

Estimates from Altimeter Data of Barotropic Rossby Waves in the Northwestern Atlantic Ocean

PHILIPPE GASPAR* AND CARL WUNSCH

*Center for Meteorology and Physical Oceanography, Department of Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Manuscript received 28 February 1989, in final form 5 July 1989)

ABSTRACT

A dynamical model representing linear barotropic Rossby waves is combined with GEOSAT data from the northwest Atlantic Ocean. The model is too simple to be very realistic in this complex area, but the problem of combining any model with real data raises a number of issues that we address. The combination method used is sequential estimation in the form of a filtering-smoothing operation. A fraction of the total signal variance (5% to 15%) in the area over several spacecraft repeat cycles is demonstrated to be consistent with five Rossby waves. That the result is robust is demonstrated by recovering known signals added synthetically to the real data.

1. Introduction

Data from altimetric satellites produces a space/time coverage of the ocean which is unachievable by any other known instrument. It has long been anticipated (e.g., TOPEX Science Working Group, 1981) that the most useful way to employ such data would be to combine it with dynamical ocean models—treating the altimetric data as boundary conditions on the model. A number of authors (e.g., Marshall 1985; Hurlburt 1986; Webb and Moore 1986; Malanotte-Rizzoli and Holland 1986, 1989; De Mey and Robinson 1987) have studied, by simulation, ways of combining altimetric data with various types of models, both steady state and time varying.

These studies have tended to focus on the model aspects of the problem, with the “data” being not only simulated, but assumed “perfect” in ways ranging from the assumption of zero observational noise, to assumptions about space/time sampling so idealized as to be physically inconsistent with orbital dynamics. The emphasis here is the opposite one; we begin exploration of the issues of combining real data with a somewhat too simple dynamical model.

Actual altimetric systems are quite complicated (e.g., Wunsch and Gaposchkin 1980) involving a long list

of error terms, which are complex functions of frequency, wavenumber and geography, as well as unusually intricate space/time sampling issues (this latter problem has been discussed specifically in Wunsch 1989a). Our purpose here is to find methods for using real data, so as to study a real oceanic question in the context of a model that is sufficiently sophisticated to be representative of much more realistic ones. But the focus is on the novel data handling and interpretation issues.

A number of previous papers have appeared (e.g., Malanotte-Rizzoli and Holland 1989) that have attempted to estimate mesoscale eddy variability from simulated altimetric measurements. Most of these papers use procedures derived from meteorological assimilation methods and are somewhat pessimistic concerning the results. Our philosophy is different. Designers of altimetric missions (e.g., TOPEX Science Working Group 1981) recognized from the beginning that the sampling characteristics dictated by orbits satisfying Newton's laws of motion were ill-suited to mapping the mesoscale, although a determination of the spectral characteristics of the mesoscale would be possible. Rather, the idea is that altimeters could define the much larger space/time scale variability where no other instrument system can even begin to produce useful observations.

We advance the idea that an eddy-resolving model—if we can force it to be consistent with large space/time scale observations—can then *compute* the mesoscale variability. A good model would be defined as one which does a good job of either: (a) getting the spectral (frequency/wavenumber) structure of the mesoscale correct as a function of geography, and (b) actually

* Permanent address: Centre National de Recherches Météorologiques, 42 avenue Coriolis, 31057 Toulouse Cedex, France.

Corresponding author address: Dr. Carl Wunsch, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

correctly computes the detailed structure of individual eddies. Accomplishing (b) is obviously much more difficult both for the modeler and the oceanographer who must use nonaltimetric methods to measure details of individual eddies. If (b) is achieved, (a) is then automatically satisfied. But for many purposes (b) may not matter a great deal if (a) is achieved.

We do not know the extent to which a model that is forced to be consistent with large-scale constraints is then improved in its ability to get the mesoscale "right"; we are not aware of any published literature on the subject. It seems possible that such constraints will be very powerful ones as models improve. Much of what follows in this paper is directed at learning how to force the large-scale constraints. The model, being linear, does not itself calculate the unobserved higher wavenumbers. But the methods we explore can be adapted to nonlinear models and we expect eventually to use these methods with such models.

2. The dynamical model and the observation equation

Our context is that of control methods—combining explicitly a dynamical evolution equation and an observation equation (see Wunsch 1989b, for a tutorial).

a. Dynamical equation

Consider the part of the ocean displayed in Fig. 1; we refer to this region as the "focus area." The region is a complex one, encompassing the Gulf Stream and its recirculations as well as portions of the interior of the subtropical gyre. We will drastically over-simplify the known dynamics by representing the signals we seek by the linearized barotropic Rossby wave equation:

$$\frac{\partial}{\partial t} \left[\frac{\partial^2 \eta}{\partial x^2} + \frac{\partial^2 \eta}{\partial y^2} \right] + \beta \frac{\partial \eta}{\partial x} = 0. \quad (1)$$

Our goal is to combine the GEOSAT altimetric observations with (1) to make estimates of the surface elevation η . One can choose better regions than the focus area to argue that (1) might be realistic; we persist in the present combination both because (i) we eventually expect to replace (1) in this area with a more complex model and would like to understand the behavior of the data before doing so, but (ii), the use of a model which is known to be inadequate allows us to address plainly difficult questions of how data/model combinations should be interpreted. Even the most sophisticated of extant oceanographic models will fail in some way; sometimes these failures will be subtle and difficult to perceive, and hence easy to disguise. In the present case, model failure will be obvious; nonetheless, the exercise is not vacuous, and we will return to the interpretation of the result at length below.

We rewrite (1) in a standard form of linear state space (control) theory. A conventional approach would

be to discretize this equation so as to obtain a suitable form. However, in this simple case, any numerical problems associated with finite differences are easily avoided by using an analytical solution. The solution adopted consists of M horizontal modes:

$$\eta(x, y, t) = \sum_{m=1}^{m=M} \alpha_m \sin(\mathbf{K}_m \cdot \mathbf{X} - \omega_m t + \theta_m) \quad (2)$$

where $\mathbf{X} = (x, y)$ is the horizontal position, $\mathbf{K}_m = (k_m, l_m)$ the wave vector, α_m and θ_m the wave amplitude and initial phase and ω_m the associated frequency. For (2) to be a solution of (1) the wave frequencies have to satisfy the dispersion relation:

$$\omega_m = -\beta k_m / (k_m^2 + l_m^2). \quad (3)$$

From basic trigonometric relations, (2) is transformed into

$$\eta(x, y, t) = \sum_{m=1}^{m=M} [q_{2m-1}(t) \cos(\mathbf{K}_m \cdot \mathbf{X}) + q_{2m}(t) \sin(\mathbf{K}_m \cdot \mathbf{X})] \quad (4)$$

where

$$\begin{aligned} q_{2m-1}(t) &= \alpha_m \sin(\theta_m - \omega_m t) \\ q_{2m}(t) &= \alpha_m \cos(\theta_m - \omega_m t) \end{aligned} \quad (5a)$$

and accordingly

$$\alpha_m^2 = q_{2m-1}^2 + q_{2m}^2. \quad (5b)$$

The evolution of the new variables (q_{2m-1} , q_{2m}) over a time step $\Delta t = t_{k+1} - t_k$ is given by the linear equation:

$$\begin{bmatrix} q_{2m-1} \\ q_{2m} \end{bmatrix}_{k+1} = \begin{bmatrix} \cos(\omega_m \Delta t) & -\sin(\omega_m \Delta t) \\ \sin(\omega_m \Delta t) & \cos(\omega_m \Delta t) \end{bmatrix} \begin{bmatrix} q_{2m-1} \\ q_{2m} \end{bmatrix}_k \quad (6a)$$

with

$$\begin{bmatrix} q_{2m-1} \\ q_{2m} \end{bmatrix}_0 = \begin{bmatrix} \alpha_m \sin \theta_m \\ \alpha_m \cos \theta_m \end{bmatrix} \quad (6b)$$

where the index k refers to values taken at time t_k . Denoting by \mathbf{A}_m the 2×2 matrix in (6a), we define the $2M \times 2M$ block-diagonal matrix:

$$\mathbf{A}(k) = \text{diag}[\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M]$$

Eq. (6a) can then be rewritten for all modes:

$$\mathbf{q}(k+1) = \mathbf{A}(k)\mathbf{q}(k) + \mathbf{w}(k) \quad (7)$$

where the vector \mathbf{q} is comprised of all variables q_i . A new term \mathbf{w} has been appended to allow, in a simple way, the system to be perturbed by external forcings, such as the wind for example. Equation (7) is a linear dynamical equation in the standard form with the "state vector" \mathbf{q} , the "state-transition matrix" \mathbf{A} , and

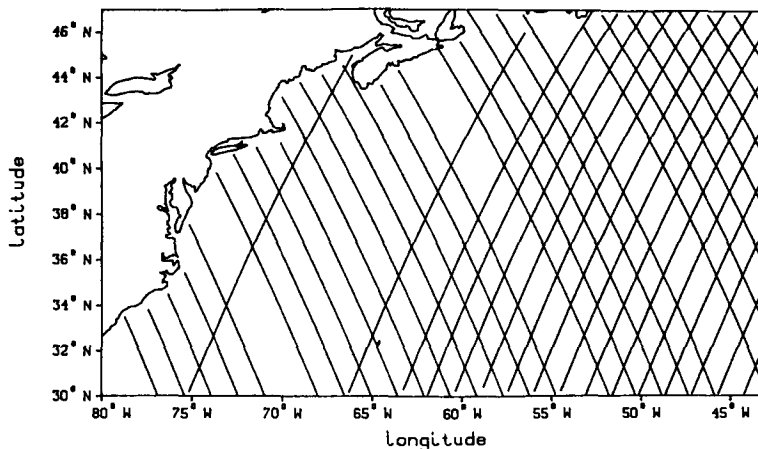


FIG. 1. The focus area with the ground tracks of the GEOSAT repeat cycle 9.

the “process noise” w ; it is known as the “state equation,” or more simply as the “dynamical model.”

b. Observation equation

The altimetric observations are of the form

$$h = \eta + n \tag{8}$$

where η is the variable appearing in (1) and n is the “measurement noise,” representing the entirety of the elevation field h not describable by (1). Written in terms of the state variables, (8) becomes

$$h(X, t) = \sum_{m=1}^{m=M} [q_{2m-1}(t) \cos(K_m \cdot X) + q_{2m}(t) \sin(K_m \cdot X)] + n(X, t) \tag{9}$$

It takes GEOSAT about 5 minutes to cover the longest arc in the focus area. Because barotropic Rossby waves have periods of several days to several months, all measurements made along a single track are treated as simultaneous. Therefore, given r measurements on a track, (9) yields:

$$\begin{bmatrix} h_1 \\ \vdots \\ h_r \end{bmatrix}_k = \begin{bmatrix} \cos(K_1 \cdot X_1) & \sin(K_1 \cdot X_1) & \vdots & \cos(K_M \cdot X_1) & \sin(K_M \cdot X_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(K_1 \cdot X_r) & \sin(K_1 \cdot X_r) & \vdots & \cos(K_M \cdot X_r) & \sin(K_M \cdot X_r) \end{bmatrix}_k \begin{bmatrix} q_1 \\ \vdots \\ q_{2M} \end{bmatrix}_k + \begin{bmatrix} n_1 \\ \vdots \\ n_r \end{bmatrix}_k \tag{10a}$$

or, in matrix form,

$$h(k) = C(k)q(k) + n(k) \tag{10b}$$

where C is the $r \times 2M$ matrix in the preceding equation. The combination of (7) with (10b) is a standard control form, of dynamical equation and observation equation. Before proceeding, we need to describe the dataset and how it was produced.

3. The data reduction

Data from ten successive 17-day repeat cycles of the GEOSAT Exact Repeat Mission are used here. These are repeat cycles 9 to 18 covering the period 24 March to 9 September 1987. After a quality check, a total of 403 tracks crossing the focus area were retained. They contain nearly 30 000 three-second average altimetric measurements.

The sea surface elevation measured along each sub-satellite arc is the absolute value relative to a reference

ellipsoid and includes the mean component owing to the deviation of the solid earth from the reference ellipsoid—a surface that we will refer to as the “geoid”; the mean elevation relative to the geoid owing to the time average ocean circulation; a time varying component owing to fluctuations in the ocean circulation; tidal contributions; and a number of mean and time varying error terms. (Some workers prefer to think of the tidal contributions as being a time dependent geoidal component.)

The data as used were produced as follows:

The mean of the data along each of the tracks in the focus area was corrected for the tides using the values produced by the GEOSAT project from the Schwiderski (1980) model. These values were then averaged for ten repeat cycles along each arc, producing a mean value $\bar{\zeta}(x_i)$ for each arc, where x_i is along-track distance. We define the deviation, h as

$$h = \zeta - \bar{\zeta}$$

It is these numbers which are referred to in the observation Eq. (8).

No corrections other than the tide were made to the data, because we have been unable to convince ourselves (Campbell 1988) that they can be made effectively. Whatever error is thus left in h is included in the noise term, n . There is one error, that owing to the orbit, which must be considered separately.

The largest of all known errors in h is that owing to the crude estimates the GEOSAT project provides for the spacecraft radial position. For this reason, ζ contains errors formally of several meters. It is well known (e.g. Tai 1988a) that these errors are mainly on the scale of the Earth's circumference, and this confinement to long wavelengths is the basis for many different error removal schemes for studying mesoscale phenomena. In the present example, we wish to demonstrate how known dynamics (1), plus stipulation of the structure of the error can be used in combination to remove the orbit error as part of procedure for estimating η . We thus leave these orbit error components in our working data. (In the long run, it is much preferable to have the major error components removed in advance, deterministically, from the data; future altimetric missions should have much reduced orbit errors. But there will always be some residual characterized by the orbital dynamics and the present methods should be even more effective then, because the error suppression algorithm will work with a much higher signal-to-noise ratio.)

The calculation of h as the deviation from the longer term mean arc is itself ultimately intolerable. Forming means from samples at 17-day intervals aliases motions of periods of 34 days and shorter into longer periods and produces an erroneous mean and erroneous estimates of the variability. These errors are tolerable for present purposes because within any given 17-day period many arcs are measured within the area, and the constancy of coefficients in (1) is consistent with the assumption underlying Eq. (10)—which has constant coefficients—that the region is statistically homogeneous. Thus as Wunsch (1989a) discusses, the aliasing is not computable as simply that from a 34-day Nyquist period, nor is it as serious. Nonetheless there *is* aliasing, and one must ultimately (when a more accurate model is being used) avoid this procedure. For the present, we believe that other errors are dominant and wish to retain the comparatively simple data reduction procedure.

4. Description of noise statistics

We can make rough estimates of the spectral characteristics of the measurement noise. One expects the correlation function to be of the general form (e.g., Wunsch and Zlotnicki 1984):

$$E[n_i(k)n_j(l)] = f[|t_k - t_l|, \Delta X_{ij}] \quad (11)$$

where E denotes an expected value, ΔX_{ij} is a measure of the distance between X_i and X_j , and f is a function to be determined. Most results of the linear state space theory are derived assuming that the measurement noise is white (i.e., not time-correlated) and has a zero mean. In that case (11) takes the form

$$E[n_i(k)n_j(l)] = R_{ij}\delta_{kl} \quad (12)$$

where δ_{kl} is the Kronecker delta and R_{ij} is a spatial correlation function with variable ΔX_{ij} . In the case where the measurement noise is not white but can be represented by a first order Markov process excited by a white noise (i.e., when the noise is red), the problem can still be recast in terms of the usual theory with white noise by using the measurement differencing technique of Bryson and Henrikson (1968). However, the actual correlation function is probably more complex than that of a simple white or red noise. Several characteristic correlation times are likely to play important roles. Wunsch (1986) uses two characteristic correlation times to model the orbit error alone. The nonmodeled fraction of the oceanic signal should also dictate some important time scales such as, for example, the characteristic propagation time of mesoscale eddies. Nevertheless, to keep this problem simple, we will make the working assumption that the measurement noise is white and use the parameterization (12). Correlations between errors in successive tracks are ignored. The major effect of this simplification is to leave a larger apparent error in the results than is actually necessary (*spatial* correlations, along any individual track are however, retained in this formalism). Wunsch and Imawaki (1989) show how between-track correlations can be used in a practical scheme. The exact form of the measurement noise covariance matrix R will be determined in section 7.

For similar reasons, we will assume that the process noise of Eq. (7) is white in time, and of zero-mean with

$$E[w_i(k)w_j(l)] = \Gamma_{ij}\delta_{kl}. \quad (13)$$

In addition, observation and process noises are supposed uncorrelated. Depending on the experiments performed, different hypotheses will be made concerning the spatial structure of the process noise covariance matrix Γ .

5. Estimation method

Given a dynamical evolution rule (7), a set of observations (10b) with known statistics (12, 13), we pose the problem of making the optimal estimate of η , where "optimal" is in the conventional sense of a minimum mean-square error criterion. η is our "state variable" and we are tackling a "state estimation" problem. This problem is a much-studied classical one (e.g., Anderson and Moore 1979). It is convenient to attack the general problem by first considering the reduced

problem of asking how best to combine the data and model when the estimate is to be made sequentially. That is, we suppose that the data are to be employed as though they were arriving in “real time” and only data preceding or at the “present time,” t , can be used to estimate the ocean state at this time. This problem is the classical “sequential filtering problem” of control theory. Our goal is to use all the data, which realistically have been stored, so that data future to time t are employed to estimate the state at t . This problem is known as a “smoothing” one. Because most smoothing algorithms are written so as to contain the optimal filter as an explicit first step, the digression through the filtering problem is useful.

a. Kalman filtering

The solution to the optimal, sequential filtering problem is the Kalman (1960) filter (see for example, Sorenson 1985; Ghil et al. 1981; Wunsch 1989b) and we will content ourselves with merely writing down the resulting algorithm. The filtering process can be divided into two steps. Assuming that the state estimate and the associated error covariance matrix are known at time t_{k-1} , the dynamical model [Eq. (7)] is first used to extrapolate these quantities from time t_{k-1} to t_k . The new estimate of the state is denoted $\hat{q}(k|k-1)$, where $(k|k-1)$ indicates that the variable is estimated at time t_k using data up till time t_{k-1} , only. At this point, the estimation error, $\tilde{q}(k)$ is the difference between the true value $q(k)$ and the estimate of it: $\tilde{q}(k) = q(k) - \hat{q}(k|k-1)$ and the corresponding error covariance matrix is $P(k|k-1) = E[\tilde{q}(k)\tilde{q}(k)^T]$. Then, the new measurements available at t_k are combined with the model forecast to produce an updated state estimate $\hat{q}(k|k)$ with a reduced error covariance matrix $P(k|k)$. The filtering process starts from known expected values of the initial state and its covariance:

$$\hat{q}(0|0) = E[q(0)] \tag{14a}$$

$$P(0|0) = E[[q(0) - \hat{q}(0|0)][q(0) - \hat{q}(0|0)]^T]. \tag{14b}$$

The state is propagated forward in time using the model, but with w set to zero (its expected value) because we have no other information yet available about it:

$$\hat{q}(k|k-1) = A(k-1)\hat{q}(k-1|k-1) \tag{15}$$

and the error of this predicted state is easily shown to be:

$$P(k|k-1) = A(k-1)P(k-1|k-1)A^T(k-1) + \Gamma(k-1) \tag{16}$$

which is the sum of the model-propagated error in the previous best-estimate plus that expected from the missing knowledge of w . The new observations are then

combined in a weighted average with the value predicted from the dynamics in the form:

$$\hat{q}(k|k) = \hat{q}(k|k-1) + G(k)[h(k) - C(k)\hat{q}(k|k-1)] \tag{17}$$

with new error,

$$P(k|k) = [I - G(k)C(k)]P(k|k-1). \tag{18}$$

I is the identity matrix and G is the “Kalman gain matrix” defined by

$$G(k) = P(k|k-1)C^T(k) \times [C(k)P(k|k-1)C^T(k) + R(k)]^{-1}; \tag{19}$$

G weights the forecast and the observations by their appropriate relative errors. This set of equations, (15) to (19), is a conventional form of the Kalman filter, requiring the inversion of a single $r \times r$ matrix in (19). Other algebraically equivalent forms exist with various computational and numerical merits. Our choice is probably not the best possible one but it proves to be sufficient for this study [a variant of (18) is used below in Eq. (41)]. For a derivation of (15 to 19) and a discussion of alternative forms of the Kalman filter the reader is referred to Anderson and Moore (1979); Wunsch (1989b) provides a heuristic derivation based upon ordinary least squares.

b. Optimal smoothing

The Kalman filter permits sequential estimation in optimum fashion. But as we noted earlier, oceanographic data are normally stored, and the data “future” to time t obviously contain information useful for estimating the oceanic state at t and earlier—information that we would like to use. “Smoothers” are estimators using such formally future data. In the present study, we are specifically interested in what is called fixed-interval smoothing: the available dataset spans a fixed interval of time $[t_i, t_N]$ and we seek optimal state estimates at all measurement times, using all available measurements, i.e., we seek estimates $\hat{q}(k|N)$, for N fixed and $\{k = 1, \dots, N\}$. Various algorithms have been proposed to solve this optimal smoothing problem. Most of them include the determination of the filtered state estimates so that the Kalman filter is part of these algorithms. Such is the case for the Rauch-Tung-Striebel (1965) algorithm we use. Having obtained all filtered state estimates up to $\hat{q}(N|N)$, smoothing is performed backwards, i.e., with the k -index going from $N-1$ to 1. The smoother equations are

$$\hat{q}(k|N) = \hat{q}(k|k) + G_s(k)[\hat{q}(k+1|N) - \hat{q}(k+1|k)] \tag{20}$$

$$P(k|N) = P(k|k) + G_s(k)[P(k+1|N) - P(k+1|k)]G_s^T(k) \tag{21}$$

where G_s is the smoothing gain matrix defined by

$$G_s(k) = P(k|k)A^T(k)P^{-1}(k+1|k). \quad (22)$$

Equation (20) can be interpreted (Wunsch 1989b) as a weighted average of the previous estimate at time k done in the filtering step, with the information gained about the true state determined from knowledge of how the ocean actually evolved later. Again, the choice of this particular smoothing algorithm may not be the best in terms of computational efficiency, but it is satisfactory. A review of different fixed-interval smoothing algorithms can be found in Meditch (1973). Attention has to be paid to issues of computational stability with some algorithms; for details, see for example, Anderson and Moore (1979).

One should be aware that smoothing is equivalent to the many forms of "adjoint" modelling now receiving much attention in both meteorology and oceanography. If equivalent information is supplied, the two procedures yield the same result for the state estimate and differ only in computational form (see Bryson and Ho 1975, p. 395; or Thacker 1986). A simple statement of the difference between the procedures is that the adjoint methods gain efficiency by suppressing the computation of the error covariances of the estimated solution, while the smoothing algorithms actually focus on these error estimates. The computational load for smoothing is dominated by the error estimates and suppressing them is very attractive. It is our belief however, that the error estimates are ultimately a necessity—both because one cannot understand oceanographic flows without a complete knowledge of the accuracy of one's results, and because the recursive updating of the oceanographic fields with newly acquired data can only be done if error estimates are available. But we emphasize that the methodologies are fundamentally equivalent, and in particular, Tziperman and Thacker (1989) show how to add error estimates to an adjoint formalism—albeit by incurring computing problems equivalent to those in smoothing.

6. Spectral description of the data

Prior to employing the optimal filter or smoother, we still have to specify the covariance matrix of the measurement noise and choose the horizontal wave modes in the solution (2). For these purposes, a spectral description of the data is most helpful.

From the SEASAT data, Fu (1983) computed and analyzed wave number spectra in different oceanic regions finding that the characteristics of the spectra are dependent on the energy level of the local mesoscale activity. Here, we use GEOSAT data from the focus area.

The spectra are computed using only sequences (tracks) having at least 100 consecutive measurements. A total of 110 tracks (37 descending and 73 ascending) meet this requirement. Thirty-seven spectra corre-

sponding to ascending tracks are averaged together to produce an "ascending track spectrum." A descending track spectrum is computed in a similar way. The two spectra are shown in Fig. 2. The wavenumber is that measured along the track and denoted k_t . $\lambda_t = 2\pi k_t^{-1}$ is the corresponding wavelength.

The two spectra appear to be very similar with their respective 95% confidence intervals overlapping at almost all wavenumbers. We thus will suppose that the spectrum is to a good first approximation horizontally isotropic. A global power density spectrum formed by averaging the spectra from all tracks is denoted Φ and shown in Fig. 3. Because most along-track data show a marked slope owing to large scale orbit errors, the spectra are markedly red. With increasing values of k_t , the power density decreases first very slowly but then drops by more than a decade between $k_t = 3 \times 10^{-3}$ cycles per kilometer (cpkm) and $k_t = 6 \times 10^{-3}$ cpkm (spectral slope ≈ -3.5). At higher values of the wavenumbers, the log-spectrum exhibits a fairly linear decrease. A linear least square fit yields a slope of -2.2 ± 0.1 .

In Fig. 4, Φ is compared to the high energy area average spectrum of Fu (1983). In Fu's spectrum, the power drops at the low wavenumber end because he detrended the data and is generally lower in value at all wavenumbers because it was computed using repeat tracks from a time span of only 24 days. The repeat tracks used to compute Φ span 170 days and as expected have more power than Fu's at all wavenumbers. Fu (1983) found a spectral slope of -4.5 ± 1.5 at wavelengths between 250 and 100 km, consistent with the -3.5 slope we find for $333 \text{ km} \geq \lambda_t \geq 166 \text{ km}$. However, at smaller wavelengths, the slope of Φ is only -2.2 . The reason for the generally steeper slope of the

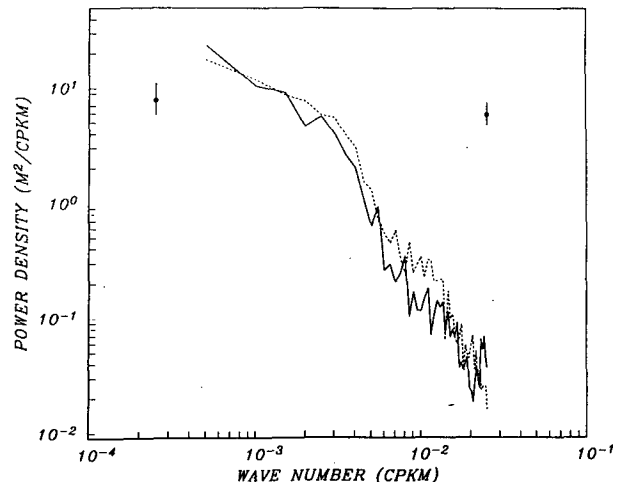


FIG. 2. Wavenumber power spectra for the ascending (solid line) and descending (dashed line) tracks. The error bar on the left represents the 95% confidence interval for the ascending track spectrum. The error bar on the right is for the descending track spectrum. The wavenumbers are in cycles per kilometer (cpkm).

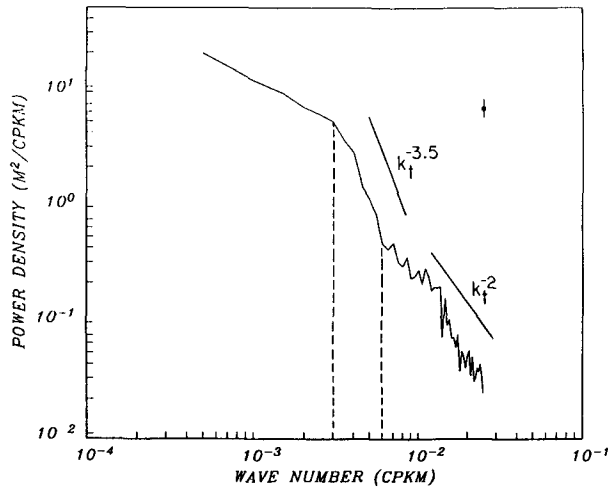


FIG. 3. The global power spectrum (Φ) with its 95% confidence interval.

Fu spectrum is unclear, but may well be related to the different frequency content of the two datasets.

Fu's spectrum reaches a white noise level at $\lambda_t = 100$ km. The power density of the white noise is close to $1.5 \cdot 10^{-2} \text{ m}^2/\text{cpkm}$ corresponding to an average noise level of about 4 cm. The GEOSAT altimeter is somewhat better with a white noise level of 2 to 3 cm for 1-second averaging (Sailor and LeSchack 1987). With the 3-second averaging we use, the noise level should be further reduced by a factor of $\sqrt{3}$. Accordingly, it should have a power density close to $2 \cdot 10^{-3} \text{ m}^2/\text{cpkm}$ which is about a decade below the signal we obtain at $\lambda_t = 40$ km.

7. Determination of the measurement noise covariance

Given (8), and assuming that the measurement noise and the barotropic wave signal are not correlated, the spatial autocovariance function of a sequence of altimetric measurements can be written:

$$R_h(x_t) = R_\eta(x_t) + R(x_t) \quad (23)$$

where R_h and R_η are the autocovariance functions of h and η , R remains the measurement noise covariance as defined in Eq. (12) and x_t is now the spatial separation along track. Taking the Fourier transform of (23), one obtains the power density spectrum of the measurements (S_h) as the spectrum of the signal (S_η) plus that of the measurement noise (S):

$$S_h(k_t) = S_\eta(k_t) + S(k_t). \quad (24)$$

The spectrum Φ is our best estimate of S_h ; k_t remains along-track wavenumber. In the next three subsections we show how to use this information to estimate the noise spectrum S and the noise covariance R .

a. Noise sources

The measurement noise is the sum of the measurement errors plus the fraction of the oceanic signal not

associated with barotropic Rossby waves. Barotropic Rossby waves at periods accessible to us have spatial scale much larger than that of typical baroclinic mesoscale activity, of order 100 km. Much of the oceanic mesoscale variability is thus to be considered as measurement noise. The wavenumbers dominated by mesoscale variability will be in the region where Φ has a slope close to -2 ($\lambda_t \leq 166$ km) and therefore the mesoscale is modelled as a red noise process. It is only at wavelengths larger than 166 km that a significant signal is seen to emerge from the red noise background, as the slope of the spectrum suddenly changes. We will take this wavelength of 166 km to be a lower limit for the wavelength of the Rossby wave signals.

At the low wavenumber end of the spectrum, the power contributed by the Rossby waves is most probably very small compared to the power contributed by the large scale orbit errors. We therefore seek signals having a maximum wavelength of 1000 km, the characteristic size of the focus area, and specify that the power appearing at larger wavelengths is due to measurement noise.

We thus have specifically identified two distinct components of measurement noise, a shortwave component identified with mesoscale eddies, and a long-wave component identified with the orbit error. As already noted above, the full observational spectrum is a complex one—involving atmospheric water vapor, seastate, errors in the tide, and many more. Quantitatively useful spectral estimates for these additional errors are lacking. In view of the simplicity of our dynamical model, no attempt is made to model them explicitly. Rather, it is assumed they are implicitly accounted for in the short- and long-wave noise components. Decompose the measurement noise covariance as

$$R(x_t) = R_s(x_t) + R_l(x_t) \quad (25)$$

where the subscripts s and l denote the short- and long-wave components of the covariance. The Fourier transform of (25) yields a similar relation for the power spectral densities:

$$S(k_t) = S_s(k_t) + S_l(k_t). \quad (26)$$

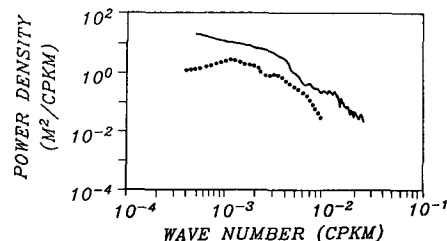


FIG. 4. Wavenumber power spectra from high-energy regions: Fu (1983, dotted line), Φ (solid line).

b. Short-wave noise

Here R_s is taken to be of the form:

$$R_s(x_i) = r_0 \exp(-|x_i/l_0|) \quad (27)$$

where r_0 is the variance of this short-wave noise and l_0 its correlation length. The corresponding power spectral density is

$$S_s(k_i) = \frac{2r_0l_0}{1 + (2\pi l_0 k_i)^2} \quad (28)$$

At large wavenumbers, the first term of the denominator becomes negligible so that one obtains the power spectral density of a red noise with a -2 slope on a log-log plot:

$$\log S_s(k_i) = \log(r_0/2\pi^2 l_0) - 2 \log k_i. \quad (29)$$

Using this linear expression¹ to fit Φ at wavenumbers larger than 6×10^{-2} cpkm one obtains the estimate:

$$r_0/l_0 = 2 \times 10^{-4} \text{ m}^2 \text{ km}^{-1}. \quad (30)$$

There is some freedom in choosing either r_0 or l_0 . The choice $l_0 = 60$ km implying $r_0 = 1.2 \times 10^{-2} \text{ m}^2$ (i.e., a rms variability of 11 cm) is consistent with the statistical properties of the mesoscale activity, even if the rms variability seems a bit small for a western boundary region. Indeed, the Gulf Stream eddies have a surface signature of about 40 cm (e.g., Cheney and Marsh 1981) but the eastern part of the selected domain is considerably quieter and the tracks crossing this region have a smaller variability. The variance of the "mesoscale noise" is meant to be an average over the whole domain (spatially variable statistics can be employed if one wishes).

Because r_0 and l_0 are chosen according to (30), a good fit of Φ is obtained at short wavelengths (Fig. 5).

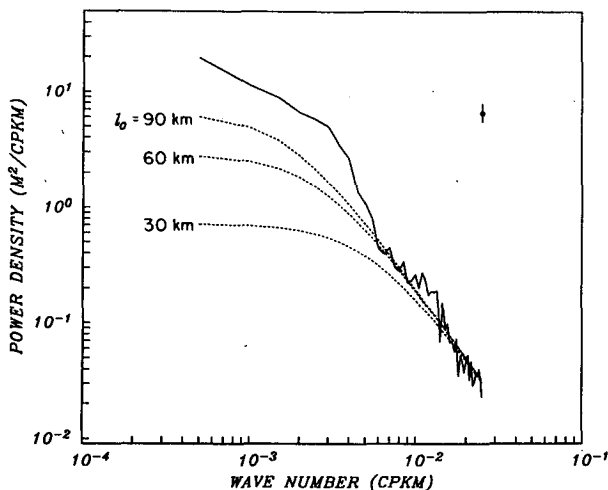


FIG. 5. The data spectrum Φ (solid line) with red noise spectra (dashed lines) computed according to (28) for different values of l_0 and a constant ratio $r_0/l_0 = 2 \times 10^{-4} \text{ m}^2 \text{ km}^{-1}$.

If the ratio r_0/l_0 is kept constant but a smaller value of l_0 is selected, the quality of the fit decreases. If one uses larger values of l_0 , larger values of the variance are necessary and the additional power appears at long wavelengths, a region where one wants to minimize the effect of the short-wave component of the noise.

c. Long-wave noise

One can think of the large-scale orbit error as a slowly varying sine with a predominant wavelength comparable to the Earth's circumference (about 40 000 km) (e.g., Wunsch and Zlotnicki 1984; Tai 1989). The conventional way to eliminate this error is to approximate it by a simple bias or a linear or quadratic or more structured function (e.g., Tai 1988) and then subtract it from the data (e.g., Thompson et al. 1983). Unfortunately, this adjustment removes oceanic signal too—as we will demonstrate later. Here we do not explicitly compute the orbit error. Instead, we filter it out on the basis of its spectral characteristics specified by R_l , related intimately to the method used by Wunsch and Zlotnicki (1984). The variance of the orbit error (r_l) is close to 1 m^2 (Smith et al. 1987) and one expects to have a correlation length (l_l) the order of 40 000 km and longer, depending upon how long orbit errors persist. Also, the long-wave noise spectrum should match Φ at the largest wavelengths. Interestingly, the simple choice of a red noise:

$$R_l(x_i) = r_l \exp(-|x_i/l_l|) \quad (31)$$

with $r_l = 1.7 \text{ m}^2$ and $l_l = 40$ 000 km yields a spectrum with suitable power densities at long wavelengths (Fig. 6). As a drawback to this choice, S_l still has non-negligible power at short wavelengths and this interferes with the previously computed short-wave noise spectrum. Therefore, to maintain the fit between Φ and $S_s + S_l$ at short wavelengths, we (somewhat artificially) decrease the variance of R_s from 0.012 to 0.01 m^2 (Fig. 6).

8. Selection of the horizontal modes

The horizontal modes in the solution (2) must be specified. From the dispersion relation for Rossby waves, attention can be restricted to wavenumbers in the left half-plane i.e.,

$$k < 0 \quad (32)$$

¹ In all spectra presented here, the power densities corresponding to negative wavenumbers have been added to their positive counterparts in such a way that the integral of the plotted densities over the positive wavenumbers exactly yields the variance of the signal. Accordingly, the power densities appearing in Φ (or any other spectrum) have to be divided by two before being compared to theoretical values like (25). Also, the theoretical values are multiplied by 2 when plotted.

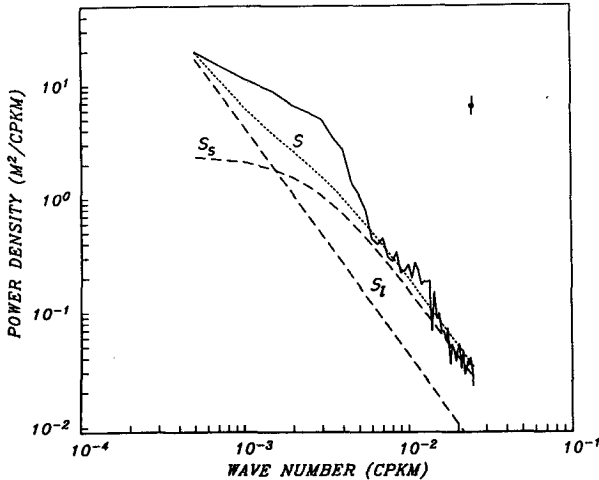


FIG. 6. The data spectrum Φ (solid line) compared with the specified short- and long-wave noise spectra (S_s and S_l , dashed lines) and the global noise spectrum ($S = S_s + S_l$, dotted line). The spectra's parameters are: $r_0 = 0.01 \text{ m}^2$, $l_0 = 60 \text{ km}$, $r_1 = 1.7 \text{ m}^2$, $l_1 = 40\,000 \text{ km}$.

corresponding to waves with positive frequencies. Only waves in the range $\lambda_i \geq 166 \text{ km}$ are sought. Owing to the curvature of the ground-track in the beta-plane, there is a difference between the wavelength measured along the track (λ_r) and the actual wavelength (λ). However, in the small domain of current interest, this difference is negligible ($\leq 0.2\%$). Therefore, the minimum wavelength constraint is easily expressed in terms of (k, l) :

$$[k^2 + l^2]^{1/2} \leq 2\pi\lambda_{\min}^{-1}. \tag{33}$$

Here $\lambda_{\min} = 166 \text{ km}$. Similarly, we do not expect to estimate Rossby wave amplitudes for wavelengths larger than $\lambda_{\max} = 1000 \text{ km}$, and select waves such that

$$[k^2 + l^2]^{1/2} \geq 2\pi\lambda_{\max}^{-1}. \tag{34}$$

Finally, since the mean tracks have been computed from data spanning a period p_{\max} of 170 days, the wave periods will be made shorter than this value. Using the dispersion relation (3), this requirement implies:

$$\frac{2\pi(k^2 + l^2)}{-\beta k} \leq p_{\max} \tag{35}$$

a relation indicating that acceptable wave vectors lie within a circle of radius $d = \beta p_{\max}/(4\pi)$, with center at $(-d, 0)$.

The constraints (32) to (35) define a locus (Fig. 7) within which we select a limited number of wave vectors: we take wavenumber vectors which are harmonics of the basin scale of 1000 km. These 32 wave vectors are represented by heavy dots in Fig. 7. Accordingly, the state vector has 64 components.

9. Numerical implementation

a. Initialization

The Kalman filter requires an initial value of the state estimate and its covariance matrix (14a, b). All initial condition estimates are for an ocean at rest, i.e.

$$\hat{q}(0|0) = \mathbf{0}. \tag{36}$$

In the complete absence of information, the error covariance of this initial estimate would be infinite. But even without altimeter data, such a state of complete ignorance is hardly realistic. Assuming that the state estimates yield independent estimates of the wave phases and amplitudes, (5b) yields

$$\begin{aligned} E(\alpha_m - \hat{\alpha}_m)^2 &= E[q_{2m-1} - \hat{q}_{2m-1}]^2 + E[q_{2m} - \hat{q}_{2m}]^2 \\ &= P_{2m-1,2m-1} + P_{2m,2m}. \end{aligned} \tag{37}$$

Estimates of the wave amplitudes can thus be used to set an upper bound on the variances of the state estimation errors. In the spectral region where we expect Rossby waves ($166 \text{ km} \leq \lambda \leq 1000 \text{ km}$), the variance appearing in Φ is 120 cm^2 above that of the specified measurement noise. If the entirety of that power was equally distributed over the 32 wave modes, their variance would be smaller than 4 cm^2 , i.e. the wave amplitudes would be smaller than 3 cm. Therefore, a conservative assumption is that the initial variance of each state is no larger than $p_0 = 400 \text{ cm}^2$ and

$$P(0|0) = p_0 I. \tag{38}$$

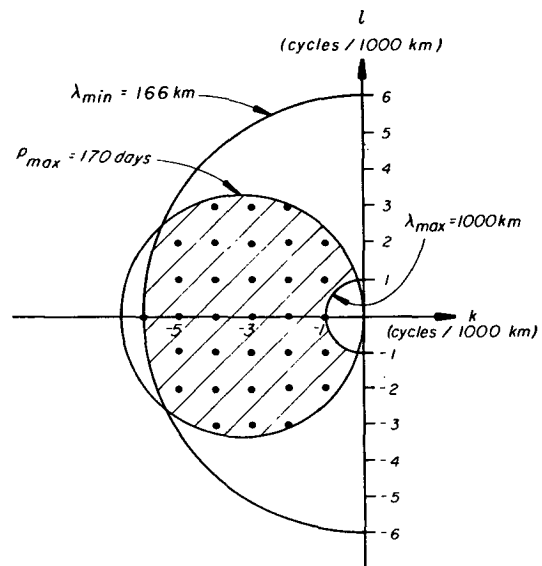


FIG. 7. Locus of the suitable wave vectors (shaded area) including the 32 vectors selected for simulations (heavy dots).

b. Process noise

In the absence of precise information about the noise that directly perturbs the state, in the dynamical equation (7), we will assume that the components of the noise vector are uncorrelated, i.e. that the Γ matrix is diagonal:

$$\Gamma = \sigma^2 I \quad (39)$$

where the process noise variance σ^2 has still to be determined. As shown in the previous section, one expects average wave amplitudes of a few centimeters. To be able to detect such a signal with some confidence, we need amplitude estimates whose errors have standard deviations smaller than about 1 cm, a requirement setting an upper bound on the acceptable magnitude of the process noise. As will be seen, the Kalman filter used is stable and the variances of the amplitude errors reach quasi-steady values. These values can be computed independently of the data from (16), (18) and (37). Such computations indicate that the standard deviation of the process noise cannot be larger than a few millimeters if one wants to reach amplitude error variances as small as 1 cm^2 .

From a more physical point of view, the noise affecting the state variables has to be comparable to the physical perturbations affecting the observed waves. The oceanic barotropic Rossby wave field can always be decomposed into a sum of free and forced modes. The present model explicitly contains only free modes so that forcing should be regarded as a noise perturbation. Consider modes generated by the surface wind stress. Assuming that the stress is purely periodic with frequency ω_0 , the amplitude of the wind-forced modes ($\delta\alpha_m$) is given by (Longuet-Higgins 1965):

$$\delta\alpha_m = \frac{(f/\rho g H) \text{curl}\tau}{\beta k_m + \omega_0(k_m^2 + l_m^2)} \quad (40)$$

where f is the Coriolis parameter, ρ the sea-water density, g the gravity, H the water depth and τ the surface wind stress. This solution fails in the case when the wind stress is in resonance with a free mode ($\omega_0 = \omega_m$). Away from the resonance, Eq. (40) can be used to estimate typical values of $\delta\alpha_m$. Assume a wave with wavenumber $(k, l) = (-3, 0)$ cycles/ 10^3 km, an average value of the wind stress curl (10^{-7} N m^{-3}), a water depth of 4000 m and typical middle latitude values of f and β . For very slow variations of the stress ($\omega_0 \rightarrow 0$), Eq. (40) yields amplitudes close to 1 mm. For a more rapidly moving wind system with a characteristic period of a few days ($\omega_0 \approx 2 \times 10^{-5} \text{ s}^{-1}$), the amplitudes are an order of magnitude smaller ($\delta\alpha_m \approx 0.1 \text{ mm}$). Therefore, it appears that a process noise with a standard deviation of order 1 mm ($\sigma^2 \approx 10^{-6} \text{ m}^2$) can crudely represent wind-forced disturbances. (We are ignoring other possible sources of wave-driving, for example, radiation from the Gulf Stream, e.g., Hogg 1988).

10. Preliminary experiment without process noise

In a preliminary experiment, hereafter referred to as "Reference-32," filtering of the altimetric data is performed in the absence of process noise ($\sigma^2 = 0$). The filter starts from the initial conditions (36, 38) and is run with the 32 horizontal modes selected in section 8. The results will be described and interpreted in terms of the estimated wave amplitudes and phases deduced from (5a) and (5b).

a. Estimated wave amplitudes at the end of the observation period

Final estimates of the wave amplitudes and the standard deviation of the estimation error are presented in the (k, l) plane (Fig. 8). The amplitudes rarely exceed 1 cm, with the largest values being observed for small values of k . In most cases the estimated amplitudes do not exceed one standard deviation of the error, i.e. they are not significantly different from 0. Only 5 out of the 32 selected waves have a final amplitude larger than one standard deviation and larger than 1 cm. These waves will be referred to as W1 to W5. Their wave-number vectors are listed in Table 1.

At first sight (Fig. 8b), the standard deviation isolines exhibit a near circular symmetry with formal errors decreasing when the modulus of the wave vector increases. In other words, the smallest error variances are associated with the waves having the smallest wavelengths. To explain that behavior, we consider an alternative form of the Kalman filter. In this form, the state estimate $\hat{q}(k|k)$ is obtained by combining the state forecast $\hat{q}(k|k-1)$ directly with the least-square estimate of the state deduced from the data $h(k)$. Each of these state estimates is given a weight proportional to the inverse of its error covariance matrix. These are $P^{-1}(k|k-1)$ and $C^T(k)R^{-1}(k)C(k)$, respectively (e.g., Liebelt 1967). The error covariance matrix of the new estimate $\hat{q}(k|k)$ is obtained by combining these inverses as follows.

$$P(k|k) = [P^{-1}(k|k-1) + C^T(k)R^{-1}(k)C(k)]^{-1}. \quad (41)$$

This equation is an alternative form of (18) and can be derived directly from Eqs. (17-19). The first term on the right-hand side characterizes the information carried by the forecast state vector at time t_k while the second term characterizes the information brought into the system by the new measurement made at this time. The model forecast does not change the wave amplitudes and the process noise variance is the same for all waves. There is thus no mechanism in the *model* that could explain why the error in some wave amplitudes decreases faster than for others. But the spatial distribution of the *measurements* could permit better estimation of certain waves. To isolate such an effect we consider the case where the model forecast has a very

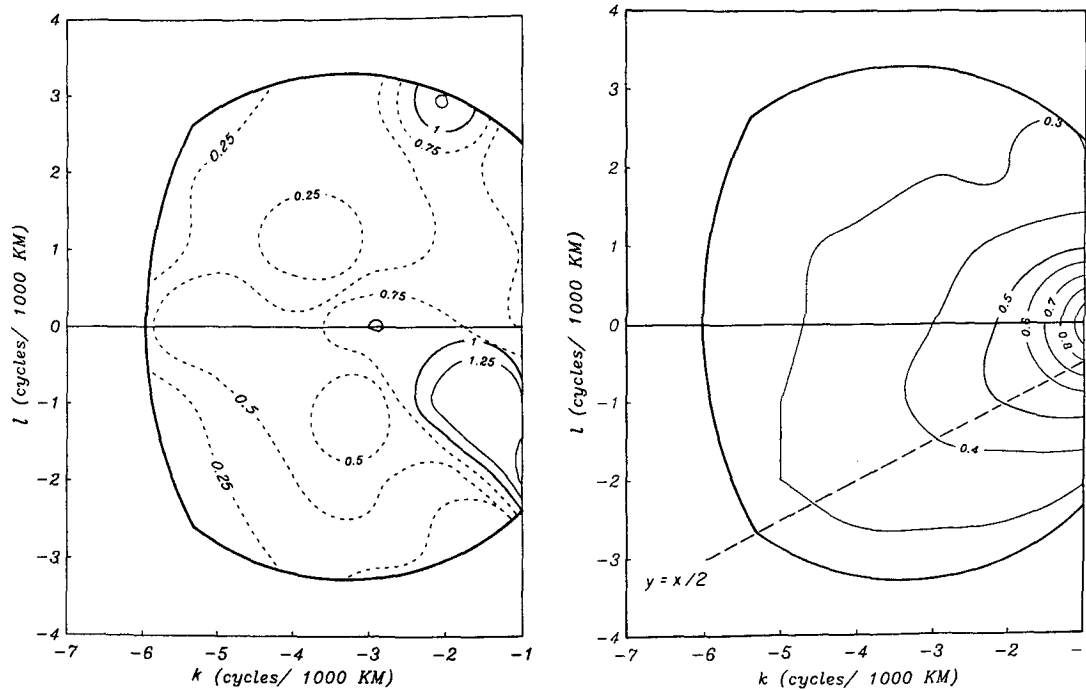


FIG. 8. Estimated wave amplitudes (in cm) after ten repeat cycles (a) and the standard deviation of the estimation error (in cm) (b) for Reference-32 ($\sigma^2 = 0$).

large error covariance ($P^{-1}(k|k-1) \rightarrow 0$, in a suitable matrix norm) so that (41) reduces to

$$P(k|k) = [C^T(k)R^{-1}(k)C(k)]^{-1}. \quad (42)$$

This matrix can be evaluated for a simple model with a single wave mode and two data points on a track, at X_1 and X_2 . According to (25), (27) and (31), $R_{11} = R_{22}$ and $R_{12} > 0$. Then using (10a), (37) and (42), one obtains after some algebra

$$P_\alpha = E[\alpha - \hat{\alpha}]^2 = 2 \frac{R_{11} - R_{12} \cos(\mathbf{K} \cdot \Delta \mathbf{X})}{\sin^2(\mathbf{K} \cdot \Delta \mathbf{X})} \quad (43)$$

where \mathbf{K} and α are the wave vector and amplitude of the model's single mode, and $\Delta \mathbf{X} = \mathbf{X}_1 - \mathbf{X}_2$. This vector is parallel to the track and its length is typically 20 km, the distance between two successive 3-second average measurements of the GEOSAT altimeter. All waves of interest have wavelengths considerably larger than 20 km so that \mathbf{K} and $\Delta \mathbf{X}$ are always at a rather small angle with a maximum value near 35° . In most cases, the angle is only a few degrees. For such small angles, (43) indicates that P_α is a decreasing function of $|\mathbf{K} \cdot \Delta \mathbf{X}|$ and with $|\Delta \mathbf{X}|$ fixed, P_α is a decreasing function of $|\mathbf{K}|$ as observed. The scalar product $\mathbf{K} \cdot \Delta \mathbf{X}$ also depends on the angle between \mathbf{K} and $\Delta \mathbf{X}$. When \mathbf{K} is perpendicular to $\Delta \mathbf{X}$ (i.e. when the wave crests are parallel to the track), the scalar product is zero and P_α becomes infinite. In reality, this angular effect is considerably attenuated by the presence of ascending and descending tracks that are far from being parallel.

Indeed, a wave that is poorly estimated along an ascending track will be much better estimated along a descending one and vice versa. Thus, on average, little angular anomaly should appear. In our dataset however, about two-thirds of the tracks are ascending ones so that the waves having crests (nearly) parallel to these tracks should still have relatively large amplitude errors. These waves have their wave vectors along the line $y = x/2$ and, indeed, one observes in Fig. 8b that this line roughly corresponds to the axis of a high standard deviation ridge.

Finally, it should be mentioned that all off-diagonal terms of the state error covariance matrix are at least one order of magnitude smaller than the diagonal terms, indicating that the state estimates are essentially independent.

b. Time evolution of the amplitudes and phases

Restrict attention now to the dominant waves W1 to W5. The estimates, shown in Fig. 9, first exhibit a large variability before progressively stabilizing. During that initial period, the state error variances decrease very rapidly (Fig. 10). After about 40 days, the estimated amplitudes and phases become less variable while the error variances keep decreasing, though very slowly. The estimated phases of W1 and W2 become nearly constant while the phases of W4 and W5 exhibit a slow, almost linear increase. The amplitudes of these two waves are quite stable after day 60. On the other hand, the amplitude of W1 drops by about 1 cm be-

TABLE 1. The five waves which are found to carry a significant amount of energy and are simultaneously consistent with both data and model.

| Wave | Wave vector (cycles/1000 km) | Period (days) |
|------|---------------------------------|------------------|
| W1 | (-3, 0) | 77.1 |
| W2 | (-2, 3) | 167.0 |
| W3 | (-2, -1) | 64.2 |
| W4 | (-1, -1) | 51.4 |
| W5 | (-1, -2) | 128.4 |

tween day 60 and day 80 and the amplitude of W3 exhibits a slow trend towards decreasing values. Though none of these changes is drastic, it should be

remembered that, strictly speaking, the homogeneous wave model is only valid for Rossby waves of constant amplitude. In section 11, the response of the model to deviations from those specifications is analyzed. This analysis proves to be useful in explaining some of the features observed in Fig. 9.

c. Explained variance

The present simple model is not expected to explain a large fraction of the surface variability. A perfect model would explain all of the observed variance not due to noise, about 120 cm^2 (see Fig. 6). To quantify the performance of the present model, define the residual signal after time-update as

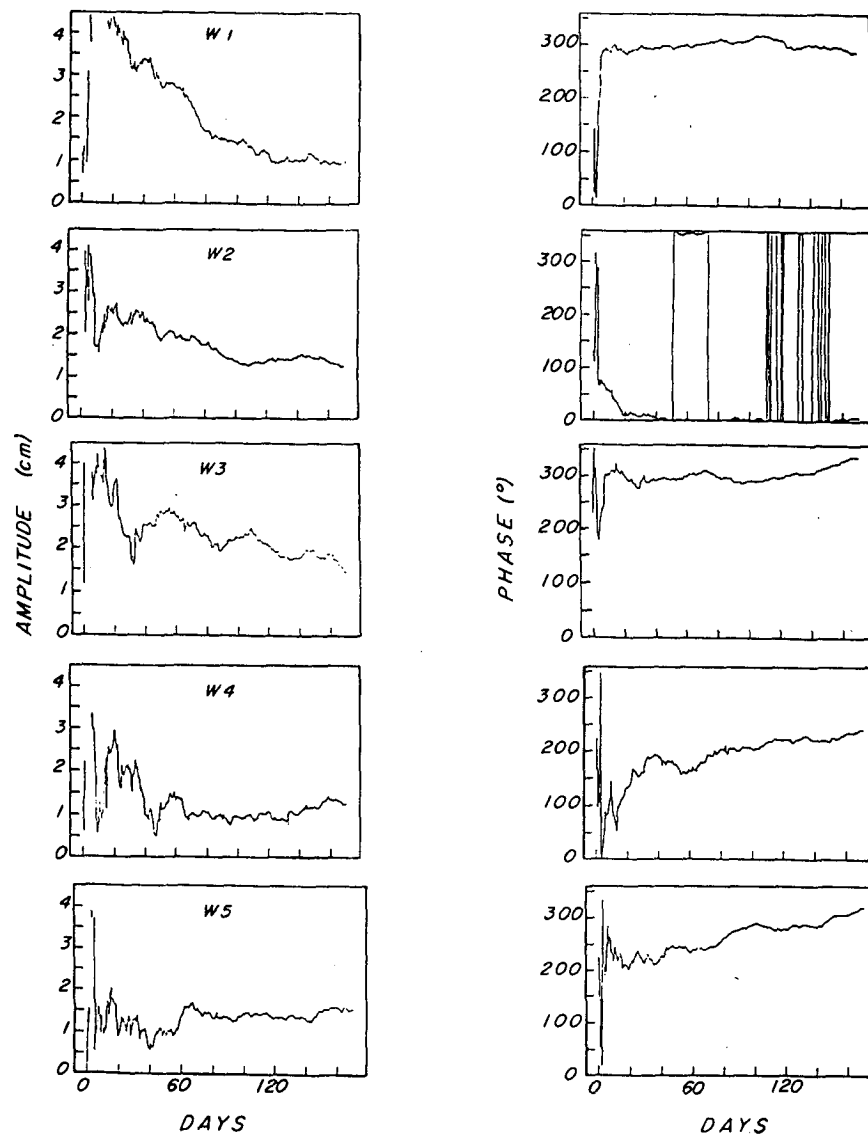


FIG. 9. Evolution of the estimated amplitudes and phases of the five dominant modes in Reference-32. The values shown are those obtained after each measurement-update. The time origin (day 0) is 24 March 1987.

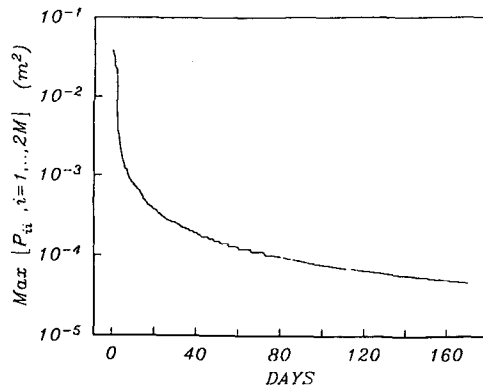


FIG. 10. Evolution of the largest state error variance in Reference-32.

$$n_{tu} = h - \hat{\eta}(k|k-1)$$

where h is any altimetric measurement made at time t_k and $\hat{\eta}(k|k-1)$ is the Rossby wave signal at the measurement location deduced from the predicted state estimate $\hat{q}(k|k-1)$. Similarly, the residual signal after the measurements are included in the update is

$$n_{mu} = h - \hat{\eta}(k|k).$$

Denote the variances of h , n_{tu} and n_{mu} by V , V_{tu} and V_{mu} . The difference $V - V_{mu}$ is the variance explained by the model data combination made by the Kalman filter. This quantity can be split in two parts: $V - V_{tu}$ the variance explained by the model forecast alone, and $V_{tu} - V_{mu}$, the variance explained by the measurement update, i.e. the variance explained by the observations when combined with the forecast.

In Fig. 11, the means of these quantities are presented for each repeat cycle. A large transient is observed during the first repeat cycle. During this cycle, the model forecast is very poor—as shown by the negative value of $V - V_{tu}$. On the other hand, the measurement-update step is very efficient in correcting the forecast. What happens during this first cycle is that the measurement-update process drives the state estimate to fit the data, track by track, using the available horizontal modes. As shown in the previous section, the estimated amplitudes and phases are highly variable. The forecast has little weight because of the large value of the initial state error variance, so that the Rossby wave physics play a minor role in the state estimate. Accordingly, each measurement-update step is very efficient in explaining the variance along each individual track but the forecast along the next track is of poor quality. As progressively more data are included, the state error variance decreases very rapidly (Fig. 10), and the forecast gains more weight. The state estimates and the values of the explained variances eventually stabilize. Some numerical experiments (not shown here) were performed with reduced values of the initial state error covariance p_0 . As expected, they exhibit a

transient of smaller amplitude during the first repeat cycle. Only minor differences subsist after that.

During the nine last repeat cycles, the measurement-update process is always effective in explaining some variance. The difference $V_{tu} - V_{mu}$ keeps values between 10 and 30 cm^2 , the mean being 18 cm^2 . The performance of the model's forecast, as measured by the difference $V - V_{tu}$ oscillates between positive and negative values, the average over the nine last repeat cycles being negative (-9 cm^2), indicating that the model forecast propagates more noise than information. We suspect that most of this noise is carried by the 27 nonsignificant modes. To test that hypothesis, we performed a new experiment (Reference-5) in which only the five dominant modes are retained in the model.

A first interesting result from this experiment is that the estimated phase and amplitude of the waves are almost identical to those obtained in Reference-32. During the last repeat cycle, the differences between Reference-32 and Reference-5 are smaller than 1 mm for the amplitudes and 3° for the phases. This result is consistent with the fact that all state estimates are essentially independent.

The differences $V - V_{tu}$ and $V_{tu} - V_{mu}$ still exhibit a transient behavior during the first repeat cycle. The average values of $V - V_{tu}$, $V_{tu} - V_{mu}$ and $V - V_{mu}$ over the last nine repeat cycles are now 4, 3 and 7 cm^2 , respectively. The model/data combination explains only 6% of the total observed variance. Compared with Reference-32, the variance explained by the time-update is better (at least, it explains some variance) and that of the measurement update is worse. The measurement-update process consists of adjusting each

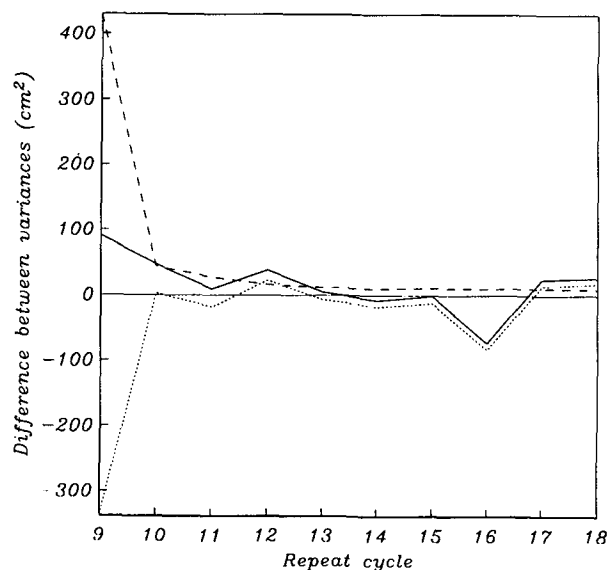


FIG. 11. Mean values of $V - V_{tu}$ (dotted line), $V_{tu} - V_{mu}$ (dashed line) and $V - V_{mu}$ (solid line) for each repeat cycle of Reference-32.

forecast state to obtain a “best fit” of the data. Unsurprisingly, a better fit can be obtained using 32 adjustable modes rather than 5. In addition, it is now clear that the 27 non-significant modes of Reference-32 are used to fit noise rather than Rossby waves. The propagation of these noisy modes according to the governing equation of the Rossby waves acts as an additional noise source that actually damages the forecast performance.

d. Smoothing

Optimal smoothing in the absence of process noise is a special case. Introducing (16) into (22) one obtains:

$$\mathbf{G}_s(k) = \mathbf{A}^{-1}(k) \quad (44)$$

so that (20) and (21) reduce to

$$\hat{\mathbf{q}}(k|N) = \mathbf{A}^{-1}(k)\hat{\mathbf{q}}(k+1|N) \quad (45)$$

$$\mathbf{P}(k|N) = \mathbf{A}^{-1}(k)\mathbf{P}(k+1|N)\mathbf{A}^{-T}(k). \quad (46)$$

Equation (45) indicates that the optimal smoothed state estimate is simply obtained by integrating the dynamical equation (7) backwards in time from the final filtered estimate $\hat{\mathbf{q}}(N|N)$. The reason the backwards integration is optimal is easy to understand. Indeed, the final filtered state estimate $\hat{\mathbf{q}}(N|N)$ is known to have a smaller error covariance matrix than all previous filtered estimates (see Fig. 10). In addition, the state is assumed to evolve in a strictly deterministic manner because no noise is present in the state equation. Therefore, the backward integration of that perfectly deterministic model from the final, best estimated, state should provide the best possible state estimates at all times. More generally, Fraser (1967) showed that the optimal smoothed estimate of the state is better than the estimate obtained with the backwards integration technique only for those states which are controllable by the process noise (see also Gelb 1973). In the absence of process noise, none of the state components is controllable by that noise. The stability of the matrix inversion in (44, 45) will be sensitive to small system eigenvalues, and can be dealt with in a number of conventional ways.

Further simplifications arise with the present dynamical model. The state-transition matrix \mathbf{A} is a rotation matrix and is therefore orthogonal ($\mathbf{A}^{-1} = \mathbf{A}^T$). As a consequence, (6a) and (45) yield

$$\hat{\alpha}_m(k|N) = \hat{\alpha}_m(N|N) \quad (47)$$

$$\hat{\theta}_m(k|N) = \hat{\theta}_m(N|N) \quad (48)$$

while (37) and (46) give

$$E[\alpha_m(k) - \hat{\alpha}_m(k|N)]^2 = E[\alpha_m(N) - \hat{\alpha}_m(N|N)]^2. \quad (49)$$

Thus, the optimal smoother yields wave estimates having an amplitude, initial phase and amplitude error variance that are constant and equal to those provided

by the Kalman filter at the final time t_N . This result is simple, but not robust, as it heavily relies on the unrealistic assumption that the system dynamics are noise free. Even in the absence of driving noise, one expects the dynamical model equation (7) to be an incomplete description of the Rossby wave physics so that substantial deviations from that simple evolution law can occur. Unlike the filter, which accommodates itself to successive observations, the smoother (45) is unable to track such deviations because the model is temporally stationary.

In spite of this, it is still interesting to run the smoother, at least to determine how much of the observed variance can be explained by a few barotropic Rossby waves with constant amplitude. A smoothed estimate was thus made using only the five most significant wave modes identified by the filter. As for the filter, we define n_s the residual signal after smoothing [$n_s = h - \hat{\eta}(k|N)$] and V_s the variance of n_s . The variance explained by the smoother is $V - V_s$. This difference does not exhibit a transient behavior as observed with the filter because smoothing starts from the best filtered estimate $\hat{\mathbf{q}}(N|N)$. On average over the whole dataset, the variance explained by the smoother is 4.5 cm², a satisfying result as this variance is almost exactly equal to the variance of the five smoothed waves.

11. Sensitivity experiments

Although the model only explains a very small fraction of the variance of the altimetric signal, it does succeed in identifying a few Rossby waves whose amplitudes appear to be significantly different from zero. Can we trust the estimated phases and amplitudes of these waves—as we know that the design of the linear wave model and the corresponding Kalman filter is based on several crude assumptions? More specifically, we would like to answer the following questions:

1) A crude description of the measurement noise is used (see section 6) and the data distribution in space/time is very irregular. Are the resulting wave parameter estimates reliable?

2) The orbit error is not directly corrected in the data but rather removed by filtering. Would a direct correction yield better results?

3) The wave model is formulated for waves of constant amplitude while observed waves do grow or decay. Are the state estimates reliable when the amplitudes vary?

4) The frequencies of the waves are determined by the simple dispersion relation (3). Departures from these theoretical values are likely. How do they affect the estimated phases and amplitudes?

5) The preliminary experiment was performed without process noise. How sensitive are the results to that noise? Questions 5 and 3 are directly connected.

To answer these questions, one would prefer to compare the estimated phase and amplitudes of some waves with their known values.

Consider (1). The characteristics of the Rossby waves present in the altimetric data are not known a priori so that a direct evaluation of the results is impossible. However, a known signal associated with a specified Rossby wave can be added to the altimetric data and then (perhaps) retrieved using the filter or the smoother. To this end, the wave vector of the added wave is chosen from among the 32-wave basis set. Denote the wavenumber and corresponding frequency by \mathbf{K} and ω . Choosing an amplitude (α_a) and an initial phase (θ_a), we compute the surface signal associated with that wave, $\alpha_a \sin(\mathbf{K} \cdot \mathbf{X} - \omega t + \theta_a)$, sample it along the tracks and add it to the GEOSAT data. We call this sum the "augmented dataset." In that set, the total signal corresponding to the specified wave vector is

$$\alpha \sin(\mathbf{K} \cdot \mathbf{X} - \omega t + \theta) = \alpha_d \sin(\mathbf{K} \cdot \mathbf{X} - \omega t + \theta_d) + \alpha_a \sin(\mathbf{K} \cdot \mathbf{X} - \omega t + \theta_a) \quad (50)$$

where α_d and θ_d are the amplitude and phase of the Rossby wave signal already present in the original altimetric data. Filtering/smoothing of the original and augmented datasets provides estimates of the original and total Rossby wave signals, respectively. By taking the difference between these two signals for the specified wave vector, one obtains the estimate of the added wave signal. Applying (50) to the estimated signals, basic trigonometric relations yield:

$$\begin{aligned} \hat{\alpha}_a^2 &= (\hat{\alpha} \sin \hat{\theta} - \hat{\alpha}_d \sin \hat{\theta}_d)^2 + (\hat{\alpha} \cos \hat{\theta} - \hat{\alpha}_d \cos \hat{\theta}_d)^2 \\ \sin \hat{\theta}_a &= (\hat{\alpha} \sin \hat{\theta} - \hat{\alpha}_d \sin \hat{\theta}_d) / \hat{\alpha}_a \\ \cos \hat{\theta}_a &= (\hat{\alpha} \cos \hat{\theta} - \hat{\alpha}_d \cos \hat{\theta}_d) / \hat{\alpha}_a. \end{aligned} \quad (51)$$

The comparison of these estimates with the known characteristics of the added wave directly informs us about the performance of the filter/smoother.

Each of the experiments with added waves requires two simulations: a reference simulation with the original altimetric dataset and another one with the augmented dataset. All experiments have been repeated for different values of the process noise variance. For $\sigma^2 = 0$, Reference-32 serves as the reference simulation. Similar reference simulations were done with different values of σ^2 . Smoothing was performed only for $\sigma^2 \neq 0$. The results of the experiments with nonzero process noise will be discussed at length in section 12. Here we concentrate exclusively on the results obtained for the added waves.

a. Waves with constant amplitudes

In a first experiment, the added wave has a constant amplitude of 2 cm and a phase of 270° . Its wave vector is $(-2, -1)$ cycles/ 10^3 km. The chosen amplitude is

typical of the dominant modes found in the preliminary experiment. The amplitude and phase of the added wave retrieved according to (51) is shown in Fig. 12, for the case with no process noise. The result is very satisfying. After 3 days, the estimation error is already smaller than 1 mm for the amplitude and smaller than 1° for the phase. Similar experiments were repeated for different values of the process noise variance ($0 \leq \sigma^2 \leq 10^{-4} \text{ m}^2$) and for different added waves having amplitudes as small as 0.5 cm. The results obtained, for both the filtered and the smoothed estimates, are of the same quality as those shown in Fig. 12. Significant differences between the smoothed and the filtered estimates appear only during the very first days when the filtered estimates are not yet stabilized. This clearly shows that, in spite of the unusual sampling distribution of the altimetric data, the optimal filter and smoother are able to accurately identify Rossby waves having (constant) amplitudes as small as, or even smaller than, 1 cm.

It is important to point out that we do not see here the large initial variability of the estimated amplitudes or phases that we saw in the results of the preliminary experiment (Fig. 9) because the effects of the measurement noise are implicitly removed from the results shown in Fig. 12. Indeed, when estimating the characteristics of the added waves, we compute the difference between the signal estimated from the original altimetric data and the signal estimated from an augmented dataset. The noise is the same in the two sets so that it disappears when the difference is taken. In fact, Fig. 9 indicates that the filter may need several weeks worth of data to eliminate most of the noise. Figure 12 shows that the Rossby wave signal itself is rapidly obtained in the state estimates. That result is only weakly sensitive to the specified variance of the measurement noise as revealed by additional tests (not shown here) in which the variances of the short and long wave measurement noises were doubled and halved. Equation (41) shows that the variance of the estimator error is directly affected by a change in the measurement noise level; nonetheless, it appears (as is usually the case in this type of problem), that the signal so estimated is much less sensitive than its formal error.

b. Effect of the orbit error correction

By specifying a long-wave measurement noise representing the orbit errors, we avoid using an explicit correction scheme. The results of the previous section indicate that this procedure provides excellent estimates of a constant Rossby wave signal. Nevertheless, it is interesting to evaluate the performance of the usual correction technique that consist of removing the bias and tilt of the tracks. To this end, we repeat the experiment described in the previous section, the only difference being that the bias and tilts are subtracted from all tracks and that r_1 is taken equal to zero to

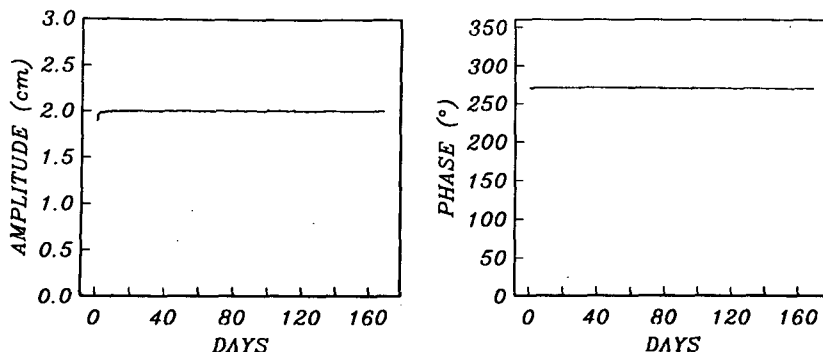


FIG. 12. Estimated amplitude and phase of an added wave with a specified constant amplitude of 2 cm and a phase of 270° .

indicate that there is no long-wave noise left. The results obtained for $\sigma^2 = 0$ are shown in Fig. 13. Other experiments performed with nonzero values of σ^2 yield similar results. The retrieved amplitude of the added wave is close to 1 cm, that is only 50% of the exact signal. The phase estimate is quite good, with an error smaller than 2° . The added wave we used is somewhat special as its crests are (almost) parallel to the ascending tracks, so that by subtracting the bias of the ascending tracks, the signal associated with this wave is suppressed. The characteristics of this wave can thus only be inferred from the descending tracks which represent only one third of the data. In fact, among all selected waves, this one is most affected by the orbit error correction. Nevertheless, the results shown in Fig. 13 are not exceptional. Experiments performed with other added waves yield amplitudes that are typically underestimated by 20% to 40%. In general, the worse estimates are obtained for the waves having the largest along-track wavelengths. Indeed, these long waves contribute significantly to the bias and slope of the tracks so that their signal is substantially altered by the bias and slope removal, justifying the earlier comment (section 7c) that direct orbit error removal procedures may be inadvisable (at least on this spatial scale).

c. Waves with variable amplitudes

Real Rossby waves are likely to deviate from the simple evolution equation (1). In particular, we expect their amplitudes to vary with time. In the following set of experiments, we examine the ability of the filter and smoother to estimate waves with changing amplitudes. The simulations are performed with an added wave whose amplitude first takes a constant value of 2 cm during 60 days. Then, on day 60, the amplitude is abruptly decreased to 1 cm, a value that is maintained through the rest of the experiment. The phase keeps a constant value of 270° and the wave vector is $(-2, 0)$ cycles/ 10^3 km. The estimated amplitude of the added wave is shown in Fig. 14 for three different experiments with $\sigma = 0, 10^{-6}$ and 10^{-4} m^2 , respectively. The estimated phase is not shown as its error is smaller than 1° in all experiments.

In the absence of process noise (Fig. 14a), the estimated amplitude reacts rather slowly to the change, with an e -folding time of about 70 days. The slowness of the response is essentially due to the fact that, in the absence of process noise, the state error covariance matrix becomes rapidly very small (see Fig. 10) so that, in the measurement-update step, the state forecast is

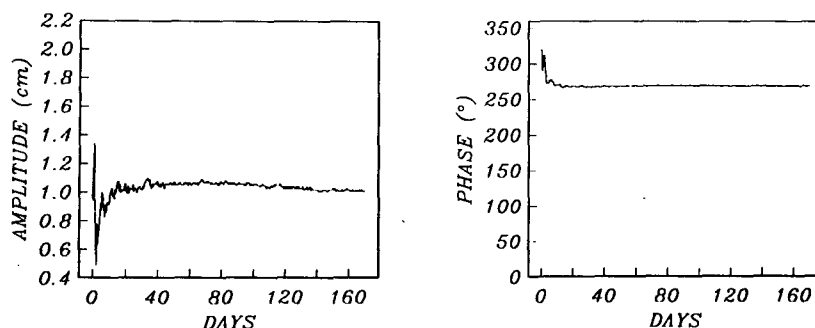


FIG. 13. As in Fig. 12 but the estimates have been made after removal of the bias and tilt of the tracks.

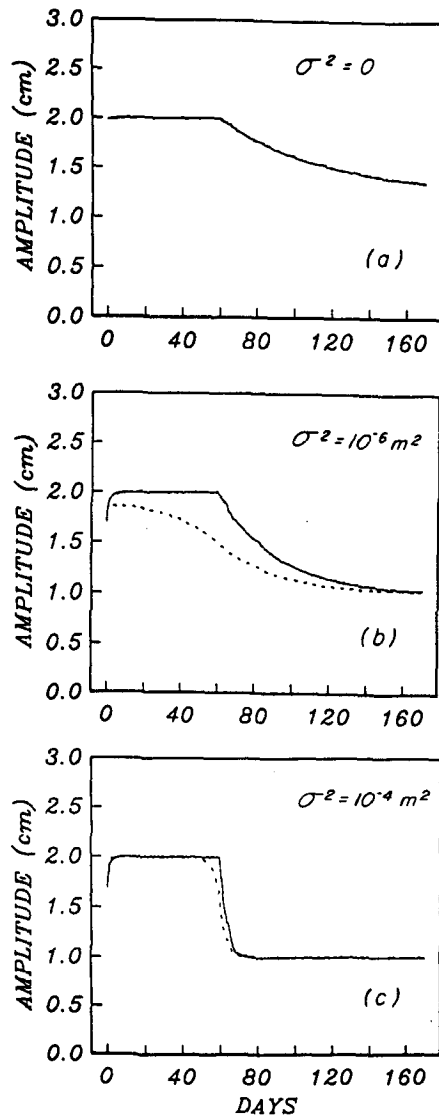


FIG. 14. Estimated amplitude of an added wave with varying amplitude for three different values of the process noise variance (σ^2). The filtered (solid line) and smoothed (dashed line) estimates are presented together. The exact value of the amplitude is 2 cm from day 0 to 60 and 1 cm afterwards.

given a very large weight relative to the observations. Since the state forecast maintains the amplitude constant, it takes a long time for the successive measurement-updates to bring the estimate back to the observed value. When process noise is present, the state error covariance decreases less rapidly as the process noise covariance is systematically added to it (16). Accordingly, the state forecast is given less weight relative to the data so that the filter reacts more quickly to changes in the observations (Fig. 14b, c). For similar reasons, the response of the smoother is faster when σ^2 increases. The response of the smoother is always less abrupt than that of the filter.

d. Waves with perturbed frequencies

The frequencies of Rossby waves in the ocean are likely to differ somewhat from the theoretical values given by (3). Altimetric data also contain many wave-like motions that are not of the Rossby type. Some of these undoubtedly have wavenumber vectors among the selected ones, but their frequencies do not obey (3). In this section, we analyze how the filter and smoother react in the presence of waves whose frequencies deviate from the dispersion relation (3)—i.e., “non-Rossby waves.”

Several experiments were performed with added waves having a perturbed frequency. These waves are of the form

$$\alpha_a \sin[\mathbf{K} \cdot \mathbf{X} - (\omega + \omega')t + \theta_a]. \quad (52)$$

The optimal filter and smoother try to identify these waves under the form

$$\hat{\alpha}_a \sin[\mathbf{K} \cdot \mathbf{X} - \omega t + \hat{\theta}_a] \quad (53)$$

which means that the estimated phase should ideally evolve like

$$\hat{\theta}_a = \theta_a - \omega' t. \quad (54)$$

Two series of experiments were performed with $\omega' = -0.1\omega$ and $\omega' = 0.5\omega$. The first case is representative of Rossby waves with rather small frequency perturbations. The second is more typical of waves that are not of the Rossby type. In all experiments, the added wave has a constant amplitude of 2 cm, an initial phase of 270° and its wave vector is $(-2, 0)$ cycles/ 10^3 km. Here again, the experiments were performed for three different values of σ^2 .

The results for $\omega' = 0.5\omega$ are shown in Fig. 15. Because this added wave is far from the Rossby type, we would like to see its signal filtered out. For $\sigma^2 = 0$ (Fig. 15a), this signal is effectively eliminated as the estimated amplitude goes to zero with damped oscillations of period $2\pi/\omega'$ (here, 103 days). This period is the time needed by the added wave to create a phase difference of 2π relative to the wave with the exact Rossby wave frequency. The estimated phase exhibits well-marked discontinuities occurring with the same period. Between these jumps, the phase decreases almost linearly at a rate smaller than predicted by (54). Estimates of the other waves present are insensitive to the presence of the added wave.

For $\sigma^2 = 10^{-6} \text{ m}^2$, filtering is less efficient in eliminating the signal of the added wave. The smoother clearly performs better as it succeeds in eliminating as much as 85% of the unwanted signal in the middle of the observation period. The observed response of the smoother is consistent with the usual property that, for fixed interval-smoothing, the error covariance of the state is minimum near the middle of the observation interval and maximum at the end points (see the example in Fig. 18). The estimated phase evolves almost

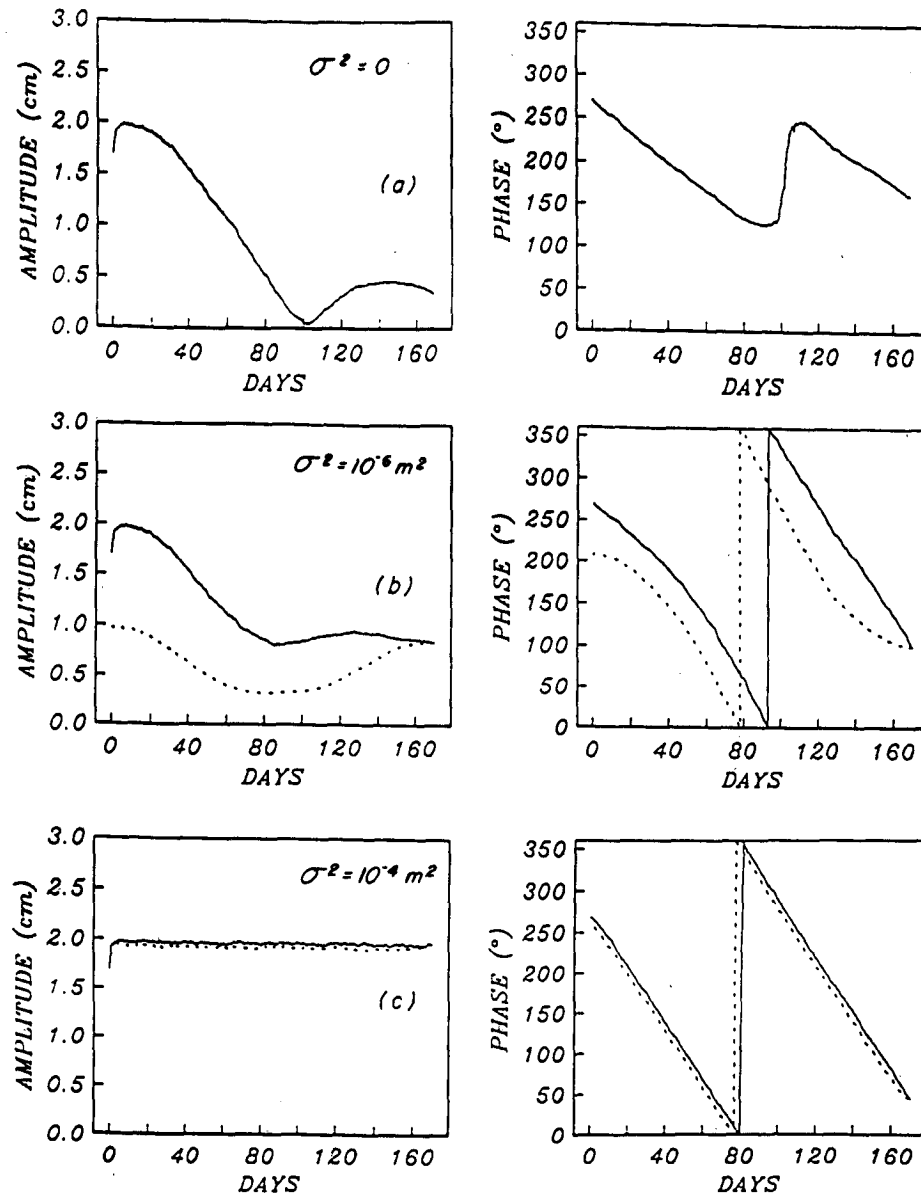


FIG. 15. Estimated amplitude and phase of an added wave with a frequency perturbation $\omega' = 0.5\omega$. The results are shown for three different values of σ^2 . The filtered estimate is the solid line. The smoothed estimate is the dashed line.

linearly as predicted by (54) though the rate of decrease is slightly less than ω' .

For $\sigma^2 = 10^{-4} \text{ m}^2$, neither the filter nor the smoother eliminate a significant fraction of the unwanted signal. In fact, that signal is almost perfectly estimated as the retrieved amplitude is correct within 1 mm and the phase closely follows (54). By increasing the variance of the process noise, one decreases the weight given to the model forecast in the measurement-update process. Accordingly, the dynamical constraint (1) is less strictly enforced and hence less efficient in filtering out the "non-Rossby" waves. For a process noise as large as 10^{-4} m^2 , the state estimates are essentially least-square

fits of the spatial modes to the observations unconstrained by the dynamics. These results are analogous to those obtained in the previous section.

The results obtained for $\omega' = -0.1\omega$ are qualitatively the same as those we have just discussed. The major difference is that the characteristic period $2\pi/|\omega'|$ is considerably longer (515 days) so that, for all values of σ^2 , the estimated amplitude of the added wave decreases more slowly than shown in Fig. 15. For $\sigma^2 = 0$, the final amplitude estimate is still 1.7 cm. For $\sigma^2 = 10^{-6} \text{ m}^2$, the smallest smoothed estimate is 1.8 cm. However, the presence of a wave with a perturbed frequency is clearly indicated by the phase estimate which

exhibits a well-marked linear trend. In this case, the estimated phase changes by more than 90° during the observation period. The results of these sensitivity experiments will be very useful for the interpretation of the results presented in the next section.

12. Standard experiment with process noise

We now turn to the standard experiment (Standard-32) performed with the 32 selected wave modes and a process noise with variance $\sigma^2 = 10^{-6} \text{ m}^2$. As indicated in section 9b, this value of the process noise variance is typical of wind-forced wave perturbations and is adequate for numerical purposes. The results of the sensitivity experiments also show that such a variance is a reasonable compromise between very small values that give an exaggerated weight to the model forecast and larger values that almost completely relax the dynamical constraint.

Because computational load is perhaps the major issue in the practical application of methods such as the ones we are using, it is worth noting the amount of computer time required in this standard case. The entire filtering-smoothing computation, run without any serious attempt at numerical optimization, required about six hours on a SUN-360 workstation. Execution time is dominated by the matrix inversions in (19) and (22) and there are many ways to make these more efficient.

a. Amplitudes and their estimation errors in the (k, l) plane

The filtered amplitudes at the end of the observation period are presented in Fig. 16a. They are larger than in the preliminary experiment (compare with Fig. 8a) but their distribution is similar. Large values appear mainly in the leftmost part of the domain, the largest being in the vicinity of $(-1, -2)$ cycles/ 10^3 km. The standard deviations of the estimation errors are shown in Fig. 16b. As in Fig. 8b, the error variance decreases for increasing values of the wave vector modulus and a ridge of relatively higher values is present along the line $y = x/2$. Notice however that at short wavelengths, the standard deviations of the estimation errors are as small as 0.3 cm in Reference-32. Here they are close to 1 cm and the filter cannot achieve a better accuracy, the limiting factor being the process noise.

The smoothed amplitude estimates at the beginning of the observation period are shown in Fig. 17. They are quite different from the final estimates displayed in Fig. 16a. The variability of the wave amplitudes will be discussed in the next section. The standard deviations for the smoothed estimates are almost identical to those shown in Fig. 16b. This accuracy represents the best attainable for the initial state with the available dataset. Other, later states however are better estimated. Indeed, Fig. 18 shows that the smoother yields the smallest error variances near the middle of the observation interval. These variances are about a factor of

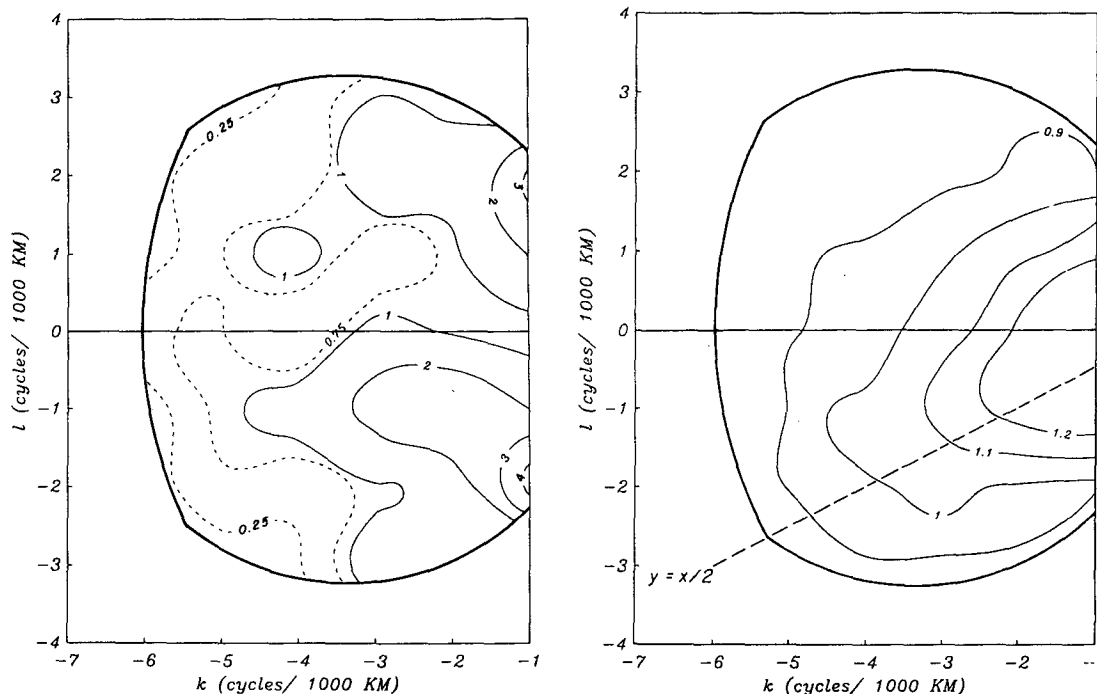


FIG. 16. Final filtered wave amplitudes in cm. (a) and the standard deviations (in cms) of the estimation errors (b) for Standard-32 ($\sigma^2 = 10^{-6} \text{ m}^2$).

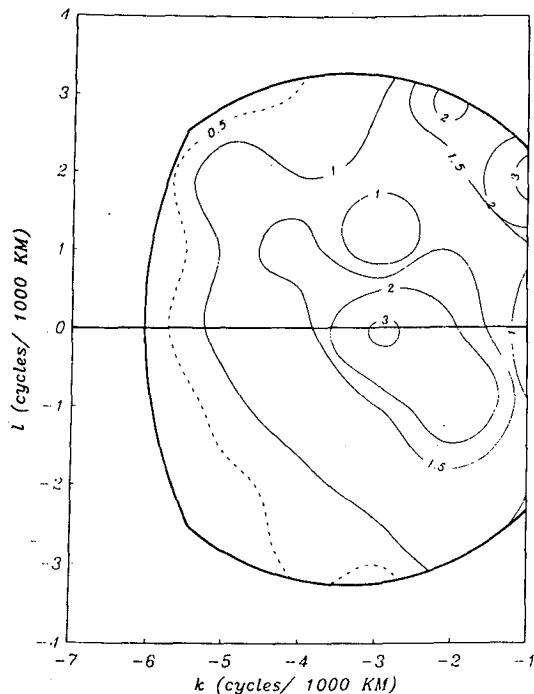


FIG. 17. Estimates of the initial wave amplitudes (cms) given by the smoother for Standard-32.

two smaller than at the end points of the observation interval. Going backwards from the final filtered estimate, we observe that the smoother needs about 50 days to reduce the error variance to a (nearly) minimum value that stays almost unchanged until it finally grows again during the last 50 days. This time scale suggests that state estimates with the smallest possible error variances can be obtained with less data than we used. For example, it is sufficient to filter the data over about 100 days and then smooth backwards during 50 days to obtain a state estimate with nearly the smallest possible error variance (see Fig. 18). In general, minimum variance estimates can be obtained by smoothing over about 50 days, starting from any filtered estimate after day 100. That type of smoothing systematically applied over the same length of time is called fixed-lag smoothing. Efficient algorithms exist to perform that task (see Anderson and Moore 1979). Figure 18 also shows that an estimate of the initial state, as good as the one presented in Fig. 17, is obtained by filtering over the first 100 days and then smoothing backwards to the initial time.

b. Time evolution of the amplitudes and phases

As in Reference-32, the estimated (both filtered and smoothed) amplitudes of most waves do not exceed one standard deviation of the error. They typically oscillate between 0 and 1 cm. However, the identification of dominant modes is not as easy as it was for Refer-

ence-32. Indeed, the amplitude estimates exhibit larger variations in Standard-32 than in Reference-32. Some waves maintain significant amplitudes for extended periods of time but then decay below the significance limit, and eventually grow again (real wave amplitudes are more likely to be variable than constant). Nine waves sustain smoothed amplitudes larger than one standard deviation or 1 cm (whichever is the largest) for more than 50 consecutive days. The five dominant waves identified in Reference-32 are among those. The evolution of their estimated phases and amplitudes is shown in Fig. 19. The smoothed estimates for both the phases and the amplitudes do not exhibit the small short-term fluctuations observed in the filtered estimates. The filtered and smoothed estimates of large amplitude changes are in good agreement when these changes are slow enough. Such is typically the case for W2. When the changes are more rapid (e.g., see W5 around day 140) the response of the smoother is less abrupt than that of the filter, as we noticed in the sensitivity experiments.

Comparing Figs. 9 and 19, we clearly see that, when $\sigma^2 = 0$ (Reference-32), most amplitude variations are strongly damped, in accordance with the results of the sensitivity experiments. Here again, a typical example is the large amplitude increase of W5 near day 140 that is barely noticeable in Reference-32. When the phase estimates are stable and the amplitudes significant, all phase estimates (filtered with $\sigma^2 = 0$, filtered and smoothed with $\sigma^2 = 10^{-6} \text{ m}^2$) are nearly coincident (for example, see W3 before day 100). When the phases exhibit a trend, the rate of change is larger for the estimates obtained with $\sigma^2 = 10^{-6} \text{ m}^2$ than for those obtained with $\sigma = 0$ (see W3 after day 100, W4 and

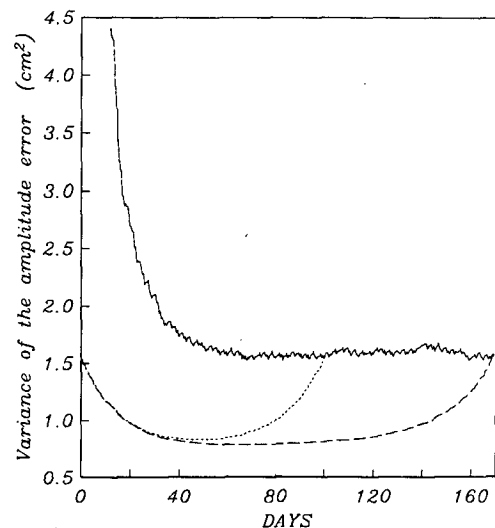


FIG. 18. Evolution of the largest amplitude error variance in Standard-32 for the filter (solid line) and the smoother, smoothing being performed from day 100 (dotted line) or day 170 (dashed line).

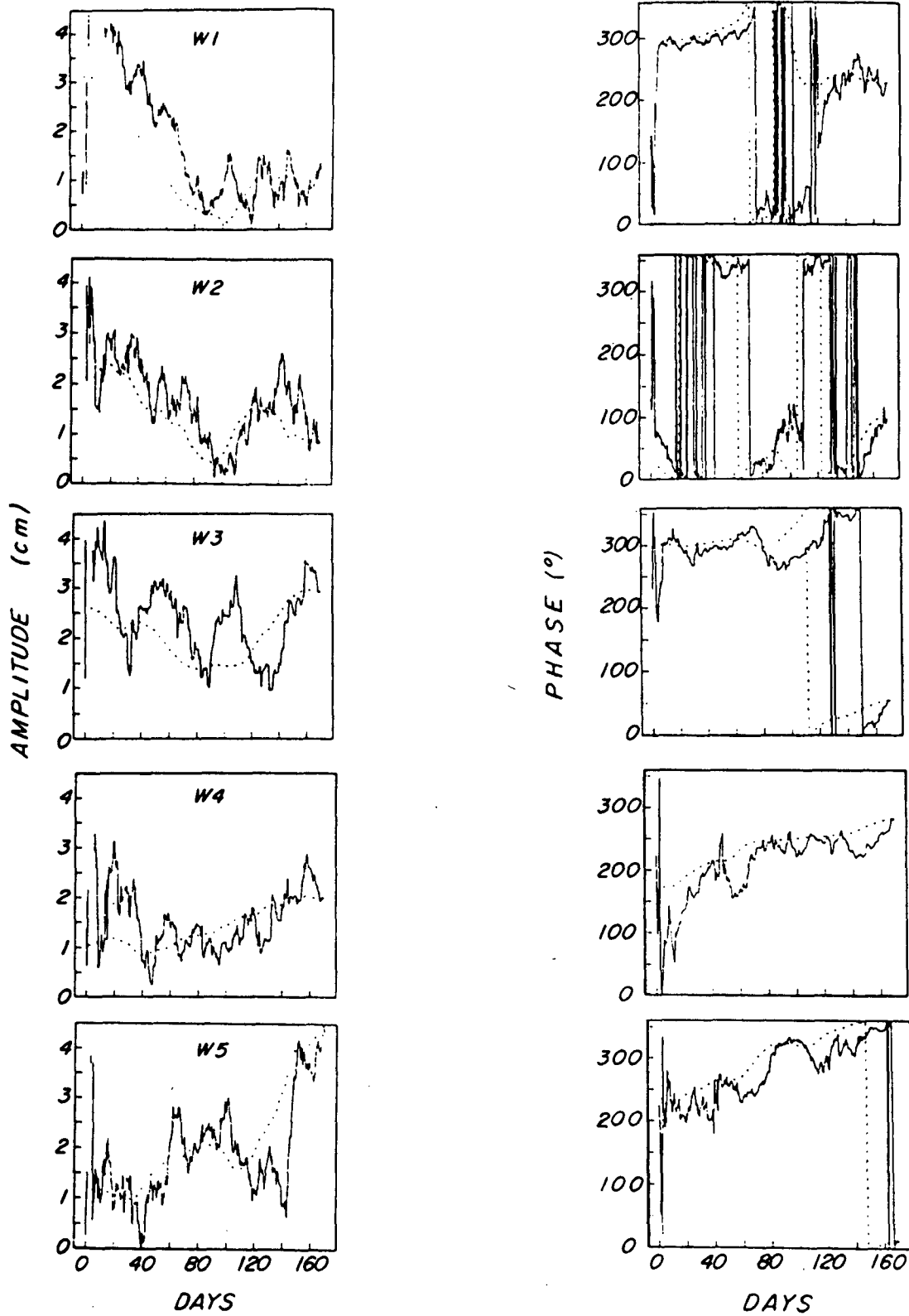


FIG. 19. Evolution of the estimated amplitudes and phases of the five dominant modes in Standard-32. The filtered estimate is the solid line. The smoothed estimate is the dashed line.

W5). For W4 and W5 the trend of the smoothed phases is positive throughout the observation period. According to (54), this result indicates that the frequency perturbation ω' is negative (i.e. the observed waves have longer periods than the corresponding Rossby waves). The estimates of $|\omega'|$ deduced from the rate of change of the smoothed phases of W4 and W5 represent roughly 10% and 30%, respectively, of the corresponding Rossby wave frequencies. The sensitivity experiments show that with $\sigma^2 = 10^{-6} \text{ m}^2$, the rate of change of the smoothed phase is somewhat smaller than the actual value of $|\omega'|$. Therefore, the values of $|\omega'|$ stated above are probably slight underestimates.

c. Explained variance

As in Reference-32, the variance explained by the time-update is negative on average over the nine last repeat cycles (-12 cm^2). Similarly, the results improve when the model is reduced to the five dominant modes (W1 to W5). In that case, (Standard-5) the average values of $V - V_{tu}$ and $V_{tu} - V_{mu}$ are both 9 cm^2 . The total $V - V_{\mu}$, represents 15% of the 120 cm^2 to be explained. On average over the whole dataset, the variance explained by the smoother is 7 cm^2 .

Compared with Reference-5, the variances explained in Standard-5 by the forecast, the measurement-update and the smoother are about doubled. Additional experiments were thus performed with the reduced model to determine if it was possible to further increase the explained variances by tuning the value of the process noise variance. The results are summarized in Fig. 20. When σ^2 increases, the dynamical constraint is progressively relaxed so that the available wave modes can be more easily adjusted to fit the data during the measurement-update step. Accordingly, we observe that the variance explained by the measurement-update is an increasing function of σ^2 . The variance explained by the forecast (time-update) first increases with σ^2 to reach a maximum at about $\sigma^2 = 10^{-6} \text{ m}^2$. Then, it decreases rapidly and becomes negative for $\sigma^2 \geq 2 \times 10^{-5} \text{ m}^2$. Interestingly, the value of σ^2 that maximizes the variance explained by the model forecast is typical of wind-forced perturbations (section 9b). For smaller values of the process noise variance the data have a too small weight compared to the forecast so that they cannot correct for the defects of the model. For larger values, the dynamical constraint is not given enough weight so that little noise is filtered out. The noise is then propagated with the phase speed of the Rossby waves and this very rapidly reduces the forecast performance. As shown by (20), the smoother combines the state estimates of the time and measurements update with the previous smoothed estimate. As for the time and measurement update, the variance explained by the smoother first increases with σ^2 and then stabilizes for $2 \times 10^{-7} \text{ m}^2 \leq \sigma^2 \leq 5 \times 10^{-6} \text{ m}^2$. The forecast starts deteriorating for the largest values

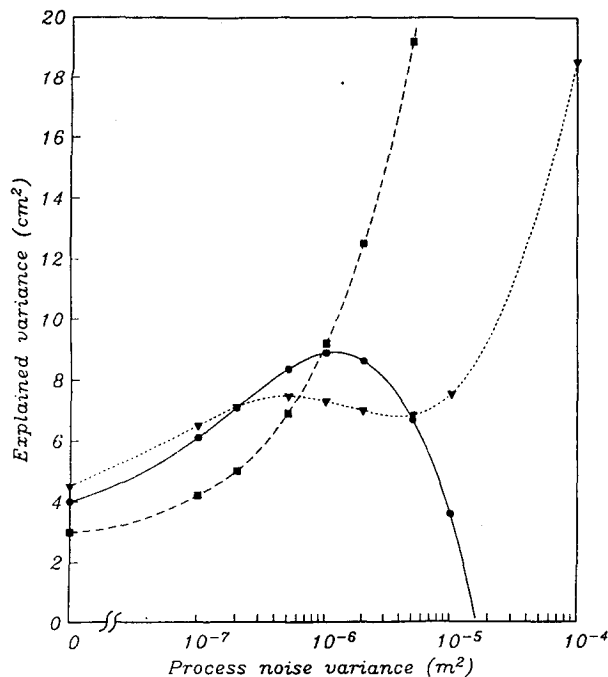


FIG. 20. Evolution of the variance explained by the time-update (solid line), the measurement-update (dashed line) and the smoother (dotted line) as a function of the specified process noise variance σ^2 .

of σ^2 in that range while the variance explained by the measurements update increases rapidly. The two effects compensate until σ^2 is large enough ($\geq 5 \times 10^{-6} \text{ m}^2$) for the dynamical constraint to become almost ineffective. Then, the smoother essentially behaves like the measurement-update. In summation, it appears that the choice $\sigma^2 = 10^{-6} \text{ m}^2$, that maximizes the variance explained by the model forecast, is appropriate for both the filter and the smoother.

13. Summary and final discussion

GEOSAT altimeter data from the western North Atlantic have been processed for the period 24 March to 9 September 1987. The mesoscale variability essentially appears as a red noise (spectral slope close to -2) at wavelengths shorter than 166 km.

These altimetric data are combined with a simple linear barotropic Rossby wave model using optimal estimation techniques such as Kalman filtering and fixed-interval smoothing. The data are essentially used as a large-scale constraint, the mesoscale signal being treated as a measurement noise. Though the model is primitive, it does permit us to pose the following question: "Is there any fraction of the oceanic variability in the northwest Atlantic which is consistent with linear barotropic Rossby waves?" By "consistent," we mean that the observed variability so described is indistinguishable from what the dynamical equation (1) demands. In the present case, the answer is "yes" that a

small fraction of the surface signal to be explained (about 6% to 15%, depending on the estimator used) appears to be this type of mode—a result of some interest.

The model only explains a small fraction of the surface variability, but succeeds in identifying a few significant barotropic Rossby modes with amplitudes of a few centimeters. To estimate these waves, we do not correct explicitly for the orbit error in the data but simply consider it as a long-wavelength measurement noise. Sensitivity experiments show that, using that technique, Rossby waves with constant amplitudes are almost perfectly estimated. Instead, using the widespread “bias and tilt removal technique” to correct for the orbit errors, one can eliminate as much as 50% of the oceanic signal associated with some Rossby modes.

The model itself has some process noise. The problem of specifying the variance of that noise σ^2 covers many aspects that have been analyzed in different experiments. When a very small σ^2 is chosen the estimators (filter or smoother) give a large weight to the model estimate and a relatively small weight to the observations. In that case, the dynamical constraint (1) is strictly enforced so that waves which frequencies do not obey the dispersion relation (3) are quite rapidly filtered out. On the other hand, since the model always maintains constant wave amplitudes, both the filter and the smoother then react very slowly to changes in the observed wave characteristics. Accordingly, the variability of the estimated amplitudes is weak. If a larger σ^2 is chosen, the weight of the model forecast diminishes and the estimators react more rapidly to observed changes. However, the filter is less efficient in eliminating waves with incorrect frequencies. The smoother is more efficient provided that σ^2 is not too large.

The variance explained by the model forecast is also sensitive to the chosen value of σ^2 . For very small values of σ^2 , the observations have such a small weight that the measurement-update can hardly correct for the model defects. The forecast explains little variance. For large values of σ^2 , the dynamical constraint is not efficient enough to filter out noisy wave modes. The model propagates that noise with the phase speed of Rossby waves and this quickly reduces the model performance. Between these two situations, an optimal value of σ^2 exists for which the variance explained by the model forecast is maximum. Interestingly, that value of σ^2 (10^{-6} m^2) is typical of wind-forced perturbations.

Compared with the filter, the smoother generally leads to a reduction of the estimation error variance by a factor of two. A considerably larger error reduction is obtained near the initial observation time. Fixed-interval smoothing has been used here and produces useful information about the data needed to achieve the best possible error reduction. Such an information can be used to devise a fixed-lag smoother.

The model employed is extremely primitive, and one can argue that its use was misguided. Indeed, a referee has suggested that we should have run the experiments described here again, with the sign of β reversed, on the grounds that we might explain more data variance than the small amount we do account for. Given that the focus region includes the Gulf Stream, and is probably energetically dominated by Gulf Stream meanders moving eastward, such a supposition may indeed be correct. We do not think that much would be proven by such a demonstration, however. The more interesting questions are raised by models such as the present one which fail to agree with the observations. No oceanic model is ever likely to be in complete accord with the data, and one must interpret the failure. In the present case, we have shown quantitatively that a measurable fraction of the apparent oceanic motions are consistent with linear barotropic Rossby waves. We would be the first to agree that this consistency is not the same as a proof that the motions are barotropic Rossby waves. But consistency is all that can ever be proven—even had we ascribed it to 90% of the variance rather than only 10%. We do think the results of these simple experiments are sufficiently encouraging to justify future experiments with more complex and realistic dynamical models and more rigorous data reduction procedures. Such efforts are now underway.

Acknowledgments. Supported in part by the Jet Propulsion Laboratory through Contract 958125, and the Office of Naval Research.

REFERENCES

- Anderson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering*. Prentice-Hall, 357 pp.
- Bryson, A. E., and L. J. Henrikson, 1968: Estimation using sampled-data containing sequentially correlated noise. *AIAA J. Spacecraft Rockets*, **5**, 662–665.
- , and Y.-C. Ho, 1975: *Applied Optimal Control*. Hemisphere, 481 pp.
- Campbell, J. W., 1988: Evaluation of GEOSAT data and application to variability of the Northeast Pacific Ocean. S.M. thesis, MIT/WHOI Joint Program in Oceanography, 163 pp.
- Cheney, R. E., and J. G. Marsh, 1981: SEASAT altimeter observations of dynamic topography in the Gulf Stream region. *J. Geophys. Res.*, **86**, 473–483.
- De Mey, P., and A. R. Robinson, 1987: Assimilation of altimeter eddy fields in a limited-area quasi-geostrophic model. *J. Phys. Oceanogr.*, **17**, 2280–2293.
- Fraser, D. C., 1967: A new technique for the optimal smoothing of data. D.Sc. thesis, Massachusetts Institute of Technology, 192 pp.
- Fu, L.-L., 1983: On the wavenumber spectrum of the oceanic mesoscale variability observed by the SEASAT altimeter. *J. Geophys. Res.*, **88**, 4331–4341.
- Gelb, A., Ed., 1974: *Applied Optimal Estimation*. MIT Press, 374 pp.
- Ghil, M., S. Cohn, J. Tavantzis, K. P. Bube and E. Issacson, 1981: Applications of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods*. L. Bengtsson, M. Ghil and E. Kallen, Eds., Springer-Verlag, 330 pp.

- Hogg, N. G., 1988: Stochastic wave radiation by the Gulf Stream. *J. Phys. Oceanogr.*, **18**, 1687–1701.
- Hurlburt, H. E., 1986: Dynamic transfer of simulated altimetric data into subsurface information by a numerical ocean model. *J. Geophys. Res.*, **91**(C2), 2372–2400.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng., Trans. ASME*, **82D**, 35–45.
- Liebelt, P. B., 1967: *An Introduction to Optimal Estimation*. Addison-Wesley, 273 pp.
- Longuet-Higgins, M. S., 1965: The response of a stratified ocean to stationary or moving wind-systems. *Deep-Sea Res.*, **12**, 923–973.
- Malanotte-Rizzoli, P., and W. R. Holland, 1986: Data constraints applied to models of the ocean general circulation. Part I: The steady case. *J. Phys. Oceanogr.*, **16**, 1665–1682.
- , and —, 1989: Along-track assimilation of altimeter data into an ocean circulation model: Space versus time resolution studies. *J. Phys. Oceanogr.*, **19**, 1507–1534.
- Marshall, J. C., 1985: Determining the ocean circulation and improving the geoid from satellite altimetry. *J. Phys. Oceanogr.*, **15**, 330–349.
- Meditch, J. S., 1973: A survey of data smoothing for linear and non-linear dynamic systems. *Automatica*, **9**, 151–162 (reprinted in Sorenson 1985).
- 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).
- Sailor, R. V., and A. R. LeSchack, 1987: Preliminary determination of the GEOSAT radar altimeter noise spectrum. Johns Hopkins APL Tech. Digest, **8**, 182–183.
- Schwiderski, E. W., 1980: On charting global tides. *Rev. Geophys. Space Phys.*, **18**, 243–268.
- Smith, S. L., G. B. West and C. W. Malyevac, 1987: Determination of ocean geodetic data from GEOSAT. Johns Hopkins APL Tech. Digest, **8**, 197–200.
- Sorenson, H. W., Ed., 1985: *Kalman Filtering: Theory and Application*. IEEE Press, 457 pp.
- Tai, C.-K., 1989: Error assessments of widely-used orbit error approximations in satellite altimetry. *J. Atmos. Oceanic Technol.*, **6**, 147–150.
- , 1988: Geosat crossover analysis in the Tropical Pacific. Part 1: Constrained sinusoidal crossover adjustment. *J. Geophys. Res.*, **93**, 10 621–10 629.
- Thacker, W. C., 1986: Relationship between statistical and deterministic methods of data assimilation. *Variational Methods in Geosciences*, Y.-K. Sasaki, Ed., Elsevier, 173–179.
- Thompson, J. D., G. H. Born and G. A. Maul, 1983: Collinear track altimetry in the Gulf of Mexico from SEASAT: Measurements, models and surface truth. *J. Geophys. Res.*, **88**, 1625–1639.
- TOPEX Science Working Group, 1981: Satellite altimetric measurements of the ocean. NASA-JPL Report, March 1.
- Tziperman, E., and W. C. Thacker, 1989: An optimal control/adjoint equations approach to studying the oceanic general circulation. *J. Phys. Oceanogr.*, **19**, 1471–1485.
- Webb, D. J., and A. Moore, 1986: Assimilation of altimeter data into ocean models. *J. Phys. Oceanogr.*, **16**, 1901–1913.
- Wunsch, C., 1986: Calibrating an altimeter: How many tide gauges is enough? *J. Atmos. Oceanic Technol.*, **3**, 746–754.
- , 1989a: Sampling characteristics of satellite orbits. *J. Atmos. Oceanic Technol.*, in press.
- , 1989b: Using data with models. Ill-posed and time-dependent ill-posed problems. *Proceedings Ecole de Physique Theorique, on Geophysical Tomography*, Les Houches France, Y. Desaubies, A. Tarantola and J. Zinn-Justin, Eds., in press.
- , and E. M. Gaposchkin, 1980: On using satellite altimetry to determine the general circulation of the oceans with application to geoid improvement. *Rev. Geophys. Space Phys.*, **18**, 725–745.
- , and V. Zlotnicki, 1984: The accuracy of altimetric surfaces. *Geophys. J. Roy. Astron. Soc.*, **78**, 795–808.
- , and S. Imawaki, 1989: An altimetric mean sea surface and geoid for the western North Atlantic constructed by recursive methods, unpublished manuscript.