

The Effects of Assimilation on the Physics of an Ocean Model. Part I: Theoretical Model and Barotropic Results*

REBECCA A. WOODGATE

Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany

PETER D. KILLWORTH

Southampton Oceanography Centre, Empress Dock, Southampton, United Kingdom

(Manuscript received 2 October 1995, in final form 15 October 1996)

ABSTRACT

Although data assimilation is now an established oceanographic technique, little work has been done on the interaction of the assimilation scheme and the physics of the underlying model. The way in which even a simple assimilation scheme (here nudging) can significantly alter the response of the model to which it is applied is illustrated here.

Using analytic and semianalytic models, the assimilation of sea surface height, density, and velocity is studied. It is shown that the assimilation can act to alter the high inertia-gravity wave frequency to be the order of the Coriolis parameter, a result that is of relevance to the problems of initialization. The theory also predicts an optimum strength of nudging, normally dependent on wavelength, wave speed, and latitude, which can give convergence of the assimilation on a timescale as short as a day. The results are verified by identical twin experiments using a full primitive equation model, the Free Surface Cox Code, both in barotropic spinup (results presented here) and in a more realistic baroclinic situation (results presented in Part II).

1. Introduction

Data assimilation is used daily by meteorological agencies across the world. In oceanography, however, it is relatively new, mainly due to lack of data. This lack of data produces interesting new problems not found in meteorological assimilation, and so many different schemes have been suggested and implemented, with varying degrees of success. Special volumes of journals and books have been dedicated to the subject, namely *Dynamics of Atmospheres and Oceans* (Haidvogel and Robinson 1989), *Dynamic Meteorology* (Bengtsson et al. 1981), and *Ocean Circulation Models: Combining Data and Dynamics* (Anderson and Wilbrand 1989). Substantial review papers may also be found by Ghil and Malanotte-Rizzoli (1991) and Talagrand (1993), or, for the tropical oceans, by Arnault and Perigaud (1992).

Assimilation uses, in essence, two methods. The mathematically elegant adjoint approach (e.g., LeDimet

and Talagrand 1986; Thacker and Long 1988) normally retains the physics of the underlying model and adjusts initial conditions to minimize some cost function (e.g., the distance between model and observations). While many studies have been made of the adjoint approach using simplified physics, the potential computational and storage costs of the method are immense, and the approach has yet to be tested for a full nonlinear ocean model. The second method we shall refer to generically as “nudging” (Anthes 1974); this includes the range of methods from the full Kalman filter (Ghil et al. 1981) to simple approaches. In all cases the model prediction at some time step is adjusted by a statistical procedure to give a “corrected” assimilation prediction that is used as the basis for the next time step.

The motivation for this work was to gain a better understanding of how to assimilate sea surface height data into a fully three-dimensional, free-surface, general circulation model. There is a substantial body of literature on methods of assimilation themselves (Malanotte-Rizzoli and Holland 1986; Moore et al. 1987; Bennett 1992). There is also a smaller set of studies concerning the use of underlying dynamics within, or to improve, an assimilation method (e.g., Webb and Moore 1986; Haines 1991; Cooper and Haines 1996). As far as we are aware, however, little attention has been paid to the change in model physics (or dynamics) by the act of assimilating. Models, like reality, possess natural fre-

*Alfred-Wegener Institute Contribution Number 1235.

Corresponding author address: Dr. Rebecca A. Woodgate, Alfred-Wegener-Inst. für Polar- und Meeresforschung, Physik II, Handelshafen, Am Handelshafen 12, 27570 Bremerhaven, Germany.
E-mail: woodgate@awi-bremerhaven.de

quencies and modes of oscillation. These frequencies, in turn, yield wave speeds and hence length scales: the width of a Kelvin wave—the deformation radius—is determined by the wave speed of inertia–gravity waves; the spinup of the western boundary layer is determined by the speed of long planetary waves; and so on. Such frequencies are derivable for models—at least for linear dynamics—and may be observable for models.

However, a continuously assimilated model does not obey the same dynamics as the unassimilated underlying model. In the limit of very weak assimilation (trusting the model far more than the observations), the assimilated model would have essentially the same dynamics as the original. In the opposite limit of very strong assimilation (direct insertion of all fields; trusting the observations far more than the model), the dynamics of the model has no relation to the original since the latter has been entirely suppressed.

How, we wondered, did a model behave with a more normal assimilation scheme that gave weight to both model and data? Under theoretically optimal schemes (e.g., the Kalman filter), the appropriate weighting scheme is specified in terms of error covariances to minimize the error at the next time step. However, the model that includes such an assimilation will have, as we noted, frequencies and space scales differing from the original model, and these differences will create (modified) inertia–gravity waves that could cause the assimilated model to diverge from reality over time, for example, if boundary layers become too thick or thin under such schemes.

We consider first, in section 2, a standard simplification of the primitive equations, namely the shallow-water equations. This permits analytical and semianalytical treatment of the effects of assimilation, here modeled as a relaxation term in the predictive equation for the variable to be assimilated, which can be sea surface height, density, or velocity. The analysis predicts the frequency response of the system under assimilation and yields optimum values for the nudging coefficient and convergence times of the assimilation depending on length scale relative to the Rossby radius.

The results are verified in section 3 using a full primitive equation model, the Free Surface Cox Code (Killworth et al. 1991), henceforth FSCC. Identical-twin experiments using coarse resolution are performed for barotropic and baroclinic models. The details of the latter and comparison to other primitive equation assimilation experiments (e.g., Moore et al. 1987; Derber and Rosati 1989; Carton and Hackert 1990) are discussed in Part II (Woodgate 1997). We discuss the limitations of the approach in section 4.

2. A theoretical assimilation model

a. Construction of the theoretical model

We construct a theoretical model of assimilation by considering the nudging assimilation of sea surface

height into the barotropic shallow-water equations. This is then easily generalized to the baroclinic case.

In the linear case of a flat-bottom, uniform density ocean, the equations of motion for η , the sea surface height, and u , v , the depth-integrated velocities, can be written in the form

$$u_t = fv - gH\eta_x \quad (1)$$

$$v_t = -fu - gH\eta_y \quad (2)$$

$$\eta_t = -(u_x + v_y), \quad (3)$$

where f is the Coriolis parameter (assumed constant), g the acceleration due to gravity, and H the ocean depth. Subscripts represent derivatives. These are the unforced¹ shallow-water equations.

For assimilation of sea surface height, an extra Newtonian relaxation term is added to (3), which then becomes

$$\eta_t = -(u_x + v_y) + \alpha(\eta_{\text{obs}} - \eta), \quad (4)$$

where α measures the strength of the forcing and can be thought of as the reciprocal of some forcing time-scale. We assume for the moment that assimilation occurs continuously. With reference to the Kalman filter, it can be shown (see the appendix) that α is the equivalent of the weighting term, which reflects the relative errors of data and model. Clearly α must be positive for convergence. When a time-stepping scheme is introduced, α is usually constrained additionally by numerical stability to be less than $1/\Delta t$, where Δt is the time step.

Analytic elimination of u and v by cross differentiation leaves a third-order forced partial differential equation for η ,

$$\eta_{ttt} + f^2\eta_t - gH\nabla^2\eta_t = \alpha\left(\frac{\partial^2}{\partial t^2} + f^2\right)(\eta_{\text{obs}} - \eta). \quad (5)$$

The solution of (5) consists of a forced response to η_{obs} and a transient, unforced response. The latter is produced by the “shocking” of the system as it undergoes assimilation. It is the decay of this transient response that gives the rate of convergence of the assimilation. Thus, all forcings, including η_{obs} , are neglected. Since the equations are linear, we assume superposition of Fourier modes of the form $e^{ikx+ily}e^{i\omega t}$, where k and l are wavenumbers in x and y , respectively. Thus, (5) becomes a complex cubic:

$$\omega^3 - i\alpha\omega^2 - \omega(f^2 + gH\kappa^2) + i\alpha f^2 = 0, \quad (6)$$

where κ is the total wavenumber, given by $\kappa^2 = k^2 + l^2$. This has three roots, giving the three modes of response of the system. When α is zero, these are the standard two inertia–gravity waves of frequency

¹ The forcing terms are neglected here as later we will seek the transient solution.

$$\omega = \pm(f^2 + c^2\kappa^2)^{1/2} \quad (7)$$

and one steady geostrophic mode. We study here the behavior of the real (oscillatory) and imaginary (decay) parts of ω when α is nonzero.²

Model discretization in space and time is easy to include. Wavenumbers k and l can be replaced with their finite-difference analog and the time-stepping scheme modeled by a matrix operator (see Woodgate 1994 for details).

Nondimensionalization allows us to extend the analysis to the baroclinic case via the baroclinic analog of the shallow-water equations, namely (Gill 1982),

$$u_t = fv - \frac{1}{\rho_0}p_x \quad (8)$$

$$v_t = -fu - \frac{1}{\rho_0}p_y \quad (9)$$

$$p_t = -\rho_0 c_n^2(u_x + v_y), \quad (10)$$

valid for each vertical modal component. The wave speed c_n is now the baroclinic wave speed, where n is the mode number and the Rossby radius R_o is given by c_n/f . Nondimensionalizing on f for time, H for vertical distance, R_o for horizontal distance, and ρ_0 for density, we attain for both systems a nondimensional form of (6), that is,

$$(\omega')^3 - i\alpha'(\omega')^2 - \omega'(1 + (\kappa')^2) + i\alpha' = 0, \quad (11)$$

with a prime representing a nondimensional quantity.

Figure 1 shows a contour plot of the three solutions to (11), for ranges of α' and $\lambda' = 2\pi/\kappa'$. Over most of the region shown (the behavior does not differ beyond the values of the parameters shown) there are two roots with the same imaginary part (decay) and equal and opposite real parts (oscillation), together with a third, purely decaying, mode. These correspond to damped inertia-gravity and geostrophic modes, respectively. In the shaded triangular area, however, all three roots are purely imaginary, corresponding to simple decay. This region only exists for α above a critical value. It is straightforward to find asymptotic solutions to (11) for various parameter ranges, and these will be referred to without proof when necessary.

b. Length scales less than R_o —Barotropic case

We consider first the barotropic response. Typical midlatitude values for a barotropic wave speed and hence R_o are 190 m s^{-1} and 2600 km (approximately 27°), respectively. The ocean model to be used in section

3 is a 30° box ocean with 1° resolution. Thus, the length scales under consideration are less than or of the order of the barotropic Rossby radius.

1) THE EFFECT OF VARYING α ON FREQUENCY RESPONSE

The effect of varying α on the frequency response is illustrated in Fig. 2, with a dimensional section from the nondimensional solution of Fig. 1, using parameters from the model of section 3. In particular, the wavelength $\lambda (= 2\pi/\kappa)$, here of order $0.8R_o$, is typical of a low-order basin mode, and the time step taken as the model barotropic time step of 150 s . Variation with λ will be discussed below.³

The real part of frequency represents oscillations of the system. As α is increased from zero, the frequency of the inertia-gravity waves decreases until at a point B , $\alpha = \alpha_b$, there are no oscillations. As α is increased further, beyond a point A , $\alpha = \alpha_c$, oscillations return, but now at the much lower frequency of f , the Coriolis parameter. Continuous assimilation has removed the high-frequency waves. (For any κ , there is a range of α 's for which this holds. Conversely, for a given α there will be a wavenumber κ sufficiently large that the real part of the frequency is unaltered, though the mode will possess a decay. However, numerical resolution limits prohibit an arbitrarily large κ , so that in practice this latter possibility cannot occur.)

The convergence of the assimilation is determined by the decay or otherwise of the transients. These are the errors in the assimilation, and thus all modes must decay for a successful assimilation. The rate of convergence is controlled by the smallest of the imaginary parts of the three roots.⁴ This is largest at point A , where it can be shown that ω_{imag} is of order f for all λ , and ω_{real} is zero. Thus, the assimilation should converge in approximately $1/f$, that is, about 1 day at midlatitudes, but rather longer at lower latitudes. For $\alpha < \alpha_c$, the smallest imaginary root tends to zero, making the convergence of any assimilation slower. Note that this critical relaxation strength, α_c , is located at the boundary between pure decay and oscillations at f .

2) THE EFFECT OF VARYING λ ON FREQUENCY RESPONSE

Figure 3 shows, nondimensionally, how α_c varies with wavelength. At longer wavelengths, both the value of α_c and the range of α with pure decay are decreased; that is, the fastest convergence occurs with weaker assimi-

² The Coriolis parameter f is assumed constant since its variation has little effect on most inertia-gravity waves (which control the convergence of the assimilation here) and adds a simple decay to the planetary waves (the extension of the geostrophic mode to nonzero β).

³ Note that it is κ^2 that appears in Eq. (6). Thus, it is the magnitude, not the direction of the wavenumber, that is important. This model does not include the β effect and thus has no preferred direction.

⁴ The nonzero intercept in the imaginary part corresponds to the numerical damping of the inertia-gravity waves.

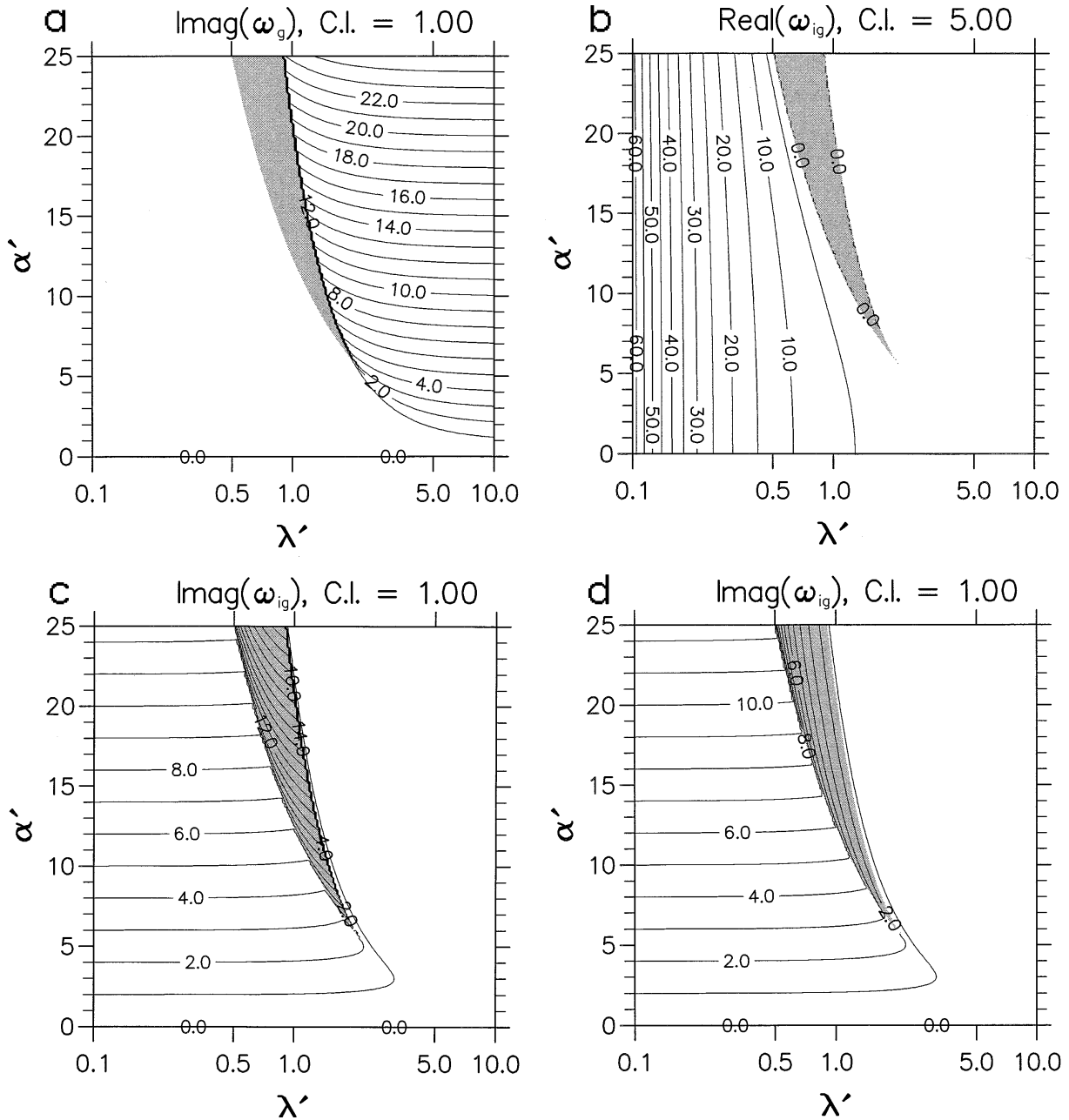


FIG. 1. Contours of real (oscillatory) and imaginary (damped) parts of the nondimensional frequency of response of a nudged system, as a function of nondimensional wavelength λ' and damping coefficient α' . There are two roots with the same real part and equal and opposite imaginary parts (inertia-gravity modes) and one root that is purely imaginary (geostrophic mode) over most of the region. In the shaded area, there are three purely imaginary roots. (a) The imaginary part of the frequency for the geostrophic mode, whose real part is zero. (b) The real part of the inertia-gravity wave frequency; the other such mode has the same frequency but of opposite sign. (c), (d) The imaginary part of the two inertia-gravity modes; these are identical except within the shaded region. Note that the geostrophic mode cannot be followed smoothly through the shaded region (its mode representation changes across the shaded region) and so an apparent discontinuity must appear somewhere—in our choice on the right of (a) and (c).

lation. Note that from (7) an increase in wavelength increases the natural inertia-gravity periods of the system. We have been unable to produce a good physical argument for the dependence of α on wavelength; recall that the action of assimilating is not physical, but statistical. However, the observed decrease in α_c , that is, the increase

in the natural period of the assimilation, may be physically justified if the relationship between the two periods is important. Such a relationship is indicated by the work of Webb and Moore (1986) for Rossby waves. At shorter wavelengths, the numerical stability limit ($\alpha < 1/\Delta t$) typically prevents α_c being attained.

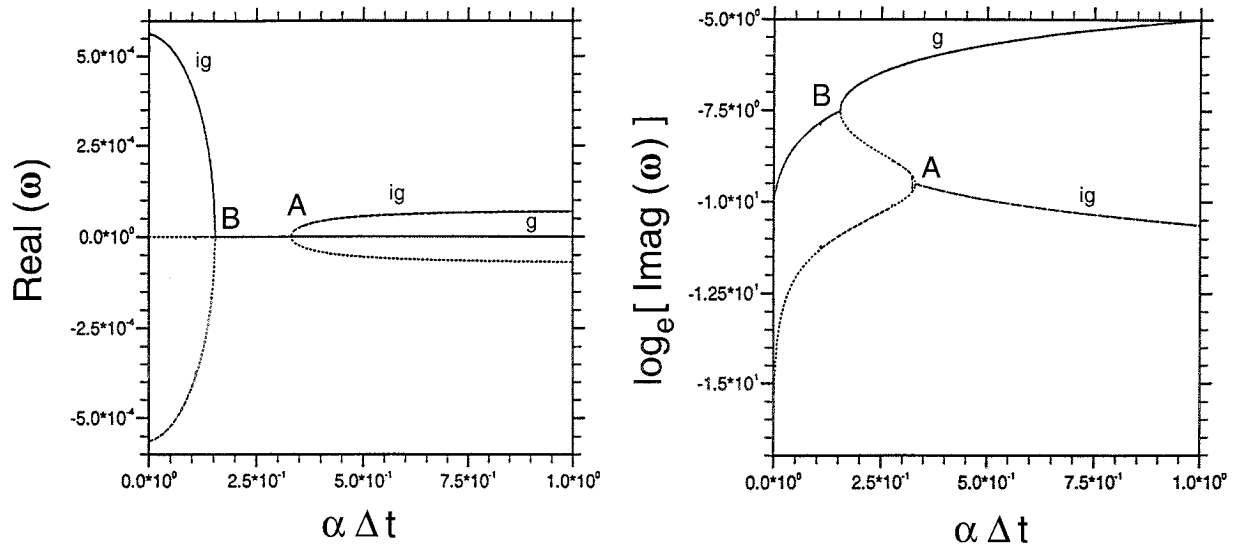


FIG. 2. Dimensional section from Fig. 1 showing real and imaginary parts of frequency for varying α for a wavelength of 21° , approximately $0.8R_o$. The x axis, $\alpha\Delta t$, where Δt is the barotropic time step of 150 s, ranges from 0 to 1, the stability limit for α of $1/\Delta t$. This corresponds to a range of α' of 0–91. Point A marks the branch point at α_c , and B marks that at α_b . The imaginary part of frequency is a log-plot for clarity: ig—inertia–gravity; g—geostrophic. At $\alpha\Delta t = 1$, the real part of the ig frequency is f .

Interestingly, these results suggest that, given knowledge of the predominant wavelength of the system, it is possible to choose an optimum value for α to give the minimum decay time of $1/f$ at that wavelength. An

additional consideration is the frequency response of the assimilating system. Too weak an assimilation will allow high-frequency inertia–gravity waves at other wavelengths that may cause numerical stability problems due to nonlinear linkage. For example, for a wavelength of order R_o , the fastest convergence is expected for an α_c of order $20f$. This, however, allows inertia–gravity waves for wavelengths of less than $0.5R_o$, which would be avoided if a larger value was used for α , that is, $54f$. It should be noted this results in a much stronger nudging coefficient than is normally used in the literature.

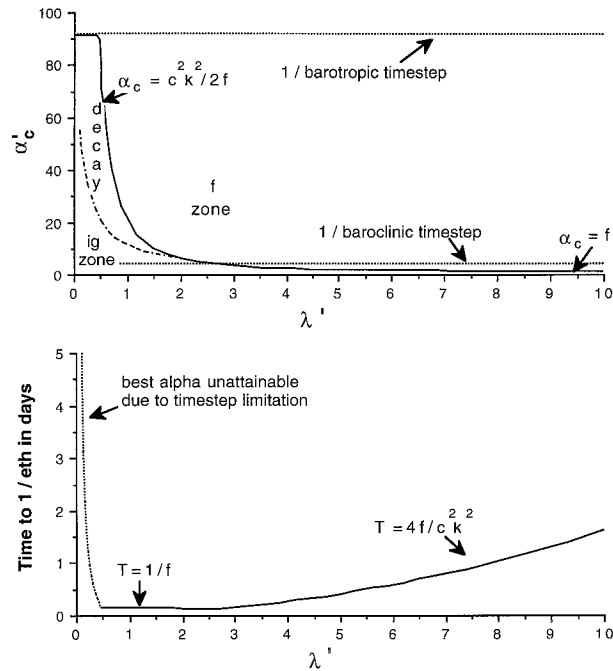


FIG. 3. Variation of the nondimensional α'_c and corresponding minimum decay time with wavelength nondimensionalized on R_o . Solid line— α_c ; dash-dot line— α_b ; dashed line—time-step limitations. Frequency response of different regions is also marked: ig—inertia–gravity.

3) A SOLUTION TO THE INITIALIZATION PROBLEM?

This result has consequences for the initialization problem, best known in meteorology (e.g., Barwell and Bromley 1988), where the introduction of unbalanced data produces unphysically large high-frequency inertia–gravity waves that give rise to forecast errors and data rejection. A common solution is the use of an approximate “slow manifold” (Daley 1981), which dynamically balances the data so as to project almost entirely onto the slow modes of the system. The analogy here would be to project only onto the geostrophic mode. In ocean models, initialization has not been found to be such a problem (e.g., Malanotte-Rizzoli et al. 1989), but most recent studies use either quasi-geostrophic or rigid-lid models. The latter explicitly exclude the fast barotropic inertia–gravity waves. In a free-surface code, however, these inertia–gravity modes are important. In a fully nonlinear system (e.g., the FSCC), initialization waves can have a catastrophic effect on

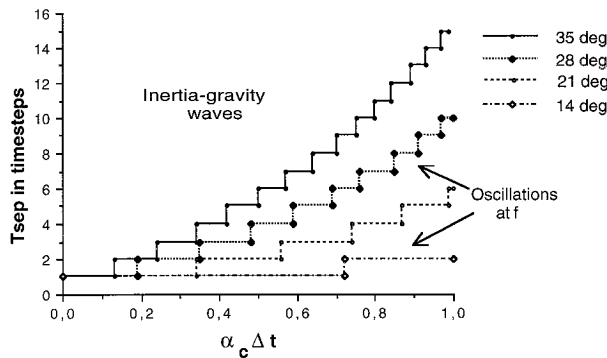


FIG. 4. The variation of T_{sep} with $\alpha_c \Delta t$ for different wavelengths for discrete assimilation. Above the curve, the oscillations are at the inertia-gravity frequency; below the frequency is f ; $\Delta t = 150$ s. The range of $\alpha_c \Delta t$ reflects the numerical stability limit of $\alpha < 1/\Delta t$.

the baroclinic part of the solution, which is stepped with a much longer time step and is unstable to high-frequency waves (Part II).

The modification of the inertia-gravity mode by assimilation to a much slower frequency suggests that nudging assimilation of sea surface height removes much of the initialization problem. In addition, in cases where the high frequency of the inertia-gravity waves limits the time step of the model, the application of continuous assimilation may make it possible to run with a longer time step because the damping present may counteract any numerical (CFL) instability. This would offer a significant computational advantage in, for example, the case of spinning up a model to a known state.

4) ASSIMILATION DISCRETE IN TIME

To make the above analysis analytical, we have assumed an assimilation continuous in time. This proves to be superior to assimilation applied only intermittently, as shown by nudging η into a simple numerical model of the shallow-water equations, every T_{sep} time steps (i.e., if $T_{sep} = 2$, only every other time step is assimilating).

The response of the system to this assimilation is shown in Fig. 4 for specific wavelengths and a range of α , with T_{sep} varied to give the fastest convergence to the new solution. If T_{sep} is too high (i.e., too few data), high-frequency inertia-gravity waves cause an overshoot of the solution. If T_{sep} is too small, the assimilation still converges, but at a slower rate. In this case, the oscillations of the system are at f as predicted above. Each value of α thus has a critical T_{sep} ; conversely, each T_{sep} has a corresponding range of critical values of α (a range occurs since T_{sep} is discretized, giving rise to the sawtooth nature of the graph).

Results show α_c increases rapidly with T_{sep} . Thus, since there is a numerical stability limit of $\alpha \leq 1/\Delta t$, T_{sep} must be small for convergence, for example, less

than $16\Delta t$ (40 min), for wavelengths up to 35° . For a larger T_{sep} , the model both rejects the data and suffers from initialization shocking. This is confirmed by direct assimilation experiments with the FSCC (not shown).

c. Length scales greater than R_o —Baroclinic case

For the baroclinic case, typical values for the wave speed and R_o of the first baroclinic mode are 2.8 m s^{-1} and 38 km (approximately 0.4°), respectively. Thus, the length scales present in the 1° basin-scale model are greater, or much greater, than R_o ($2.5R_o$ to $75R_o$).

1) ASSIMILATION OF DENSITY

The analog of assimilating sea surface height is the assimilation of density data (giving pressure and assuming sea surface height). The relevant solution is again given by Fig. 3, now in the region of $\lambda > R_o$, that is, $\lambda' > 1$. The numerical bound of $\alpha < 1/\Delta t$ is now more restrictive since the baroclinic time step is greater (3600 s), but α_c , which is now independent of wavelength and is of order f , can still be attained for wavelengths greater than about $3R_o$. The minimum decay time, also shown, now increases with λ^2 , and the frequency of response of the model is again at f rather than the inertia-gravity frequency. For $\lambda < 3R_o$, the results are as in section 2b, that is, α_c decreases with increasing wavelength and the time of convergence is bounded by f^{-1} .

It is interesting to note the different dependence on f in the different regions of this graph. When $\lambda < 2.5R_o$, the decay time is of order $1/f$ and increases with decreasing latitude. For $\lambda > 2.5R_o$, the decay time varies as $4f/c^2\kappa^2$, that is, decreasing with decreasing latitude. The model, being f plane, is not valid in the limit of $f \rightarrow 0$. This behavior does, however, suggest different mechanisms are responsible for the assimilation convergence at different spatial scales.

2) ASSIMILATION OF VELOCITY ALONE

For assimilation of velocity, the relaxation term must be added instead to (8) and (9). This leads to a cubic for ω of the form

$$\omega^3 - 2i\alpha\omega^2 - \omega(\alpha^2 + f^2 + c_n^2\kappa^2) + i\alpha c_n^2\kappa^2 = 0. \tag{12}$$

Remarkably, the solution of (12) is simply $\omega_v = i\alpha - \omega_\eta$, where the subscripts denote velocity and η assimilation, respectively, so that the properties of the solution can be deduced from the preceding discussion.

Figure 5 shows how the roots vary with α for a specific wavelength of 1.4° of order $3.5R_o$.⁵ For such parameters and a reasonable nudging coefficient, there is

⁵ Parameters correspond to the model to be used in Part II. Specifically, the time step is now the baroclinic.

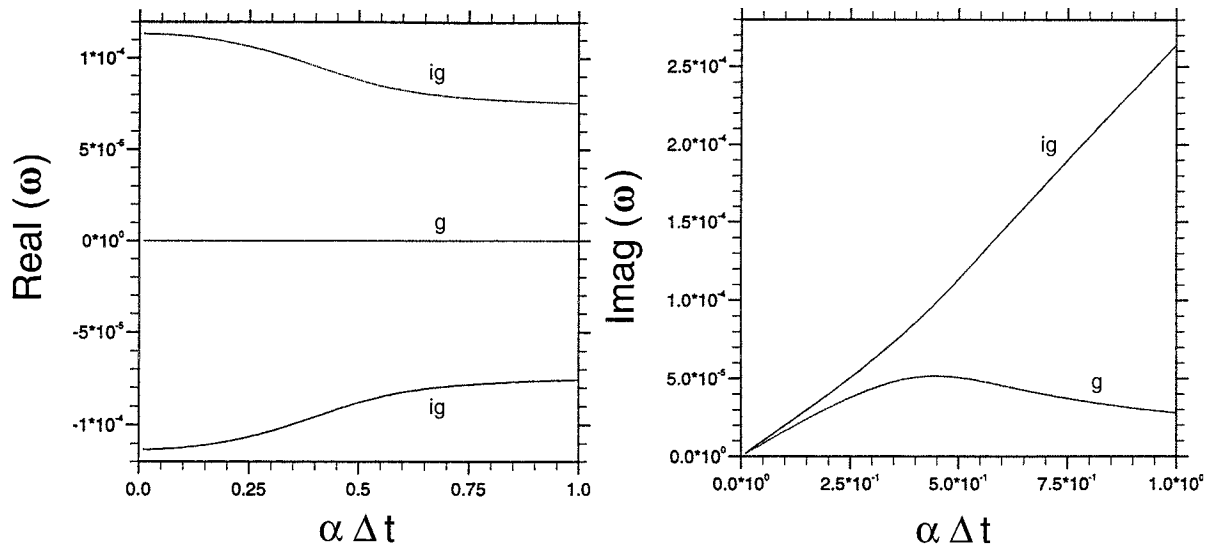


FIG. 5. Real and imaginary parts of frequency for assimilation of velocity for varying α , for wavelength of 1.4° , approximately $3.5R_p$. The x axis, $\alpha\Delta t$, where Δt is now the baroclinic time step of 3600 s, ranges from 0 to 1, the stability limit for α of $1/\Delta t$. This corresponds to a range of α' of 0–4. The change from the inertia–gravity frequency to f is now not as marked. Note also the presence of the maximum in the smallest imaginary root in the modified geostrophic mode (g).

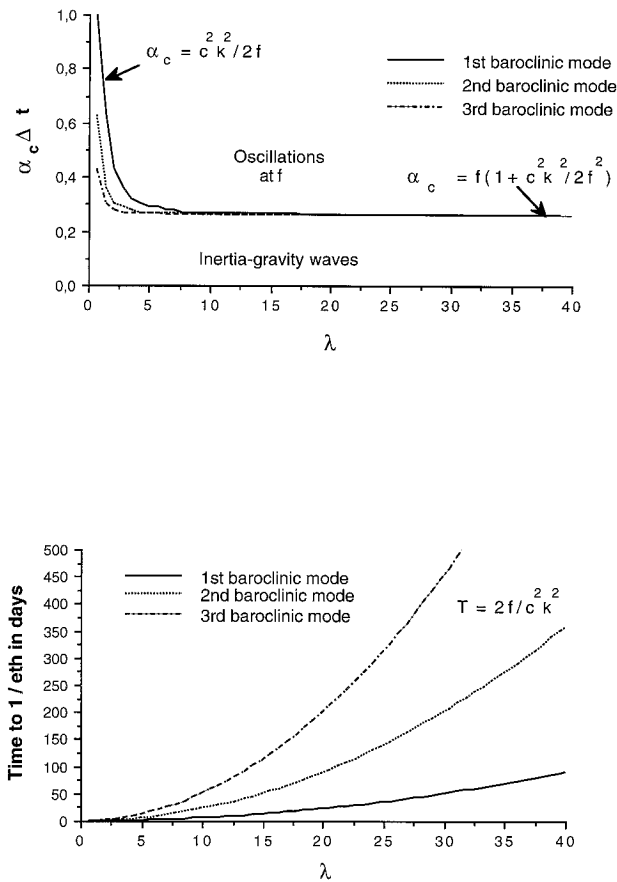


FIG. 6. Variation of α_c (represented as $\alpha_c\Delta t$ as in Fig. 5) and corresponding minimum decay time with dimensional wavelength for the assimilation of velocity for the first three baroclinic modes, corresponding to wave speeds of 2.76, 1.38, and 0.92 m s^{-1} , respectively.

no transition region present. The oscillatory frequencies (real part) of the modes with velocity assimilation are precisely those for η assimilation, although differing parameters conceal this slightly (e.g., the smaller frequency change for a given wavenumber change is due to the smaller wave speed c_n used). The role of the modes in the decay of the transients has now been reversed—it is now the geostrophic mode that limits the decay rate. This again has a maximum for some value of α , α_c , corresponding to the minimum decay time. This maximum, however, is not very sharp and ω_i decreases only slowly as α is increased. For the inertia–gravity modes, the decay of the transients increases roughly linearly with α , suggesting that although stronger assimilation is likely to shock the system more, the transients produced will decay more rapidly than with a weaker assimilation.

Figure 6 shows the variation of α_c for different wavelengths and different baroclinic modes. This α_c is found analytically from (12) to be $f(1 + c_n^2\kappa^2/2f^2)$. Thus, at large wavelengths, α_c is independent of wavelength and of order f , as found for the assimilation of density. The decay time, given by largest value of $1/\omega_i$ at this α_c , does vary with wavelength, as $2f/c_n^2\kappa^2$, as quantified in the bottom plot of Fig. 6 for the first three baroclinic modes. Thus, for short length scales, baroclinic disturbances should be fairly rapidly corrected by assimilation, whereas for longer length scales, convergence time may be on the order of months. It should be noted the timescale is proportional to f and thus would be expected to decrease at lower latitudes.

The same analysis can be performed for the assimilation of both η and velocity (results not presented here). In this case, there is no region of pure decay, though

an α_c still exists. The smallest imaginary root corresponds again to the inertia-gravity modes and gives a convergence time of less than 1 day.

d. Eddy length scales

The analysis has been phrased in terms of a predominantly linear (and hence large-scale) flow regime. Assimilation typically involves adjustment on the eddy scale. If we grant the limitation of linear dynamics and assume a wavelength of order R_o , it is interesting to see what the above analysis predicts for an eddy-resolving situation. For assimilation of density or sea surface height, from Fig. 3, we predict values for the optimum nudging coefficient α_c and the convergence time τ of

$$\alpha_c = \frac{c^2 \kappa^2}{2f} \sim 2\pi^2 f \quad \tau \sim \frac{1}{f}$$

For velocity assimilation, using Fig. 6, we obtain

$$\alpha_c \sim 2\pi^2 f \quad \tau = \frac{2f}{c^2 \kappa^2} \sim \frac{1}{2\pi^2 f};$$

that is, convergence is faster than the assimilation of density.

This is in agreement with results from geostrophic adjustment theory (as discussed by Moore 1986), which predicts that, for length scales of order R_o , the final state of some disturbance is dependent predominantly on the initial flow field.

The model verification performed here is limited to a 1° , non-eddy-resolving model. However, studies by Oschlies (1994) use a nudging technique with an eddy-resolving primitive equation model. He presents results using a nudging coefficient far less than the optimum predicted here, with a value of α_c of about $0.08f = 1/(2 \text{ days})$, which show a convergence of errors in times of order 20–30 days. From (11), such an α would give a convergence rate of about 80 days for a wavelength of the Rossby radius (not well resolved in their study) and 22 days for a wavelength of twice that, which is well resolved. This latter timescale is probably relevant and agrees well with his findings.

3. Model verification

a. Verification using the Free Surface Cox Code

We verify the above results using the FSCC. The linear shallow-water equations used above are a simplification of the barotropic equations solved in this model. The FSCC uses a split-explicit method, separating the velocity response of the system into barotropic (defined as depth integrated) and baroclinic parts and stepping the two parts separately, the barotropic time step being smaller than the baroclinic (see Killworth et al. 1991). To test the barotropic theory, the model is set up with uniform (potential) temperature. (For baroclinic experiments, see Part II.) Thus, the major deviations

from the assumptions used for the theoretical shallow-water equation model are the relaxation of both the f -plane assumption and linearity.

The experiments are of the “identical-twin” form often used in oceanographic data assimilation. Each experiment consists of three runs. The first, the “truth run,” mimics the role of the real ocean. The model is integrated forward in time from some initial state, and the data from that run are considered as the observations to be assimilated. This gives an advantage over using real data in that all the fields of the truth run are known and can be compared with those from the assimilation. Obviously, since the model is not a perfect representation of reality, these results will certainly overestimate the success of the assimilation when using real data. The second run is the “assimilation run.” Here the assimilating model is started from a different initial state than that used for the truth run (for choice of this state, see below) and integrated forward assimilating the data from the truth run. A fully successful assimilation run will reproduce the final state of the truth run. To quantify this success, a third run, the “control run,” is performed. This starts from the same initial state as the assimilation run but is integrated without assimilation, thus giving a measure of how well the nonassimilating model would have reproduced the final state of the truth run. In general, the model used for the assimilation run may be the same as the model used for the truth run or may be degraded in some manner to mimic the inadequacies of representing the real world with a numerical model. Here, the same model is used for all runs.

A subtlety of the identical-twin method is the choice of initial conditions for the runs. For the truth run, a spunup initial state is chosen (where by “spunup” we imply the point at which rate of change of the fields of the model has slowed). For the assimilation and control runs, the choice of initial conditions is rather more critical. If the initial state chosen is very different from the truth run, then the error convergence will appear to be good (in that the model has improved rapidly), but the final fields may still be far distant in some sense from the truth. The magnitude of this problem can be assessed by how fast the control run converges to the truth; this is taken into account when planning experiments.

b. Barotropic results

The model used is a $30^\circ \times 30^\circ$ square box model, stretching from 20° to 50°N with five levels in the vertical, a flat-bottom, uniform density, 1° grid spacing, and a steady east–west wind stress that varies sinusoidally north–south. The top level of 100 m gives a reasonable representation of the Ekman layer, while the other level thicknesses are 1000, 1000, 500, and 500 m, to allow the introduction of a 1000-m-high ridge. This did not significantly alter the results presented here and is not considered further. Time steps used are 150 s for

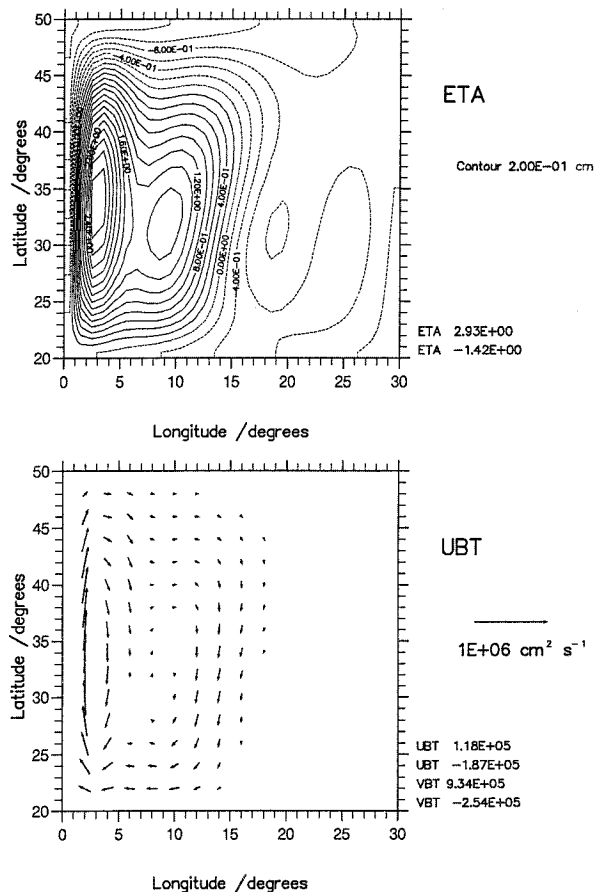


FIG. 7. Spunup state of box ocean ETA—sea surface height, UBT—depth integrated velocity. Maximum and minimum values are noted.

the barotropic mode and 3600 s for the baroclinic. The model has no bottom friction and no external fluxes.

The model is spun up from rest for 20 days, by which time a barotropic gyre circulation is established (Fig. 7). The predominant response is the first barotropic basin mode (e.g., Le Blond and Mysak 1978), which for this setup has a period of order 10 days, as is clearly dominant in the rms results presented below.

The assimilation is applied only to η , the sea surface elevation, by the addition of an extra term to the time derivative,

$$\eta_t = (\dots \text{physics} \dots) + \frac{\gamma}{\Delta t}(\eta_{\text{obs}} - \eta). \quad (13)$$

The coefficient γ is the $\alpha\Delta t$ of the preceding figures. It is a dimensionless number, bounded by numerical stability to be positive and less than or equal to 1. The barotropic time-stepping scheme in the model is the full Euler backward scheme and has two stages. The assimilation forcing is applied on both passes of the time-stepping scheme.

The truth run is defined as the model run from 20 to 37 days, the experiment length being chosen so as to not coincide with any of the low-order barotropic basin

mode frequencies. The final state of the truth run is taken as the initial state of the assimilation run, which then runs from day 37 to 54. All runs are forced by the same wind stress.

1) ASSIMILATION CONTINUOUS IN TIME

Given that we lack a continuous data supply, some interpolation in time must be performed on the data. As a first approach, for times when there are no data, the model is forced to the next measurement in the future. Thus, the data to be assimilated become stepped functions, with the changes in value coming immediately after the time corresponding to the true data.⁶

Figure 8 shows the time variation of rms fields and rms errors for the three barotropic components of the flow, that is, η , the sea surface height, and UBT and VBT, the depth-integrated velocities. The assimilation is successful, with good convergence in all fields. The predictions of the shallow-water model have also been verified; the fields converge on the timescale of about 1 day and the frequency of waves excited by the assimilation is f . So in the full model, nudging assimilation of η does remove the initialization problem.

Further experiments confirm the concept of the optimum value for the assimilation forcing and the predicted variation of the response with varying α . When $\alpha < \alpha_b$ [section 2b(1)], the inertia-gravity waves return. The value of α required for this corresponds to a wavelength of order 60° , showing that the dominant wavelength is that of the lowest basin mode. The assimilation parameters, α and T_{sep} , may be varied to improve the assimilation. Decreasing α from α_c gives better final error convergence for the velocity fields but worse convergence for η . (Note that α_c was for the *fastest* convergence.) The magnitude of the adjustment waves excited is less, but the time taken to converge is greater in line with model predictions.

The further evolution of the model after assimilation has been applied shows that the assimilation has converged. If the correct wind forcing is applied, the model continues to follow the truth run with some small inertia-gravity wave adjustment immediately after the assimilation ceases. Thus, the assimilation has been entirely successful in correcting the model state to the truth state. A better time interpolation scheme (e.g., linear interpolation in time) removes the sawtooth appearance of Fig. 8 but has no effect on the rate of convergence or the final error value.

⁶ Other methods of interpolating in time can be envisaged (for example, linear interpolation in time between data points) but do not significantly affect the results. Indeed, this stepped assimilation scheme demonstrates the frequency response best.

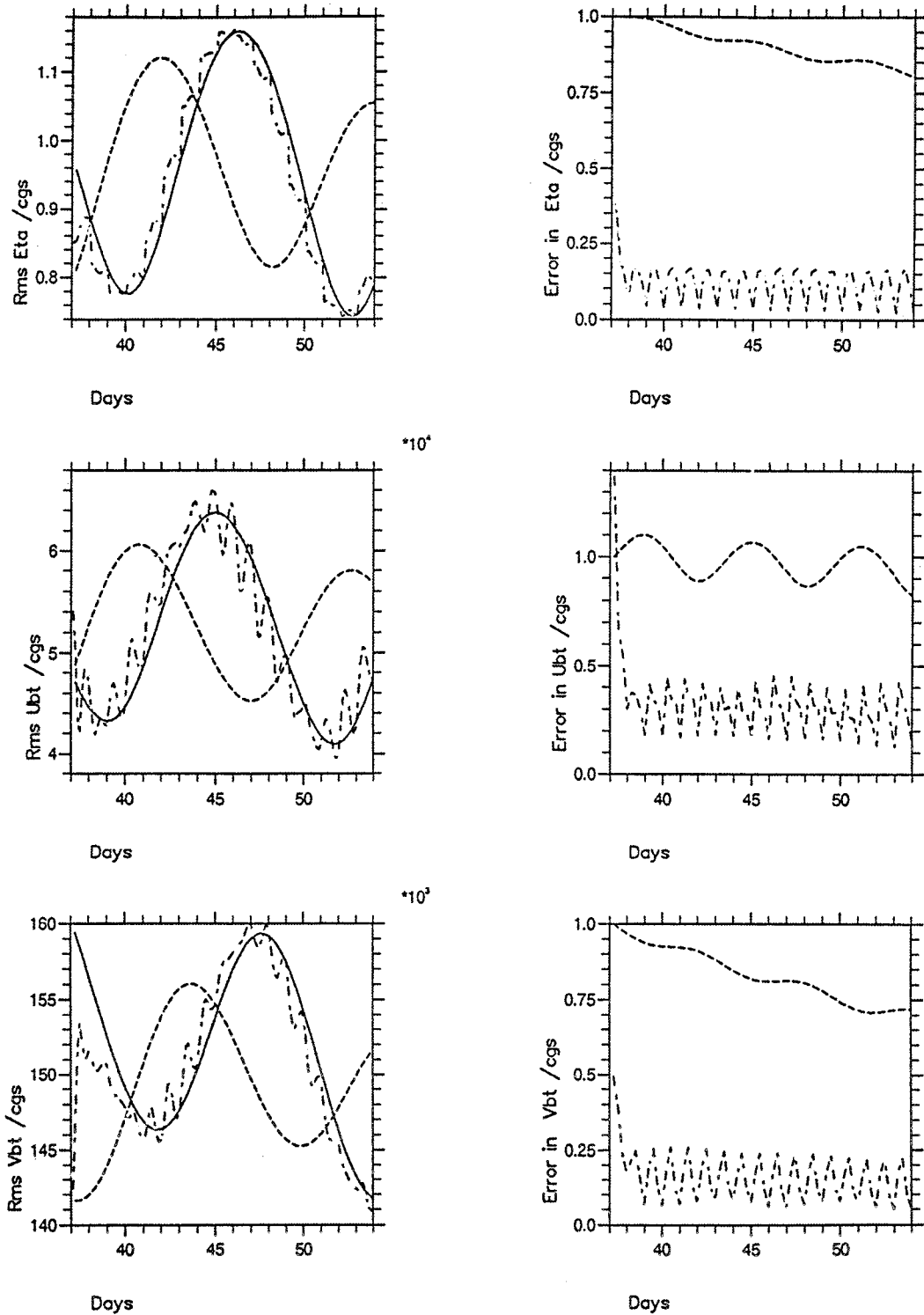


FIG. 8. Variation with time of basin averages of rms (left) and error (right) fields for the three barotropic components, ETA, sea surface height, and UBT, VBT, depth-integrated velocities for continuous assimilation of η (the rms error is scaled on initial control error); $\gamma = 0.75$. New data are provided every day. Solid line—truth, dashed line—control, dash-dot line—assimilation. Frequency of excited oscillation is f . ETA is in centimeters. UBT and VBT are in square centimeters per second.

2) A LIMIT TO THE ACCURACY OF THE ASSIMILATION

As with more complicated assimilation methods, the error range stabilizes during the run. This implies a limit to the success of the assimilation. For the assimilation of η , this is of the order of 5%–10% of the initial control error and arises from the interaction of the assimilation with the time-stepping scheme used. Consider again the nudging equation (13),

$$\frac{\partial \eta}{\partial t} = (\dots \text{physics} \dots) + \frac{\gamma}{\Delta t}(\eta_{\text{obs}} - \eta_a),$$

where the assimilation term now contains η_a , a value that must represent the model η at the time in question. For leapfrog and other centered schemes, η_a can be defined as the predicted η at that time level. For the two-step Euler backward scheme used here, the value needs to be chosen carefully.

In the first step, the fields η_n , etc., at time t_n are used to predict the time derivatives at t_n and hence intermediate fields, η' , a time step later, t_{n+1} . Here there is only one choice for η_a , that is, η_n . In the second step, the intermediates are used to predict new time derivatives that then step the model from the first fields, η_n at t_n , to their final values η_{n+1} at t_{n+1} . Here there are two options for η_a , namely η' or η_n . Use of η' gives the rather counterintuitive result that, for $\gamma = 1$, the η field is hardly affected and the assimilation effectively cancels itself out. Thus, η_n is used. The result of this two-step assimilation procedure is then

$$\eta_{n+1} = \eta_n(1 - \gamma) + \gamma\eta_{\text{obs}} + \Delta t \frac{\partial \eta'}{\partial t},$$

where $\partial \eta' / \partial t$ is the time derivative of η as calculated from the intermediate velocity fields equivalent to η' . When $\gamma = 0$, this reverts to the full Euler backward scheme. When $\gamma = 1$, the error in η is given by

$$\eta_{n+1} - \eta_{\text{obs}} = \Delta t \frac{\partial \eta'}{\partial t}.$$

Order of magnitude estimates of this term agree with the computational result, giving here a minimum error in η of order 10%.

Thus, due to the interaction of the assimilation and the model, there is a limit on the success of the assimilation scheme. This is a result of the time-stepping scheme chosen and as such appears unavoidable. However, in practice it is found that this few percent error is acceptably small, and indeed as $\partial \eta' / \partial t$ becomes smaller, it can be taken as negligible.

4. Discussion

This paper has used a theoretical shallow-water equation model to show how nudging assimilation can affect the physics of a model. The results indicate a resolution of the initialization problem and yield predictions for

the optimum strength of nudging and the rate of convergence of the assimilation. In the nondimensional form, this can be used to explain both barotropic and baroclinic systems, with behavior depending on length scale relative to the Rossby radius. Results are tested and verified using identical twin experiments in a full primitive equation model. Formally, the theories on which the results are based are by nature linear and f plane, which immediately excludes high nonlinear current regimes and equatorial regions. Also, by use of normal-mode theory to attain orthogonal modes, we have restricted ourselves to flat-bottom and small perturbation from a background density. However, tests with the FSCC have shown the results to be relevant to a fully nonlinear (although not eddy-permitting) model on a spherical earth, including topography, and further work with an Atlantic model spanning the equator (Woodgate 1994) shows that these predictions also work reasonably well close to the equator.

Although predictions for α_c in a multiwavelength system are obviously misleading, the results for rate of convergence and predictions for α_c are also relevant if taken in a rather more qualitative manner. We have already seen α should be chosen according to required frequency response since the convergence rate is often not so critically dependent on the value for α . To predict an exact convergence rate from a system made of many different modes is obviously ridiculous. However, an estimate of the expected timescales for convergence can be attained, and these estimates appear reasonable despite deviations from normal-mode dynamics. Changes in background density will change wave speeds but not by orders of magnitudes. Interestingly, it is found that, due to the nature of the time stepping of the barotropic mode, there is a theoretical minimum to the error in sea surface height, and this is confirmed numerically. (In practice this error is negligibly small.)

The result of primary importance from this work is the alteration of the natural frequencies of the system by continuous assimilation. In the barotropic case, the change is dramatic—for a sufficiently strong assimilation, the high inertia–gravity frequency is reduced to the Coriolis parameter. This has far-reaching consequences for the initialization problem in the free surface code and might even be relevant to atmospheric schemes. It is normal in data assimilation to choose the strength of the assimilation according to the relative errors in model estimate and data, as, for example, in the Kalman filter. The appendix shows that nudging can be thought of as a time-lagged Kalman filter, with the coefficient α taking the role of the ratio of the error coefficients. Here instead we use theory to predict an optimal strength of assimilation. A hidden assumption of identical-twin experiments is that of zero error in the data and infinite error in the model solution. The existence of an optimum strength of nudging shows that the most effective manner of making the model con-

verge is to blend the data and model, even if the data are considered completely correct.

We assume that not all fields are known completely, of course, so that modified inertia-gravity oscillations can exist. With complete data coverage (assumed error-free), direct insertion would clearly be the best route. With less than complete data coverage, however, but retaining error-free data, error minimization involves damping the inertia-gravity waves, and as we have shown, the optimal approach involves a blending of model and data.

Acknowledgments. Our thanks go to Jeff Blundell for his help with the graphics.

APPENDIX

Nudging as a Time-Lagged Kalman Filter

Optimal interpolation (Eliassen 1954), statistical interpolation (Gandin 1963), and Kalman filter techniques (e.g., Ghil et al. 1981; Webb 1989) are now well established. They all consider a combination of model estimate and data, taking into account the errors prevalent in both systems. They predict an analysis increment to be added to the first-guess field, which is made up of a weighted sum of the differences between the model prediction and the data. The analysis increment, $\alpha(r) - f(r)$, is given by

$$\alpha(r) - f(r) = \sum W_i(r)(o_i - f_i), \tag{A1}$$

where $\alpha(r)$ represents the analyzed value (the final combination of model and data at point r), $f(r)$ is the first-guess field from the model at point r , o_i is the data or observation at the i th point, and f_i is the first-guess field at the i th point. The matrix W_i gives the weights and can be shown to be

$$W_i = [\langle f(r)f_i \rangle]^T \mathbf{P}^{-1} (\mathbf{P} + \mathbf{D})^{-1}, \tag{A2}$$

where $\langle f(r)f_i \rangle$ is the covariance of the first-guess field at the point in question; \mathbf{P} is the prediction error covariance matrix, $\langle f_i f_i \rangle$; and \mathbf{D} is the data error covariance matrix, $\langle o_i o_i \rangle$. This can be thought of in two parts:

- 1) $\mathbf{P}(\mathbf{P} + \mathbf{D})^{-1}$, the filter, which makes the best combination of the data and the first-guess field at the observation point; and
- 2) $[\langle f(r)f_i \rangle]^T \mathbf{P}^{-1}$, the interpolator, which interpolates the filtered value to the required grid point.

In practice neither \mathbf{P} nor \mathbf{D} is perfectly known, and the differential forms of statistical interpolation, successive correction, and Kalman filter all look at evaluating these terms.

The form of assimilation, nudging (Anthes 1974), used in this work can be related to this optimal interpolation theory. In nudging, a correction term proportional to the difference between the model and the ob-

servation is added to the time derivative of the variable to be assimilated, X , that is,

$$\frac{\partial X}{\partial t} = (\dots \text{physics} \dots) + \alpha(X_{\text{obs}} - X),$$

where X_{obs} is the observation, for example, sea surface height. The method is also known as Newtonian relaxation, robust diagnostic relaxation, or, incorrectly, Haney relaxation. The important parameter is α , the nudging coefficient, which is constrained by numerical stability to be less than the reciprocal of the time step of the model, Δt . Though in these studies α will be taken as a constant, for a more advanced scheme α will vary in space and time. For physical insight, α can be thought of as the reciprocal of an assimilation timescale, though in practice the rate of convergence of the assimilation is different to this, as has been shown.

To relate nudging to the statistical theory, consider one time step of the model. Without assimilation,

$$\begin{aligned} X_{t+1}^f &= X_t + \Delta t \frac{\partial X}{\partial t} \\ &= X_t + \Delta t (\dots \text{physics} \dots), \end{aligned}$$

where subscripts t and $t + 1$ give time step and superscript f denotes ‘‘forecast,’’ that is, model guess. For the assimilation, the time derivative has the extra forcing term, that is,

$$\begin{aligned} X_{t+1}^a &= X_t + \Delta t \frac{\partial X}{\partial t} \\ &= X_t + \Delta t (\dots \text{physics} \dots) + \alpha \Delta t (X_{\text{obs}} - X) \end{aligned}$$

$$X_{t+1}^a = X_{t+1}^f + \alpha \Delta t (X_{\text{obs}} - X_t^a),$$

where superscript a indicates ‘‘analyzed,’’ that is, final combination of model and data. Thus, with reference to (A1), nudging can be thought of as a time-lagged Kalman filter, with α taking the place of the ratio of the error covariances, that is, the W_i of (A2).

REFERENCES

Anderson, D. L. T., and J. Willebrand, 1989: *Oceanic Circulation Models: Combining Data and Dynamics*. NATO ASI Series, Vol. 284, Kluwer Academic, 605 pp.

Anthes, R., 1974: Data assimilation and initialization of hurricane-predicting models. *J. Atmos. Sci.*, **31**, 702–719.

Arnault, S., and C. Perigaud, 1992: Altimetry and models in the tropical oceans: A review. *Oceanol. Acta*, **15** (5), 411–430.

Barwell, B. R., and R. A. Bromley, 1988: The adjustment of numerical weather prediction models to local perturbations. *Quart. J. Roy. Meteor. Soc.*, **114**, 665–689.

Bengtsson, L., M. Ghil, and E. Källén, 1981: *Data Assimilation Methods*. Vol. 36, *Dynamical Meteorology*, Springer-Verlag, 330 pp.

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

Carton, J. A., and E. C. Hackert, 1990: Data assimilation applied to the temperature and circulation in the tropical Atlantic, 1983–1984. *J. Phys. Oceanogr.*, **20**, 1150–1165.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.*, **101** (C1), 1059–1077.

- Daley, R., 1981: Normal mode initialization. *Dyn. Meteor.*, **36**, 77–109.
- Derber, J., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347.
- Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil, and E. Källén, Eds., Springer-Verlag, 319–330.
- Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields*. Gidrometeoizdat, 242 pp.
- Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics*, Vol. 33, Academic Press, 141–266.
- , S. Cohn, J. Travantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. *Dyn. Meteor.*, **36**, 139–224.
- Gill, A. E., 1982: *Atmosphere–Ocean Dynamics*. Academic Press, 662 pp.
- Haidvogel, D. B., and A. R. Robinson, 1989: Special issue on assimilation. *Dyn. Atmos. Oceans*, **13**, 171–518.
- Haines, K., 1991: A direct method for assimilating sea surface height data into ocean models with adjustments to the deep circulation. *J. Phys. Oceanogr.*, **21**, 843–868.
- Killworth, P. D., D. Stainforth, D. J. Webb, and S. M. Paterson, 1991: The development of a free-surface Bryan–Cox–Semtner ocean model. *J. Phys. Oceanogr.*, **21**, 1333–1348.
- Le Blond, P. H., and L. A. Mysak, 1978: *Waves in the Ocean*. Elsevier Scientific, 602 pp.
- LeDimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- 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.
- , R. E. Young, and D. B. Haidvogel, 1989: Initialization and data assimilation experiments with a primitive equation model. *Dyn. Atmos. Oceans*, **13**, 349–378.
- Moore, A. M., 1986: Data assimilation in ocean models. Ph.D. thesis, University of Oxford, 174 pp. [Available from Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Oxford OX1 3PU, United Kingdom.]
- , N. S. Cooper, and D. L. T. Anderson, 1987: Initialization and data assimilation in models of the Indian Ocean. *J. Phys. Oceanogr.*, **17**, 1965–1977.
- Oschlies, A., 1994: *Assimilation of Satellite Altimeter Data into an Eddy-Resolving Primitive Equation Model of the North Atlantic Ocean*. Vol. 256, Institut für Meereskunde, Kiel, Germany, 130 pp.
- Talagrand, O., 1993: Data assimilation problems. *Energy and Water Cycles in the Climate System*, NATO ASI Series, E. Raschke and D. Jacob, Eds., Vol. 15, Kluwer Academic, 187–213.
- Thacker, W. C., and R. B. Long, 1988: Fitting dynamics to data. *J. Geophys. Res.*, **93**, 1227–1240.
- Webb, D. J., 1989: Assimilation of data into ocean models. *Ocean Circulation Models: Combining Data and Dynamics*, D. L. T. Anderson and J. Willebrand, Eds., NATO ASI Series, Vol. 284, Kluwer Academic, 233–256.
- , and A. Moore, 1986: Assimilation of altimeter data into ocean models. *J. Phys. Oceanogr.*, **16**, 1901–1913.
- Woodgate, R. A., 1994: Data assimilation in ocean models. Ph.D. thesis, University of Oxford, 237 pp.
- , 1997: The effects of assimilation on the physics of an ocean model. Part II: Baroclinic identical-twin experiments. *J. Atmos. Oceanic Technol.*, **14**, 910–924.