained by the optimal state estimate, E_a, accounts for a significant portion of the signal variance, it is primarily due to the data update (E_d). The cause for the poor model performance lies in the fact that most of the 40 wave components have amplitudes that are not significantly above the noise level; these wave components propagate more noise

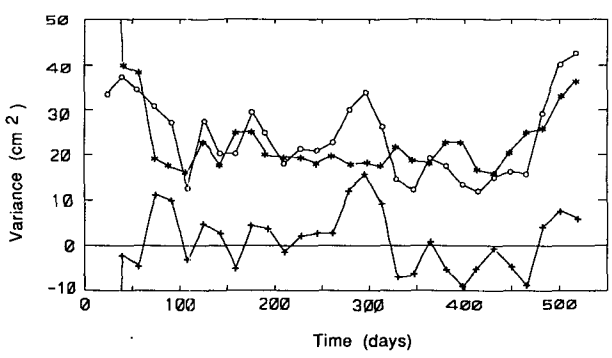

FIG. 6. Variances accounted for by the 40-wave model: circles—the variance explained by the optimal state estimate (E_a); pulses—the variance explained by the model forecast (E_f); asterisks—the variance explained by the data update (E_d). The data plotted are 17-day averaged values.

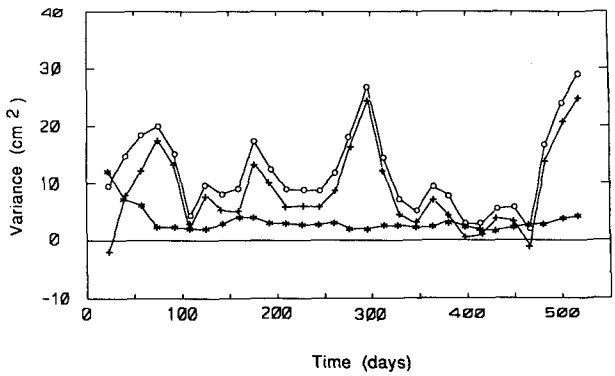

FIG. 7. Same as Fig. 6 except for the four-wave model.

than signal. This phenomenon was also noted by Gaspar and Wunsch (1989) in their study.

b. The solution with only 4 waves

To exclude those undesirable wave components, we have repeated the calculation using only the four dominant waves as identified earlier. The amplitude and phase of these four waves resulting from the new calculation are indistinguishable from what is shown in Fig. 5, indicating that the estimates for the various wave components are independent of each other. By retaining only the four dominant waves, the model performance is significantly improved (Fig. 7). The value of E_f is positive almost all the time with an average value of 8.5 cm^2, which represents 18% of the signal variance. The variance explained by the optimal state estimate is smaller than that in the case of the 40 wave model, however. This is not surprising because a better fit to the data can be achieved with the additional 36 adjustable wave components. The model forecast, instead of the data update, is now making the primary contribution to E_a.

As mentioned in section 5, the model performance is dependent on the specification of the model error variance σ_f. The choice for σ_f is determined such that the model performance is maximized. To search for the optimal σ_f, we have calculated the time-averaged E_f, denoted by $\bar{E}_f$, using different values for σ_f. The result is displayed in Fig. 8. As noted earlier the optimal value for σ_f is 0.003 cm^2. The dependence of the model performance on σ_f has been interpreted in section 5. We have essentially utilized the observations to quantify the model error with a single parameter σ_f. Note that σ_f represents the error variance in the model forecast for one time step, which is the interval between two consecutive altimeter passes. This interval is normally only a few hours. This is why σ_f is so small. See section 7 for the discussion on the accumulation of σ_f over time and its eventual balance with the effects of data update.

We have also investigated the effects of varying the Kelvin wave phase speed, c, on the results discussed

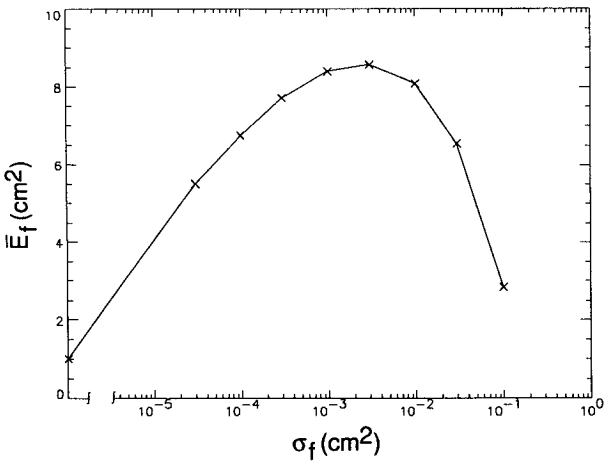

FIG. 8. The time averaged E_f, denoted by $\bar{E}_f$, as a function σ_f.

above. It is found that the results are insensitive to the choice of c within the range of 2.2–3.0 m s^{-1}, which encompasses most of the reported values for c.

Using the four-wave solution we have constructed the sea level anomaly field with the use of (1). The time–longitude presentations of the results at the equator and 4°N are displayed in Fig. 9 (to be compared with Fig. 2). Due to the symmetry of the solution with respect to the equator, the result at 4°S is identical to that at 4°N and hence is not shown. At the equator, the 1986–87 ENSO event is reproduced by the model with a reduced magnitude. The model, in general, has qualitatively reproduced the main features in the data. To quantify the comparison of the model with the observation, a correlation coefficient for the two fields was computed. At the equator, the correlation coefficient is 0.32, which is significantly different from zero at the 95% confidence level. At 4°N the model seems to have done a better job in reproducing the observation with a correlation coefficient of 0.62. The reason for the higher correlation at 4°N is that the model has successfully reproduced the main westward propagating feature in the middle of the observation period. This feature accounts for a major portion of the observation variance. The model's ability to reproduce the observed zonal variabilities is severely limited by the limited number of wave components. The pronounced alternating highs and lows in the model results clearly reflect the dominant zonal wavelength of 15 000 km.

7. Discussions

The total variance of the Geosat sea level observation in the study area is 75 cm^2. The error analysis discussed in section 4 suggests that the altimeter measurement errors account for 28 cm^2, leaving 47 cm^2 to be accounted for by the oceanographic signals. By using a Kalman filter approach, a simple model containing four basin-scale equatorially trapped waves (two Kelvin waves and two Rossby waves) has been used to describe the observations. Using more waves actually degrades the model performance because the additional waves propagate more noise than signal. It is found that the model performance is dependent on the specification of the model error variance, σ_f, in the estimation process. The optimal value for σ_f is found to be 0.003 cm^2. With this σ_f, the one-step model forecast (i.e., E_f) can account for a variance of 8.5 cm^2 on average (18% of the signal variance). With data update, the optimal state estimate (i.e., E_a) can account for 11 cm^2 on average (23% of the signal variance).

The model error variance σ_f plays an important role in the estimation process. This parameter is determined so as to allow the model to propagate the maximum amount of information. It parameterizes the effects of the missing physics of the model, allowing the amplitude and phase of the waves to vary with time to achieve maximum forecast capability. This parameter, in conjunction with the data error variance σ_d, also determines the error variance of the state estimate, denoted by δ. As noted earlier, δ initially decreases rapidly when the data are progressively included in correcting the large initial state error. At this stage δ behaves like σ_d/N, where N is the number of independent measurements. Because the dominant measurement errors are correlated over the distance of an altimeter pass, N is essentially the number of passes that have been assimilated into the process. In the meantime the forecast error grows like $N\sigma_f$ due to the accumulation of model errors as the Kalman filter process proceeds. Eventually

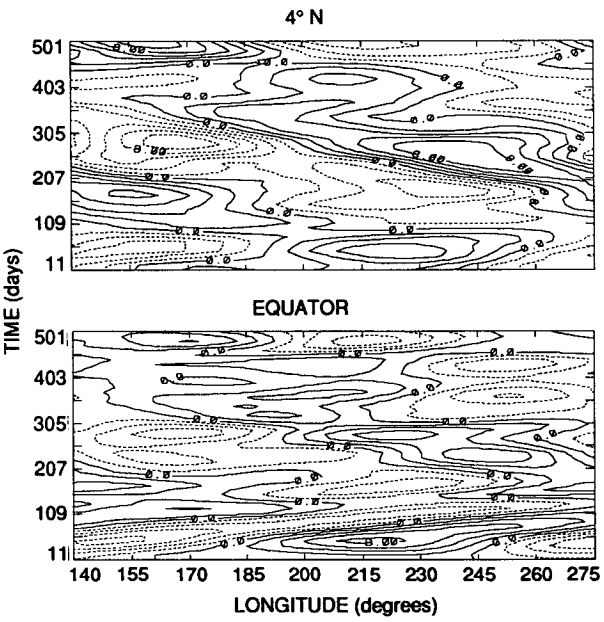

FIG. 9. Time–longitude plots of the sea level anomalies represented by the four-wave model solutions at 4°N (upper) and the equator (lower).

when the decrease in the state estimation error due to data update is balanced by the increase in the forecast error, the state estimation error variance then reaches a steady value δ_s. The number of time steps taken to reach this stage, denoted by N_s, can be roughly estimated from the following relation (assuming that $\sigma_f \ll \sigma_d$; see Appendix):

$$\frac{\sigma_d}{N_s} = N_s \sigma_f \tag{12}$$

from which we obtain $N_s = (\sigma_d/\sigma_f)^{1/2}$ and $\delta_s = (\sigma_d\sigma_f)^{1/2}$. Taking $\sigma_d = 28$ cm^2 and $\sigma_f = 0.003$ cm^2, we obtain $\delta_s = 0.29$ cm^2, and $N_s = 97$ (or about ten days). The root-mean-square error for the wave amplitude is then given by $(\delta_s)^{1/2}$, which yields 0.54 cm. The order of magnitude of this error estimate, which is derived from a rather crude, heuristic argument, is in agreement with the result of the Kalman filter calculation shown in Fig. 5. Note also from Fig. 5 that it takes about ten days to reach the steady error estimate, consistent with the estimate for N_s.

Another implication of (12) is that the data that are older than N_s time steps (ten days in this case) have very little weight in influencing the current state estimate because the associated forecast error can no longer be compensated by the information gained from new data. Even though the Kalman filter in theory carries all the prior information in estimating the current state, only the most recent data are useful when the model is defective.

The Kalman filter uses all prior information to make a state estimate. If the purpose is to make a forecast, then one has used the maximum amount of information available. But if one wants to make an optimal estimate of the state in a hindcast mode, then the posterior information should also be used. The optimal estimation approach to accomplishing this goal is known as optimal smoothing (e.g., Gelb 1974; Bennett and Budgell 1989; Gaspar and Wunsch 1989). A variety of algorithms exist for implementing the approach. The Rauch–Tung–Striebel (1965) algorithm allows one to start with the final state estimate made by the Kalman filter and run the smoothing backward in time to obtain the optimal state estimate using all the data available. We have applied this algorithm to the four-wave model. Based on the same reasoning given in the preceding paragraph, only the data within about ten days after the point of estimation provide new information to improve the estimate. Thus, the smoothing has only a local (in time) effect and has not changed the previous solutions in any significant manner. In the interior of the observation period, the error variance of the smoothed estimate is about 50% of that of the unsmoothed estimate because the amount of information available for each estimate is doubled. However, the error variance at the beginning and the end of the observation period is twice that in the interior due to the missing half of the information. This finding is in agreement with that of Gaspar and Wunsch (1989).

The model performance is ultimately limited by the overly simplified physics of the model. Nevertheless, it is interesting to note that most of the prominent features in the data have been qualitatively described by the four-wave model solution. One might be able to improve the model performance by further tuning the model error covariance matrix, for example, by assigning different weights to the various wave components. However, we feel that such efforts may not be worthwhile considering the simplicity of the model. The more interesting approach is to use more sophisticated models. Our objective is to use the Geosat data to progressively test models of increased complexity in describing the observed sea level anomalies. An attempt at applying the model of Miller and Cane (1989) to a Kalman filter analysis of the Geosat data in the tropical Pacific is underway and the results will be reported in Part 2 of this paper.

Acknowledgments. Discussions with Professor Carl Wunsch of MIT and Dr. Ichiro Fukumori are gratefully acknowledged. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Partial support from the TOPEX/POSEIDON Project funded under the NASA TOPEX/POSEIDON Announcement of Opportunity is also acknowledged.

APPENDIX

Temporal Evolution of the Error Variance

We hereby provide a derivation for (12) using an extremely simplified case. We consider a one-dimensional linear model in which both the state transition matrix **A** and the observation matrix **C** are reduced to unity. At time step k, the error variance of the state estimate, denoted by δ_k, is given by (e.g., Liebelt 1967)

$$\delta_k^{-1} = (\delta_{k-1} + \sigma_f)^{-1} + \sigma_d^{-1}. \tag{A1}$$

Note that (A1) can be shown to be equivalent to (11). Assuming that δ_k reaches a steady value, δ_s, as k becomes very large, then one obtains the following solution for δ_s by substituting δ_s for δ_k and δ_{k-1} in (A1):

$$\delta_s = (\sigma_d\sigma_f)^{1/2}\left[\left(1 + \frac{\epsilon}{4}\right)^{1/2} - \frac{\epsilon}{2}^{1/2}\right]$$

where $\epsilon = \sigma_f/\sigma_d$. If $\epsilon \ll 1$ as in the present case, then

$$\delta_s \approx (\sigma_d\sigma_f)^{1/2}, \tag{A2}$$

as obtained in section 7. Because $\epsilon \ll 1$, one could infer from (A2) that $\delta_s \gg \sigma_f$, which implies that $\delta_k \gg \sigma_f$. Therefore (A1) can be written

$$\delta_k^{-1} \approx \delta_{k-1}^{-1} + \delta_d^{-1}$$

which leads to

$$\delta_k \approx \frac{\sigma_d}{k}.$$

Therefore, N_s, the number of time steps taken to reach the steady state, is given by

$$N_s \approx \frac{\sigma_d}{\delta_s} \approx \left(\frac{\sigma_d}{\sigma_f}\right)^{1/2}. \qquad \text{(A3)}$$

Note that (A3) provides only an order-of-magnitude estimate instead of an accurate estimate, for N_s.

REFERENCES

Bennett, A. F., and W. P. Budgell, 1989: The Kalman smoother for a linear quasigeostrophic model of ocean circulation, *Dyn. Atmos. Oceans,* **13,** 219–267.

Bratseth, A. M., 1986: Statistical interpretation by means of successive corrections, *Tellus,* **38,** 439–447.

Cheney, R. E., and L. Miller, 1988: Mapping the 1986–1987 El Niño with GEOSAT altimeter data, *Eos Trans., Amer. Geophys. Union,* **69,** 754–755.

——, B. C. Douglas, R. W. Agreen, L. L. Miller, D. L. Porter, 1987: GEOSAT altimeter geophysical data record user handbook. NOAA Tech. Memo. NOS-NGS-46, NOAA/National Ocean Service, Rockville, MD.

Delcroix, T., J. Picaut and G. Eldin, 1991: Equatorial Kelvin and Rossby waves evidenced in the Pacific Ocean through GEOSAT sea level and surface current anomalies. *J. Geophys. Res.,* **96** (suppl.) 3249–3262.

Fu, L.-L., and V. Zlotnicki, 1989: Observing oceanic mesoscale eddies from Geosat altimetry: preliminary results. *Geophys. Res. Lett., 16,* 457–460.

——, D. B. Chelton, and V. Zlotnicki, 1988: Satellite altimetry: observing ocean variability from space, *Oceanogr. Mag.,* **2,** 4–11.

Gaspar, P., and C. Wunsch, 1989: Estimates from altimeter data of barotropic Rossby waves in the northwestern Atlantic Ocean. *J. Phys. Oceanogr.,* **19,** 1821–1844.

Gelb, A., Ed., 1974: *Applied Optimal Estimation,* MIT Press, 374 pp.

Ghil, M., S. Cohn, J. Tavantzis, K. P. Buke and E. Issacson, 1981: Applications of estimation theory to numerical weather prediction, in *Dynamic Meteorology: Data Assimilation Methods,* L. Bengtsson, M. Ghil and E. Kallen, Eds., Spring-Verlag, 330 pp.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics,* Academic Press, 662 pp.

Liebelt, P. B., 1967: *An Introduction to Optimal Estimation,* Addison-Wesley, 273 pp.

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.

Rauch, H. E., F. Tung and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AIAA J.,* **3,** 1445–1450 (reprinted in Sorenson 1985).

Schwiderski, E. W., 1980: Global ocean tides, 1: A detailed hydrodynamical interpolation mode, *Mar. Geod.* **3,** 161–217.

Sorenson, H. W. (Editor), 1985: *Kalman Filtering: Theory and Application.* IEEE Press, 457 pp.

Tai, C.-K., 1989: Accuracy assessment of widely used orbit error approximations in satellite altimetry, *J. Atmos. Oceanic Technol.,* **6,** 147–150.

Tapley, B. D., G. H. Born and M. E. Parke, 1982: The SEASAT altimeter data and its accuracy assessment, *J. Geophys. Res.,* **87,** 3179–3188.

Tripoli, G. J., and T. N. Krishnamurti, 1975: Low-level flows over the GATE area during summer 1972, *Mon. Wea. Rev.,* **103,** 197–209.

White, W. B., N. Graham and C.-K. Tai, 1990: Reflection of annual Rossby waves at the maritime western boundary of the Tropical Pacific, *J. Geophys. Res.,* **95,** 3101–3116.

Zlotnicki, V., A. Hayashi and L.-L. Fu, 1990: The JPL-Oceans-8902 Version of Geosat Altimetry Data, JPL Tech. Rep. D-6939, Jet Propulsion Laboratory, Pasadena, (internal document).