Advances in Cyclogenesis Theory: Toward a Generalized Theory of Baroclinic Development

BRIAN F. FARRELL
Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, USA

1. Introduction

Baroclinic cyclones are the primary weather-producing agents in midlatitudes, and the passing of lows and highs on the synoptic chart has long been recognized as heralding the change from storm to clearing skies. The association between a "falling glass" and deteriorating conditions gave impetus to the development of a variety of cyclogenesis theories, and the quest for a physical understanding of the cyclone has challenged some of the finest minds from the previous century to our own day. While all of the theories preceding our present understanding in terms of quasigeostrophic instability theory enjoyed an interval of ascendency, each predecessor ultimately suffered a more or less thorough eclipse; a historical observation that should at the least serve to forestall complacency.

In the "New World," atmospheric science enjoyed an early acceptance, and a surprising number of the pioneers in meteorology hailed from the brash provincialism and intellectual ferment of the post-colonial United States, including William Redfield, James Espy, Elias Loomis, and William Ferrel (Kutzbach 1979). Evidently, this interest was not parochial in character, as Vilhelm Bjerknes himself was quickly recognized across the Atlantic for his generalization of the circulation theorem and awarded a Carnegie Institution grant after his U.S. lecture tour in 1905, which supported some of his work in Bergen until 1941.

The first scientific theory of cyclogenesis is arguably that of Espy (1841) who maintained that the heating and resulting expansion of air when water vapor is condensed to form clouds must give rise to a warm core depression in a region of intense precipitation. Moreover, in Espy's view the wind field must converge upon the depression—a theory that contrasted with the centrifugal balanced winds circling the low pressure center envisioned in the contemporary theory of Redfield (1831). These opposing predictions were put to the test by Loomis (1841), who was an early advocate of the case study approach and claimed perceptively that "meteorology is to be promoted, not so much by taking the mean of long continued observations, as by studying the phenomena of particular storms." The results of his investigation of the Great Storm of 1836 did not support Espy's theory, and he was forced to conclude that "although a fall of the barometer is usually accompanied by an elevation of temperature, the reverse is sometimes the case."

In addition to the contradictory common observation that copious rainfall is not invariantly associated with cyclogenesis, the prediction that winds blow into the center of the low was not supported by the case study undertaken by Loomis, although it must be admitted that the alternative of Redfield also failed full support. The debate between Espy and Redfield on the structure and origin of cyclones is instructive in that although evidence against the theory of Espy was immediately produced, the theory itself persisted even in the writings of the Bergen school (J. Bjerknes 1919). This persistence of Espy's thermal theory in spite of overwhelming observational evidence that midlatitude cyclones are primarily cold core provides additional confirmation that a logical and appealing theoretical construct can withstand a protracted siege of contradictory observation without apparent damage.

In the latter part of the nineteenth century, upper-level observations from mountain stations, kites, and balloons had a role in establishing the thermal structure of cyclones and bringing a weight of contradictory evidence to bear upon the predictions of the condensation theory of their formation; in the middle of the twentieth century, a parallel role occurred when extensive upper air observations collected after World War II began to show that the surfaces of discontinuity in temperature and wind envisioned by the frontal cyclone model did not extend into the upper atmosphere. By this time it was recognized from the work of Margules (1903) that the energy source for the cyclone was in fact potential energy associated with geostrophically balanced density contrasts arising in association with thermal contrasts. Drawing on the ideas of Helmholtz (1888), Margules (1906a) postulated an extended zero-order front making an angle α with the horizontal satisfying: .

$$\tan(\alpha) = \frac{f}{g}\frac{T_c v_w - T_w v_c}{T_w - T_c} \tag{1}$$

where $T_{c,w}$ and $v_{c,w}$ refer to the separate warm and cold layer temperature and front parallel velocity, f is the Coriolis parameter, and g is gravitational acceleration. This represented a lasting advance in that the baroclinic energy source for the cyclone had been identified, and Margules' advance constituted a revolutionary departure from the thermal theory of Espy. However, Margules (1906b) recognized that problems remained in that the mechanism by which this potential energy is realized to form the cyclone was as yet unclear. He perceptively indicated a possible resolution of this difficulty in terms of finite amplitude instability of the jet:

> Horizontal temperature gradients and masses with a great store of potential energy are everywhere but the storm is relatively rare. The masses, associated with moderate velocities, remain in a near-stationary state, which maintains itself. . . . Are there larger disturbances of the dynamic equilibrium, which grow on their own, feeding from the potential energy . . . (?)

The Bergen school sought a physical model of the process by which the release of potential energy identified by Margules takes place. Under the guidance of V. Bjerknes, a conceptual mechanism was advanced in which a process of transient frontal deformation produced structural evolution, ultimately taking the form of occlusion and lifting of the warm air from the surface. This frontal cyclone theory comprised an essentially transient finite amplitude description that made no necessary reference to linear perturbation theory. However, the well-known linear instability study of shear layers by Helmholtz (1886) strongly implied that geostrophically balanced frontal discontinuities should be perturbatively unstable. Conceptually useful as the Norwegian cyclone model may have been and continues to be, it ultimately failed to ground itself in a convincing perturbative baroclinic instability. In addition, events overtook the theory after World War II when upper air soundings made it increasingly clear that frontal discontinuities failed to penetrate into the midtroposphere but instead gave way to continuous jets. However, being superseded theoretically has not visibly diminished the influence of the polar front theory as a descriptive device, as a glance at any synoptic chart testifies.

While the thermal and polar front theories of cyclone formation were descriptive and phenomenological, the baroclinic instability theories of Eady (1949) and Charney (1947) adopted the program of Rayleigh (1880), according to which the explanation for the occurrence of waves at finite amplitude in a fluid was to be understood as the logical consequence of the existence of exponentially growing modes of the dynamic equations linearized about a representative background flow. With a flow chosen to model the recently discovered continuous baroclinic jets and with use of the recently formulated quasigeostrophic equations, Charney and Eady succeeded in finding unstable modes in highly simplified model problems. The normal mode paradigm envisioned modes growing exponentially from infinitesimal beginnings over a large number of e-foldings so that the mode of greatest growth emerged finally as a finite amplitude wave. This assumption of undisturbed growth was necessary to assure the asymptotic dominance of the most rapidly growing exponential normal mode and its emergence as the dominant structure at finite amplitude, which permitted the theory to make predictions concerning the structure of observed weather systems. In the words of Eady (1949): .

> An arbitrary disturbance (corresponding to some inhomogeneity of an actual system) may be regarded as analyzed into "components" of a certain simple type, some of which grow exponentially with time. In all the cases examined there exists one particular component which grows faster than any other. [and] by a process analogous to "natural selection" this component becomes dominant.

Despite this invocation of Darwin, baroclinic instability theory seems to have shared with its predecessor thermal theory an immunity to natural selection. Difficulties appeared immediately, and Eliassen (1956) was forced to conclude that:

> The present instability theories, while useful for explaining the broad features of the formation of growing disturbances, are of little use in synoptic analyses, except as background information.

This was, as Petterssen (1955) realized, due to the fact that:

> Cyclogenesis results not from the release of an infinitesimal perturbation by a dynamically unstable state, but rather from the release of some kind of instability by finite perturbations which can be identified with the wave-shaped motion patterns in the middle and upper troposphere.

In fact, the genesis and growth of midlatitude cyclones takes place primarily in association with preexisting disturbances that take the form of fronts or shallow centers at the surface and of short waves or jet streaks aloft. Forecasters have long recognized the importance of the interaction between these disturbances by pointing to correlates of this interaction, such as "positive vorticity advection over the low center" as associated with development. By contrast, the modal theory of cyclogenesis developed in the past 40 years deals exclusively with exponential growth of infinitesimal waves of normal mode form. As we have seen from Petterssen's remarks, the widespread acceptance of this paradigm by the theoretical community was never accompanied by equal enthusiasm among synoptic meteorologists. These reservations were accurately stated by Palmén (1951):.

> The disturbance caused by the supposed instability of the zonal current has been the subject of a large number of theoretical studies in the last twenty-five years. The instability of "infinitely small" disturbances superimposed upon the westerly current has been the usual starting point in these investigations. A study of these attempts to solve the cyclone problem, however, gives the impression that there still does not exist any theoretical solution fully applicable to the cyclone problem. Therefore the question arises whether it is, in principle, permissible to start from infinitely small perturbations in discussing the cyclone problem. If such an infinitely small perturbation could cause the development of strong cyclones, it would indicate that the atmosphere is extremely unstable. How such an instability associated with storing of useful potential energy could develop in an atmosphere where strong perturbations of all kinds are always present seems difficult to understand.
>
> If we consider this, it seems more likely that extratropical cyclones are induced by rather large migrating disturbances which were already in existence. These large disturbances must always be present. Since cyclones and anticyclones are probably cells for transforming potential energy into kinetic energy, the pre-existence of a large amount of useful potential energy is necessary for a strong cyclonic development. However, the pre-existing situation must correspond to some kind of potential instability that can be released only by the influence of finite perturbations.

Nor was this dissatisfaction with the modal theory of development confined to the meteorological community; in a discussion of results from modal stability analyses applied to the canonical engineering shear flow problems of Couette and pipe Poiseuille flow, which were well known to exhibit transition to turbulence when perturbed at sufficiently high Reynolds number, Sommerfeld (1946) voiced his doubts as follows:

> In every case the transcendental equation associated with the problem leads to the conclusion that the flow is stable. Thus we have a striking contradiction between theory and experiment. What conclusions should we draw? Should we suspect the method of small oscillations that has proved reliable in all other domains of mechanics including astronomy? Should we rather assume that for such a proof of instability finite instead of (infinitely) small disturbances ought to be resorted to just because the object of investigation is the laminar motion? Or should we blame the Navier-Stokes equations as inadequate for our problem? This does not seem justifiable either, particularly since in the last analysis they form the foundation for all previous theoretical statements.

The failure of modal instability theory to account for observations of perturbation growth in the atmosphere and in laboratory flows was twofold: on the one hand, the postulated modes often failed to exist, especially in flows more realistic than those chosen to illustrate the theory, and even when instabilities could be found, modal growth rates were often far too slow to account for observed development timescales. However, an equally fundamental problem was that identified by Petterssen—observed development occurs in association with time-varying structures that do not resemble the fixed structure of normal modes.

Once the $t \rightarrow \infty$ asymptotic underpinnings of modal theory are abandoned, in light of observed temporal decorrelation time scales of synoptic disturbances on the order of six days, it is necessary to inquire if there are alternative transient structures with competing growth rates on the synoptically relevant time scales of a day to a week. It turns out that such disturbances can be found (Farrell, 1989a), and the most rapidly growing of these structures exhibit the characteristic development time and structural evolution (upper trough overtaking surface depression) that is the hallmark of midlatitude cyclogenesis. This transience of structure during baroclinic development and corresponding greater potential growth had been explained to generations of synoptic students in the insightful heuristic model of Sanders (1971).

Recognition that a large subspace of perturbations grows by exploiting both the barotropic and baroclinic energy of the meanflow, including the energy associated with downstream variation of jets in diffluence and confluence (Farrell 1989b), allows the great variety of transient development phenomena referred in synoptic parlance by colorful descriptions such as "digging troughs" and "phasing troughs" to be included under the aegis of stability theory. Freed of concentrating on the $t \rightarrow \infty$ asymptote, instability theory transcends the most telling objections of the synoptician: that the observed structures are not modal but highly variable both temporally and structurally.

There emerges from the study of transient development an instability theory affirming the revolutionary insight of Margules that the cyclone's energy source is primarily baroclinic, as well as that of Eady and Charney that quasigeostrophic perturbation theory is valid in flows with high shear such as midlatitude jets, but that incorporates these ideas in a synthetic extension of traditional modal stability theory. This generalization of stability theory explains individual examples of cyclogenesis and the variety of other transient development phenomena while also providing a theory for the growth of small forecast errors (Farrell 1990).

Moreover, the existence of a subspace of growing disturbances in baroclinic flow, as revealed by generalized stability theory, suggests a further generalization is possible. It can be shown that the net source of energy to the perturbation field attributable to nonlinear interactions among waves vanishes. From this it follows that the source of perturbation energy that arises from extraction of mean flow energy by the set of growing disturbances identified by generalized stability theory must be responsible for maintaining the fully developed turbulent state. Pursuing this line, it follows that generalized stability theory, valid for initial-value problems, can be extended by subjecting the flow to continuous perturbative forcing to provide a theory for the fully developed flow. The appropriate methodology is that of the stochastic dynamics of non-normal linear systems. The resulting theory for maintenance of variance in baroclinic jets is summarized below.

2. Generalized Stability Theory

Consider the general linear dynamical system that includes the governing equation for perturbation dynamics of the atmosphere: .

$$\frac{d}{dt}u = \mathbf{A}u, \tag{2}$$

in which $u(t)$ is the state vector of the system at time t, and $\mathbf{A}$ is the linear dynamical matrix. If the background state is steady, then $\mathbf{A}$ is not a function of time and the solution of Eq. (2) is explicit:

$$u(t) = e^{\mathbf{A}t}u(0). \tag{3}$$

In any case, there is a propagator matrix $\mathbf{R}_t$ that advances the system in time:

$$u(t) = \mathbf{R}_t u(0). \tag{4}$$

Only the existence of a propagator is required for the following development and stationarity is assumed for convenience only.

The central distinguishing attribute of $\mathbf{A}$ that determines its transient dynamics is its normality, that is, whether or not $\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+\mathbf{A}$. If $\mathbf{A}$ commutes with its Hermitian transpose, here indicated by the plus sign superscript, then $\mathbf{A}$ has a complete set of orthogonal eigenvectors and perturbation growth rate is bounded by the member of the eigenspectrum of $\mathbf{A}$ with maximum real part, indicated by $\lambda_{\max}(\mathbf{A})$, explicitly: $\| e^{\mathbf{A}t} \| = e^{\lambda_{\max}(\mathbf{A})t}$. The stability of such a normal dynamical system is determined for all time by its eigenspectrum. However, the dynamics of baroclinic and barotropic jet flows is governed by highly non-normal $\mathbf{A}$'s, so it is necessary to generalize ideas of perturbation growth by considering the growth, σ, of an arbitrary perturbation $u(0)$ over a time interval t as measured by the norm associated with the inner product represented by (u,v):

$$\begin{aligned}\sigma &= \frac{(u(t), u(t))}{(u(0), u(0))} \\ &= \frac{\left(e^{\mathbf{A}t}u(0), e^{\mathbf{A}t}u(0)\right)}{(u(0), u(0))} \\ &= \frac{\left(e^{\mathbf{A}^+t}e^{\mathbf{A}t}u(0), u(0)\right)}{(u(0), u(0))}.\end{aligned} \tag{5}$$

where in the last step the definition of the adjoint has been used (Courant and Hilbert, 1962). It follows that a complete set of orthogonal perturbations $u(0)$ can be ordered in growth over time t by eigenanalysis of the matrix: .

$$e^{\mathbf{A}^+t}e^{\mathbf{A}t}. \tag{6}$$

In particular, the greatest growth over time t is given by $\lambda_{\max}(e^{\mathbf{A}^+t}e^{\mathbf{A}t})$.

Alternatively, a singular value decomposition of the propagator reveals the unitary initial (columns of U) and final (columns of $\mathbf{V}$) states as well as the growth (along the diagonal of Λ) realized by the complete set of disturbances:

$$e^{\mathbf{A}t} = \mathbf{V}\Lambda\mathbf{U}^+. \tag{7}$$

Two limits of this decomposition of the propagator are worth noting. In the limit $t \rightarrow \infty$ maximum growth occurs for the eigenfunction with associated eigenvalue of maximum real part as normal mode theory would suggest. To see this, consider the similarity transformation by the matrix $\mathbf{E}$, which has the eigenvectors of $\mathbf{A}$ arranged as columns in order of growth rate, and the diagonal matrix, Σ, of their associated eigenvalues:

$$e^{\mathbf{A}t} = \mathbf{E}e^{\Sigma t}\mathbf{E}^{-1}. \tag{8}$$

In the limit $t \rightarrow \infty$, the first column of $\mathbf{E}$ and the first row of $\mathbf{E}^{-1}$ exponentially dominate with amplification factor $e^{\mathrm{Re}(\Sigma_{11}t)}$:

$$\lim_{t\rightarrow\infty} e^{\mathbf{A}t}_{\alpha\beta} = \mathbf{E}_{\alpha 1}e^{\Sigma_{11}t}\mathbf{E}^{-1}_{1\beta}. \tag{9}$$

The initial condition of unit norm producing maximum growth over time t is the complex conjugate of $\mathbf{E}^{-1}{}_{1\beta}$, which is the conjugate of the biorthogonal of the leading eigenvector rather than the leading eigenvector itself:

$$\lim_{t \to \infty} e^{\mathbf{A}t}_{\alpha\beta}\left(\mathbf{E}^{-1}_{1\beta}\right)^* = \mathbf{E}_{\alpha 1} e^{\Sigma_{11} t}. \qquad (10)$$

Although intuition is vindicated in that in the limit $t \to \infty$, that eigenfunction dominates that has an associated eigenvalue with maximum real part; not so obvious is the fact that the optimal initial condition with which to excite that mode should be the conjugate of the biorthogonal of the dominant mode rather than the mode itself. In highly non-normal systems such as the atmosphere, a mode and its biorthogonal differ greatly and the perturbation that optimally excites a mode bears little resemblance to the mode itself (Farrell 1989a).

Given the observed time scale for cyclogenesis of 24 hr (Roebber 1984), the $t \to \infty$ asymptotic is not likely to provide an accurate description of the cyclogenesis process; of greater utility is analysis of the less familiar $t \to 0$ limit of Eq. (6), which controls the initial growth of perturbations. It is this limit that determines the maximum possible instantaneous growth and the structure that produces this maximum growth, in addition to supplying other information such as the rate of expansion of the error ellipse in the short-time forecast and those structures that contribute most to the short-time error growth.

The limit of (6) as $t \to 0$ is easily obtained by Taylor expansion:

$$\begin{aligned} e^{\mathbf{A}^+ t} e^{\mathbf{A}t} &\approx (\mathbf{I} + \mathbf{A}^+ t)(\mathbf{I} + \mathbf{A}t) \\ &= \mathbf{I} + (\mathbf{A} + \mathbf{A}^+)t + O(t^2). \end{aligned} \qquad (11)$$

A tight upper bound on instantaneous growth rate and identification of the structure producing this limiting growth can be determined by eigenanalysis of the matrix $(\mathbf{A} + \mathbf{A}^+)$. Eigenanalysis of $(\mathbf{A} + \mathbf{A}^+)$ typically reveals that high growth rates over short times can be realized in baroclinic flows even though all normal modes of A are damped. Because these rapidly growing structures dominate development over the short time scales typical of cyclogenesis (12–48 hrs), the eigenfunctions of $(\mathbf{A} + \mathbf{A}^+)$ more accurately model the rapid development stage of cyclogenesis than do the eigenfunctions of $\mathbf{A}$.

The most relevant time scales for development in the atmosphere lie between the asymptotic limits $t \to 0$ and $t \to \infty$, and for these intermediate time scales the initial and final structures are found most easily from the SVD analysis of (6). Given that both asymptotic limits are subsumed, it is appropriate to refer to this analysis as the generalized stability analysis of the system (2).

3. Generalized Stability Theory Applied to a Baroclinic Jet

As an example of generalized stability analysis, consider stochastic excitation of perturbations on a zonal baroclinic flow. Harmonic scaled streamfunctions of the form $\phi(z,t) = \varepsilon^{-1/2} e^{z/2} \psi(z,t)$ with zonal wavenumber k and meridional wavenumber l in a β-plane channel obey a linearized quasigeostrophic potential vorticity equation with the form of Eq. (2):

$$\frac{d\phi}{dt} = \mathbf{B}\phi, \qquad (12)$$

where

$$\mathbf{B} = \frac{e^{z/2}}{\sqrt{\varepsilon}}\left(\Delta^{-1}(-ikU - R)\Delta - ik\Delta^{-1}Q_y\right)\frac{\sqrt{\varepsilon}}{e^{z/2}}, \qquad (13)$$

with

$$\Delta = \frac{d^2}{dz^2} - \left(\frac{\alpha^2}{\varepsilon} + S^2 - \frac{dS}{dz}\right). \qquad (14)$$

Here U is the mean zonal wind, ε is the square ratio of the Coriolis parameter to the Brunt–Väisälä frequency, α is the total wavenumber, and R is a linear potential vorticity damping coefficient. In the above equations, time has been nondimensionalized by the reciprocal of the Coriolis parameter (f_0), vertical distance by the scale height (H), and horizontal distance by $H(\varepsilon_0)^{-1/2}$. The mean potential vorticity gradient is given by

$$Q_y = \frac{\beta}{\varepsilon} + 2S\frac{dU}{dz} - \frac{d^2U}{dz^2}. \qquad (15)$$

and the stability parameter is defined as .

$$S = -\frac{1}{2}\left(\frac{1}{\varepsilon}\frac{d\varepsilon}{dz} - 1\right). \qquad (16)$$

The operator Δ^{-1} is rendered unique by incorporation of the usual boundary conditions at the ground resulting from the vertical velocity produced by Ekman pumping associated with a coefficient of vertical diffusion ν. Values of parameters are chosen to be appropriate for the midlatitude atmosphere: $f_0 = 10^{-4}$ s^{-1}, $N = 10^{-2}$ s^{-1}, $H = 10$ km, and $\beta = 1.6 \times 10^{-11}$ m^{-1} s^{-1}. These values result in horizontal wavenumber $k = 1$ corresponding to 6300 km. The boundary condition at the top of the atmosphere is taken to be a vanishing vertical velocity at 4 scale heights. A realistic stratification and zonal wind distribution are also included (Fig. 1; see Farrell and Ioannou (1993) for details).

The continuous dynamical system is reduced to a finite dynamical system using standard finite differencing of Eq. (12).

The energy density is chosen as a perturbation measure:

$$E^t = \phi^+ \mathbf{M} \phi \tag{17}$$

in which the energy metric is given for a grid of width δ by:

$$\mathbf{M} = -\frac{\delta}{8}(\mathbf{E}\mathbf{T}^2\mathbf{E} - \alpha^2\mathbf{P}), \tag{18}$$

where T is the discretized d/dz operator, $\mathbf{P}_{ij} = \rho_i\, \delta_{ij}$, $\mathbf{E}_{ij} = (\rho_i \varepsilon_i)^{1/2} \delta_{ij}$, and ρ_i is the mean density at the ith grid.

The generalized stability analysis of this system can be compactly displayed by plotting the log of the maximum energy increase obtained by any initial perturbation over a given time interval as a function of that time interval (Fig. 2). The slope of this curve near $t = 0$ gives the maximum instantaneous energy growth rate, and the asymptotic slope at large times gives the rate at which the energy of the least damped eigenmode decays. The large increase in energy at intermediate times illustrates the potential for transient amplification realized in observed cyclogenesis occurring on these time scales.

4. Extension of Generalized Stability Theory to Constitute a Theory for the Statistical Equilibrium

The very great increase in energy obtained by a substantial subspace of perturbations suggests a further generalization of stability theory. The conventional goal of stability theory has

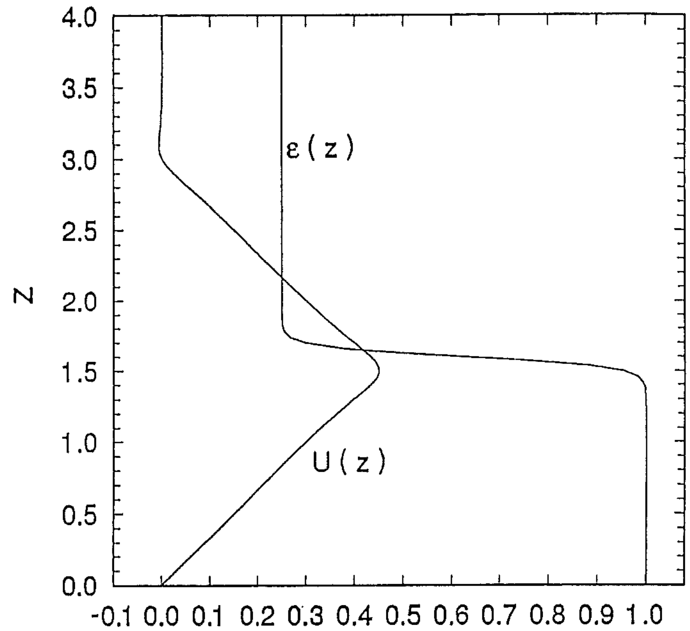

FIG. 1. The distribution with height of the zonal wind and static stability chosen for the examples. The nondimensional shear is $s = 0.3$, corresponding to a dimensional maximum zonal velocity of 45 m s^{-1}; the tropospheric value $\varepsilon = 1$ corresponds to a Brunt-Väisälä frequency $N = 10^{-2}$ s^{-1}.

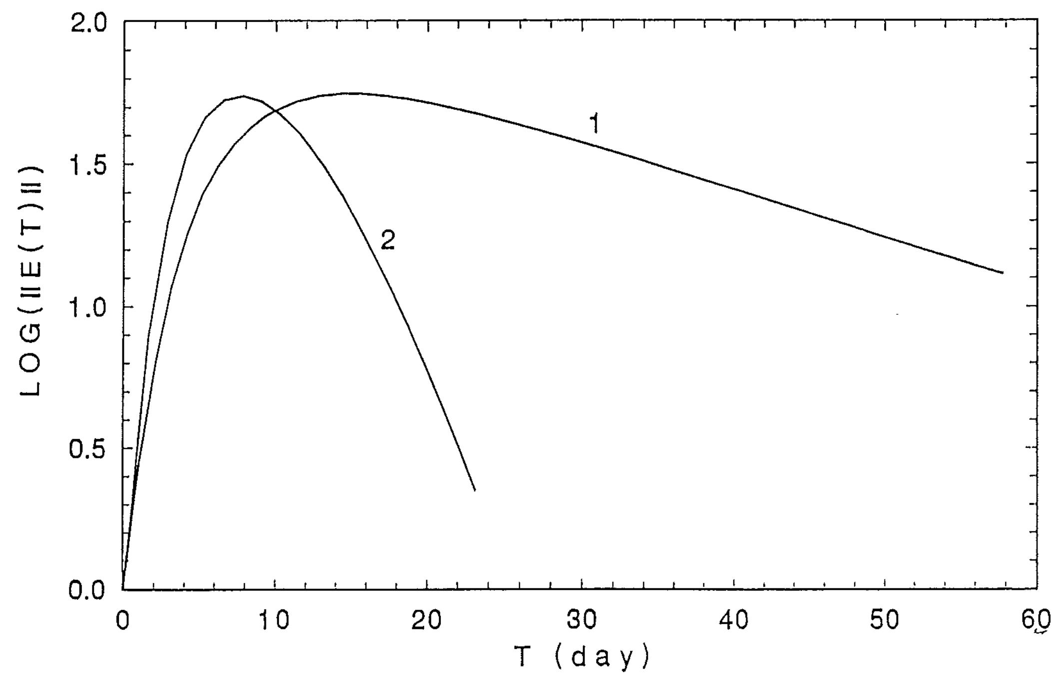

FIG. 2 Generalized stability analysis of the baroclinic jet shown in Fig. 1. The maximum energy obtained by any vertical structure is plotted as a function of the time, t, over which the growth is optimized. Asymptotic growth in the limits $t \to 0$ and $t \to \infty$ is clearly seen, as is the large growth at intermediate times. Ekman damping corresponding to $\nu = 20$ m^2 s^{-1} and Rayleigh damping with $R = 1/6$ day^{-1} has been used. (1) $k = 0.821$, $l = 1$; (2) $k = 4$, $l = 2$.

been to account for the existence of wavelike perturbations to the background flow field. However, in strongly sheared flows such as the midlatitude jet, the linearized equations embody the entire energetics by which the perturbation field is maintained, there being no net integrated contribution to perturbation energy from the nonlinear interaction of perturbations among themselves. Rather, the interaction among perturbations scatters energy, augments dissipation, and replenishes the subspace of growing perturbations as these are depleted over their transient growth cycles. Advantage may be taken of this pivotal role of the linear dynamical operator in maintaining the turbulent state to formulate a theory for the maintenance of turbulence together with its structures and transporting heat and momentum making use of the statistical dynamics of the non-normal dynamical operator governing perturbation dynamics. This is accomplished by stochastically forcing the dynamical system (2) (Farrell and Ioannou 1993, 1994a, 1994b, 1995; DelSole and Farrell 1995, 1996). This stochastic forcing is to be understood as a parameterization of the scattering of wave energy, which, together with an augmentation of dissipation to account for the disruption of waves by their mutual interaction, constitutes a parameterization of the effects of the nonlinear terms in the complete dynamics. To the extent that the statistically maintained variances and fluxes in the system are determined primarily by the transfer function of the strongly sheared jet, the exact distribution in space and time of the stochastic forcing is of secondary importance.

In order to solve for the stochastic dynamics of our system, it is convenient to transform (2) into generalized velocity variables $u = \mathbf{M}^{1/2}\phi$. In these generalized velocity variables, the stochastically forced perturbation potential vorticity equation takes the form:

$$\frac{du_i}{dt} = \mathbf{A}_{ij}u_j + \mathbf{F}_{ij}\xi_j , \tag{19}$$

where $\mathbf{A} = \mathbf{M}^{1/2}\mathbf{B}\mathbf{M}^{-1/2}$ and ξ is the random forcing assumed to be a δ-correlated Gaussian white-noise process with zero mean and with unit variance:

$$\left\langle \xi_i(t)\xi_j^*(t')\right\rangle = \delta_{ij}\delta(t-t'), \tag{20}$$

where $< \bullet >$ denotes the ensemble average. Note that the stochastic forcing excites independently and with equal unit magnitude each spatial forcing distribution as specified by the columns $f^{(j)}$ of the matrix F_{ij}. We want to determine the evolution of the variance sustained by (19), which in physical variables is the ensemble averaged energy $<E^t> = <u_i^*(t)u_i(t)>$.

The random response, u, is linearly dependent on ξ and consequently is also Gaussian distributed. Therefore, the statististics of the response of the dynamical system are fully characterized by the first two moments. The first moment vanishes for large times if $\mathbf{A}$ is asymptotically stable. The expression for the second moment, the ensemble average energy density, can be reduced to:

$$\left\langle E^t\right\rangle = \left\langle u_i^*(t)u_i(t)\right\rangle = \mathrm{trace}(\mathbf{F}^{+}\mathbf{K}^{t}\mathbf{F}), \tag{21}$$

where

$$\mathbf{K}^t = \int_0^t e^{\mathbf{A}^{+}(t-s)}e^{\mathbf{A}(t-s)}ds. \tag{22}$$

The evolution equation for $\mathbf{K}^t$ with initial condition $\mathbf{K}^0 = 0$ can be derived by direct differentiation of $\mathbf{K}^t$. It is

$$\frac{d\mathbf{K}^t}{dt} = \mathbf{I} + \mathbf{A}^{+}\mathbf{K}^t + \mathbf{K}^t\mathbf{A}, \tag{23}$$

in which $\mathbf{I}$ is the identity matrix. When the potential vorticity damping, R, is chosen so that $\mathbf{A}$ is asymptotically stable, the asymptotic value of $\mathbf{K}^\infty$ can be determined by solving the asymptotic form of Eq. (23), which is the Liapunov equation:

$$\mathbf{A}^{+}\mathbf{K}^\infty + \mathbf{K}^\infty\mathbf{A} = -\mathbf{I}\,. \tag{24}$$

Note that with an orthonormal set of forcing functions such that $\mathbf{F}\mathbf{F}^{+} = \mathbf{I}$, the expression for the energy density simplifies to $<E> = \mathrm{trace}(\mathbf{K}^\infty)$ and the variance is independent of the specific forcing distribution.

It is also useful to determine the ensemble average correlation matrix of the response $C_{ij}^t = \left\langle u_i(t)u_j^*(t)\right\rangle$. The correlation matrix can be equivalently written as:

$$\mathbf{C}^t = \int_0^t e^{\mathbf{A}(t-s)}\mathbf{F}\mathbf{F}^{+}e^{\mathbf{A}^{+}(t-s)}ds, \tag{25}$$

which satisfies in steady state the Liapunov equation:

$$\mathbf{A}\mathbf{C}^\infty + \mathbf{C}^\infty\mathbf{A}^{+} = -\,\mathbf{F}\mathbf{F}^{+}. \tag{26}$$

Note that because the operator is non-normal (i.e., $\mathbf{A}^{+}\mathbf{A} \neq \mathbf{A}\mathbf{A}^{+}$), Eq. (26) differs from Eq. (24) even for the case of unitary forcing. In any case, the ensemble average energy for full rank unitary forcing is:

$$<E> = \mathrm{trace}(\mathbf{C}^\infty) = \mathrm{trace}(\mathbf{K}^\infty)\,. \tag{27}$$

If $\mathbf{A}$ were normal, as is the case in the absence of basic state shear, then all of the matrices, $\mathbf{A}$, $\mathbf{C}^\infty$, and $\mathbf{K}^\infty$, commute, and Eq. (26) can be immediately solved to yield, for unitary forcing, the steady-state ensemble average energy:

$$\begin{aligned}\langle E\rangle &= \mathrm{trace}(\mathbf{C}^\infty)\\ &= \mathrm{trace}\left(-(\mathbf{A}+\mathbf{A}^{+})^{-1}\right)\\ &= \sum_i \frac{1}{2\,\mathrm{Re}\left(-\lambda_i(\mathbf{A})\right)}\end{aligned}, \tag{28}$$

where $\lambda_i(A)$ are the eigenvalues of **A**. This expression for the variance maintained by normal dynamical operators has been extensively investigated in the past (cf. Wang and Uhlenbeck 1945). In the case of a normal dynamical operator, the motion can be resolved into orthogonal normal modes with the total variance found as the sum of contributions from each individual mode. Moreover, the variance contributed by each mode is inversely proportional to its damping rate. In such systems, the forcing is the only energy source, and the maintained variance is accumulated from the forcing, resulting in high variance if damping is small.

For non-normal dynamical operators, the non-orthogonality of the modes is indicative of the possibility for the perturbations to extract energy from the background flow field despite the absence of exponential instability. The energy balance in such a system is between the stochastic driving together with the induced extraction of energy from the background flow, on the one hand, and the dissipation on the other. Tapping the energy of the mean flow can consequently lead to levels of variance orders of magnitude larger than would have been expected to result in a normal system if the estimation of variance were made from the rate of dissipation of each mode

5. Determination of the Statistical Equilibrium Forcing and Response Functions

We showed that the ensemble average energy for a full-rank unitary forcing distribution can be derived either from the trace of $\mathbf{C}^\infty$ or, equivalently, from the trace of $\mathbf{K}^\infty$. Both $\mathbf{K}^\infty$ and $\mathbf{C}^\infty$ are by construction positive definite Hermitian forms with positive real eigenvalues associated with mutually orthogonal eigenvectors. Each eigenvalue equals the variance accounted for by the pattern of its corresponding eigenvector, and the pattern that corresponds to the largest eigenvalue contributes most to the variance. The decomposition of the correlation matrix into its orthogonal components is often referred to as the EOF decomposition.

The EOF decomposition of $\mathbf{C}^\infty$ determines the structures that contribute most to the ensemble average variance of the statistically steady state. These are the primary response structures of the dynamical system. They are determined by solving the eigenvalue problem:

$$\mathbf{C}^\infty u^{(i)} = \lambda^{(i)} u^{(i)}, \qquad (29)$$

in which the variance accounted by the structure $u^{(i)}$ is given by $\lambda^{(i)}$.

The EOF's can be interpreted dynamically using the stochastic theory developed here, thus providing a link between observed atmospheric statistics and dynamical theory. The leading EOF can be determined as the eigenfunction corresponding to the maximum eigenvalue of the operator:

$$\mathbf{C}^\infty = \lim_{t\to\infty} \int_o^t e^{\mathbf{A}(t-s)} e^{\mathbf{A}^+(t-s)} ds\,. \qquad (30)$$

Note that due to the non-normality of **A** the eigenfunctions of Eq. (30) do not coincide with the eigenfunctions of the dynamical operator **A**. Only for normal dynamical operators, for which **A** and $\mathbf{A}^+$ commute, do the eigenfunctions of Eq. (30) coincide with the modes of the system.

Eigenanalysis of $\mathbf{K}^\infty$, on the other hand, allows ordering of the forcing distributions according to their contribution to the variance of the statistically steady state. This follows from (21) and the observation that the eigenvalues of $\mathbf{K}^\infty$ are stationary values of the Rayleigh-Ritz quotient:

$$I[f] = \frac{f^+\mathbf{K}^\infty f}{f^+ f}\,. \qquad (31)$$

Consequently, the forcings, $f^{(i)}$, obtained from eigenanalysis of

$$\mathbf{K}^\infty = \lim_{t\to\infty} \int_o^t e^{\mathbf{A}^+(t-s)} e^{\mathbf{A}(t-s)} ds \qquad (32)$$

can be ordered according to their relative contribution to the stochastically maintained variance. The optimal stochastic excitation consists of the eigenfunction corresponding to the largest eigenvalue of Eq. (32). Note once again that because of the non-normality of **A** the eigenfunctions of **A**, $\mathbf{K}^\infty$, and $\mathbf{C}^\infty$ do not coincide. Examples of the differences in these structures are given in Fig. 3.

According to the theory presented here, eigenanalysis of (29) is expected to yield the observed EOF's if **A** is taken to correspond to the linearized operator of an observed atmospheric state. The set of primary forcing functions derived from eigenanalysis of $\mathbf{K}^\infty$ can in turn be found by solving Liapunov equation (24). It is clear from this analysis that the attempt to associate the EOF's with the normal modes of the system can succeed only when the dynamical system is governed by a normal dynamical operator.

6. Determination of the Statistical Equilibrium Spectrum

The Fourier transform of Eq. (19) readily yields that

$$\left\langle E^\infty \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) d\omega\,, \qquad (33)$$

where the frequency response

$$F(\omega) = \mathrm{trace}(\mathbf{R}^{+}(\omega)\mathbf{R}(\omega))\,, \tag{34}$$

follows from the resolvent

$$\mathbf{R}(\omega) = (i\omega\mathbf{I} - \mathbf{A})^{-1}\,. \tag{35}$$

We may say that the power spectrum of the response to white noise forcing is given by the Frobenius norm of the resolvent $\mathbf{R}(\omega)$. On the other hand, $\|R(\omega)\|_2$ gives the maximum possible response at this frequency, and the optimal forcing structure to produce such a response can be obtained by singular value decomposition of the resolvent.

When $\mathbf{A}$ is normal, the magnitude of the response can be calculated by the distance of the eigenvalues from the forcing frequency ω. In fact, in that case we have exactly (Kato 1976):

$$\|\mathbf{R}(\omega)\| = \frac{1}{\mathrm{dist}(i\omega, \Lambda(A))}\,, \tag{36}$$

where $\Lambda(\mathbf{A})$ is the set of eigenvalues of $\mathbf{A}$. When $\mathbf{A}$ is non-normal, the maximum response may be seriously underestimated by the value predicted by the proximity to the eigenvalues of $\mathbf{A}$. Examples showing the difference between the response of normal and non-normal operators are discussed in Farrell and Ioannou (1993, 1994a, 1995, 1996).

7. Determination of the Ensemble Average Heat Flux

The ensemble average heat flux is defined as

$$\begin{aligned} H &= c_p \int_0^{\infty} \rho \overline{v\theta} dz \\ &= \frac{c_p T_g L_d}{gH} E_{in} \frac{h\,\mathrm{trace}(\mathbf{H}^{\infty})}{N_f} \end{aligned} \tag{37}$$

$$\mathbf{H}^{\infty} = \frac{k}{2} \mathrm{Im}\left(\mathbf{ETM}^{-1/2}\mathbf{C}^{\infty}\mathbf{M}^{-1/2}\mathbf{E}^{+}\right), \tag{38}$$

where h is the grid interval, N_f is trace($\mathbf{FF}^{+}$), C_p is the heat capacity, T_g is a representative temperature, H is the scale height, g is gravitational acceleration, $L_d = NH/f_0$ is the

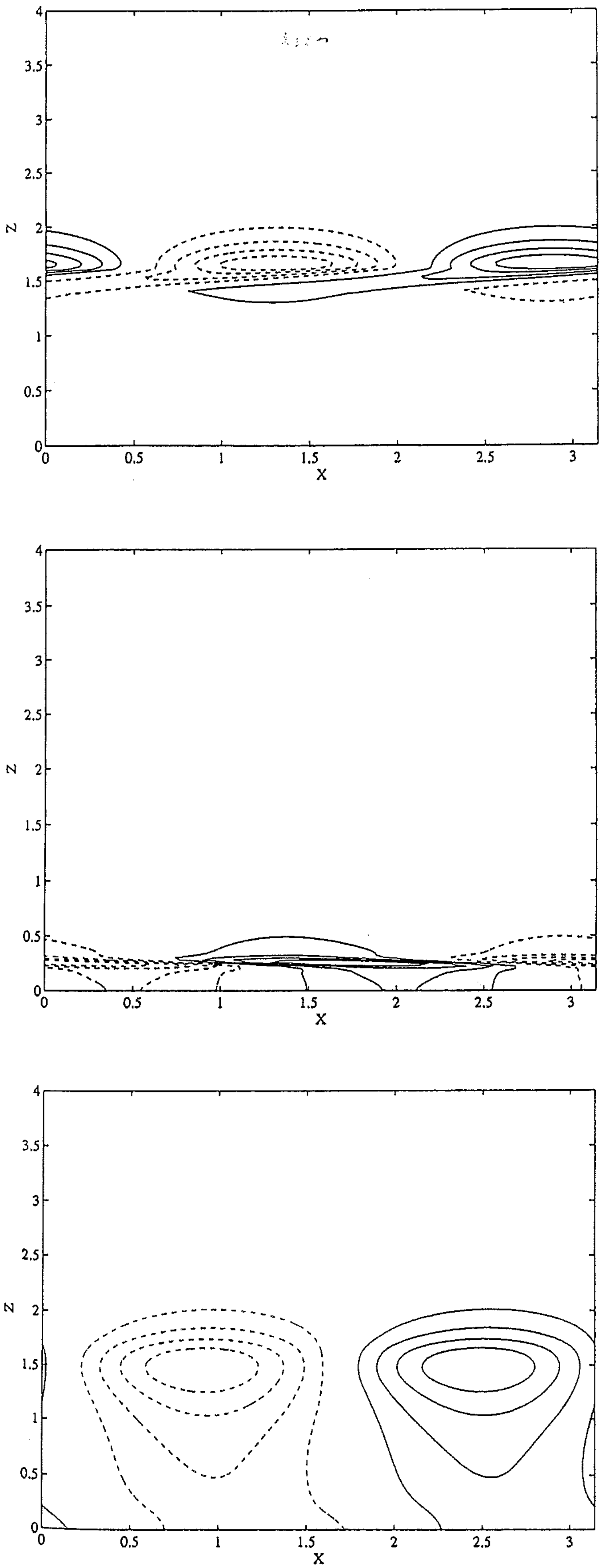

FIG. 3 (left): (a) The first eigenfunction of A; (b) the first eigenfunction of $\mathbf{K}^{\infty}$, corresponding to the first forcing function; (c) the first eigenfunction of $\mathbf{C}^{\infty}$, corresponding to the first response function. The wavenumbers are $k = 1$, $l = 1$. Ekman damping corresponding to $\nu = 10$ m^2 s^{-1} and Rayleigh damping with $R = 1/12$ day^{-1} have been used. The variable shown is the streamfunction f associated with the respective eigenfunctions.

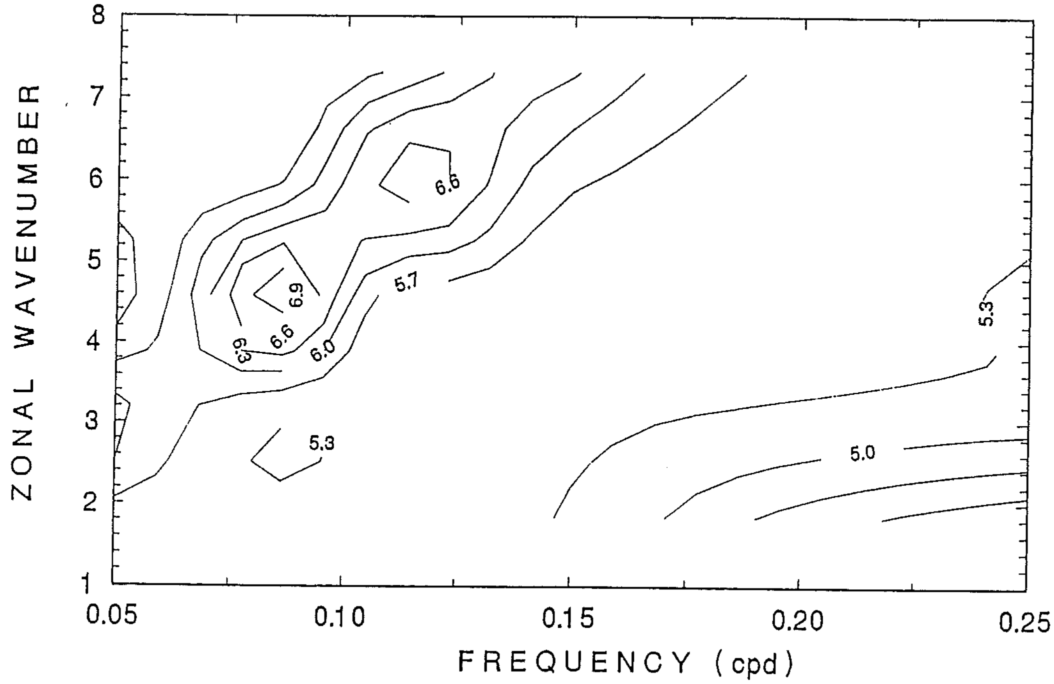

FIG. 4 The variance as a function of period and zonal wavenumber for the jet shown in Fig. 1. Ekman damping corresponding to $\nu = 20\,\mathrm{m^2\,s^{-1}}$ and Rayleigh damping with $R = 1/8\,\mathrm{day^{-1}}$ have been used.

Rossby radius of deformation, and $\mathbf{C}^\infty$ is the correlation matrix found by solving Eq. (26).

The correlation matrix of the generalized velocities determines the heat flux and, in particular, the diagonal elements of $\mathbf{H}^\infty$ give the height distribution of the heat flux.

8. The Statistical Equilibrium Energy Spectrum and Heat Flux in a Baroclinic Jet

Consider a baroclinic jet with stochastically forced waves of meridional wavenumber $l = 1$ (the meridional confinement implied by $l = 1$ is chosen to be consistent with the atmospheric state given by Salby (1982) (cf. Farrell and Ioannou 1993; Ioannou and Lindzen 1986)).

The full spectrum of the response as a function of period and zonal wavenumber is shown in Fig. 4. Comparison with observations (i.e., Hansen et al. 1989) reveals that the derived spectra are in remarkable agreement with the observed spectra.

Dimensional values of ensemble average energy and associated heat flux as a function of zonal wavenumber are shown in Fig. 5 for stochastic forcing of 1 W m^{-2} (for the dimensional forms of these quantities, refer to Farrell and Ioannou (1994a)). The maximum response is concentrated between zonal wavenumbers 5 and 6 (note also the reduction of the variance and the fluxes as dissipation increases). The fact that the atmosphere often exhibits enhanced variance at these scales suggests further that the atmosphere is stochastically maintained in a state of near neutrality. An example of the distribution with height of the ensemble average energy and heat flux is shown in Fig. 6.

9. Conclusion

We have described a generalization of stability theory that includes the transient development process recently studied using initial value problem approaches. This generalized stability theory emphasizes the central role of the non-normality of the linear dynamical operator in determining stability. Generalized stability theory unites the $t \to 0$ asymptotic of the propagator that controls initial error growth and bounds the instantaneous growth rate of disturbances with the $t \to \infty$ asymptotic that is addressed by modal theory. These limits are connected by the singular value decomposition of the propagator, and it is the transient development at intermediate time scales of a day or two that is most relevant to cyclogenesis.

In addition, we have advanced a further generalization of stability theory building on the great potential for transient growth of perturbations in the highly non-normal system underlying dynamics of the midlatitude jet. In this theory, ensemble equilibrium statistics of the turbulent state are found with nonlinear terms in the dynamical equation parameterized as a combination of stochastic forcing to account for scattering of energy among waves and an augmentation of dissipation to account for disruption of wave coherence by wave/wave interaction. The spectral distribution of transient wave energy and the spatial distribution of variance and sensible heat flux obtained using this theory are found to be in remarkable agreement with observations.

The brief summary of generalized stability theory contained in this chapter necessarily leaves many issues open; some of these are addressed in a recent review (Farrell and Ioannou 1996). In addition, efforts to apply the stochastic

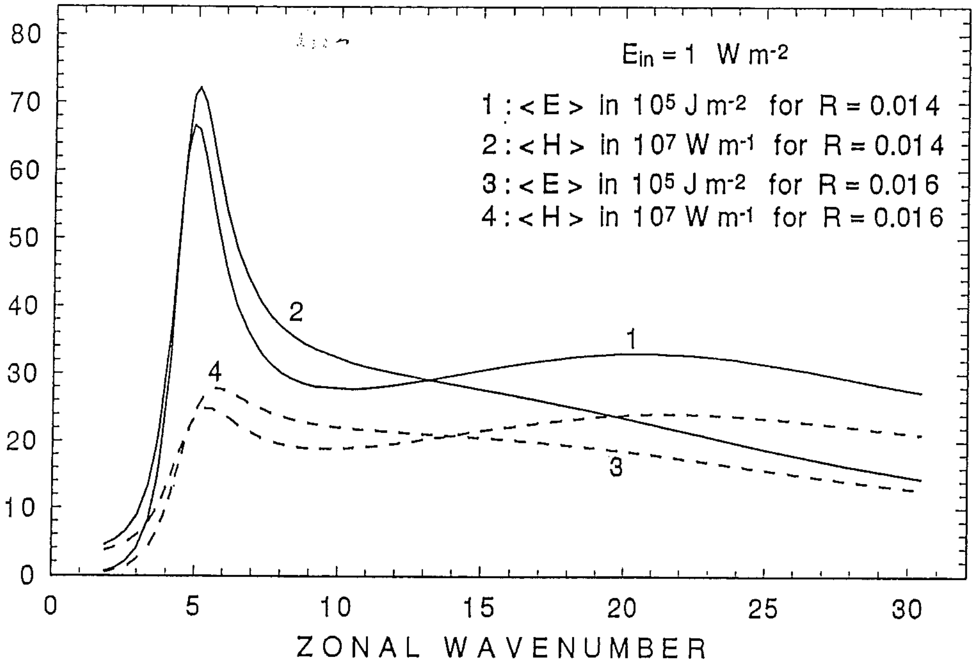

FIG. 5 The ensemble average energy and heat flux as a function of zonal wavenumber for $l = 1$. Ekman damping corresponding to n = 20 m^2 s^{-1} and Rayleigh damping with R = 1/8 day^{-1} (solid line) and R = 1/7 day^{-1} (dashed line) have been used.

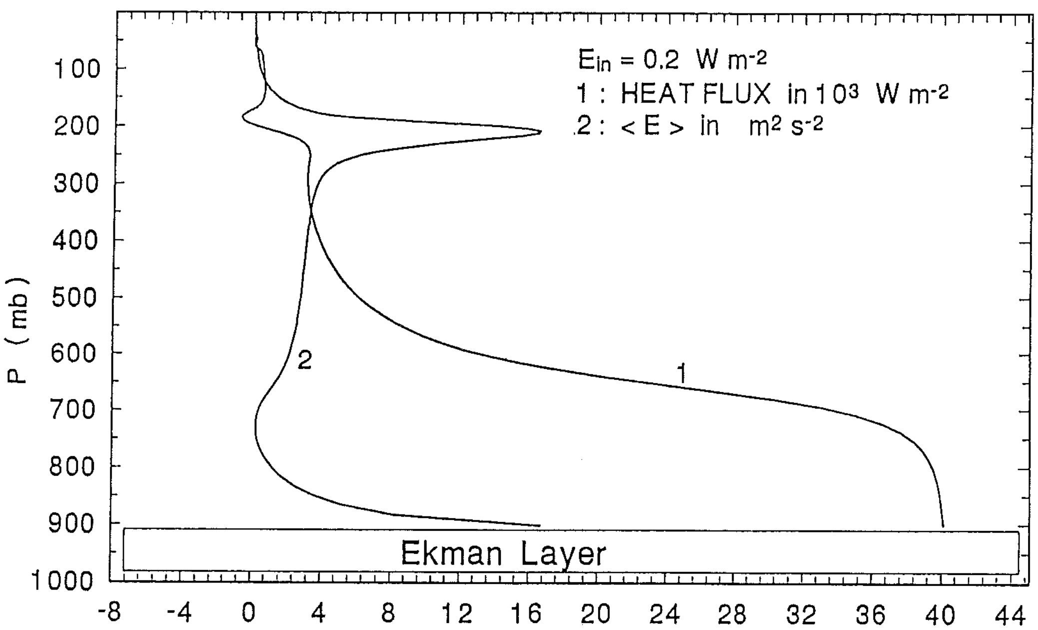

Fig. 6 The distribution with height of the ensemble average energy and heat flux for global zonal wavenumber 5 waves (k= 0.821 and l= 1). Ekman damping corresponding to n = 20 m^2 s^{-1} and Rayleigh damping with R = 1/8 day^{-1} have been used.

theory for equilibrium statistics to more realistic systems are ongoing (DelSole 1996).

Acknowledgment

The author wishes to acknowledge the esential contribution of Petros Ioannou to the theoretical results recounted in this paper. This work was supported in part by NSF ATM-92-16813 and NSF ATM-96-23539, and in part by the U.S. Department of Energy's (DOE) National Institute for Global Environmental Change (NIGEC) through the NIGEC Northeast Regional Center at Harvard University. Financial support does not constitute an endorsement by DOE of the views expressed in this chapter.

References

Bjerknes, J., 1919: On the structure of moving cyclones. *Geofys. Publ.*, **1**, 8.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135–163.

Courant, R., and D. Hilbert, 1962: *Methods of Mathematical Physics V. II.* John Wiley & Sons, New York.

DelSole, T. M., 1996: Can quasigeostrophic turbulence be modelled stochastically? J. *Atmos. Sci.*, **53**, 1617–1633.

———, and B. F. Farrell, 1995: A stochastically excited linear system as a model for quasigeostrophic turbulence: Analytic results for one- and two-layer fluids. *J. Atmos. Sci.*, **52**, 2531–2547.

———, and ———, 1996: The quasilinear equilibration of a thermally maintained, stochastically excited jet in a quasigeostrophic model. *J. Atmos. Sci.*, **53**, 1781–1797.

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

Eliassen, A., 1956: Instability theories of cyclone formation. *Weather Analysis and Forecasting Vol. 1*, McGraw-Hill.

Espy, J. P., 1841: *The Philosophy of Storms*. Little and Brown.

Farrell, B. F., 1989a: Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, **46**, 1193–1206.

———, 1989b: Transient development in confluent and diffluent flow. *J. Atmos. Sci.*, **46**, 3279–3288.

———, 1990: Small error dynamics and the predictability of atmospheric flows. *J. Atmos. Sci.*, **47**, 2409–2416.

———, and P. J. Ioannou, 1993: Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.*, **50**, 4044–4057.

———, and ———, 1994a: A theory for the statistical equilibrium energy and heat flux produced by transient baroclinic waves. *J. Atmos. Sci.*, **51**, 2685–2698.

———, and ———, 1994b: Variance maintained by stochastic forcing of a non-normal dynamical system arising in fluid shear flow. *Phys. Rev. Lett.*, **72**, 1188–1191.

———, and ———, 1995: Stochastic dynamics of the midlatitude atmospheric jet. *J. Atmos. Sci.*, **52**, 1642–1656.

———, and ———, 1996: Generalized stability theor. Part I: autonomous operators. *J. Atmos. Sci.*, **53**, 2025–2040.

Fraedrich, K., and H. Bottger, 1978: A wavenumber-frequency analysis of the 500 mb geopotential at 50°N. *J. Atmos. Sci.*, **35**, 745–750.

Hansen, A. R., A. Sutera, and D. E. Venne, 1989: An examination of midlatitude power spectra: Evidence for standing variance and the signature of El Niño. *Tellus*, **41A**, 371–284.

Helmholtz, H., 1886: *Uber discontinuirliche Flussigkeitsbewegungen*. Berlin Akad. Monatsver.

———, 1888: *Ueber atmosphärische Bewegungen*. Sitzber. Akad. Berlin.

Ioannou, P. J., and R. S. Lindzen, 1986: Baroclinic instability in the presence of barotropic jets. *J. Atmos. Sci.*, **43**, 2999–3014.

Kato, S., 1976: *Perturbation Theory for Linear Operators*. Springer-Verlag.

Kutzbach, G., 1979: *The Thermal Theory of Cyclones*. Amer. Meteor. Soc., 254 pp.

Loomis, E., 1841: On the storm which was experienced throughout the United States about the 20th December, 1836. *Trans. Amer. Philos. Soc.*, **7**, 125.

Margules, M., 1903: *Uber die Energie der Sturme*. Jahrb. Zentralanst. Meteorol. Wien., 1–26.

———, 1906a: Ueber Temperaturschichtung in stationär bewegter und ruhender Luft. *Meteor. Z.*, **23**, 243–254.

———, 1906b: Zur Stormtheorie. *Meteor. Z.*, **23**, 482–483.

Palmén, E., 1951: The aerology of extratropical disturbances. *Compendium of Meteorology*, T. F. Malone, Ed., Amer. Meteor. Soc., 599–620.

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

Rayleigh, Lord, 1880: *Theory of Sound*. 2d ed. Macmillan.

Redfield, W. C., 1831: Remarks on the prevailing storms of the Atlantic coast of the North American States. *Amer. J. Sci.*, **20**, 29.

Roebber, P. J., 1984: Statistical analysis and updated climatology of explosive cyclones. *Mon. Wea. Rev.*, **112**, 1577–1589.

Sanders, F., 1971: Analytic solutions of the nonlinear omega and vorticity equations for a structurally simple model of disturbances in the baroclinic westerlies. *Mon. Wea. Rev.*, **99**, 393–407.

Salby, M. L., 1982: A ubiquitous wavenumber-5 anomaly in the Southern Hemisphere during FGGE. *Mon. Wea. Rev.*, **110**, 1712–1720.

Schafer, J., 1979: A space-time analysis of tropospheric planetary waves in the Northern Hemisphere geopotential heights. *J. Atmos. Sci.*, **36**, 1117–1123.

Sommerfeld, A., 1946: *Lectures in Theoretical Physics Vol. II*. Academic Press.

Wang, M. C., and G. E. Uhlenbeck, 1945: On the theory of the Brownian motion II. *Rev. Modern Phys.*, **17**, 323–342.