Optimal Excitation of Baroclinic Waves

BRIAN F. FARRELL

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

(Manuscript received 6 June 1988, in final form 13 October 1988)

ABSTRACT

Development of perturbations in a baroclinic flow can arise both from exponential instability and from the transient growth of favorably configured disturbances that are not of normal mode form. The transient growth mechanism is able to account for development of neutral and damped waves as well as for an initial growth of perturbations asymptotically dominated by unstable modes at significantly greater than their asymptotic exponential rates. Unstable modes, which are the eigenfunctions of a structure equation, are discrete and typically few in number. In contrast, disturbances favorable for transient growth form a large subset of all perturbations. To assess the potential of transient growth to account for a particular phenomena it is useful to obtain from this subset the initial condition that gives the maximum development in a well-defined sense. These optimal perturbations have a role in the theory of transient development analogous to that of the normal modes in exponential instability theory; for instance they are the structures that the theory predicts should be found to precede rapid development.

In this work optimal perturbations for the excitation of baroclinic stable and unstable waves are found. The optima are obtained for the formation of synoptic scale cyclones as well as for the development of planetary scale stationary and transient baroclinic Rossby waves. It is argued from these examples that optimal perturbations are likely to limit predictability on time scales relevant to the short and medium range forecast problem and that unstable modes, if present, dominate the long range forecast.

1. Introduction

The ultimate source of energy for atmospheric motion is heating from solar radiation. At synoptic and planetary scales the differential insolation between low and high latitudes maintains an approximately geostrophically balanced zonal flow concentrated into narrow jet streams. The kinetic energy of the jets and the potential energy associated with their geostrophically balanced density contrasts provide the energy for the growth and maintenance of the large scale variance field. Of these the primary store of energy available to the waves is in the potential form (Lorenz 1955).

Understanding of the mechanism by which the potential energy is tapped by synoptic scale waves is gained by use of the quasi-geostrophic equations which can be cast in the form of conservation following the geostrophic flow of the quasi-geostrophic potential vorticity. Restricted to dynamics associated with waves drawing on the available potential energy this mechanism of growth is referred to as baroclinic instability, a terminology which arose because with certain restrictions the linearized conservation equation together with appropriate boundary conditions admits solutions of normal mode form that grow exponentially in time (Charney 1947; Eady 1949). However, it is also known that perturbations not of normal mode form undergo transient development on time scales appropriate to synoptic and planetary space scales and that the growth rate of these favorable perturbations can greatly exceed that of the unstable modes when these are supported and, furthermore, that transient development proceeds robustly in flows that are not exponentially unstable (Farrell 1985). While study of exponential instability focuses attention on the eigenmodes and in particular on the mode with maximum growth rate, transient development arises from a large subset of all perturbations (Farrell 1987). These are characterized in the case of baroclinic dynamics by structures producing heat fluxes down the mean temperature gradient, that is by possessing or producing during their development a westward tilt with height of the geopotential field.

Previous work involved with the problem of cyclogenesis sought to deal with the ambiguity of choosing a perturbation by appeal to observation. One structure that produces development is well known in the synoptic literature to correspond to the advection of an upper level vorticity center such as a short wave over a preexisting surface vorticity concentration such as a front. In Farrell (1985) this was modeled by a westward tilting plane wave confined between the ground and a scale height in the classic Charney (1947) problem. This simple model succeeded in producing a robust cyclogenesis on the advective time scale that was remarkably similar to observations of cyclone formation.

Corresponding author address: Dept. of Earth and Planetary Sciences, Harvard University, Pierce Hall, 29 Oxford Street, Cambridge, MA 02138.

Furthermore, the development process proceeded regardless of whether the flow supported instabilities.

Armed with a few general principles and a little experience one becomes adept at producing disturbances that develop rapidly in a given flow. These disturbances resemble in their diversity the growing perturbations that arise in turbulent fluids such as the atmosphere, perturbations that are associated with energetic events stochastically distributed in space and time (Salmon 1980, his Fig. 12). Because it is unlikely that the most rapidly developing disturbance will arise by chance, it is necessary in general to make do with suboptimal growth. Nevertheless there are theoretical reasons to seek the optimal excitation.

For example, one can find the optimal excitation over a period of time characteristic of rapid cyclogenesis without regard for modal projection and compare it with the optimal excitation of the unstable mode over the same period and compare these with the growth of the unstable mode alone to gain an appreciation for the role of the initial value problem in cyclogenesis. As a second example, consider a neutral mode such as one of the retrograde waves in the Charney problem. This eigenmode excited in isolation cannot draw on the APE of the mean flow. However, a composite disturbance including the mode can draw on this principle store of energy. In effect the neutral mode behaves as if it were an unstable mode for a restricted time period as far as its energetics are concerned. It turns out (Farrell 1988) that modes differ greatly in their potential to draw on mean flow energy and this potential is most easily assessed by finding the optimal excitation.

Finally, consider the problem of predictability. It is well known that flow regimes differ greatly in predictability in the sense that perturbations cause the trajectory of the solution in its phase space to deviate more readily for some flows than for others (Hoffman and Kalnay 1983; Palmer 1988). It is sometimes maintained that this deviation can be assessed by examining the eigenvalues of the tangent linear differential equation, the Lyapunov exponents, but for time scales associated with synoptic forecast this is not so. In fact, there are perturbations that grow much faster than the first Lyapunov exponent rate and the perturbation that grows the fastest is the optimal excitation. For time scales of interest in forecast, assessment of the predictability of different flow regimes should be made by comparing the development of optimal perturbations to the flows and not by comparing Lyapunov exponents (Lacarra and Talagrand 1988).

Optima for the excitation of baroclinic flows are found here in the L_2 and energy norms including the optimal excitation of the neutral wave and the most unstable eigenmode. Optima are also found in the same norms for the excitation of disturbances over a fixed time interval without regard for projection. This is the explosive growth solution that is also relevant to assessing predictability.

2. Baroclinic quasi-geostrophic model with Ekman damping

Restriction to a pure baroclinic basic state with the zonal wind a function of height only isolates the dynamics to that of the primary energy reservoir. We expect that optimal perturbations in basic flows with horizontal and vertical shear exploit the kinetic energy of the jet by adopting a component of the tilted trough configuration familiar in synoptic observations. Restriction to horizontal shear alone isolates this mechanism (Farrell 1988) but the mixed shear problem is so much more demanding of computational resources as to encourage preliminary study of a pure baroclinic model. It is also assumed here that the initial growth of small but finite perturbations can be adequately modeled with linear dynamics.

In order to maintain the flexibility to simulate more realistic basic states the model includes variation of the static stability and density with height as well as variation of the Coriolis parameter with latitude. For simplicity, the density scale height is taken to be constant $H^{-1} = -\rho^{-1}(\partial\rho/\partial z)$ and the Coriolis parameter is linearized $f = f_0 + \beta y$.

With these assumptions the nondimensional perturbation quasi-geostrophic potential vorticity equation in the scaled streamfunction:

$$\Phi = \psi(\tilde{z}, \tilde{t})\frac{e^{\tilde{z}/2}}{\sqrt{\epsilon}} e^{i\tilde{k}\tilde{x}} \cos(\tilde{l}\tilde{y}) \qquad (2.1)$$

is

$$\left(\frac{\partial}{\partial \tilde{t}} + R(z) + i\tilde{k}\tilde{U}\right)\left[\psi_{\tilde{z}\tilde{z}} - \left(\tilde{S}^2 - \tilde{S}_{\tilde{z}} + \frac{\tilde{\alpha}^2}{\epsilon}\right)\psi\right] + i\tilde{k}\left(\frac{\tilde{\beta}}{\epsilon} + 2\tilde{S} - \tilde{U}_{\tilde{z}\tilde{z}}\right)\psi = 0 \qquad (2.2)$$

with boundary conditions including a vertical velocity induced through Ekman convergence at the lower boundary:

$$\frac{\partial}{\partial \tilde{t}}(\psi_{\tilde{z}} + \psi\tilde{S}) - i\tilde{k}(\tilde{U}_{\tilde{z}} - \tilde{\Gamma})\psi = 0, \quad \tilde{z} = 0 \qquad (2.3a)$$

$$\psi \rightarrow 0, \quad \tilde{z} \rightarrow \infty. \qquad (2.3b)$$

In addition (2.1) allows for a gravest meridional mode with $\psi = 0$ at $y = \pm\pi/(2\tilde{l})$, and (2.2) for the imposition of linear potential vorticity damping.

The nondimensional quantities are

$$\tilde{t} = t\Lambda\sqrt{\epsilon_0}$$

$$\tilde{k} = \frac{kH}{\sqrt{\epsilon_0}}$$

$$\tilde{z} = \frac{z}{H}$$

where $\tilde{\alpha} = \sqrt{\tilde{k}^2 + \tilde{l}^2}$ is the total horizontal wavenumber, and $\epsilon_0 = f_0^2/N_0^2$ is the square ratio of the Coriolis parameter to a characteristic Brunt–Väisälä frequency. Additionally we define: $\epsilon_0 = f_0^2/N_0^2$, $\epsilon = \hat{\epsilon}/\epsilon_0$. The problem is characterized by the nondimensional beta parameter $\tilde{\beta} = \beta H/\Lambda\epsilon_0$, the Ekman parameter:

$$\tilde{\Gamma} = \frac{iN_0}{\Lambda H}\left(\frac{\nu}{2f_0}\right)^{1/2}\frac{\tilde{\alpha}^2}{\tilde{k}}$$

and the stability parameter:

$$\tilde{S} = -\frac{1}{2}\left(\frac{\epsilon_{\tilde{z}}}{\epsilon} - 1\right)$$

where ν is the vertical eddy viscosity coefficient and Λ is a characteristic shear. The last bracketed term in (2.2) is referred to as β_{eff}.

Tildes are dropped in sequel.

The above may be reduced to the canonical Charney problem by taking $R = 0$, $\epsilon = 1$ and $U(z) = \Lambda z$:

$$\left(\frac{\partial}{\partial t} + ikz\right)\left[\psi_{zz} - \left(\alpha^2 + \frac{1}{4}\right)\psi\right] + ik(\beta + 1)\psi = 0 \qquad (2.4)$$

$$\frac{\partial}{\partial t}\left(\psi_z + \frac{\psi}{2}\right) - ik(1 - \Gamma)\psi = 0, \quad z = 0 \qquad (2.5a)$$

$$\psi \to 0, \quad z \to \infty. \qquad (2.5b)$$

The Charney problem has been extensively studied and many facts including its dispersion relation and eigenfunction structure are well known and available in standard references, e.g. Pedlosky (1987). For this reason it is appropriate to introduce ideas with reference to the Charney problem and to solve the more general problem only where greater realism is indicated to serve a particular purpose.

We choose values for parameters appropriate to the midlatitude troposphere: $f_0 = 10^{-4}\ \text{s}^{-1}$, $N = 10^{-2}\ \text{s}^{-1}$, $H = 10$ km, $\Lambda = 3\ \text{m s}^{-1}/\text{km}$, $\beta = 1.6 \times 10^{-11}\ \text{m}^{-1}\ \text{s}^{-1}$. This results in $\beta = 0.53$. The meridional wavenumber $l = 2.0$ corresponds to a 3100 km wavelength, typical of midlatitude cyclogenesis, while $k = 0.40$ corresponds to 15 500 km, a planetary wave scale. A unit of nondimensional time is 9.3 h.

It has been remarked previously (Card and Barcilon 1982; Farrell 1985, 1989; Hoskins and Valdes 1988) that Ekman damping suppresses baroclinic instability in this model problem. As an example of the rapid decrease in the growth rate of the unstable wave with increasing vertical diffusion, the synoptic scale wave with $k = l = 2.0$ at $\nu = 2.5\ \text{m}^2\ \text{s}^{-1}$ retains only 28% of the undamped growth rate and at $\nu \approx 4.5\ \text{m}^2\ \text{s}^{-1}$ growth disappears entirely. Even the larger of these values is modest compared with what is typical in midlatitudes.

3. Numerical solution of the baroclinic problem

The Charney problem (2.4) together with its boundary conditions (2.5) can be written in operator notation assuming the solution form $\psi(x, y, z, t) = \hat{E}(z)e^{ik(x-ct)}\cos(ly)$ as

$$L\hat{E}(z) = c\hat{E}(z). \qquad (3.1)$$

With

$$L \equiv \Delta^{-1}[z\Delta + \beta_{eff}],$$

where $\Delta \equiv \partial_{zz} - [\alpha^2 + \frac{1}{4}]$ and $\beta_{\text{eff}} = \beta + 1$.

Here L is expressed using centered differences on N collocation points; z_i, $i = 1, N$ so that the N eigenfunctions in the finite difference approximation are each represented by a vector (Green 1960; Lindzen et al. 1982):

$$\psi_j = \hat{\mathbf{E}}_j e^{ik(x-c_jt)}\cos(ly).$$

The vertical structure of the physical streamfunction modes are related to the scaled variable by (2.1),

$$\mathbf{E}_j = \mathbf{P}\hat{\mathbf{E}}_j e^{ik(x-c_jt)}\cos(ly)$$

where,

$$\mathbf{P} = \frac{e^{z_i/2}}{\sqrt{\epsilon(z_i)}}\delta_{ij}.$$

Assuming a fixed wavenumber k, the evolution in time of an initial physical perturbation $\Phi_0 e^{ikx}\cos(ly)$ can be expressed as

$$\mathbf{\Phi} = \sum_{j=1}^{N}\gamma_j\mathbf{E}_j e^{ik(x-c_jt)}\cos(ly). \qquad (3.2)$$

Where γ is the spectral projection of the perturbation on the eigenvectors, obtained using $\mathbf{E}$, the matrix having the physical eigenvectors as columns:

$$\gamma = \mathbf{E}^{-1}\mathbf{\Phi}_0. \qquad (3.3)$$

4. Formulation of the optimization problem

The concept of an optimum requires a measure of perturbation magnitude for which an obvious choice is the rms amplitude of the streamfunction. Another and perhaps more physical measure is the square root of the total perturbation energy. The former will be referred to as the L_2 norm and the latter as the energy norm. It was remarked in the Introduction that while unstable normal modes imply a structure for the developing disturbance, transient growth arises from a large subset of perturbations. An advantage gained by using two norms is that examples of optimal perturbations in these norms can be compared to obtain an impression of the variability within the set of rapidly developing disturbances.

Because the structure of an isolated normal mode is invariant, all its norms increase at the exponential growth rate and one norm is as good a measure of amplitude as another. The greater freedom of choice permitted by the initial value problem with perturbations allowed to change structure with time is reflected in the freedom to choose perturbations that increase most rapidly in a given norm. For example, the L_2 norm optimum tends to favor exploiting temperature perturbations in the lower troposphere to produce large amplitude disturbances in geopotential in the upper troposphere and stratosphere when compared to an energy norm optimum at the same scale. This is because the derivative with height of the geopotential, proportional to temperature perturbations, is not directly penalized in the L_2 norm as it is in the energy norm while the decrease in density with height favors the propagation of a disturbance upward where, other effects being equal, a greater geopotential deflection will be produced. The freedom to choose a measure of perturbation amplitude has physical significance. It may be argued for instance that the most destructive cyclone corresponds to maximizing the wind at ground level.

We now consider some example problems, the first posed as follows: find the minimum initial disturbance required to excite a chosen mode at unit amplitude. The motivation could be to locate potential vorticity sources arising from topography and diabatic heating so as to give rise to the strongest driving of the external mode in a study of the planetary wave pattern. On the other hand, the geopotential configuration preceding the most rapid set up of the unstable Charney mode would be relevant to understanding the cyclogenesis problem. It might be supposed that the best way to excite one of these modes would be to put the available amplitude or energy, as the case may be, directly into the mode. This is not so; in fact, driving the desired mode directly is highly suboptimal. It is much better in flows with nonorthogonal modes to distribute the initial disturbance so that the interaction between the nonorthogonal modes and the mean flow transfers energy from the mean to the perturbation. The result of such a choice of perturbation is an increase of disturbance energy even for a stable or damped problem. This disturbance energy is drawn from the mean flow despite the possible absence of instability.

With experience one becomes skilled at producing perturbations that are energetically active and the choice is well guided by synoptic observations of configurations leading to development. However, the limiting transient growth is determined by finding the optimal perturbation. If, for instance, no perturbation could produce significant growth in a given flow then the energy of the mean is not available in the linear limit. This is a quite different result from the absence of exponential instabilities placing a limit on $t \to \infty$ asymptotic growth, and the existence of such instabilities has no clear relationship with the potential for transient growth, as should be clear from examining the overdamped Charney problem (Farrell 1985). In practice transient development results from properly configured perturbations whenever there is available potential energy in the flow, as required by energy integral relations (Pedlosky 1987).

The best perturbation for exciting a given mode is found by solution of a variational problem. In the L_2 norm the functional to be minimized is $\overline{\mathbf{\Phi}^* \cdot \mathbf{\Phi}}$ at $t = 0.0$, where the bar denotes a normalized volume integral over a wavelength. It is proportional, using the spectral projection 3.3, to:

$$(\mathbf{E}\gamma)^* \cdot (\mathbf{E}\gamma) = \gamma^* \mathbf{E}^* \mathbf{E}\gamma = \gamma^* \mathbf{A}\gamma. \quad (4.1)$$

Choosing as a constraint that the nth mode be of unit magnitude, it is necessary to render stationary the functional:

$$F = \gamma^* \mathbf{A}\gamma + \lambda(\gamma \cdot \epsilon_n - 1)$$

where ϵ_n is the unit column vector. Setting the first variation in γ to zero gives the relation that the optimum γ satisfies which is, recognizing $\mathbf{A}$ to be Hermitian:

$$\mathbf{A}\gamma = -\lambda\epsilon_n. \quad (4.2)$$

Solution for the optimal spectral projection for the nth mode,

$$\gamma = -\lambda \mathbf{A}^{-1}\epsilon_n \quad (4.3)$$

is completed by choosing λ so that $\gamma_n = 1.0$.

There is a relationship between this optimum and the eigenvectors of the matrix adjoint to $\mathbf{L}$. Whereas the eigenfunctions of a self-adjoint operator are orthogonal and consequently dynamically independent, a nonself-adjoint operator such as $\mathbf{L}$ has an adjoint operator, $\mathbf{L}^*$ which for the L_2 norm is its Hermitian transpose. The eigenvalues of $\mathbf{L}^*$ are the complex conjugates of the eigenvalues of $\mathbf{L}$ and an eigenvector of an eigenvalue in the adjoint matrix is orthogonal to all eigenvectors of the original matrix except for the one with eigenvalue conjugate to its own, provided only that the eigenvalues are distinct as can be verified for the examples to follow (Nobel 1969). It is clear from the definition of $\mathbf{A}$ in (4.1) and the condition of stationarity (4.2) that $\mathbf{E}\gamma$ must be orthogonal to all the columns of $\mathbf{E}$ except the nth. Therefore, the L_2 optimal initial condition for exciting the mode $\mathbf{\Phi}_n = \mathbf{P}\psi_n$ is

$$\mathbf{\Phi}_{\text{opt}} = \mathbf{P}^{-1}\phi_n, \quad (4.4)$$

where ϕ_n is the eigenfunction of $\mathbf{L}^*$ with eigenvalue c_n^*.

A practical consequence of this observation is that the optimal excitation in the L_2 norm can be found from eigenanalysis of the adjoint problem (Farrell 1988). It will be shown that this result carries over to solution of the differential operator as well with differential adjoints replacing the matrix adjoints of the finite difference approximation.

An alternative interpretation of the L_2 optimum arises from considering the problem of maximizing the projection of an arbitrary physical perturbation $\mathbf{\Phi}_0$ written as

$$\mathbf{\Phi}_0 = \sum_{i=1}^{N} \gamma_i \mathbf{\Phi}_i.$$

Exploiting the biorthogonality of $\mathbf{\Phi}_i$ and its adjoint $\hat{\mathbf{\Phi}}_i$ the projection is

$$\gamma_i = \frac{(\mathbf{\Phi}_0 \cdot \hat{\mathbf{\Phi}}_i^*)}{(\mathbf{\Phi}_i \cdot \hat{\mathbf{\Phi}}_i^*)}. \quad (4.5)$$

Clearly γ_i is maximized by the choice $\mathbf{\Phi}_0 = \hat{\mathbf{\Phi}}_i$ as we have already seen but in addition it is now apparent that the magnitude of γ_i depends on the projection of $\mathbf{\Phi}_i$ on its adjoint. In effect, modes with dissimilar adjoints are more able to exploit the transient growth mechanism than are modes with similar adjoints (Farrell 1988). This sensitivity of the projection of the initial perturbation on the modes implies a loss of predictability arising because the inevitable uncertainties in observing the initial state are magnified by the ill-conditioning resulting when the denominator in (4.5) is small.

Optima in the energy norm are obtained using the expression for perturbation energy, $\bar{K}$:

$$\bar{K} = \frac{\rho(0)}{2} \overline{[\alpha^2\{(\mathbf{D\Phi})^* \cdot (\mathbf{D\Phi})\} + (\mathbf{P}^{-1}\mathbf{\Phi}_z)^* \cdot (\mathbf{P}^{-1}\mathbf{\Phi}_z)]},$$

where $\mathbf{D} \equiv e^{-z_i/2}\delta ij$.

The area average energy density can be expressed using the spectrum $\boldsymbol{\gamma}$ as $\bar{K} = \boldsymbol{\gamma}^*\mathbf{B}\boldsymbol{\gamma}$ with

$$B \equiv \frac{\rho(0)}{8} [\alpha^2(\mathbf{DE})^*(\mathbf{DE}) + (\mathbf{P}^{-1}\mathbf{E}_z)^*(\mathbf{P}^{-1}\mathbf{E}_z)], \quad (4.6)$$

where $\mathbf{E}_z$ is the matrix with the eigenvector derivatives as columns.

The optimization proceeds as before, stationarity requiring of the optimal $\boldsymbol{\gamma}$:

$$\mathbf{B}\boldsymbol{\gamma} = -\lambda\boldsymbol{\epsilon}_n, \quad (4.7)$$

giving the optimal spectral projection:

$$\boldsymbol{\gamma} = -\lambda\mathbf{B}^{-1}\boldsymbol{\epsilon}_n,$$

with λ chosen to make the projection on the nth mode unity.

5. Optimal excitation of the Charney modes

The unstable mode of the Charney problem at $k = l = 2.0$ is shown in Fig. 1a. The optimal initial condition to excite this mode in the L_2 norm is shown in Fig. 1b and the optimal for the energy norm in Fig. 1c. Evolution of the streamfunction for these are given in Fig. 2 and Fig. 3 respectively. In the figures the maximum of the streamfunction as well as the norm are indicated

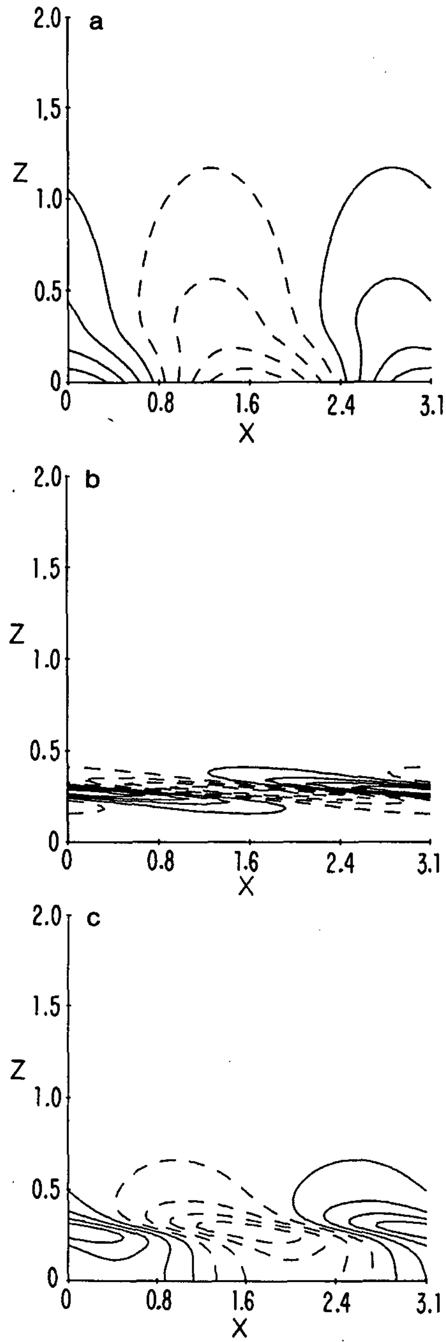

FIG. 1. (a) Unstable mode in the Charney model with $\beta = 0.53$ corresponding to midlatitude parameter values and wavenumber $k = l = 2.0$ corresponding to a zonal and meridional wavelength of 3100 km. The eigenvalue $c = (0.28, 0.068)$. Optimal excitation of the unstable mode in (b) the L_2 norm and (c) the energy norm.

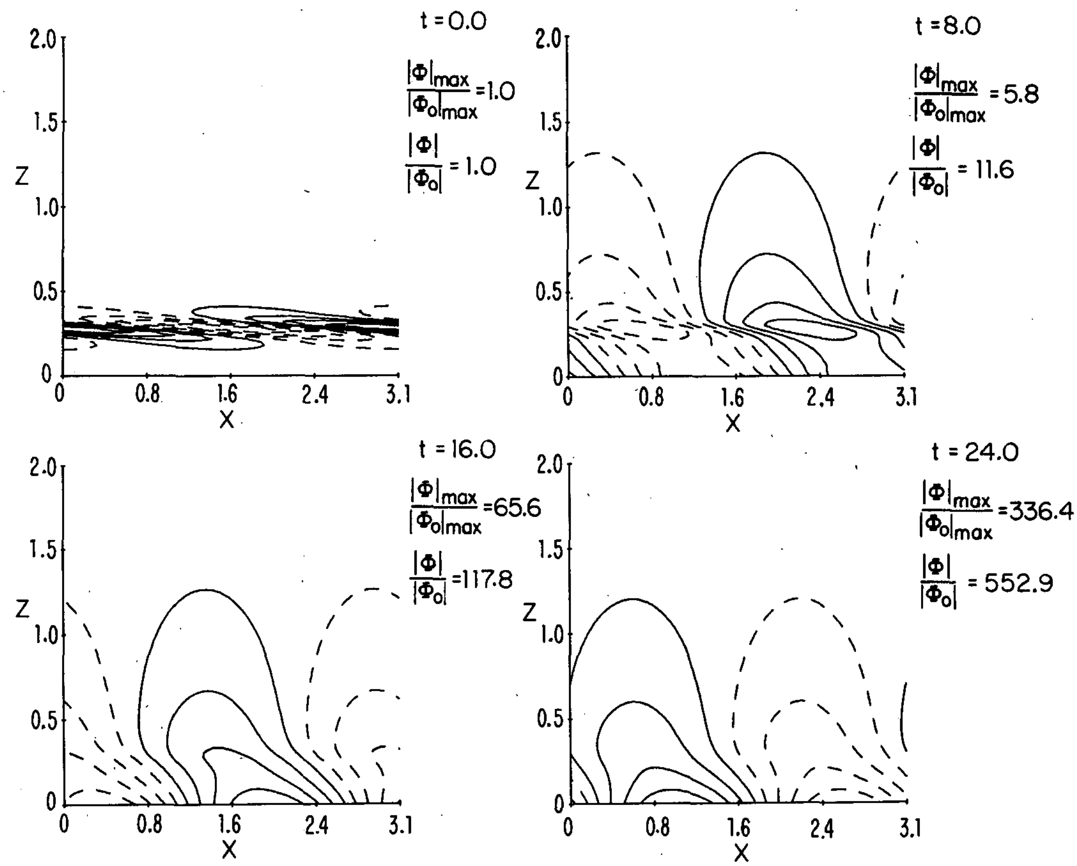

FIG. 2. Development of the L_2 norm optimal excitation of the unstable Charney mode with $k = l = 2.0$. The maximum of the perturbation streamfunction and its L_2 norm are given at the indicated times, each normalized by its initial value.

at each time so that the resolution of the plot need not be compromised by a constant contour interval.

If the unstable mode alone had been excited, the amplitude and energy would have increased by a factor of 26.2 compared with the optimal 552.9 and 282.5 respectively from Fig. 2 and Fig. 3. However, even this growth is far from the maximum obtainable in either norm as the perturbation has been highly and arbitrarily constrained to project optimally on the unstable mode and an unconstrained optimal perturbation grows even faster as we shall see. Furthermore, if a modest amount of Ekman dissipation is included, corresponding to a vertical coefficient of diffusion of 4.5 $m^2 s^{-1}$, the model does not support exponential instability at this wavenumber but, even so, transient growth on time scales appropriate to synoptic development is still robust (Farrell 1985).

The external mode in the Charney problem arises at the wavenumber for which $r = 1$ solves (Pedlosky 1987):

$$r = \frac{\beta + 1}{(1 + 4\alpha^2)^{1/2}}$$

with $\beta = 0.53$ this implies $\alpha = 0.58$. At total wavenumbers less than this there exists a neutral mode with small retrograde phase speed and the large scale equivalent barotropic structure of stationary planetary waves (Held et al. 1985). Assuming the mean surface wind in midlatitudes is small and westerly the most favorable driving of the stationary wave pattern can be studied as the optimal excitation of the external mode for a total wavenumber slightly below $\alpha = 0.58$, the wavenumber of the mode with zero doppler shifted phase speed. Here $k = l = 0.40$ is chosen corresponding to a dimensional wavelength of 15 500 km. A lid placed at $z_T = 8.0$ for computational convenience results in only a small modification of the external mode which is found to have phase speed -6.7 m s^{-1} in the model. Figure 4 shows the external mode and the optimal excitation in the L_2 and energy norms. The optimal excitation is equivalent barotropic and surface concentrated. The development of the optimum in the L_2 norm is shown in Fig. 5, and similarly the development of the energy norm optimum is shown in Fig. 6.

Return now to the biorthogonality relation between eigenvectors of L and its adjoint. A parallel relation exists between the differential equation:

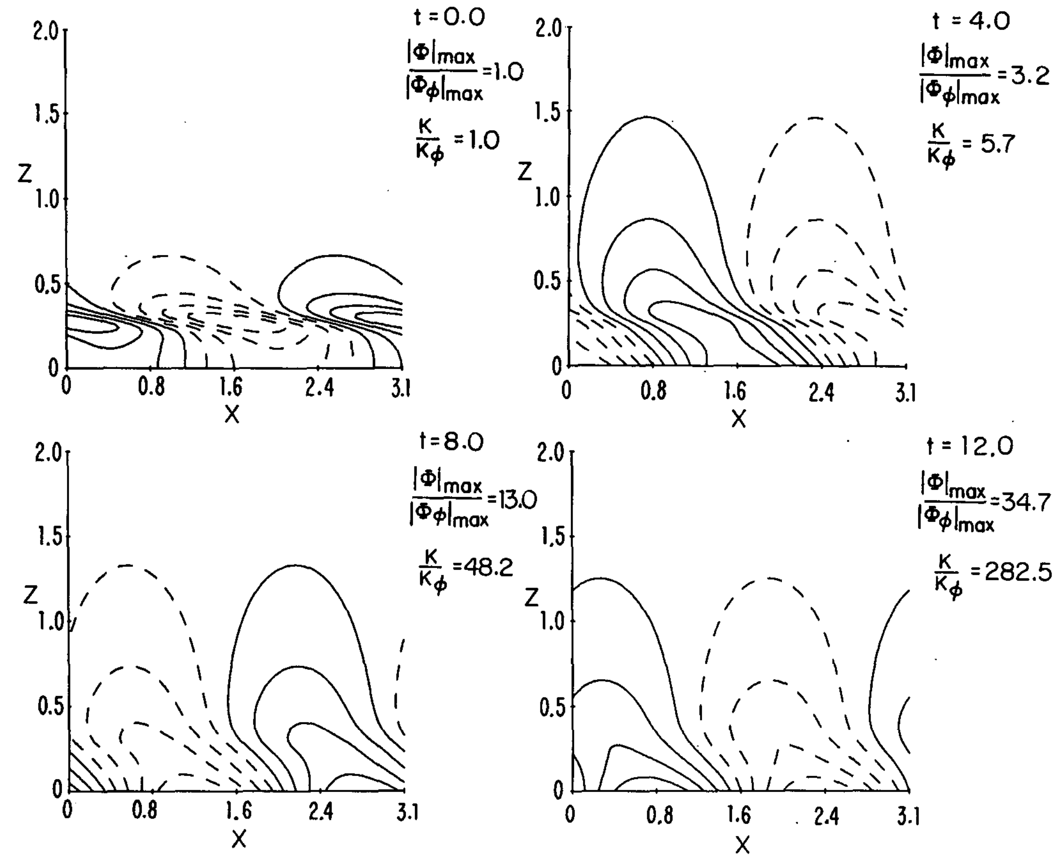

FIG. 3. Development of the energy norm optimal excitation of the unstable Charney mode with $k = l = 2.0$. The maximum of the perturbation streamfunction and the perturbation energy are given, each normalized by its initial value.

$$(U - c_i)\Delta\psi_i + \beta_{eff}\psi_i = 0 \tag{5.1}$$

and its adjoint:

$$\Delta(U - c_j)\phi_j + \beta_{eff}\phi_j = 0. \tag{5.2}$$

Use is made of the Bretherton (1966) δ adjustment of the zonal wind to $U_z = 0$ in a thin layer adjacent to the boundaries which simplifies the boundary condition to

$$\psi_z + \frac{\psi}{2} = \phi_z + \frac{\phi}{2} = O_2, \quad z = 0$$

$$\psi \to 0, \quad \phi \to 0, \quad z \to \infty.$$

Remarks to follow apply equally to the barotropic model (Farrell 1988) for which $\psi = \phi = 0$ is appropriate at channel walls.

Inspection of (5.1) and (5.2) suffices to establish $\phi_i = \psi_i/(U - c_i)$. If (5.1) is multiplied by ϕ_j and (5.2) by ψ_i and the difference taken, integration by parts results in the orthogonality relation:

$$(c_i - c_j)\int_0^\infty \psi_i\Delta\phi_j dz = 0. \tag{5.3}$$

The argument leading to expression (4.3) for the optimal spectral projection in the vector product $\psi_i^* \cdot \psi_j$ can be made in a similar fashion for continuous functions by replacing the vector product with the analogous expression:

$$\int_0^\infty \psi_i^*\psi_j dz.$$

Examination of 5.3 reveals that properly normalized eigenfunctions, ψ_i with eigenvalue c_i and adjoint eigenfunctions ϕ_j with eigenvalue c_j are biorthogonal in a way similar to that of their matrix analogues,

$$\int_0^\infty \psi_i\Delta\phi_j = \delta_{ij}. \tag{5.4}$$

An examination of (4.4) and (5.4) shows that the analogy between the matrix and differential optima requires the physical matrix adjoint mode $\mathbf{\Phi}_i = \mathbf{P}^{-1}\phi_i$ be identified with the Laplacian of the conjugate of the differential adjoint mode $P^{-1}\Delta\phi_i^*$ where $P^{-1} = e^{-z/2}\sqrt{\epsilon}$. The conceptual advantage of this observation is that the optimal excitation in the L_2 norm can be found almost by inspection knowing the target mode and its complex phase speed. The Laplacian of

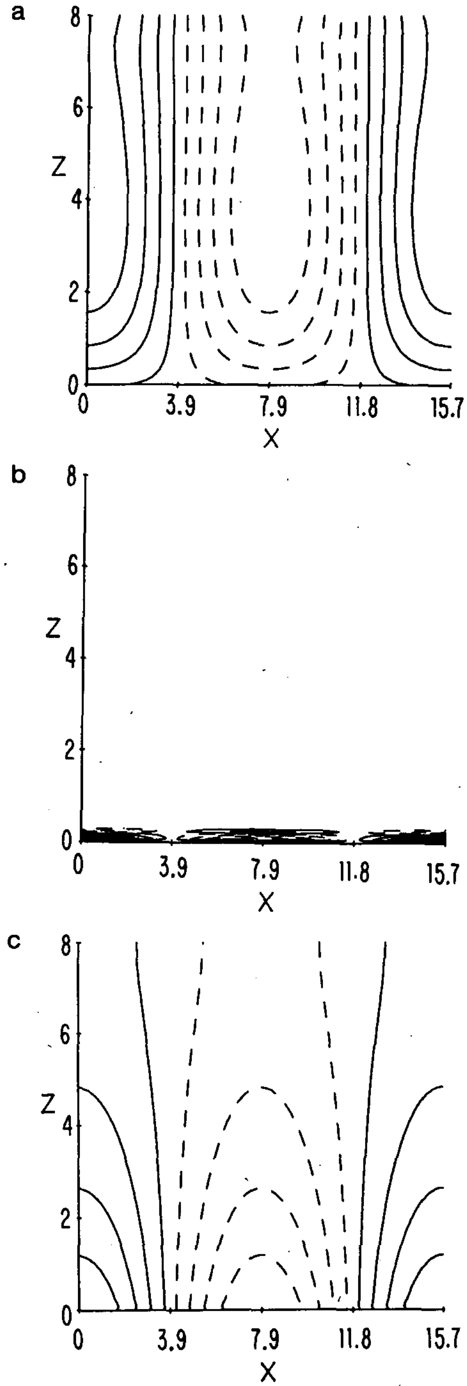

FIG. 4. (a) The external mode in the Charney problem with $k = l = 0.40$. Optimal excitation of the external mode in (b) the L_2 norm and (c) the energy norm.

the streamfunction amplified by the factor of the inverse of $(U - c)$ produces a highly structured wave field concentrated near the steering level. For example $[\psi/(U-c)]^*$ and its Laplacian scaled by P^{-1} are shown in Fig. 7 for the Charney mode at $k = l = 2.0$; Fig. 7b should be compared with Fig. 1b.

In the case of the barotropic problem and in the baroclinic if the density and stability variation can be ignored an even more direct relation exists between the optimal excitation in the energy norm and the differential adjoints. To see this set $\Delta \equiv \nabla^2$ and integrate (5.4) twice by parts:

$$\int_0^\infty (\nabla^2\psi_i)\phi_j dz = \delta_{ij}.$$

This is the integral equivalent of (4.7) if the identifications $\mathbf{B} \rightarrow (\nabla^2\mathbf{E}^*)\mathbf{E}$ and $\phi_n^* \rightarrow \mathbf{E}\gamma_n$ are made. The requirement for stationarity in the discrete case, that $\mathbf{E}\gamma_n$ be biorthogonal to $\nabla^2\mathbf{E}_i^*$, is parallel to the requirement that ϕ_n be biorthogonal to $\nabla^2\psi_i$ in the continuous case. We can identify $\phi_n^* = [\psi_n/U - c_n]^*$ and the optimal perturbation in the energy norm. For instance in the Charney problem with $k = l = 2.0$ stratification plays a minor role in the Laplacian and Fig. 7a is nearly identical to Fig. 1c.

A different perspective on projections and excitation of modes is afforded by a discussion complementary to the above framed in terms of pseudomomentum orthogonality (Held 1985).

6. Optimal growth over fixed time

Optimal excitations demonstrate directly by example that the available potential energy of the mean flow may be as available to neutral modes on time scales appropriate to synoptic and planetary space scales as it is to unstable modes. Furthermore, the optimal excitation of the unstable Charney mode shows that the unstable mode is far from the most favorable perturbation even for the task of exciting itself. These observations serve to clarify the relation of the neutral and unstable modes in the initial value problem. However, a problem of greater physical interest is to find the most rapidly growing perturbation without arbitrary restrictions on its spectral projection. For instance, a large amplitude planetary wave pattern can be more effectively excited by an optimal initial condition that does not also require the external mode be maximally driven because this additional restriction results in a suboptimal total growth. As another example consider the cyclogenesis problem. We have seen that the Charney mode can be excited optimally by perturbations which do not initially resemble it and that grow much faster. But the problem of cyclogenesis is best addressed without modal prejudice. It is stated as follows: find the disturbance that grows maximally in energy over 48 h.

The functional to be rendered stationary is

$$F = \gamma^*\mathbf{B}_t\gamma + \lambda(\gamma^*\mathbf{B}_0\gamma^{-1}).$$

Here $\mathbf{B}_t$ is given by (4.6) where $\mathbf{E}$ is replaced by $\mathbf{E}_t$, the matrix of eigenvectors each advanced in time as in (3.2).

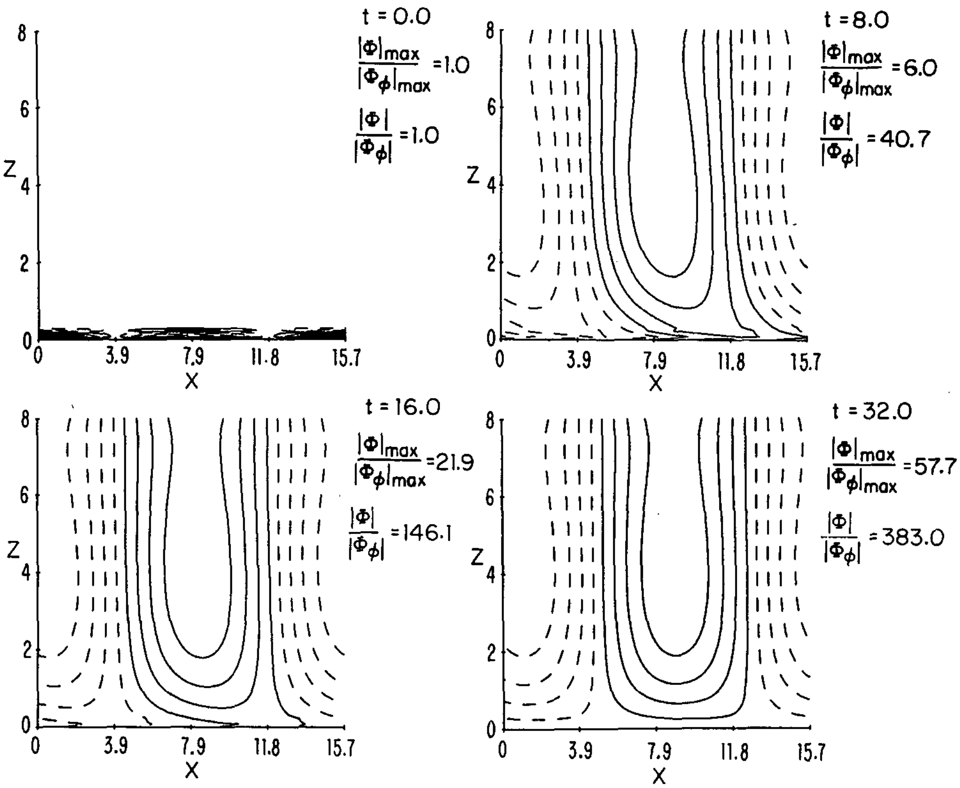

FIG. 5. Development of the L_2 norm optimal for the external mode in the Charney problem with $k = l = 0.40$.

The requirement of stationarity is that γ satisfy

$$\mathbf{B}_t\gamma + \lambda\mathbf{B}_0\gamma = 0,$$

or:

$$(\mathbf{B}_0^{-1}\mathbf{B}_t + \lambda\mathbf{I})\gamma = 0.$$

The eigenvectors of the above matrix are the spectral projections of the stationary solutions and one of these is the desired optimum.

For the cyclogenesis problem, a zonal wind and static stability distribution is chosen to model the troposphere and lower stratosphere. The zonal wind rapidly approaches a constant above a scale height and static stability increase by a factor of 4 at this simulated tropopause:

$$U(z) = z - (z - 1.5)\left[\frac{1 + \tanh\left(\frac{z - 1.5}{0.15}\right)}{2}\right]$$

$$\epsilon^{-1}(z) = 1 + 3\left[\frac{1 + \tanh\left(\frac{z - 1.5}{0.15}\right)}{2}\right].$$

Otherwise parameter values are the same as in the midlatitude cyclogenesis example in section 3 except that Ekman damping corresponding to $\nu = 10\ \mathrm{m^2\,s^{-1}}$ has been included ($\Gamma = 0.075$) to eliminate unstable modes. The perturbation that results in maximum growth in energy over 5 nondimensional time units, corresponding to 47 h, is shown in Fig. 8 and its growth rate in Fig. 9. This perturbation resembles the classic configuration of an upper tropospheric trough approaching a surface center which was modeled by a plane wave tilting westward with height in Farrell (1985). It is remarkable that the optimum perturbation resembles so closely a disturbance chosen to model observations of a common precursor to cyclogenesis considering that this is an unconstrained optimum. It appears that the observed wave field is rich enough that synopticians have accurately identified the form of the most rapidly growing perturbation.

A final example makes use of the realistic basic state at wavenumber $k = l = 0.40$ corresponding to a planetary wave length of 15 500 km with the addition of a Rayleigh damping in the upper stratosphere:

$$R(z) = \frac{1 + \tanh(z - 6.0)}{2}$$

in addition to Ekman damping as in the previous example ($\Gamma = 0.075$). The energy optimum over 14 time units corresponding to 5.4 d is shown in Fig. 10. A

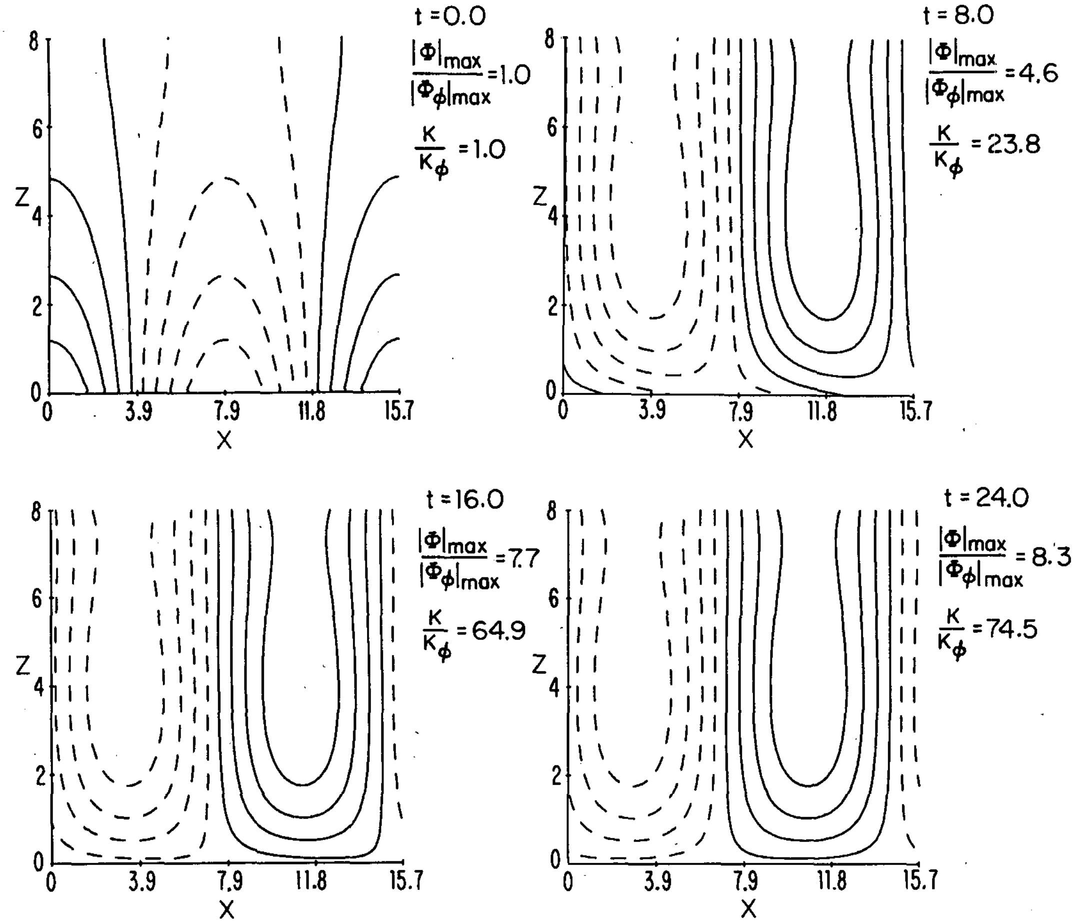

FIG. 6. Development of the energy norm optimal for the external mode in the Charney problem with $k = l = 0.40$.

rapid growth of perturbation amplitude results from a low level wave being overtaken by an upper level disturbance in a manner similar to that operating at the synoptic scale in the previous example but here at planetary scale. It can be seen that this development excites a strong response in the stratosphere. Episodic development has been observed to occur at these space and time scales accompanied by similar transient baroclinic structures (Shilling 1988).

7. Discussion and conclusions

The primary source of energy for synoptic and planetary scale waves is the APE associated with the baroclinic vertically sheared zonal wind. Understanding mechanisms for the excitation and maintenance of the wave field is central to theoretical progress on the forecast and predictability problem on both the cyclone and planetary scales. Previous work (Farrell 1985) has shown that perturbations chosen to roughly model those associated with development in the synoptic literature (Petterssen 1955; Palmen and Newton 1969) result in growth much more rapid than that of unstable modes in flows that support such modes and that even when no unstable modes exist properly chosen perturbations undergo a period of transient growth on time scales appropriate to synoptic development that is little diminished by the absence of instability. These results imply a redirection of attention from the most unstable mode and its structure to the combination of modes neutral, damped and perhaps unstable which together produce rapid development. Because developing perturbations comprise a fairly large subset of all perturbations there are a number of variations on the theme of development. Having suspended the search for an exponential instability behind every energetic event, one begins to see the evolving wave field in a new light. There are many scenarios on display, some of which are identified in the synoptic literature by such names as Petterssen's type B cyclogenesis, tilted trough development, Sutcliffe's self-development and trough phasing. These can be understood as indicative of the presence of perturbations favorable to transient development.

There is no necessity for an energetic event to be associated with an optimal excitation as the examples

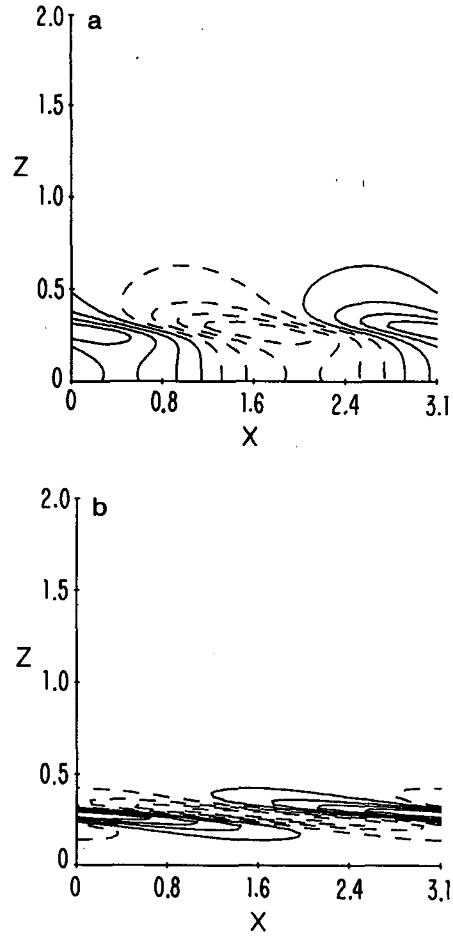

FIG. 7. (a) The complex conjugate of the differential adjoint of the unstable Charney mode with $k = l = 2.0$. This is an approximation to the energy norm optimal excitation in Fig. 1c. (b) The optimal excitation in the L_2 norm of the unstable Charney mode with $k = l = 2.0$ given by the Laplacian of the complex conjugate of the differential adjoint. This is to be compared to the discrete optimal in Fig. 1b.

in Farrell (1985) show. The suboptimal perturbations used in that work were meant to approximately correspond to one common precursor of development. A little experience integrating a model builds up an intuition and one becomes skilled at obtaining a variety of growth scenarios. However, obtaining the optimal perturbation for a given flow has the theoretical advantage of bringing order to the search and identifying the limits of the transient growth. While it is easy to find perturbations that grow robustly, the examples above of optimal perturbations are more highly structured in some cases than one might be bold enough to try without encouragement.

Additional topics of discussion follow:

(i) A generalization of transient development that is of heuristic value arises from considering a stationary solution to the nonlinear equations with allowance for a steady progressive wave. In the frame of reference of the wave the condition of stationarity is

$$J(\bar{\psi}, \bar{Q}) = 0,$$

where $\bar{\psi}$ is the stationary streamfunction and $\bar{Q}$ is the potential vorticity, both perhaps a function of z. Consider a perturbation to this solution $\psi = \bar{\psi} + \psi'$. If the perturbation is sufficiently small its development is governed by a linear equation resulting from linearizing about the stationary solution. This equation is analogous to (1) but is of higher dimension if the stationary wave field varies in other directions. Assuming the stationary solution has a nonzero deformation field associated with it, there are perturbations that decay and those that grow. The optimal perturbation over a given time interval is the one that grows the most and it can be found by means analogous to those used above. This most dangerous perturbation produces the maximum disruption of the stationary solution and this optimum is central to determining the predictability. A flow is unpredictable on the given time scale in proportion to the growth of the optimal excitation not, as is commonly assumed, in proportion to its first Lyapunov exponent which is the growth rate of the most unstable exponential normal mode. In the limit $t \to \infty$ these growth rates are the same if the flow supports an instability but as examples above make clear, over synoptic time scales the $t \to \infty$ asymptotic is inappropriate and predictability is likely to be much worse than the first Lyapunov exponent would suggest.

This result may be generalized to a fully developed nonlinear wave field by appeal to the tangent differential equation which is the linear equation linearized about the time developing flow field. This equation governs the growth of small perturbations to the nonlinear, possibly turbulent flow. It admits optimal perturbations which control the predictability on synoptic time scales as well as Lyapunov exponents to which the optimum is asymptotic as $t \to \infty$. These issues are further developed in the context of a barotropic model by Lacarra and Talagrand (1988).

(ii) The transient growth mechanism and the optimal perturbations have been discussed with reference to the quasi-geostrophic equations here and using the barotropic vorticity equations in Farrell (1988). This was done for analytic and computational convenience and it is clear from the analysis above that the mechanism depends only on the fact that the modes of the problem are not orthogonal. This in turn results from the differential dynamical equation being nonself-adjoint. Any linear evolution equation of the form:

$$\frac{\partial \psi}{\partial t} = L\psi$$

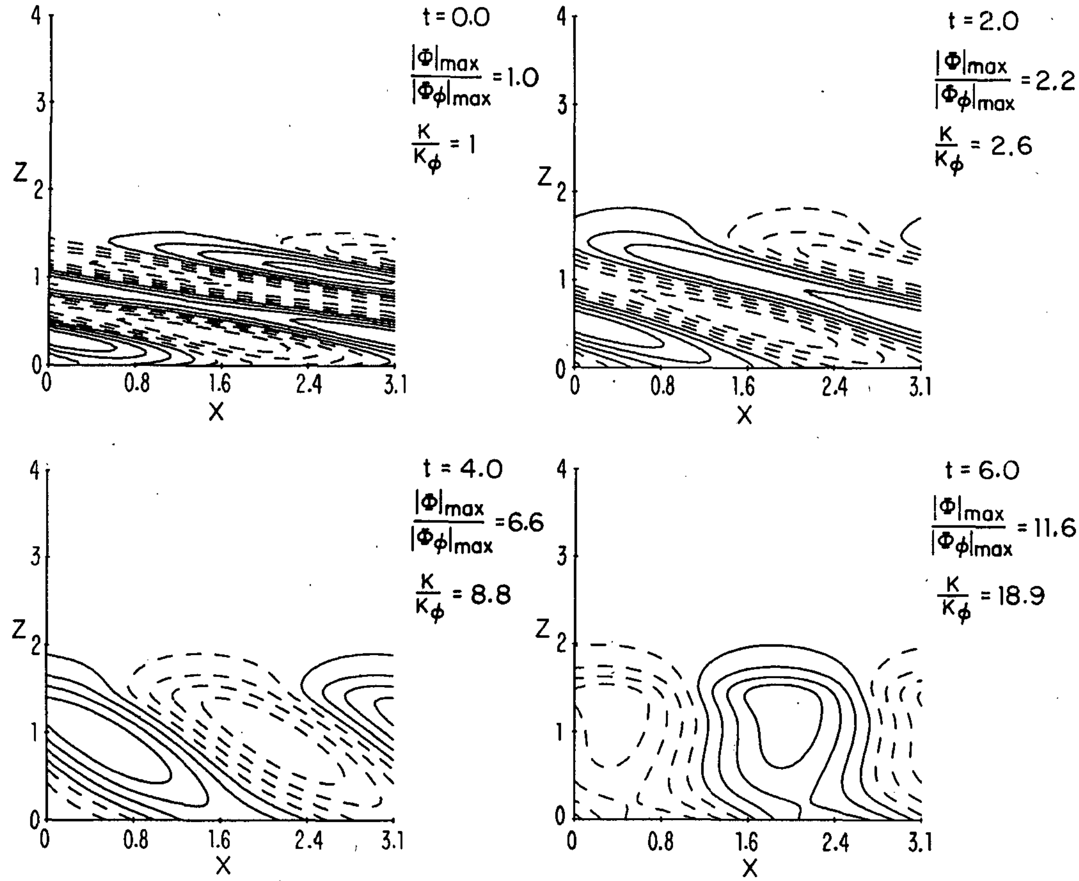

FIG. 8. The optimal excitation in the energy norm over five units of nondimensional time of the realistic basic state with supercritical damping and $k = l = 2.0$. No requirements are imposed on the modal composition.

supports the mechanism if L is not self-adjoint, including the primitive equations. The only common exceptions are dynamical equations for a channel with a constant zonal wind and analogously solid body rotation on the sphere. It is clear in these cases without a mean deformation field that there can be no energetic interactions.

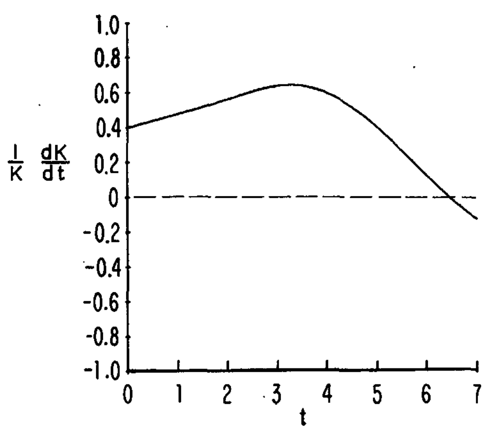

FIG. 9. The energy growth rate for the overdamped realistic basic state optimal excitation in Fig. 8.

(iii) Perturbations favorable for development can arise from the superposition of interior potential vorticity centers as in the example of cyclogenesis in Fig. 8. In that case it is clear that a development will ensue because the perturbation has a streamfunction field that tilts westward with height, the energetically favorable configuration (Pedlosky 1987). However, developing perturbations are not limited to such configurations. For example, the stationary wave in Fig. 6 develops from an equivalent barotropic perturbation in the classic self-development scenario, acquiring its tilt as the disturbance evolves.

(iv) Disturbances that give rise to high amplitude smooth, low vertical wavenumber modes are themselves often of small amplitude with complex, high vertical wavenumber structure. If observations are collected using an averaging technique it is likely that the large amplitude final state will dominate the perceived structure and the crucial highly variable perturbation it came from will be lost in the noise. This can have

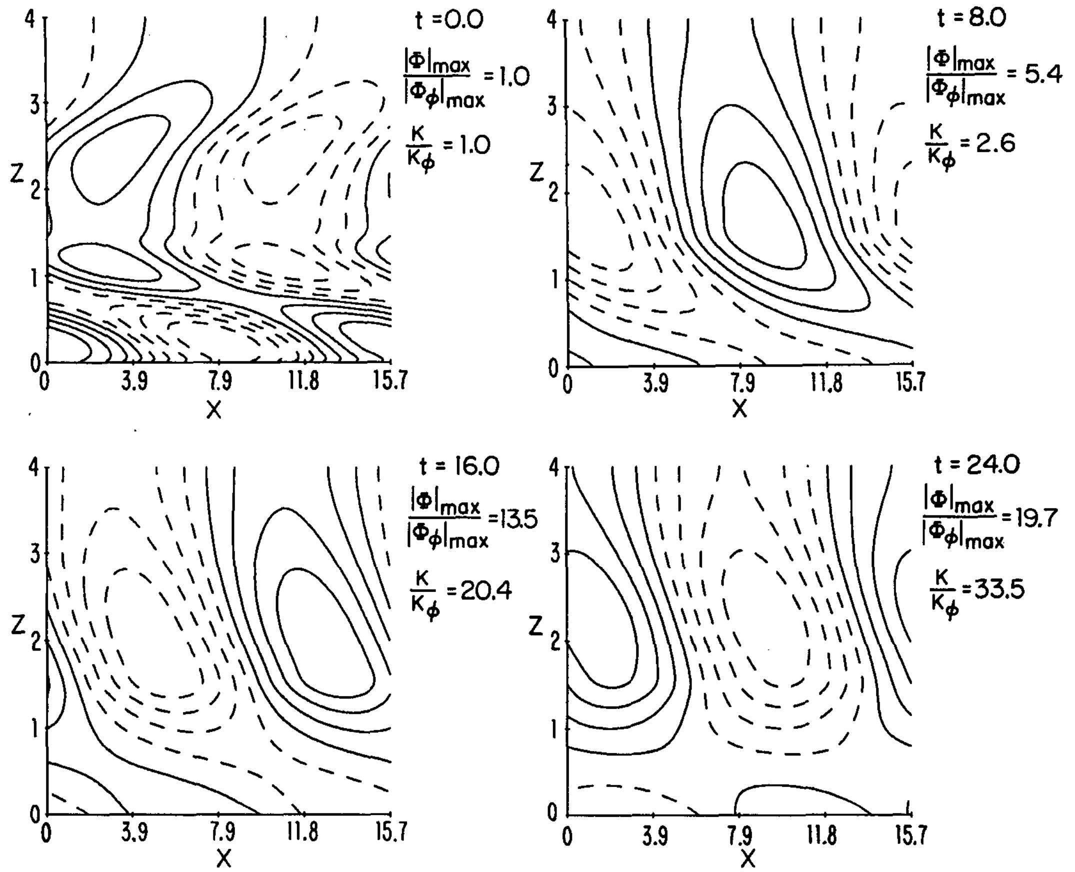

FIG. 10. The optimal excitation in the energy norm over 14 time units corresponding to 5.4 d of the realistic basic state with $k = l = 0.40$ corresponding to a planetary scale wavelength of 15 500 km.

the effect of diverting attention from the dynamically relevant small scales to the comparatively inert large scale final states.

A similar comment applies to models where it is sometimes asserted that because the energy bearing scales are large, coarse resolution can be justified. As we have seen, these large scale structures may be related on advective time scales to much higher wavenumber disturbances which must be resolved in order to accurately forecast development.

(v) Transient development provides a powerful mechanism for amplifying small perturbations in the forcing of the planetary wave field. An intermittent excitation such as a cold surge over Southeast Asia that has the form of a near optimal excitation would generate a rapidly developing baroclinic wave that draws on the APE of the Pacific jet and reaches a high amplitude and equivalent barotropic structure in about a week (Figs. 5 and 6). The stochastic character of the formation would be due to the chance occurrence of a favorable configuration which requires both that the background flow admit a robust optimum and that the forcing approximate it. This scenario is similar to that advanced by Branstator (1985) using barotropic dynamics.

(vi) Consider the role of planetary waves in the local and global climate. A substantial fraction of the global heat flux is associated with these waves which are neutral in their modal structure. As we have seen, however, proper excitation can result in a large baroclinic flux as the waves amplify. This flux is directly related to the fine structure of the forcing which in turn is related to the details of orography and heating implying a great sensitivity of the planetary wave amplitude and heat flux to what may appear to be modest changes in the forcing.

Acknowledgments. This work was supported by NSF Grant ATM-8712995.

REFERENCES

Branstator, G., 1985: Analysis of general circulation model sea surface temperature anomaly simulations using a linear model. Part II: Eigenanalysis. *J. Atmos. Sci.*, **42**, 2242–2254.

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Quart. J. Roy. Meteor. Soc.*, **92**, 325–345.

Card, P. A., and A. Barcilon, 1982: The Charney stability problem with a lower Ekman layer. *J. Atmos. Sci.,* **39,** 2128–2137.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.,* **4,** 135–162.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus,* **1,** 33–52.

Farrell, B., 1985: Transient growth of damped baroclinic waves. *J. Atmos. Sci.,* **42,** 2718–2727.

——, 1987: Developing disturbances in shear. *J. Atmos. Sci.,* **44,** 2191–2199.

——, 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.,* **45,** 163–172.

——, 1989: Unstable baroclinic modes damped by Ekman dissipation. *J. Atmos. Sci.,* **46,** 397–401.

Green, J. S. A., 1960: A problem in baroclinic stability. *Quart. J. Roy. Meteor. Soc.,* **86,** 237–251.

Held, I. M., 1985: Pseudomomentum and the orthogonality of modes in shear flows. *J. Atmos. Sci.,* **42,** 2280–2288.

——, R. L. Panetta and R. T. Pierrehumbert, 1985: Stationary external Rossby waves in vertical shear. *J. Atmos. Sci.,* **42,** 865–883.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting. *Tellus,* **35,** 100–118.

Hoskins, B. J., and P. J. Valdes, 1988: Baroclinic instability of the zonally averaged flow with boundary damping. *J. Atmos. Sci.,* **45,** 1584–1593.

Kuo, H. L., 1979: Baroclinic instabilities of linear and jet profiles in the atmosphere. *J. Atmos. Sci.,* **36,** 2360–2378.

Lacarra, J., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus,* **40A,** 81–95.

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus,* **7,** 157–167.

Nobel, B., 1969: *Applied Linear Algebra.* Prentice-Hall, 523 pp.

Palmen, E., and C. Newton, 1969: *Atmospheric Circulation Systems.* Academic Press, 603 pp.

Palmer, T. N., 1988: Medium and extended range predictability and stability of the Pacific/North American mode. *Quart. J. Roy. Meteor. Soc.,* **114,** 691–713.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* Springer-Verlag, 710 pp.

Petterssen, S., 1955: A general survey of factors influencing development at sea level. *J. Meteor.,* **12,** 36–42.

Salmon, R., 1980: Baroclinic instability and geostrophic turbulence. *Geophys. Astrophys. Fluid Dyn.,* **15,** 167–211.

Shilling, H. D., 1988: Baroclinic outbreak episodes in wavenumber domain. *Tellus,* **40A,** 188–204.