Assimilation of Altimeter Data into Ocean Models

D. J. WEBB

Institute of Oceanographic Sciences, Wormley, Godalming, Surrey GU8 5UB, United Kingdom

A. MOORE

Department of Atmospheric Sciences, Clarendon Laboratory, Oxford OX1 3PU, United Kingdom

(Manuscript received 5 September 1985, in final form 19 May 1986)

ABSTRACT

The problem of assimilating satellite altimeter data into an ocean model is considered for the case in which the ocean currents are weak, so that they can be represented by a superposition of linear Rossby waves, and the altimeter measurements are exact and available everywhere. The state of the model at each instant is represented by a state vector, and the process of assimilating data is represented by the projection of this vector onto the surface made up of all the model states consistent with the observations. The projection and the evolution of the model between assimilating each batch of data may be represented by a matrix operator, whose eigenvalues characterize the convergence properties of the scheme.

The possibility of using altimeter observations of the ocean surface to determine the deeper structure of the ocean is investigated. It is found to be limited by the phase separation that develops over each assimilation cycle between modes of the ocean with the same horizontal wavenumber but differing vertical structure. If the phase separation is small, as occurs with baroclinic Rossby waves when the assimilation period is 20 days, then the convergence rate may be improved by increasing the assimilation period.

Detailed calculations are made for a midlatitude ocean using a model with a barotropic and two baroclinic modes. Using a period of 100 days between assimilating new data, good phase separation between the vertical modes is achieved when the horizontal scale of the modes is on the order of the Rossby radius (~30 km). The altimeter data is inefficient at separating modes with shorter horizontal scales, modes with a predominant north–south wavenumber, and baroclinic modes with a large horizontal scale. If the assimilation period is reduced to 20 days, the altimeter is better at separating the barotropic mode from the baroclinic modes at large scales. However, in all other respects, the use of a short assimilation period is less effective.

1. Introduction

The problem of assimilating data into ocean models is of interest at present because of the anticipated launch of a number of ocean monitoring satellites (Robinson, 1985; Duchossois, 1983). These satellites will provide much useful information from the altimeters, scatterometers and other instruments carried, and it is planned to make use of the data in experiments such as the World Ocean Circulation Experiment planned for the 1990s (Gautier and Fieux, 1984; Allan, 1983).

Of particular importance for studies of ocean circulation are the radar altimeters, because by measuring the position of the sea surface to an accuracy of a few centimeters they make possible the worldwide study of tides, eddies and other features affecting the surface topography of the ocean. Features with a period of more than a few days are expected to be in geostrophic balance, and so for these the altimeters also give information on the surface current field. Measurements made by the Seasat radar altimeter enabled the Gulf Stream and nearby cold core rings to be detected (Cheney and Marsh, 1981; Bernstein et al., 1982). In the future, improvements in instrument design and satellite tracking should enable weaker features to be observed as well as the large-scale changes in the strength of the ocean gyres (Tapley, 1982; Robinson, 1985).

Given this wealth of information on the surface of the ocean, it is natural to ask whether it can be used to deduce the currents deeper in the ocean. The development of surface features certainly depends on the deeper structure; for example, the speed of Rossby waves depends on their vertical mode number. It might therefore be possible to reverse the process and use the time development of the surface field to deduce the deep structure.

One way to do this is to assimilate the data into a numerical model. If the assimilation method and the model used are good, then as more data is obtained the model should ultimately converge on the real world. In statistics the theory of such data assimilation is usually treated under the heading of estimation theory (Deutsch, 1965), and a number of the techniques developed have been applied successfully to atmospheric (Bube et al., 1981; Ghil et al., 1981) and oceanographic

problems (Marshall, 1985; Timchenko, 1984; Cornuelle et al., 1985). In the present paper, however, although use is made of the projection methods of estimation theory, it is assumed that the measurements and the model are error free. This is done so that the complexities due to random noise do not detract from the other important limits on the assimilation process. A similar assumption was made by Talagrand (1981) in a study of data assimilation into a one-layer reduced-gravity model.

We also assume that the velocities are small enough so that the ocean can be represented by the superposition of linear Rossby waves, each of which propagates independently. For this approximation to be valid, the current speeds in the ocean should be less than the phase speed of the Rossby waves. In practice the two speeds are often of the same magnitude, so the approximation is a poor one. However, it is a good point from which to start analyzing the problem of assimilating oceanic data, and in addition, there are some regions of ocean, for example those near the equator, where the phase speeds of the planetary waves are large enough for the approximation to be valid. Elsewhere in the ocean the effect of topography may also increase the speed of the planetary waves.

At a given instant the numerical model used to assimilate the data may be represented by a state vector in a many dimensional space, whose axes correspond to the different degrees of freedom of the model. The height measurements made by the altimeter define a surface in this space made up of model states that have the same ocean surface elevation. The problem of data assimilation then reduces to finding a projection, from the point representing the model ocean onto the surface defined by the data, which will ultimately give convergence between the model and the real world. If there is more than one suitable projection, we would like to know which one will give the most rapid convergence.

Section 3 of this paper shows how this scheme of assimilating data can be represented by a matrix operator. The rate of convergence of the scheme depends on the eigenvalues of the matrix, which for stability must have moduli less than or equal to one. Eigenvalues with a modulus of one correspond to degrees of freedom for which the assimilation scheme can neither reduce the errors nor make them worse. It is found that even if a scheme is otherwise suitable, such behavior always occurs when two of the oceanic modes with similar horizontal wavenumbers but different vertical structure have phase increments that differ by an integral value of 2π over each assimilation cycle.

It is shown that when the difference from $2\pi n$ is small, the modulus of the eigenvalue, $|\lambda|$, remains near one, and as a result, the efficiency of the assimilation scheme is poor. In practice this problem most often occurs with the higher-order Rossby waves, which have similar frequencies. For these,

$$|\lambda| \approx 1 - \text{const}(\omega_i - \omega_j)^2 \delta t^2,$$

where ω_i and ω_j are the angular velocities of the two modes and δt is the time interval between model updates. This result is used in section 5 to show that it is more efficient to update infrequently, so that the phase separation $(\omega_i - \omega_j)\delta t$ is on the order of one radian, than to update frequently.

In section 6 these results are illustrated for a model that includes the barotropic and the first two baroclinic modes. With an assimilation period of 100 days, the assimilation scheme is best at separating vertical modes with a horizontal scale on the order of the Rossby radius. The assimilation scheme is not good at separating modes with much shorter horizontal scales, modes with wavenumbers predominantly in the north–south direction, or baroclinic modes with much larger horizontal scales. If the assimilation cycle time is reduced to 20 days, barotropic modes with a large horizontal scale are resolved more rapidly, but otherwise the convergence properties of the scheme are worse despite the extra data used.

Finally, section 7 discusses the main results of the paper and some of their implications.

2. Expansion in normal modes

For simplicity we neglect the gravitational part of the ocean spectrum and assume that the ocean current field can be represented by the superposition of a set of linear Rossby waves. Expanding the pressure field in terms of the Rossby waves gives

$$P(\mathbf{x}, z, t) = \int d\mathbf{k} \sum_n \Pi_n(z) P_n(\mathbf{k}) \times \exp[i\mathbf{k}\cdot\mathbf{x} - i\omega_n(k)t] + \text{c.c.}, \quad (2.1)$$

where P is the perturbation pressure, $\mathbf{x}$ the horizontal and z the vertical coordinate, t time, ω the angular velocity of the wave, and c.c. the complex conjugate. The expansion is in terms of an integral over the horizontal wavenumbers $\mathbf{k}$ and a sum over the vertical modes $\Pi_n(z)$.

The vertical modes are defined by the eigenfunction equation (cf. LeBlond and Mysak, 1978),

$$\rho_0\left(\frac{\partial}{\partial z}\frac{1}{\rho_0 N^2}\frac{\partial \Pi_n}{\partial z}\right) + \frac{1}{gh_n}\Pi_n = 0. \quad (2.2)$$

With boundary conditions at the surface and at depth,

$$\frac{\partial \Pi_n}{\partial z} + \frac{N^2}{g}\Pi_n = 0, \quad z = 0$$

$$\frac{\partial \Pi_n}{\partial z} = 0, \quad z = -H.$$

In the ocean ρ_0 is the undisturbed density profile, N the Brunt–Väisälä frequency, g gravity, and the eigenvalue h_n the equivalent depth of the mode, $(gh_n/f^2)^{1/2}$ being its Rossby radius.

For an individual wave $P_n(\mathbf{k})\Pi_n(z)\exp(i\mathbf{k}\cdot\mathbf{x} - i\omega_n(\mathbf{k})t)$, the geostrophic component of horizontal velocity $\mathbf{u}$, the vertical velocity w, and the surface displacement ζ are given by

$$\mathbf{u} = i\frac{\hat{\mathbf{n}}\times\mathbf{k}}{f\rho_0}P_n(\mathbf{k})\Pi_n(z)\exp[i(\mathbf{k}\cdot\mathbf{x}-\omega_n(\mathbf{k})t)],$$

$$w = i\frac{\omega_n(\mathbf{k})}{\rho_0 N^2}P_n(\mathbf{k})\frac{\partial\Pi_n(z)}{\partial z}\exp[i(\mathbf{k}\cdot\mathbf{x}-\omega_n(\mathbf{k})t)],$$

$$\zeta = \frac{1}{\rho_0 g}P_n(\mathbf{k})\Pi_n(z)\exp[i(\mathbf{k}\cdot\mathbf{x}-\omega_n(\mathbf{k})t)], \quad (2.3)$$

where $\hat{\mathbf{n}}$ is the unit vertical vector.

The total surface displacement seen by an altimeter is

$$\zeta(\mathbf{x},t) = \frac{1}{\rho_0 g}\int d\mathbf{k}\sum_n P_n(\mathbf{k})\Pi_n(0) \times\exp[i(\mathbf{k}\cdot\mathbf{x}-\omega_n(\mathbf{k})t] + \text{c.c.} \quad (2.4)$$

We assume that the altimeter makes measurements over the whole ocean. If the surface displacement is expanded in terms of its horizontal wavenumber components,

$$\zeta(\mathbf{x},t) = \int d\mathbf{k}H(\mathbf{k},t)\exp(i\mathbf{k}\cdot\mathbf{x}) + \text{c.c.}, \quad (2.5)$$

then from (2.4) and (2.5),

$$H(\mathbf{k},t) = \frac{1}{\rho_0 g}\sum_n P_n(\mathbf{k})\Pi_n(0)\exp[-i\omega_n(\mathbf{k})t]. \quad (2.6)$$

Hence the problem of assimilating altimeter data becomes one of using Eq. (2.6) to estimate the coefficients P from measurements of H. If the full spectrum of oceanic waves is included, a similar equation is obtained but it also contains terms for the extra gravitational waves.

3. The projection process

We assume that we have an ocean model available that correctly represents the first N vertical modes of the ocean and their evolution in time. In the present paper both the model and the ocean are assumed to be linear; however, the assimilation scheme developed can, in principal, be extended for use with a fully nonlinear model. We also assume that at time zero an initial survey is made with the altimeter and that for each wavenumber $\mathbf{k}$ a set of coefficients $P_n(\mathbf{k})$ is chosen consistent with the data. These coefficients, denoted by the vector $\mathbf{m}_0(\mathbf{k})$, are used to initialize the model. The unknown coefficients representing the true initial state of the real ocean are denoted by the vector $\mathbf{r}_0(\mathbf{k})$.

At a time δt later, the model vector $\mathbf{m}_0(\mathbf{k})$ will have been transformed to

$$\mathbf{m}^-(\mathbf{k}) = \mathbf{C}\mathbf{m}_0(\mathbf{k}), \quad (3.1)$$

where $\mathbf{C}$ is a diagonal matrix giving the change in phase of each mode,

$$C_{jl} = \exp(-i\omega_j\delta t)\delta_{jl},$$

where δ_{jl} is the Kroneker delta function. Here and in the next few sections we will consider only one value of the horizontal wavenumber at a time so the $\mathbf{k}$ variable is dropped. If $\mathbf{h}$ denotes the vector whose jth component is $\Pi_j(0)/\rho g$, then the ocean surface height predicted by the model is

$$H_m = \mathbf{h}\cdot\mathbf{m}^-. \quad (3.2)$$

Similarly, after a time δt, the real ocean vector $\mathbf{r}$ and the observed height H_{obs} are

$$\mathbf{r}(\delta t) = \mathbf{C}\mathbf{r}_0, \quad (3.3)$$

$$H_{\text{obs}} = \mathbf{h}\cdot\mathbf{r}. \quad (3.4)$$

The relationships between $\mathbf{m}^-$, $\mathbf{r}$ and $\mathbf{h}$ are illustrated in Fig. 1 for the case in which the ocean has two vertical modes. Each of the variables involved is complex, so there is also a similar diagram corresponding to their imaginary components.

The model predicts the ocean to have evolved to point $\mathbf{m}^-$, whereas in reality it is at $\mathbf{r}$. The height observation does not define $\mathbf{r}$ exactly, but it does constrain it to lie on the surface S given by

$$H_{\text{obs}} = \mathbf{h}\cdot\mathbf{s}. \quad (3.5)$$

By the Pythagorean theorem, whatever the position of the vector $\mathbf{r}$ on the surface S, the separation between the model vector and $\mathbf{r}$ can be reduced by projecting $\mathbf{m}^-$ along the direction $\mathbf{h}$ normal to S. If the model vector resulting from the projection is denoted by $\mathbf{m}^+$,

$$\mathbf{m}^+ = \mathbf{m}^- + \mathbf{h}(\mathbf{h}\cdot\mathbf{r} - \mathbf{h}\cdot\mathbf{m}^-)/(\mathbf{h}\cdot\mathbf{h}). \quad (3.6)$$

In terms of the vectors at time zero,

$$\mathbf{m}^+ = \mathbf{C}\mathbf{m}_0 + \mathbf{h}(\mathbf{h}\cdot\mathbf{h})^{-1}[\mathbf{h}\cdot\mathbf{C}\cdot(\mathbf{r}_0 - \mathbf{m}_0)]. \quad (3.7)$$

Let the error $\mathbf{e}$ be defined as equal to $(\mathbf{m} - \mathbf{r})$. If its initial value is $\mathbf{e}_0$, then at a time δt later, just before the projection is made, the error $\mathbf{e}^-$ is given by $\mathbf{C}\mathbf{e}_0$. As $\mathbf{C}$ is unitary, the length of $\mathbf{e}^-$ is the same as $\mathbf{e}_0$.

After the projection is made, the error $\mathbf{e}^+$ is

$$\mathbf{e}^+ = (\mathbf{I} - \mathbf{h}(\mathbf{h}\cdot\mathbf{h})^{-1}\mathbf{h}\cdot)\mathbf{e}^-, \quad (3.8)$$

where $\mathbf{I}$ is the unit matrix. It is convenient to write this in the form

$$\mathbf{e}^+(\delta t) = (\mathbf{I} - \mathbf{P})\mathbf{C}\mathbf{e}_0, \quad (3.9)$$

where $P_{ij} = h_i(\sum h_k h_k)^{-1}h_j$. Iterating the assimilation cycle n times, the error after a time $n\delta t$ is given by

$$\mathbf{e}^+(n\delta t) = [(\mathbf{I} - \mathbf{P})\mathbf{C}]^n\mathbf{e}_0. \quad (3.10)$$

From 3.8, the reduction in the error variance during each assimilation cycle is

$$|\mathbf{e}^-|^2 - |\mathbf{e}^+|^2 = |\mathbf{P}\mathbf{e}^-|^2. \quad (3.11)$$

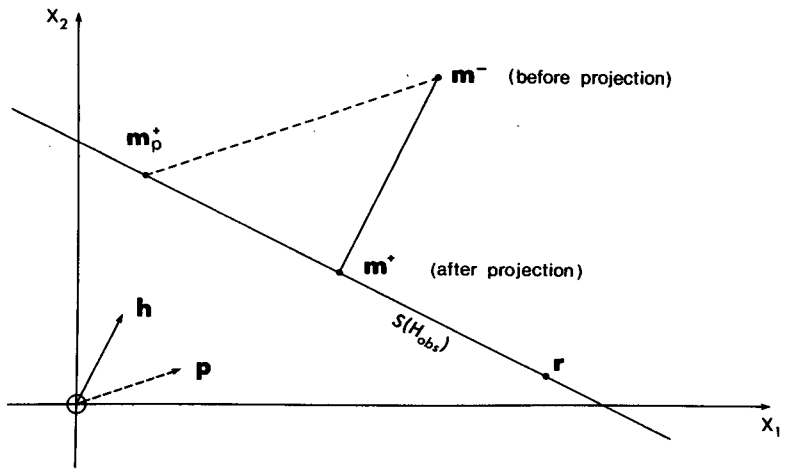

FIG. 1. The projection vector for an ocean with two vertical modes. Here $\mathbf{m}^-$ is the model ocean coordinate before projection, $\mathbf{r}$ the real ocean coordinate, S the surface defined by the observed ocean surface elevations H_{obs}, $\mathbf{h}$ the vector normal to S, $\mathbf{m}^+$ the model coordinate after projection parallel to $\mathbf{h}$, $\mathbf{p}$ an alternative projection, and $\mathbf{m}_p^+$ the resulting model coordinate. (Note: This diagram corresponds to the real part of the coordinates; there will be a similar one for the imaginary parts.)

The value $\mathbf{Pe}^-$, the length of the projection vector from $\mathbf{m}^-$ to $\mathbf{m}^+$ (Fig. 1), is also proportional to the difference between the sea surface elevation predicted by the model and that observed. The model error is therefore reduced most where this difference is at a maximum at the moment the projection is made.

It is of interest to note that Eq. (3.8) is closely related to the Kalman–Bucy and Weiner filters (Deutsch, 1965; Sorenson, 1966), which have been used for data assimilation in atmospheric models (Ghil et al., 1981). The Kalman–Bucy equations corresponding to (3.8) are

$$\mathbf{e}^+ = [\mathbf{I} - (\mathbf{Sh})(\mathbf{h}\cdot\mathbf{Sh} - \mathbf{R})^{-1}(\mathbf{Sh})^*\cdot]\mathbf{e}^-, \quad (3.12a)$$

$$\mathbf{S}^+ = [\mathbf{I} - (\mathbf{Sh})(\mathbf{h}\cdot\mathbf{Sh} - \mathbf{R})^{-1}(\mathbf{Sh})^*\cdot]\mathbf{S}, \quad (3.12b)$$

where

$$\mathbf{e}^- = \mathbf{Ce}_0, \quad \mathbf{S} = \mathbf{CS}_0.$$

Here $\mathbf{e}_0$ and $\mathbf{S}_0$ are the error vector and error covariance matrix error at the end of the previous assimilation cycle, and $\mathbf{R}$ is the covariance matrix of the observational errors. Because Eq. (3.12b) is independent of the detailed behavior of the model, the matrix $\mathbf{S}$ may converge to an asymptopic form. If this is used with Eq. (3.12a), it gives the Weiner filter.

In the assimilation scheme studied in this paper, the observational errors are assumed to be zero, so the matrix $\mathbf{R}$ is zero. The use of a fixed projection $\mathbf{h}$ is then equivalent to using the Kalman–Bucy equations, but with $\mathbf{S}$ always equal to the unit diagonal matrix.

The eigenvalue problem. Consider the eigenvalue problem

$$[(\mathbf{I} - \mathbf{P})\mathbf{C} - \lambda\mathbf{I}]\mathbf{x} = 0, \quad (3.13)$$

with eigenvalues λ_m and eigenvectors $\mathbf{x}_m$. If the initial error is expanded in the form

$$\mathbf{e}_0 = \sum_m \alpha_m \mathbf{x}_m, \quad (3.14)$$

then

$$e(n\delta t) = [(\mathbf{I} - \mathbf{P})\mathbf{C}]^n \sum_m \alpha_m \mathbf{x}_m = \sum_m (\lambda_m)^n \alpha_m \mathbf{x}_m. \quad (3.15)$$

Thus the rate at which the data assimilation process reduces the initial errors depends on the magnitude of the eigenvalues. If they are all less than one, the model will eventually converge to the real ocean state.

In general, the eigenvalues and eigenvectors depend in a complicated way on the variables h_i, ω_i and δt. However, it is easy to show that one of the eigenvalues is always zero and the corresponding eigenvector is $\mathbf{C}^{-1}\mathbf{h}$. This eigenvalue arises because at every time step one has complete information in the direction of the vector $\mathbf{h}$.

To obtain constraints on the other eigenvalues, rewrite (3.13) in the form

$$\lambda_m \mathbf{x}_m = (\mathbf{I} - \mathbf{P})\mathbf{C}\mathbf{x}_m. \quad (3.16)$$

Multiplying (3.16) by its complex conjugate, one obtains

$$|\lambda_m|^2 |\mathbf{x}_m|^2 = |\mathbf{x}_m|^2 - |\mathbf{PCx}_m|^2. \quad (3.17)$$

All of the terms in this equation are positive, so $|\lambda_m|$ must be equal to one if $\mathbf{PCx}_m$ is zero and less than one if $\mathbf{PCx}_m$ is nonzero.

Values of $|\lambda_m|$ equal to one are of interest because they correspond to degrees of freedom of the ocean whose amplitude cannot be determined from the altimeter data. For these eigenvectors, from Eq. (3.17),

$$\mathbf{PCx}_m = 0, \quad (3.18a)$$

$$\mathbf{h}\cdot\mathbf{Cx}_m = 0. \quad (3.18b)$$

Using Eq. (3.16),

$$\mathbf{C}\mathbf{x}_m = \lambda_m \mathbf{x}_m. \tag{3.19}$$

If $(\mathbf{x}_m)_j$ represents the jth component of the vector $\mathbf{x}_m$, then Eqs. (3.18) become

$$\sum_j h_j \exp(i\delta_j)(\mathbf{x}_m)_j = 0, \tag{3.20}$$

$$\exp(i\delta_j)(\mathbf{x}_m)_j = \lambda_m(\mathbf{x}_m)_j \quad \text{for all } j. \tag{3.21}$$

The phase δ_j is defined as equal to $-w_j\delta t$. From Eq. (3.21), all the components of $\mathbf{x}_m$ will be zero except for those for which λ_m equals $\exp(i\delta_j)$.

If only one component $(\mathbf{x}_m)_j$ satisfies (3.21), then from (3.20), the corresponding component of the projection vector $\mathbf{h}_j$ must equal zero. Physically, this will occur if the jth Rossby wave has no effect on the surface elevation. Then the altimeter data cannot improve the initial estimate of its amplitude.

The case of more than one component satisfying (3.21) arises either if there are two or more Rossby waves with the same angular velocity or if their separation in angular velocity produces, over the interval δt, a phase difference between them of $2\pi n$ with n integer. From (3.20) the corresponding eigenvectors must then satisfy the equation

$$\mathbf{h} \cdot \mathbf{x}_m = 0. \tag{3.22}$$

If two components satisfy (3.21), i.e.,

$$\lambda_m = \exp(i\delta_k) = \exp(i\delta_l), \tag{3.23}$$

then the corresponding eigenvector is

$$(\mathbf{x}_m)_j = \frac{1}{h_k}\delta_{jk} - \frac{1}{h_l}\delta_{jl}. \tag{3.24}$$

If there are three components so that

$$\lambda_m = \exp(i\delta_k) = \exp(i\delta_l) = \exp(i\delta_r),$$

then there will be one eigenvector of the form (3.24) and one of the form

$$(\mathbf{x}_m)_j = \frac{1}{h_k}\delta_{jk} + \frac{1}{h_l}\delta_{jl} - \frac{2}{h_r}\delta_{jr}. \tag{3.25}$$

In general, N degenerate components will give $(N-1)$ independent eigenvectors.

We conclude from these results that although a single set of altimeter observations cannot distinguish between the different vertical modes present, it is possible to use a sequence of observations to do this, as long as all the modes present affect the ocean surface elevation and have different phase changes, modulo 2π, between one set of observations and the next.

4. Useful results

a. Systems of two or three vertical modes

In general, analytic solutions of the eigenvalue equation (3.13) are not possible, except for systems of two, three or four vertical modes. If (3.13) is written as

$$(\mathbf{A} - \lambda_m \mathbf{I})\mathbf{x}_m = 0, \tag{4.1}$$

where

$$\left.\begin{aligned} A_{jk} &= (\delta_{jk} - h_j h_k/\gamma)\exp(i\delta_k) \\ \gamma &= \sum_k h_k h_k \end{aligned}\right\}. \tag{4.2}$$

Then the eigenvalues can be found by solving the equation

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{4.3}$$

For a system of two modes this gives a quadratic equation for λ. As in the general case, one of the eigenvalues is zero and the other is given by

$$\begin{aligned}\lambda_2 &= A_{11} + A_{11} \\ &= (1 - h_1^2/\gamma)\exp(i\delta_1) + (h_1^2/\gamma)\exp(i\delta_2).\end{aligned} \tag{4.4}$$

The dependence of λ_2 on h_1, δ_1 and δ_2 is illustrated in Fig. 2. The maximum value of λ_2 occurs when $\delta_1 - \delta_2$, modulo 2π, equals zero.[1] In practice one would want λ_2 to be as small as possible. This occurs when $\delta_1 - \delta_2$, modulo 2π, equals π.

When the ocean has three vertical modes, one of the eigenvalues is again zero. The determinant equation can then be reduced from a cubic equation to a quadratic, the solutions of which give the two other eigenvalues. If

$$\begin{aligned}B = -[&(1 - h_1^2/\gamma)\exp(i\delta_1) + (1 - h_2^2/\gamma)\exp(i\delta_2) \\ &+ (1 - h_3^2/\gamma)\exp(i\delta_3)]\end{aligned}$$

$$\begin{aligned}C = &(h_1^2/\gamma)\exp[i(\delta_2 + \delta_3)] + (h_2^2/\gamma)\exp[i(\delta_3 + \delta_1)] \\ &+ (h_3^2/\gamma)\exp[i(\delta_1 + \delta_2)],\end{aligned}$$

where

$$\gamma = h_1^2 + h_2^2 + h_3^2.$$

Then

$$\lambda_{2,3} = \frac{1}{2}[-B \pm (B^2 - 4C)^{1/2}]. \tag{4.5}$$

The expressions simplify if two of the phase increments δ_i are equal. Thus if δ_2 equals δ_3,

$$\left.\begin{aligned} \lambda_2 &= (1 - h_1^2/\gamma)\exp(i\delta_1) + (h_1^2/\gamma)\exp(i\delta_2) \\ \lambda_3 &= \exp(i\delta_2) \end{aligned}\right\}. \tag{4.6}$$

As expected, having two of the phase increments equal results in one of the eigenvalues having a modulus of one. However, the other eigenvalue is not affected and has the same value as in the case of two modes.

The solution (4.5) can also be expanded to show the behavior of the eigenvalues when the difference between δ_2 and δ_3 is small. This gives, for λ_3,

[1] This constraint has an analogue in time-series analysis where two frequencies cannot be distinguished if they differ by integral multiples of the Nyquist frequency.

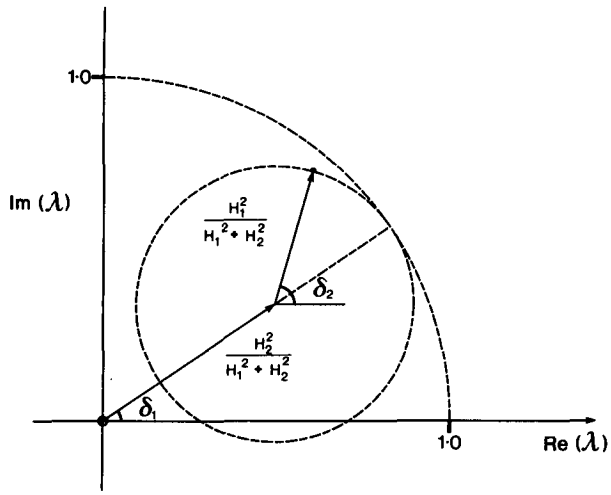

FIG. 2. Argand diagram for the nonzero eigenvalue λ_2 in a system of two vertical modes [Eq. (4.4)]. δ_1 and δ_2 are the phase increments of the two modes.

$$\lambda_3 \approx [1 - h_3^2/(h_2^2 + h_3^2)] \exp(i\delta_3) + h_3^2/(h_2^2 + h_3^2) \exp(i\delta_2). \quad (4.7)$$

If this result is plotted as in Fig. 2, it shows a similar behavior but with h_3 and δ_3 replacing h_1 and δ_1.

b. *Perturbation solutions*

A perturbation solution can be obtained for the case where the phase difference between two or more of the modes is small. Consider the case of N vertical modes, in which the phase increments of the kth and lth modes are such that $(\delta_k - \delta_l)$, modulo 2π, is small. If $\mathbf{A}$ is the matrix operator of Eq. (4.1) with eigenvalues λ_n and eigenvectors $\mathbf{x}_n$, and if $\mathbf{A}^0$ is the operator obtained when the perturbation $(\delta_k - \delta_l)$ is zero, with eigenvalues $\lambda_n{}^0$ and eigenvectors $\mathbf{x}_n{}^0$, then

$$\mathbf{A}\mathbf{x}_n = \lambda_n \mathbf{x}_n, \quad (4.8)$$

$$\mathbf{A}^0 \mathbf{x}_n{}^0 = \lambda_n{}^0 \mathbf{x}_n{}^0. \quad (4.9)$$

As $\mathbf{A}^0$ is not Hermitian, we introduce the eigenvectors $\mathbf{y}_n{}^0$ of the Hermitian adjoint matrix $\mathbf{A}^{0*}$.

$$\mathbf{A}^{0*} \mathbf{y}_n{}^0 = \lambda_n' \mathbf{y}_n{}^0. \quad (4.10)$$

It can be shown (Morse and Feshbach, 1953) that λ_n' equals $(\lambda_n{}^0)^*$ and that, when suitably normalized,[2]

$$(\mathbf{y}_n{}^0)^* \cdot \mathbf{x}_m{}^0 = \delta_{nm}. \quad (4.11)$$

Writing

$$\mathbf{A} = \mathbf{A}^0 + \mathbf{A}', \quad (4.12)$$

and then if δ_k remains fixed while δ_1 varies,

$$A_{ij}' = \delta_{jl}(\delta_{ij} - h_i h_j/\gamma)[\exp(i\delta_l) - \exp(i\delta_k)]. \quad (4.13)$$

The perturbation expansion for λ (Matthews and Walker, 1965) is

$$\left.\begin{aligned} \lambda_n = \lambda_n{}^0 + Q_{nn} + \sum_{m\neq n} \frac{Q_{nm}Q_{mn}}{\lambda_n - \lambda_m{}^0} \\ + \sum_{m\neq n}\sum_{p\neq n} \frac{Q_{nm}Q_{mp}Q_{pn}}{(\lambda_n - \lambda_m{}^0)(\lambda_n - \lambda_p{}^0)} + \cdots \end{aligned}\right\}, \quad (4.14)$$

where

$$Q_{nm} = (\mathbf{y}_m{}^0)^* \cdot \mathbf{A}' \mathbf{x}_n{}^0. \quad (4.15)$$

The values of $\mathbf{x}_n{}^0$ and $\lambda_n{}^0$, needed to calculate the first-order correction, were derived earlier in Eqs. (3.23) and (3.24). Using a similar method, $\mathbf{y}_n{}^0$ can be obtained to give

$$(\mathbf{y}_n{}^0)_j = \left(\frac{1}{h_k}\delta_{jk} - \frac{1}{h_l}\delta_{jl}\right)\left(\frac{1}{h_k{}^2} + \frac{1}{h_l{}^2}\right)^{-1}. \quad (4.16)$$

Substituting in (4.14) gives, to first order,

$$\begin{aligned} \lambda_n &\approx \lambda_n{}^0 + [h_k{}^2/(h_k{}^2 + h_l{}^2)][\exp(i\delta_l) - \exp(i\delta_k)] \\ &= [1 - h_k{}^2/(h_k{}^2 + h_l{}^2)] \exp(i\delta_k) \\ &\quad + [h_k{}^2/(h_k{}^2 + h_l{}^2)] \exp(i\delta_l). \end{aligned} \quad (4.17)$$

Comparing (4.17) with (4.4) shows that the equation for a perturbed eigenvalue is similar to that for a system of two modes. Hence, the dependence of the perturbed eigenvalue on h and δ can be portrayed by a diagram similar to Fig. 2.

c. *Other projections*

So far we have discussed only one projection, in which the model point $\mathbf{m}^-$ (Fig. 1) is projected normally onto the surface S along the direction of the vector $\mathbf{h}$. Other projections are possible, the projection in the direction of a general vector $\mathbf{p}$ giving $\mathbf{m}_p{}^+$, where

$$\mathbf{m}_p{}^+ = \mathbf{m}^- - \mathbf{p}(\mathbf{p}\cdot\mathbf{h})^{-1}(\mathbf{h}\cdot\mathbf{m}^- - \mathbf{h}\cdot\mathbf{r}). \quad (4.18)$$

The error $(m_p{}^+ - r)$ will not be reduced by all projections $\mathbf{p}$. To help identify which projections are suitable, consider a transformation from the coordinate system x_i to a new primed system x_i', in which the scaling of each axis has been changed,

$$x_i' = \beta_i x_i.$$

In the new system, contravariant vectors, such as any vector $\mathbf{s}$ on the surface S, transform as

$$s_i' = \beta_i s_i. \quad (4.19)$$

We assume that the transformation has been chosen so that in the new system $\mathbf{p}'$ is normal to the surface S. Hence,

$$\mathbf{p}' \cdot \mathbf{s}' = H_{\text{obs}}.$$

[2] Equation (4.10) can be written as $\mathbf{C}^*(\mathbf{I} - \mathbf{P})\mathbf{y}^0 = \lambda'\mathbf{y}^0$. The lowest eigenvalue λ_1' is equal to 0, and the corresponding eigenvector is $\mathbf{h}$. Then, from (4.11), $\mathbf{h}\cdot\mathbf{x}_n{}^0$ equals δ_{1n}. Thus, in Eq. (3.13) all of the eigenvectors $\mathbf{x}_n$ except the first are orthogonal to $\mathbf{h}$.

However, in the unprimed system,

$$\mathbf{h} \cdot \mathbf{s} = H_{\text{obs}},$$

so from 4.19,

$$p_j' = h_j/\beta_j. \tag{4.20}$$

The results obtained earlier in this paper can now be used in the primed coordinate system. In particular, a projection in the direction of $\mathbf{p}'$ onto the surface S will reduce the error e', defined as

$$e' = [\sum_j (m_j' - r_j')^2]^{1/2}. \tag{4.21}$$

Transforming back to the unprimed space, a projection along $\mathbf{p}$, where

$$p_j = p_j'/\beta_j = h_j/\beta_j^2,$$

will minimize the error,

$$e' = [\sum_j \beta_j^2 (m_j - r_j)^2]^{1/2}. \tag{4.22}$$

Thus a projection in any direction $\mathbf{p}$ will converge, as long as the coefficients β_j^2 relating $\mathbf{h}$ and $\mathbf{p}$ are positive.

This is a useful result because it gives a degree of freedom that can be used to improve the efficiency of the assimilation scheme. For example, Leith (1980) has proposed that when gravity waves are present and nonlinear effects are important, the projection should be made onto a slow manifold to prevent the generation of free gravity waves. This technique was used by Ghil et al. (1981) in a study using the shallow-water equations. The extra degree of freedom might also be used to satisfy other constraints, such as minimizing the total energy change produced by the projection.

5. The updating strategy

In choosing an assimilation scheme, one has a choice of the interval δt or the projection vector $\mathbf{p}$. In the rest of this paper we concentrate on the choice of δt.

Equation (4.17) shows that if two of the vertical modes have similar angular velocities, such as often occurs with baroclinic Rossby waves in the ocean, each cycle of the assimilation scheme is most efficient if the time interval δt is sufficiently large for the phase separation $\delta_k - \delta_l$ to be on the order of one radian. However, it is not clear whether, when much data is available, it is better to use many assimilation cycles with the interval δt small, or few with δt large.

To investigate this further, we consider the behavior of the modulus of the corresponding eigenvalue $|\lambda_m|$ as the phase separation is varied. When $\delta_k - \delta_l$ is zero, then, from (3.21), $|\lambda_m|$ equals one, and as the phase separation increases, then, from (3.17), $|\lambda_m|$ becomes smaller. Writing $\delta_k - \delta_l$ as $-(\omega_k - \omega_l)\delta t$, this means that for small values of δt, the terms linear in δt must be zero, and if the function is well behaved,

$$|\lambda_m| = 1 - \text{const}(\omega_k - \omega_l)^2 \delta t^2 + \text{O}(\delta t^3). \tag{5.1}$$

This behavior can be confirmed by expanding the first-order perturbation solution for λ_m [Eq. (4.17)],

$$\lambda_m = [h_l^2 \exp(-i\omega_k t) + h_k^2 \exp(-i\omega_l t)]/(h_k^2 + h_l^2),$$

$$|\lambda_m| = [h_k^4 + h_l^4 + 2h_k^2 h_l^2 \cos((\omega_k - \omega_l)\delta t)]^{1/2}/(h_k^2 + h_l^2).$$

Expanding in a power series in δt, one obtains

$$|\lambda_m| = 1 - \frac{h_k^2 h_l^2}{2(h_k^2 + h_l^2)}(\omega_k - \omega_l)^2 \delta t^2 + \text{O}(\delta t^3). \tag{5.2}$$

When the next term in the perturbation expansion (4.14) is included, there are further δt^2 contributions to $|\lambda_m|$, from the coupling to all the other modes present. Numerical calculations show that these give a further reduction in $|\lambda_m|$ similar in magnitude to that predicted by (5.2).

Using this result with (3.12), the reduction in the corresponding component of the error vector during each assimilation cycle is

$$\text{const}(\omega_k - \omega_l)^2 \delta t^2. \tag{5.3}$$

If instead of using a time step of δt, assimilation is carried out n times during this period with time step $\delta t/n$, then the error reduction during each short cycle is

$$\text{const}(\omega_k - \omega_l)^2 \delta t^2/n^2$$

and the reduction after n cycles is

$$\text{const}(\omega_k - \omega_l)^2 \delta t^2/n. \tag{5.4}$$

Comparing (5.3) and (5.4), it is seen that although a reduction in the time interval δt increases the amount of information used, it actually decreases the rate at which the error is reduced. This is an important result, and it means that when trying to distinguish modes of similar angular velocity one should lengthen the cycle time δt, if possible, to a value such that the two modes have a difference in phase on the order of one radian. A result similar to this was obtained by Bube and Ghil (1981) in a study that introduced height information into a single-layer reduced-gravity model. It should also be noted that the result is not connected with the problem of allowing time for high-frequency gravity-wave modes to decay, which limits the time interval used in some meteorological assimilation schemes (Williamson and Dickinson, 1972).

6. Application to a Rossby wave spectrum

The theory developed so far may be applied to many physical systems composed of independent oscillators. As an example of how the scheme performs with a field of Rossby waves, we will consider an ocean with constant Brunt-Väisälä frequency $N(z)$. As discussed by LeBlond and Mysak (1978), in the limit where $\epsilon(=N^2H/2g)$ is small the equivalent depths and eigenfunctions of the vertical eigenfunction equation (2.2) are, for the barotropic mode,

$$\Pi_0(z) = A_0; \quad h_0 = H, \tag{6.1}$$

and for the baroclinic modes,

$$\left.\begin{aligned}\Pi_n(z) &= A_n \cos(n\pi z/H)\\ h_n &= H(N^2H/gn^2\pi^2)\end{aligned}\right\}. \tag{6.2}$$

The dispersion relation for Rossby waves is

$$\omega_n(\mathbf{k}) = \frac{-\beta k_x}{\mathbf{k}\cdot\mathbf{k} + (f^2/gh_n)}. \tag{6.3}$$

In these equations H is the ocean depth, g gravity, f the Coriolis parameter ($=2\Omega \sin\theta$), where Ω is the Earth's angular velocity and θ the latitude; β is the northward gradient of f ($=2\Omega \cos\theta/R$) where R is the Earth's radius. The A are normalization constants.

We consider a case where data is being assimilated for the barotropic and first two baroclinic modes. The value of N is taken to be 2×10^{-4} s^{-1}, the depth H as 4000 m, and the latitude as 40°. If the constants A are all taken to equal one, then the vector $\mathbf{h}$ used in the projection (3.6) has all its components h_n equal to $1/\rho g$.

The results for an assimilation cycle time δt of 100 days are shown in Figs. 3 and 5. Figure 3 shows the values of $|\lambda|$ on the negative k_x axis, and in Fig. 5 the values are contoured for a region of the wavenumber plane.

In these figures, the changes in $|\lambda|$ arise only from the differences in phase that develop between the Rossby waves over the 100 day period. On the k_x axis at a wavenumber of $(-k_0, 0)$, the phases of the barotropic and the two baroclinic waves after this period are 236°, 118° and 47°, respectively. The differences in phase are reasonably large and result in values of $|\lambda_2|$ and $|\lambda_3|$ of 0.4 and 0.7. Inspection of the eigenvectors shows that λ_2 corresponds roughly to the separation of the barotropic mode from the two baroclinic modes. Thus for waves of about 170 km wavelength after 300 days, the barotropic mode would be determined to an accuracy of a few percent, and after 600 days the two baroclinic modes would be determined to a similar accuracy.

At shorter wavelengths, or higher wavenumbers, the angular velocities of the barotropic and first baroclinic modes are reduced and that of the second baroclinic mode increases slightly. As a result, at a wavenumber of $(-2k_0, 0)$, the phase increments of the three waves during each time step are 118°, 94° and 59°, respectively. The smaller spread in phase results in $|\lambda_2|$ and $|\lambda_3|$ having values of 0.94 and 0.97. These values are much larger than before and mean that in order to determine the amplitudes of the waves to an accuracy of a few percent, an assimilation run lasting for a period of 10 to 20 years would be needed. However, if a much larger value of δt were used, then faster convergence rates could be obtained.

At lower wavenumbers, the angular velocity of the baroclinic modes decreases, but that of the barotropic

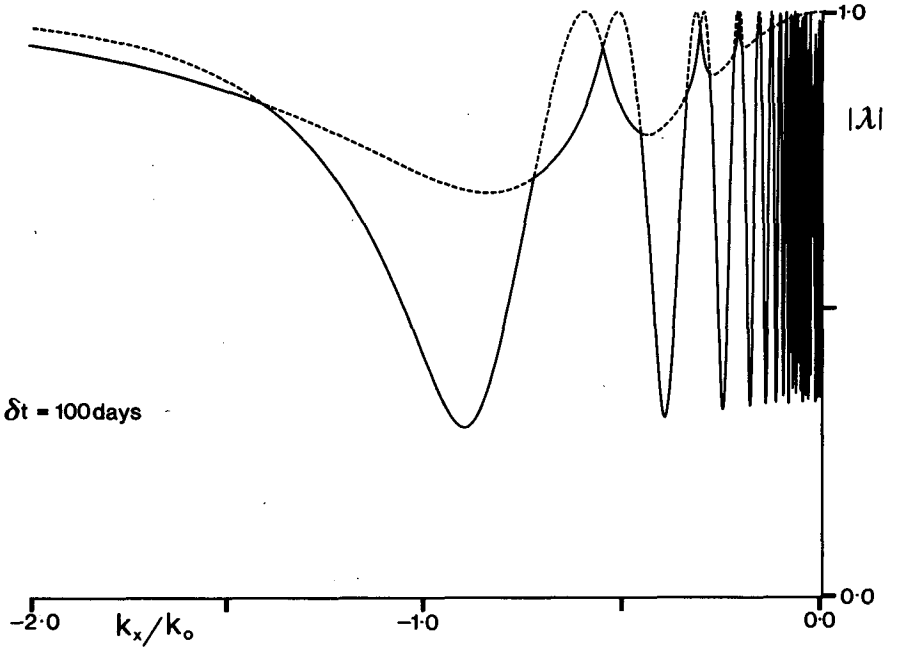

FIG. 3. Moduli of the two nonzero eigenvalues in a system of three vertical modes along the negative k_x axis. The time interval δt is 100 days. Other quantities are defined in the text. The solid (dotted) lines correspond to the eigenvalue of smallest (largest) magnitude. The wavenumber k_0 corresponds to the Rossby radius of the first baroclinic mode. Note that at small values of k_x, the finite interval used has missed some of the oscillations of $|\lambda_2|$.

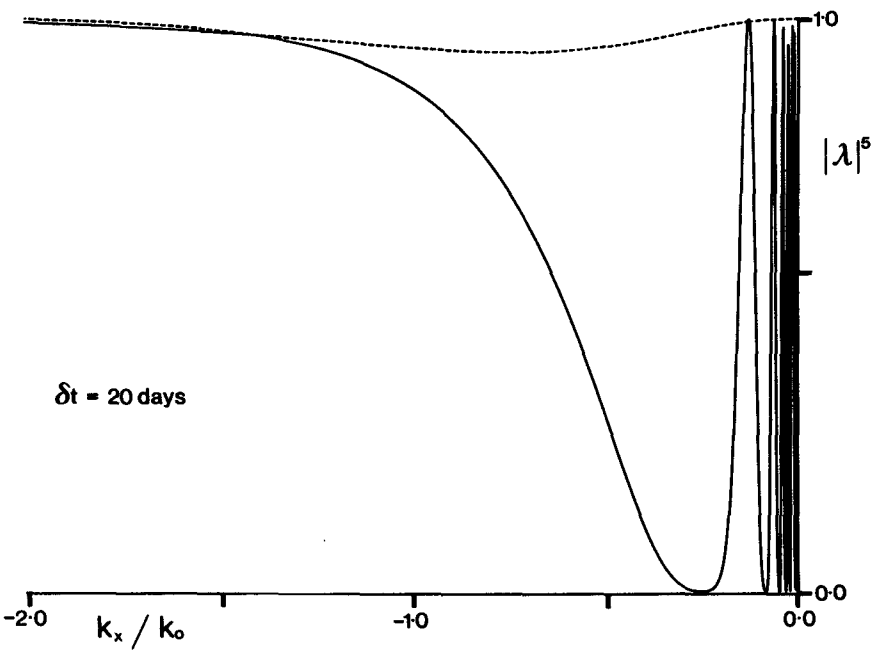

FIG. 4. Fifth power of the moduli of the two nonzero eigenvalues when a time interval of 20 days is used. It thus gives the error reduction after five assimilation cycles or 100 days. All other variables as in Fig. 3.

mode increases. The latter produces an oscillation in one of the eigenvalues, the first minimum occurring when the phase difference between the barotropic and baroclinic modes is approximately π, and further minima when it is 3π, 5π, etc. The phase difference between the two baroclinic modes steadily decreases, and so the modulus of the third eigenvalue tends to one. In practice one would probably try to circumvent the oscillating nature of the eigenvalues by using a slightly different assimilation scheme for each wavenumber and choosing δt in each case so that the eigenvalues are near a minimum.

There are two other notable features in Fig. 3. The first is that when the phase increment of the three waves is similar, there is an interaction between the two eigenvalues and they change branches, as occurs in degenerate perturbation theory. The second is that the oscillations in $|\lambda_2|$ occur over a range of between 1 and approximately ⅓. This behavior is considered in the Appendix, and it is shown that if the phases of all of the Rossby waves are similar [Eqs. (A6)], the amplitude of $|\lambda_2|$ lies between 1 and $1 - 2h_1^2/(h_1^2 + h_2^2 \ldots)$. For the three-mode case when the h_i are equal, this lower limit becomes ⅓. The perturbation due to small differences in phase between the baroclinic modes is also studied and gives a similar result [Eq. (A20)].

Figure 5 shows the continuation of Fig. 3 into the k_x, k_y plane. Except for the complexities that occur where the values of $|\lambda_2|$ and $|\lambda_3|$ are similar, the contours follow roughly circular loci. These loci correspond approximately to lines of constant angular velocity difference between two of the modes. If the two modes have Rossby wavenumbers of S_1 and S_2, their angular velocities are given by

$$\left.\begin{aligned}\omega_1 &= -\beta k_x/(\mathbf{k}\cdot\mathbf{k} + S_1^2)\\ \omega_2 &= -\beta k_x/(\mathbf{k}\cdot\mathbf{k} + S_2^2)\end{aligned}\right\}. \tag{6.4}$$

This gives an equation for the loci,

$$(\mathbf{k}\cdot\mathbf{k})^2 - \mathbf{k}\cdot\mathbf{k}(S_1^2 + S_2^2) + \left[S_1^2 + S_2^2 + \beta k_x \frac{(S_2^2 - S_1^2)}{\omega_1 - \omega_2}\right] = 0. \tag{6.5}$$

If $|k| > S_1$ (i.e., the barotropic mode) and $|k| < S_2$ (i.e., the baroclinic modes near the origin), then this simplifies to give an equation for a circle,

$$\left[k_x - \frac{\beta}{2(\omega_1 - \omega_2)^2}\right]^2 + k_y^2 = \left[\frac{\beta}{2(\omega_1 - \omega_2)}\right]^2. \tag{6.6}$$

Changing the assimilation interval. Finally, we consider the effect of changing the assimilation interval δt. In Fig. 4, $|\lambda|^5$ is plotted for δt equal to 20 days. Thus, as with Fig. 3, it shows the reduction in error after 100 days. As expected from the analysis of section 4, when $|\lambda|$ is near one the use of a smaller value of δt produces a much worse result.

However, at low wavenumbers where $|\lambda_2|$ has a reasonably small value, repeating the assimilation cycle five times is much more effective than repeating it only once. In this region the minimum values of $|\lambda_2|$ lie near

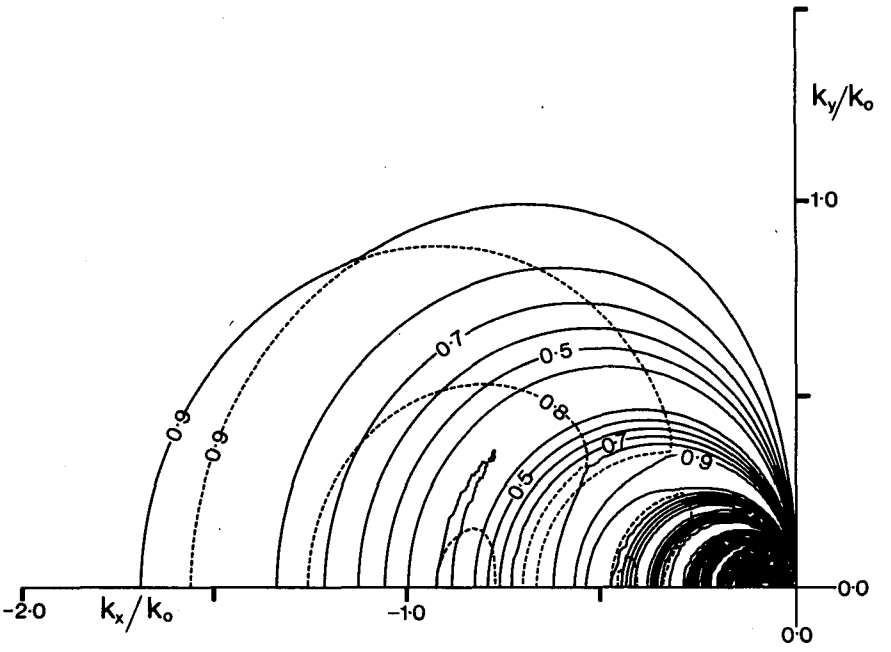

FIG. 5. Moduli of the two nonzero eigenvalues in a system of three vertical modes contoured in the wavenumber k_x, k_y phase. All variables as in Fig. 3.

0.333 and the minimum values of $|\lambda_2|^5$ near 0.004. The improvement occurs for values of k_x less than $0.5k_0$, which corresponds to waves with wavelengths greater than 350 km and periods of less than 76 days.

Thus, for these long wavelengths, a repeat period of 20 days would be successful in separating the barotropic wave from the higher baroclinic waves. However, for studies of the ocean eddy field, which has scales nearer the Rossby radius, and for studies that separate the higher baroclinic modes, much longer time intervals are required.

7. Discussion

This paper has shown that the assimilation of data into an ocean model can be considered as a projection in the appropriate vector space from a point that describes the model onto the surface defined by the data. The rate at which the model error is reduced depends on the eigenvalues of a matrix operator, which incorporates both the projection and the evolution of the ocean during each assimilation cycle.

It has also been shown that the efficiency of the assimilation scheme depends on how much the different vertical modes of the ocean, with the same horizontal wavenumber, get out of phase during each assimilation cycle. If the phase difference is too small, then the scheme is unable to distinguish between the modes, and the length of the assimilation cycle should be increased until the phase differences are approximately one radian. If the phase difference is large, then an improvement in the assimilation scheme can be obtained by reducing the time interval between observations.

Such improvements would be practical in a model where it is straightforward to treat the different horizontal wavenumbers individually, as we have done so far in this paper. However, in most models it will be easier to treat each spatial grid point in the model separately. This corresponds to using the same assimilation scheme at all wavenumbers. Although the time interval used in the scheme can then be chosen so that it is efficient at the wavelengths and modes of most interest, a problem does arise in that at other wavelengths and for other modes the efficiency of the scheme will be poor.

An alternative strategy that might help to overcome this problem is to use an adaptive scheme. This is based on Eq. (3.11), which shows that the reduction in the model error is greatest if new data is assimilated when the differences between the observations and the model are at a maximum. An adaptive scheme may therefore be proposed that waits until the observed errors reach their maximum before assimilating any new data.

When the theory is applied to a midlatitude ocean it is found that for waves with horizontal scales on the order of the Rossby radius, an assimilation cycle time of 100 days is efficient at separating the barotropic and first two baroclinic modes from each other. The scheme is not efficient with shorter wavelengths or for waves with a north–south wavenumber because for these waves the phase separation that develops between the different vertical modes over a 100 day period is too small. At longer wavelengths the scheme is efficient at

separating off the barotropic mode but is inefficient at distinguishing between the two baroclinic modes.

If a shorter time interval of 20 days is used, then the scheme is even better at separating off the long wavelength barotropic mode, but otherwise the scheme is less efficient. However, in practice an interval of 20 days may not be quite so bad because topographic effects in the ocean can increase the frequency of many of the waves present and so allow a shorter cycle time to be used.

The results show that satellite altimeter data will be of most use in determining the vertical structure of oceanic features when these have horizontal scales on the order of the Rossby radius. Such features include mesoscale eddies and frontal regions such as the Gulf Stream. At larger scales the altimeter data will also be of use if it can be used with other sources of data that give the required extra information on the large-scale baroclinic field. However, its use in the study of small-scale features or for features with dominant north–south wavenumber, such as zonal current systems, will be limited to the information it gives on the surface currents only.

In practice, both model and real ocean will be affected by nonlinearities and the data will be corrupted by noise. As a result, the convergence rate of an assimilation scheme will always be slower than the values predicted by the theory in this paper. However, a number of modifications can be made that improve the scheme. For example, the projection used can be optimized so as to give a faster convergence rate (Webb, 1986). Alternatively, if random noise is a problem, its effect can be reduced by using a greater sampling rate and averaging over neighboring samples.

Also we are not limited to integrating forward in time, and an alternative technique is to repeatedly integrate the model forward and backward through the period for which data is available (Bengtsson, 1975; Talagrand, 1981). This technique is particularly useful if other factors force the use of a short-time interval so that during each individual assimilation cycle the reduction in the model error vector is small. Over the many repeated cycles, the model error is further reduced to give a final solution where accuracy is only limited by the accuracy and extent of the data.

Better observing systems should also eventually reduce the noise present, but the problem of nonlinearities will remain a serious one. If they are small it might be possible to treat them as random noise. They will usually degrade the performance of an assimilation scheme because they move the point to which the scheme is converging. However, nonlinearities are not always a problem, and as shown by Ghil et al. (1981), in regions where information is not otherwise available the advection by nonlinearities can be beneficial.

In general, however, near surface currents are typically 20 cm s^{-1} at midlatitudes, whereas the phase speed of the Rossby waves is only a few centimeters per second. The nonlinear terms in the equations of motion are therefore large and will have an important effect on the scheme used. The time scale on which they become important will be similar to the correlation time found in ocean eddy fields (Saunders, 1983), roughly 20 days.

If a data assimilation interval of 100 days is used with such an eddy field, the nonlinearities would probably overwhelm the assimilation scheme. Instead, it may be necessary to work with an assimilation cycle time of 20 days and to compensate for its inefficiency in separating the baroclinic modes, either by using other sources of data or by using an improved method such as the forward–backward assimilation scheme. Work on this problem is continuing.

Acknowledgments. We would like to acknowledge discussions with Dr. D. Anderson.

APPENDIX

The Oscillations at Low Wavenumbers

We will consider the case that arises when all the vertical modes except one have a similar phase. In the ocean this arises with long Rossby waves when the time step δt is small, because all of the baroclinic modes have small angular velocities and only the barotropic modes have a large angular velocity.

In this case, the eigenvalue equation (4.1) has one solution with eigenvalue λ_1 equal to zero and one solution with eigenvalue λ_2 having a magnitude between zero and one. The rest of the solutions have eigenvalues with magnitudes of approximately one. In this appendix we obtain an expression for λ_2 by perturbing about the state in which the phases of all the baroclinic waves are equal.

1. The unperturbed state

Consider a system of N Rossby waves in which all the phases δ_k of Eq. (4.2) are equal to δ, except for the first δ_1. The results of section 3 show that the lowest eigenvalue λ_1 equals zero, and the higher ones λ_3, λ_4, etc., equal $e^{i\delta}$. To calculate λ_2, expand the determinent of the operator of Eq. (4.1):

$$\det(\mathbf{A}-\lambda\mathbf{I}) = [(1-h_1{}^2/\gamma)e^{i\delta_1}-\lambda] \times[(1-h_2{}^2/\gamma)e^{i\delta}-\lambda](\cdots)+O(\lambda^{N-2}), \quad \text{(A1)}$$

$$=(-\lambda)^N+(-\lambda)^{N-1}[(1-h_1{}^2/\gamma)e^{i\delta_1}+\cdots] +O(\lambda^{N-2}). \quad \text{(A2)}$$

The eigenvalues λ_n correspond to the zeros of the determinant, thus,

$$\det(\mathbf{A}-\lambda\mathbf{I}) = (\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda), \quad \text{(A3)}$$

$$=(-\lambda)^N+(-\lambda)^{N-1}(\lambda_1+\lambda_2+\cdots) +O(\lambda^{N-2}). \quad \text{(A4)}$$

Equating the coefficients of $N-1$ in (A2) and (A4) gives

$$\lambda_1+\lambda_2+\cdots = e^{i\delta_1}(1-h_1^2/\gamma) + e^{i\delta}(1-h_2^2/\gamma+1-h_3^2/\gamma+\cdots)$$

$$\lambda_2+(N-2)e^{i\delta} = e^{i\delta_1}(1-h_1^2/\gamma) + e^{i\delta}[(N-1)+(1-h_1^2/\gamma)]. \quad \text{(A5)}$$

Thus,

$$\lambda_2 = (1-h_1^2/\gamma)\exp(i\delta_1)+(h_1^2/\gamma)\exp(i\delta). \quad \text{(A6)}$$

For the perturbation expansion, the corresponding eigenfunction is also required. This is obtained by writing (4.1) in the form,

$$\lambda_2\mathbf{x} = \mathbf{A}\mathbf{x}. \quad \text{(A7)}$$

For the first row this gives,

$$\lambda_2 x_1 = (1-h_1^2/\gamma)e^{i\delta_1}x_1-(h_1h_2/\gamma)e^{i\delta}x_2-\cdots. \quad \text{(A8)}$$

Substituting for λ_2 from (A6) and rearranging the terms gives

$$\mathbf{h}\cdot\mathbf{x}=0. \quad \text{(A9)}$$

For the nth row,

$$\lambda_2 x_n = -(h_nh_1/\gamma)e^{i\delta_1}x_1+\cdots+(1-h_n^2/\gamma)e^{i\delta}x_n+\cdots. \quad \text{(A10)}$$

Multiplying (A9) by $h_ne^{i\delta}$, subtracting from (A10) and substituting for λ_2 gives

$$[(1-h_1^2/\gamma)e^{i\delta_1}+(h_1^2/\gamma)e^{i\delta}]x_n = (h_nh_1/\gamma)\times(e^{i\delta}-e^{i\delta_1})x_1+[(1-h_n^2/\gamma)e^{i\delta}+(h_n^2/\gamma)e^{i\delta_1}]x_n. \quad \text{(A11)}$$

This simplifies to give

$$x_n = -x_1\{h_nh_1/[\gamma(1-h_1^2/\gamma)]\}. \quad \text{(A12)}$$

Thus the eigenvector corresponding to λ_2 is

$$\mathbf{x}_2 = \mathbf{h}-(\gamma/h_1)\mathbf{t}, \quad \text{(A13)}$$

where $(\mathbf{t})_i = \delta_{i1}$. In a similar way, the eigenvector of the adjoint equation can be shown to be,

$$\mathbf{y}_2 = \mathbf{h}e^{-i\delta}+\mathbf{t}[(h_1-\gamma/h_1)e^{-i\delta_1}-h_1e^{-i\delta}]. \quad \text{(A14)}$$

2. The perturbation equation

As in section 4, let the superscript zero represent the unperturbed operator, eigenfunction and eigenvalue.

$$\mathbf{A}^0\mathbf{x}_n^0 = \lambda_n^0\mathbf{x}_n^0. \quad \text{(A15)}$$

Then,

$$\lambda_n = \lambda_n^0+\mathbf{y}_n^{0*}\mathbf{A}^1\mathbf{x}_n^0+\cdots. \quad \text{(A16)}$$

where A^1 is the perturbation.

Consider the case when the mth mode is perturbed from a phase δ to δ_m. Then the elements of the perturbation matrix $\mathbf{A}^1$ are,

$$(\mathbf{A}^1)_{ij} = \delta_{jm}[(\delta_{ij}-(h_ih_j/\gamma)][\exp(i\delta_m)-\exp(i\delta)]. \quad \text{(A17)}$$

Substituting for $\mathbf{x}_2^0$, $\mathbf{y}_2^0$ and $\mathbf{A}^1$ in (A16),

$$\lambda_2 \approx \lambda_2^0+\frac{[\exp(i\delta_m)-\exp(i\delta)](h_m^2h_1^2/\gamma^2)}{(1-h_1^2/\gamma)}. \quad \text{(A18)}$$

Thus,

$$\lambda_2 \approx (1-h_1^2/\gamma)\exp(i\delta_1) + (h_1^2/\gamma)[1-(h_m^2/\gamma)/(1-h_1^2/\gamma)]\exp(i\delta) + (h_1^2/\gamma)[(h_m^2/\gamma)/(1-h_1^2/\gamma)]\exp(i\delta_m). \quad \text{(A19)}$$

If more of the baroclinic waves differ slightly from δ, then their contribution can be added in a similar manner, giving

$$\lambda_2 \approx (1-h_1^2/\gamma)\exp(i\delta_1)+(h_1^2/\gamma)\exp(i\delta) \times[1-\sum_m(h_m^2/\gamma)/(1-h_1^2/\gamma)] + (h_1^2/\gamma)\sum_m\exp(i\delta_m)(h_m^2/\gamma)/(1-h_1^2/\gamma). \quad \text{(A20)}$$

This can be represented geometrically in a similar manner to Fig. 2 and shows that when Eq. (A20) holds,

$$1 \geqslant |\lambda_2| \geqslant |(1-2h_1^2/\gamma)|.$$

For the three-mode case, when all the values of h are equal this predicts that $1 \geqslant \lambda_2 \geqslant \frac{1}{3}$. This behavior is illustrated in Fig. 3 where near the origin λ_3 has a value near one, and λ_2 lies within the range predicted.

REFERENCES

Allan, T. D., Ed., 1983: *Satellite Microwave Remote Sensing.* Ellis Horwood, 526 pp.

Bengtsson, L., 1975: 4-dimensional assimilation of meteorological observations. *GARP Publication Ser.* No. 15, WMO, Geneva, 75 pp.

Bernstein, R. L., G. H. Bern and R. H. Whritner, 1983: Seasat altimeter determinations of ocean current variability. *J. Geophys. Res.*, **87**(C5), 3261–3268.

Bube, K. P., and M. Ghil, 1981: Assimilation of asynoptic data and the initialization problem. *Dynamic Meteorology: Data Assimilation Methods,* L. Bengtsson, M. Ghil and E. Kallen, Eds., Springer-Verlag, 330 pp.

Cheney, R. E., and J. G. Marsh, 1981: Seasat altimeter observations of dynamic ocean currents in the Gulf Stream region. *J. Geophys. Res.*, **86**, 473–483.

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

Deutsch, R., 1965: *Estimation Theory.* Prentice-Hall, 269 pp.

Duchossois, G., 1983: Status and future plans for ERS-1. *Satellite Microwave Remote Sensing,* T. D. Allan, Ed., Ellis Horwood, 501–514.

Gautier, C., and M. Fieux, Eds., 1984: Large-Scale Oceanographic Experiments and Satellites. Reidel, 288 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.

LeBlond, P. H., and L. A. Mysak, 1978: *Waves in the Ocean.* Elsevier, 602 pp.

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

Marshall, J. C., 1985: Determining the ocean circulations and improving the geoid from satellite altimetry. *J. Phys. Oceanogr.*, **15**, 330–349.

Matthews, J., and R. L. Walker, 1965: *Mathematical Methods of Physics*. W. A. Benjamin, 475 pp.

Morse, P. M., and H. Feshbach, 1953: *Methods of Theoretical Physics, Vol I*. McGraw-Hill, 997 pp.

Robinson, I. S., 1985: *Satellite Oceanography*. Ellis Horwood, 455 pp.

Saunders, P. M., 1983: Benthic observations on the Madeira abyssal plain: Currents and dispersion. *J. Phys. Oceanogr.*, **13**, 1416–1429.

Sorenson, H. W., 1966: Kalman filtering techniques. *Advances in Control Systems*, Vol. 3, C. T. Leondes, Ed., Academic Press, 346 pp.

Talagrand, O., 1981: Convergence of assimilation procedures. *Dynamic Meteorology: Data Assimilation Methods*. L. Bengtsson, M. Ghil, E. Kallen, Eds., Springer-Verlag, 330 pp.

Tapley, B. D., 1982: The Seasat altimeter data and its accuracy assessment. *J. Geophys. Res.*, **87**(C5), 3179–3188.

Timchenko, I. E., 1984: *Stochastic Modelling of Ocean Dynamics*. Harwood Academic Publishers, 311 pp. [Translated by E. T. Premuzic.]

Webb, D. J., 1986: On the choice of the best-projection for assimilating altimeter data into ocean models. Unpublished manuscript.

Williamson, D., and R. Dickinson, 1972: Periodic updating of meteorological variables. *J. Atmos. Sci.*, **29**, 190–193.