Towards a unified instability theory of large-scale atmospheric disturbances

Jorgen S. Frederiksen
CSIRO Division of Atmospheric Research, Private Bag No. 1, Mordialloc, Victoria 3195, Australia.

ABSTRACT

This article presents a review of developments, during the last decade, in the application of three-dimensional normal mode instability theory for understanding the dynamical origins of large-scale atmospheric disturbances. The essential properties of a wide variety of observed fluctuations can be captured within this theoretical framework. We review progress in regional and local cyclogenesis in both hemispheres, including the deflection and splitting of storm tracks by large scale anomalies such as blocks and the effects of horizontally variable static stability and nongeostrophic effects. The roles of baroclinic, barotropic and topographic instability in the formation of blocks and other low-frequency anomalies and quasi-stationary teleconnection patterns are examined. The relative importance of wave-CISK cumulus heating and three-dimensional flow instability for generating monsoon disturbances and intraseasonal oscillations is discussed. We also review the role of wave instability in the formation of stratospheric disturbances and the relevance of non-normal modes, adjoint modes and optimal perturbations to developing atmospheric disturbances.

1. INTRODUCTION

Normal mode—temporal growth instability theory has a long history in fluid dynamics with the foundations laid last century. In dynamic meteorology, Charney (1947), Eady (1949) and Phillips (1951) pioneered the now generally accepted theory of cyclogenesis, in which synoptic storms result from the baroclinic instability of the large scale atmospheric flow field, which they approximated by simple zonal mean flows on beta-planes or f-planes. The baroclinic instability of zonally averaged basic states has subsequently been thoroughly analysed on beta- and f-planes (e.g., Green, 1960; Brown, 1969; Stone, 1969; Simmons, 1974; Staley and Gall, 1977) and on the sphere (e.g., Hollingsworth, 1975; Moura and Stone, 1976; Wam, 1976; Gall, 1976a; Simmons and Hoskins, 1976, 1977; Baines and Frederiksen, 1978; Frederiksen 1978b, 1981a).

The first analytical study of the instability of idealised atmospheric flows with longitudinal variations was made by Lorenz (1972). He examined a basic state consisting of a zonal mean flow and a single planetary scale Rossby wave using a beta-plane barotropic model. He suggested on the basis of this study that barotropic instability of Rossby waves is largely responsible for the unpredictability of atmospheric flows. Further studies of the instability of Rossby waves were carried out by Hoskins and Hollingsworth (1973), Gill (1974) and Duffy (1975) using beta-plane barotropic models and by Hoskins (1973) and Baines (1976) using spherical barotropic models. Additional references to the literature are given in the review by Grothahn (1984a).

In all the above studies, the basic states were taken as being exact steady state solutions (possibly in a rotating reference frame) of the free (unforced) equations of motion. This is very restrictive in the type of basic state flows that can be examined, often allowing only very simple wave structures. Real atmospheric flows are of course forced and both instantaneous and climatologically averaged flows have complex wave structures. Climatological flows are not steady state solutions but statistical steady state solutions whose stationarity is due to the statistical effects of transients (see, for example, appendix A of Frederiksen, 1989).

Frederiksen (1978a) examined the instability properties of three-dimensional forced basic state flows consisting of baroclinic zonal mean flows and planetary waves in two-level models on a sphere. Free steady state solutions and planetary waves by themselves were also examined. In this study and in that of Frederiksen (1979b), it was found that the theoretical model could capture the locations of regional cyclogenesis poleward and slightly downstream of the jetstream maxima.
The instability characteristics of three-dimensionally varying observed Northern Hemisphere winter climatological basic states were analysed by Frederiksen (1982a) using a two-level quasigeostrophic eigenvalue model. It was found that the regions of preferential development of the faster growing monopole cyclogenesis modes were in quite good agreement with the locations of the observed Atlantic and Pacific storm track maxima. It was also argued that some of the slower growing high-low dipole instability modes that were found were in fact onset-of-blocking modes, precursors to the formation of mature blocks over the Pacific or Atlantic oceans. Frederiksen (1982a, b) suggested that three-dimensional instability theory should be able to provide a unified theoretical framework for understanding the dynamical origins of a wide range of atmospheric disturbances including localised cyclogenesis, blocking and aspects of stratospheric sudden warmings.

In this article, we review the subsequent developments in three-dimensional normal mode instability theory made over the last decade by my collaborators and myself and how these studies relate to the works of other researchers. Section 2 gives a brief formulation of normal mode instability theory while in Section 3 applications to regional and local cyclogenesis are discussed including the climatological storm tracks in both hemispheres, the deflection and splitting of storm tracks during blocking and effects of variable static stability and non-geostrophic terms. Blocking in both hemispheres is considered in Section 4 while the generation of low-frequency anomalies and quasi-stationary teleconnection patterns is reviewed in Section 5. The generation of quasi-stationary monsoon disturbances and intraseasonal oscillation modes in a two-level primitive equation model with a wave-CISK cumulus parameterization is considered in Section 6. The role of wave instability in the formation of stratospheric disturbances is discussed in Section 7 while in Section 8 the relevance of non-normal modes, adjoint modes and optimal perturbations to developing atmospheric disturbances is examined. The conclusions are summarised in Section 9.

2. NORMAL MODE INSTABILITY THEORY

The nonlinear equations describing atmospheric flows do not lend themselves to semianalytical treatments or easy understanding. However, a very useful semianalytical approach for studying the dynamics of atmospheric disturbances is the method of linearization of the flow about a basic state and expansion of the perturbation in terms of normal modes.

The method has been applied to multi-level models with three-dimensional basic states. Here we shall illustrate the essential theoretical concepts for the barotropic vorticity equation describing forced dissipative flow over topography on a rotating sphere.

(a) Linearization

The non-dimensional form of the vorticity equation is:

\[ \frac{\partial \nabla^2 \Psi}{\partial t} + J(\Psi, \nabla^2 \Psi + 2 \mu + h) + \kappa \nabla^2 \Psi = F_o(\lambda, \mu) \]  

Here the notation follows that of Frederiksen (1982c); we have taken \( a \) (earth's radius) and \( \Omega \) (earth's angular velocity) \( \times \) as length and time scales, respectively. In Eq. (2.1), \( \Psi \) is the streamfunction, \( \lambda \) longitude, \( \mu = \sin(\text{latitude}) \), \( \tau \) is time, \( J \) the Jacobian operator, \( \kappa \) a drag coefficient, and we could, of course, also have included various diffusion parameterizations. Also, \( h \) represents the topography multiplied by various factors and for simplicity we take the forcing \( F_o(\lambda, \mu) \) to be inhomogeneous (also called "constant"); viz., it does not depend on \( \Psi \).

Now suppose Eq. (2.1) has an exact steady state solution \( \Psi(\lambda, \mu) \). For example, in the absence of dissipation and forcing \( F_o \), Eq. (2.1) has exact steady state solutions \( \Psi(\lambda, \mu) \) which are solutions of

\[ \nabla^2 \Psi(\lambda, \mu) + 2 \mu + h(\lambda, \mu) = \Lambda \Psi(\lambda, \mu) \quad (2.2) \]

for arbitrary \( \Lambda \) (Frederiksen and Carnevale, 1986 and references therein). Suppose now that we perturb \( \Psi(\lambda, \mu) \) by a small perturbation \( \epsilon \Psi(\lambda, \mu, \tau) \) at \( \tau = 0 \), where \( \epsilon \) is very much smaller than 1. Then neglecting terms of order \( \epsilon^2 \), we find that \( \Psi' \) satisfies the equation

\[ \frac{\partial \nabla^2 \Psi'}{\partial \tau} + J(\Psi', \nabla^2 \Psi' + 2 \mu + h) + \kappa \nabla^2 \Psi' = 0, \quad (2.3) \]

where we have restored the dissipation term in the linearized equation. Further, the total flow field at a given time \( \tau \) resulting from the initial perturbation is given by

\[ \Psi(\lambda, \mu, \tau) = \Psi' + \epsilon \Psi'(\lambda, \mu, \tau) + O(\epsilon^2) \quad (2.4) \]

where the expansion in \( \epsilon \) is an asymptotic one which is valid as long as \( \epsilon \) is not too large.

(b) Expansion in terms of normal modes

The perturbation field \( \Psi'(\lambda, \mu, \tau) \) is then expanded in orthogonal functions (or discretised in some other way) which for spherical geometry are spherical harmonics

\[ \Psi'(\lambda, \mu, \tau) \sim \sum_{n=-R}^{R} \sum_{m=-n}^{n} \Psi_{jm}(\lambda) P_n^m(\mu) e^{im\lambda} \quad (2.5) \]
Here we have dropped the prime on the perturbation field and added the subscript $j$ which allows us to generalise the theory to multi-level fields and to allow for other fields such as temperature, divergence etc. Here $m$ is the zonal wavenumber, $n$ the total wavenumber, $\Psi_{mn}$ are spectral coefficients, $P_n^m(\mu)$ are Legendre functions and $R$ is a truncation wavenumber. With the upper limit for $n$ in Eq. (2.5) being $|m|+R$, the truncation scheme is referred to as rhomboidal. Here we shall restrict our attention to systems truncated to a finite number of spectral components ($R < \infty$) and with discrete modes.

When the orthogonality properties of the spherical harmonics are employed then it may be shown (Frederiksen, 1982b, appendix; Frederiksen and Frederiksen, 1992) that for the quasigeostrophic and primitive equation multi-level models as well as for the barotropic model, the equations for the spectral coefficients have the form:

$$\begin{align*}
[A] \frac{\partial \Psi}{\partial \tau} &= i[B] \Psi + f
\end{align*}$$

with initial condition $\Psi(t = 0) = \Psi_0$. Here $\Psi(t)$ is a column vector of spectral coefficients of length $N$ and $[A]$ and $[B]$ are $N \times N$ complex matrices obtainable from the appendix of Frederiksen (1982b) and from Frederiksen and Frederiksen (1992) for quasigeostrophic and primitive equation models respectively. The matrices $[A]$ and $[B]$ contain information about the basic state and dissipation parameterizations. We have also generalised the perturbation equation to allow for the perturbation being generated by changes in the forcing. The vector $f$ represents an anomalous streamfunction or thermal forcing or in the case of anomalous topographic forcing contains products of the basic-state and anomalous topographic spectral coefficients. Here $N$ is the total number of spectral coefficients obtained from the number of fields, levels and zonal and total wavenumbers used in the truncation scheme. We also note from Eq. (2.5) that since $\Psi(0, \mu, \tau)$ is a real function we must have

$$\Psi_{j,m}(t) = \Psi_{m,j}(t)$$

(2.7)

with our convention $P_n^m(\mu) = P_n^m(\mu)$ and the star denotes a complex conjugate.

Then it may be shown (Coddington and Levinson, 1955, chapter 3, section 4) that for non-zero and non-degenerate eigenvalues $\omega' = \omega_x^* + i\omega_y^*$, the solution to (2.6) is

$$\begin{align*}
\Psi(t) &= \sum_{\nu=1}^{N} \kappa^\nu \phi^\nu \exp(-i\omega^\nu t) \\
\quad + \sum_{\nu=1}^{N} \frac{\eta^\nu \phi^\nu}{i-\omega^\nu} \exp(-i\omega^\nu t) - 1
\end{align*}$$

(2.8)

Here $\omega^\nu$ are the eigenvalues and $\phi^\nu$ the right-hand column eigenvectors which satisfy

$$\begin{align*}
(\omega^\nu [I] - [C]) \phi^\nu &= 0, \nu = 1, \ldots, N
\end{align*}$$

(2.9)

where $[C] = [A]^{-1}[B]$ and $[I]$ is the unit matrix. Also $\omega_1^*$ is the growth rate and $\omega_x^*$ the phase frequency.

The coefficients $\kappa^\nu$ and $\eta^\nu$ in Eq. (2.8) are given by (Gantmacher (1959), chapter 9, section 8; Wilkinson (1965), section 1.4; Morse and Feshbach (1953), Vol. I, pp. 884-886).

$$\begin{align*}
\kappa^\nu &= (\alpha^\nu, \Psi_0)/(\alpha^\nu, \phi^\nu) \\
-\chi^T \Psi_0/\chi^T \phi^\nu
\end{align*}$$

(2.10a)

and

$$\begin{align*}
\eta^\nu &= (\alpha^\nu, [A]^{-1}f)/(\alpha^\nu, \phi^\nu) \\
-\chi^T [A]^{-1}f/\chi^T \phi^\nu
\end{align*}$$

(2.10b)

where $\chi^T$ is the left-hand row eigenvector of $[C]$ corresponding to eigenvalue $\omega^\nu$ and $\alpha^\nu$ is the right-hand eigenvector of the adjoint matrix $(C)^T = ([C]^T)^*$ corresponding to eigenvalue $\omega^\nu$. Here the superscripts + denote adjoint, $T$ transpose and * complex conjugation.

The parentheses in Eq. (2.10) denote the inner product

$$\begin{align*}
(y, x) &= y^T x
\end{align*}$$

(2.11a)

In particular, the $L_2$ norm for an eigenvector $x$ is defined by

$$\|x\| = (x, x)^{1/2} - (x^T x)^{1/2}$$

(2.11b)

The same approach applies in other geometries and discretizations. Frederiksen and Frederiksen (1989) for example consider the choice of expansion representations for instability theory on beta-planes.

Regarding actual atmospheric flows as close to an exact steady state solution may be a severe approximation. In this article we shall also review a number of studies of the instability of observed instantaneous and climate flows which, with certain approximations as discussed in detail in appendix A of Frederiksen (1989), lead to essentially the same eigenvalue-eigenvector equations.

(c) Perturbation expansion in truth

Nonlinearity does of course play an important role in the actual generation of large-scale atmospheric disturbances. This has been stressed in many studies, for example, by Gall (1976b), Simmons and Hoskins (1978), Frederiksen (1981a,b,c, 1983b), Young and Villiere (1985), Frederiksen
and Webster (1988) and Barnes and Young (1992).

To some extent, the roles of nonlinear and transient effects can be taken into account by performing a sequence of linear instability studies using instantaneous flows at suitable time intervals (Frederiksen 1989; Frederiksen and Bell 1990). However, particularly in the case of climatological basic states, we only regard normal mode instability theory as the lowest-order term in a so-called "perturbation expansion in truth" with perhaps non-normal modes and nonlinear effects playing the roles of higher-order terms in such an expansion. As such, it is a very useful starting point for understanding many phenomena.

3. REGIONAL CYCLOGENESIS

Since the pioneering studies of Charney (1947), Eady (1949) and Phillips (1951), there have been numerous studies of the baroclinic instability of (two-dimensional) zonally averaged basic states on beta and f-planes (e.g. Green, 1960; Brown, 1969; Stone, 1969; Simmons, 1974; Staley and Gall, 1977) and on the sphere (e.g. Hollingsworth, 1975; Moura, 1976; Moura and Stone, 1976; Warn, 1976; Simmons and Hoskins, 1976, 1977; Gall, 1976a; Baines and Frederiksen, 1978; Frederiksen 1979b, 1981a). However, it is only in the last 10 to 15 years that the problem of regional cyclogenesis has been systematically analysed on the basis of the three dimensional instability of observed flows.

(a) Northern Hemisphere studies

(i) Idealized basic states

Frederiksen (1978a) examined the instability properties of both free (unforced) and forced exact steady state solutions consisting of zonal flows and planetary waves (as well as planetary waves by themselves) in two-level models on a sphere. It was found that the regions of preferential development of cyclone scale disturbances occur slightly downstream of the position where the excess vertical shear of the zonal wind above Phillips' (1954) criterion is a maximum. Here Phillips' criterion generalized to spherical geometry is given by

\[ \frac{\tilde{\mu}^1 - \tilde{\mu}^3}{a\Omega} \frac{\bar{\sigma}(1 - \mu^2)^{1/2}}{\mu^2} \geq 0 \]  

where \( \tilde{\mu}^1 \) and \( \tilde{\mu}^3 \) denote the upper and lower level zonal winds, \( \bar{b} = (0.124) \) is a constant, \( \bar{c}_s \) is the specific heat at constant pressure and \( \bar{\sigma} \) is the static stability parameter.

Also, pure barotropic waves in the basic states were found to decrease the growth rates of the fastest growing modes while pure baroclinic waves increase them. Subsequent studies examined the effects of increasing the vertical resolution to five levels (Frederiksen, 1979b) and the location of heat and momentum fluxes within two-level models (Frederiksen, 1980) for basic states consisting of zonal flows and single planetary waves. Frederiksen (1979a) examined the regions of preferential development for a thirty degree jet plus a wavenumber 3 basic state wave which provides a gross approximation to the Northern Hemisphere winter flow in a two-level model. It was found that the model could explain the locations of regional cyclogenesis, polewards and slightly downstream of the jet stream maxima, as well as the general character of momentum and heat fluxes as found from observations by Blackmon (1976) and Blackmon et al. (1977).


(ii) Observed climatological basic states

The instability properties of observed three-dimensionally varying climatological basic states were examined by Frederiksen (1982a); see also the review by Frederiksen (1986). The Northern Hemisphere winter flow averaged over the eight winters 1963/64 to 1970/71 was used in a two-level quasigeostrophic model and the faster growing modes examined for different values of the static stability parameter or potential temperature difference between the upper and lower levels. In the most unstable case corresponding to a potential difference of 23K (typical of the atmosphere), the fastest growing mode has a growth rate of 0.4270 day\(^{-1}\), a period of 3.3 days and is a monopole cyclogenesis mode, as shown in Fig. 1a for the upper layer, with largest amplitudes in the Atlantic and Pacific as shown in Fig. 1b. Here the amplitude of the streamfunction is just \( \sqrt{2} \) times the rms random phase ensemble average (RPEA) streamfunction defined in Eq. (2.7) of Frederiksen (1982a).

The mode has a westward tilt with height (not shown) and poleward eddy heat fluxes typical of baroclinic instability modes. The regions of preferential development are slightly downstream and poleward of the jetstream maxima of the basic state. The locations of these regions are in quite good agreement with band pass-filtered observations of the storm tracks (e.g. Blackmon, 1976, Fig. 5a), although the actual relative values of the maxima in the Atlantic and Pacific vary from mode to mode.

For the two-level model the maximum amplitudes occur at the upper level. However in multi-level models the fastest growing modes tend to have their maxima at the surface and also are more strongly localized in the Atlantic and Pacific regions as shown in Fig. 9 of Frederiksen (1983a) for a five-level quasigeostrophic model. The vertical structure problem, viz., that multi-level cyclogenesis
instability modes tend to have their maxima at the surface while in the observations the amplitudes peak near the tropopause occurs with both zonally averaged and three-dimensional basic states. It is related to the fact that nonlinearity is important for the observed anomalies and considerable improvement in the vertical structure of disturbances is found when the effects of nonlinearity are taken into account (Gall, 1976b; Simmons and Hoskins, 1978; Frederiksen, 1981 a,b,c, Frederiksen and Puri, 1985).

Robertson and Metz (1990a,b) investigated the feedbacks of transient eddies generated by instability theory on three-dimensional Northern Hemisphere flows and compared storm track instability modes, for general circulation model generated basic states, with the general circulation model storm tracks. Frederiksen and Frederiksen (1993a) found that the inclusion of explicit moisture in the basic state has little effect on the structures of the storm track modes but increases their growth rates.

(iii) Local and global modes

Phillips' criterion (3.1) or the corresponding growth rate, when applied locally may provide a rough but useful measure of the regions of instability of horizontally varying baroclinic flows. As pointed out by Pierrehumbert (1984), there is however one effect that these simple criteria cannot account for even approximately. This is the fact that if the baroclinicity varies significantly in the downstream direction, a wave packet could propagate out of the unstable region before it has time to grow significantly. Therefore large mean flows will tend to produce a reduction in the growth rate of such disturbances while the Phillips' criterion is independent of the mean flow.

The WKBJ theory of spatially (and temporally) developing disturbances provides a method of analysing such flows semianalytically, in the case when the dispersion relation is relatively simple. This method was first applied in plasma physics; see, for example, Bers (1983) and Huerre and Monkewitz (1990) for reviews. It was introduced to geophysical fluid dynamics by Thacker (1976) and Merkine (1977). Subsequent semianalytical WKBJ studies have been made by Merkine and Shafranek (1980), Pierrehumbert (1984), Peng and Williams (1986, 1987) and Ebisuzaki (1989, 1989a, b) and numerical studies by Mak and Cai (1989), Cai and Mak (1990) and Crum and Stevens (1990).

Pierrehumbert (1984) showed that for local modes the growth rate is equal to the absolute growth rate determined at the point of maximum vertical shear. In contrast global modes have growth rates which depend on the average baroclinicity over the domain. For typical climatological flows in two-level models, Phillips' criterion nevertheless provides at least a useful rough estimate of where preferential development is likely to occur. Perhaps this indicates that the variations in the baroclinicity of the mid-latitude climatological flows are not too drastic.

Pierrehumbert (1984) and Merkine and Shafranek (1980) also argue that since real cyclones mature and decay before completing even a single circuit of the globe, global modes are unphysical except in the sense that a large number of them could be superimposed to represent the transient evolution of a wave packet. This would seem to be reasonable for individual synoptic scale disturbances growing on time evolving flows. The observed climatological storm tracks (e.g. Blackmon, 1976, Fig. 5a; Trenberth, 1981, Fig.

Fig.1 (a) Disturbance streamfunction and (b) amplitude of disturbance streamfunction at the upper level for mode 1, the fastest growing mode for the eight winter average Northern Hemisphere basic state (from Frederiksen, 1986).
17; Trenberth, 1982, Figs. 6 and 11; Trenberth, 1991, Figs. 6a and 7a) however are global in character, and represent the statistics of the superposition of many disturbances around each hemisphere. The dominant storm track instability modes of climatological three-dimensional basic states capture the central regions of these observed storm tracks rather than representing individual cyclones. Indeed the instability theory for climatological basic states may be interpreted as the statistical problem of the growth of an ensemble of perturbations on an ensemble of basic states sampled from the climatological basic state. Here the assumption that the variance of the ensemble of perturbations is small is needed.

(iv) Deflection and splitting of storm tracks during blocking

Frederiksen (1989) and Frederiksen and Bell (1990) examined the deflection and splitting of storm track instability modes from the usual climatological positions, produced by large scale blocks over the Gulf of Alaska and over the Atlantic Ocean. They used sequences of observed instantaneous flows during November and January 1979 for their Northern Hemisphere basic states within a five-level quasigeostrophic model. Robertson and Metz (1989), using a two-level quasigeostrophic model, studied the sensitivity of the location of cyclogenesis modes to the superposition of persistent anomaly patterns on Northern Hemisphere climatological winter flows. They found that the storm tracks are also deflected significantly from their climatological positions as a result.

Fig. 2, taken from Frederiksen and Bell (1990), shows an example of the deflection and splitting of a storm track instability mode caused by a block over southern Greenland on 24 January 1979. Shown is the 500mb streamfunction for the third fastest growing mode. In fact during the whole period from 14 January to about 30 January such modes occur in which the Atlantic storm track splits with one branch deflected poleward and one equatorward around the observed block. During this period the splitting also retrogresses with the block from Russia across the North Atlantic to southern Greenland. In Fig. 2, the splitting occurs near the east coast of North America.

We note in particular that one branch of the storm track is deflected polewards across Greenland and then south across Europe. As discussed by Bengtsson (1981), during the period around 20 January, Europe experienced severe snowstorms and cold outbreaks consistent with disturbances such as shown in Fig. 2.

The changes in the storm track modes during this period may of course be related to changes in the location of the jetstreams, and in particular to the "excess shear" (cf., Eq. (3.1)) during this period. That is, the splitting and deflection of the storm-track modes follows the splitting and deflection of the jetstreams from their climatological positions. Blackmon et al. (1986), Lau (1988) and Dole (1989) have noted similar behaviour of the storm tracks during blocking in their observational studies.

(v) The effects of variable static stability and nongeostrophic effects

The previous studies have used the quasigeostrophic equations with a static stability parameter which is constant over the globe or hemisphere. Frederiksen and Frederiksen (1992) formulated the three-dimensional instability theory for a two-level primitive equation model in spherical geometry and with variable static stability parameter. The static stability parameter, as shown in Fig. 1c of Frederiksen and Frederiksen for January 1979, has quite substantial variations over the Northern Hemisphere. This, like the planetary waves structure, could produce significant shifts in the location of the regions of preferential cyclogenesis.

Fig. 3 shows the rms random phase ensemble average (RPEA) mean streamfunction for the fastest growing (a) primitive equation and (b) quasigeostrophic eigenmodes within two-level models. Here the quasigeostrophic model has a globally constant static stability of 11.5K while for the primitive equation model it is as depicted in Fig. 1c of Frederiksen and Frederiksen (1992).

Although the anomaly correlation, \( A_e = 0.818 \), is quite large there are nevertheless significant differences, in some
The difference between the primitive equation and quasigeostrophic eigenmodes can be understood in terms of the basic state static stability and Phillips' criterion for the different models shown in Figs. 2a and 2b of Frederiksen and Frederiksen (1992). There is a region of very low static stability with a minimum of 10.3 K just off the northwest coast of the United States. Such a region is conducive to cyclogenesis and would have the effect of locating the storm track farther upstream than for the case of a globally constant static stability of 11.5 K. That the region of preferential development should occur farther upstream is also consistent with the stability criterion depicted in Fig. 2a of Frederiksen and Frederiksen. In particular, this diagram shows a maximum coincident with the location of the maximum in streamfunction RPEA for the fastest growing PE mode. In contrast, Fig. 2b of Frederiksen and Frederiksen shows a maximum farther downstream over the North American continent and also in the west Pacific; these same features are reflected in the RPEA pattern of Fig. 3b for the corresponding quasigeostrophic mode.

The instability criteria are also sensitive to the inclusion or exclusion of the divergent wind shear, especially over the North American continent. Nongeostrophic effects also appear to play a significant role in influencing the location of the storm track in this region.

None of the quasigeostrophic storm track modes studied by Frederiksen and Frederiksen have maxima over the western United States. This suggests that the local static stability of a flow field and the nongeostrophic terms can, at least in situations of large-scale anomalous flow such as during January 1979, significantly affect the geographical location of preferred cyclogenesis. In the North American case, the westerly location of the storm track is consistent with the description by Wagner (1979) of the synoptic situation during much of the month. In particular, Wagner describes severe storms along the west north-west coast and in the Midwest of the United States in the first half of January.

In contrast, for the eight-winter (1963/64 - 1970/71) climatological flow (Frederiksen 1982a) and for the January 1978 flow (Frederiksen 1983a), quasigeostrophic instability theory gives close correspondence with observations (Blackmon et al. 1977, 1984; Metz and Lu 1990) as far as the geographical locations of the storm tracks are concerned, including the Atlantic storm track. Metz and Lu (1990), in particular, point out the striking similarity between the observed storm-track complex empirical orthogonal function patterns and the fastest growing normal modes from quasigeostrophic three-dimensional instability theory.

Figure 3 The rms random phase ensemble average of the vertical mean streamfunction for the fastest growing (a) primitive equation and (b) quasigeostrophic eigenmodes with the monthly mean Northern Hemisphere January 1979 basic state (from Frederiksen and Frederiksen, 1992).

geographical locations, between these two modes. Most obvious is that the quasigeostrophic mode has local maxima in amplitude farther downstream from the region of maximum shear in the American-Atlantic and western Pacific regions and these tend to be more poleward.

The differences are most noticeable for the North American storm track. Here the primitive equation mode has maximum RPEA amplitude centered over the western part of the United States in contrast to the extreme eastern location in the quasigeostrophic mode.

Frederiksen and Frederiksen (1992) have also examined Phillips' criteria for instability and the instability modes, both within primitive equation and quasigeostrophic models,
for the 6-year average January flow (1979 - 84). For this longer-term average there is again a very close correspondence between the maxima in the instability criteria and the regions of preferential development of the dominant storm track modes in the Northern Hemisphere, in both the primitive track and quasigeostrophic models. Thus, while quasigeostrophic models may produce some displacement of the storm tracks, for Northern Hemisphere flows this appears to be primarily a problem in situations of large anomalies, such as occurred during the persistent blocking during January 1979.

Recently, Mak (1993) has also examined the role of horizontally varying static stability in determining the regions of localized cyclogenesis. He used a two-level linear balance equation model on a beta-plane with simple idealized basic states and studied the fastest growing normal modes.

(b) Southern Hemisphere studies

(i) Climatological storm track modes

The roles of planetary scale basic state waves in locating the Southern Hemisphere storm tracks were examined by Frederiksen (1985a) and Frederiksen and Frederiksen (1993b). Frederiksen (1985a) studied the storm track modes in a 5-level quasigeostrophic model for a three-dimensional climatological January basic state while Frederiksen and Frederiksen (1989b) used a two-level primitive equation model and January and July climatological basic states.

In the study of Frederiksen and Frederiksen (1993b), the fastest growing mode for the January basic state is a typical monopole cyclogenesis mode. In Fig. 4(a), we show, for this mode, the upper level streamfunction. This mode has the typical westward tilt with height of disturbances growing due to baroclinic instability. It has largest amplitudes in the eastern hemisphere across the southern Indian Ocean and to the southeast of Australia between about 45-50°S. This structure is consistent with modes of this type making contributions to the observed storm tracks in the Southern Hemisphere for January. The amplitude of the upper level streamfunction for this mode, shown in Figure 4c of Frederiksen and Frederiksen (1993b), compares favourably with Fig. 7(a) of Trenberth (1991) which shows the standard deviation of the meridional wind in the 2-8 day band for January at 300mb. There is a general agreement between the theoretical and observational results, although the maxima in mode 1 appear to be slightly downstream of those of the observations. The maximum amplitude of mode 1 occurs slightly downstream of where the generalised Phillips’ criterion, in Fig. 5(a) of Frederiksen and Frederiksen (1993b) has its maximum value in the southern Indian Ocean and where the maximum jet strength and, in particular, shear occur.

The cyclogenesis modes found by Frederiksen and Frederiksen with the two-level primitive equation model are quite similar, as far as the geographical location of the storm tracks are concerned, to the cyclogenesis modes obtained by Frederiksen (1985a; Figs. 5,6 and 7) using a five-level quasigeostrophic model and the 1972-1976 January climatology. There are, however, some differences in the structures of the two- and five-level modes. The five-level modes tend to be more localized. In addition, the five-level modes suffer from the vertical structure problem, that is, their maximum amplitudes occur near the surface while for observed storm track disturbances they occur near the tropopause. This difference is due to the importance of nonlinear effects for the observed disturbances and one sees a corresponding improvement with observations when initial instability modes are modified by nonlinearity (Frederiksen 1981a, b, c and references therein). In contrast, the two-level model tends to filter out many of the shallow disturbances which tend to be quickly modified by the nonlinearity. For example, for mode 1 the largest amplitude occurs at the upper level with max (amp(ψ^1)) = 248 and max (amp(ψ^0)) = 143 arbitrary units.

Fig. 4b shows the mean streamfunction for the fastest growing mode with the July basic state of Frederiksen and Frederiksen (1993b). In contrast to the January cyclogenesis modes, which were located purely on the polar jet, we now see the formation of elongated eddies, downstream of Australia, where the mode grows on both jet streams. The largest amplitudes, however, again occur in the region of the polar jet. Maxima in the amplitude of the streamfunction occur in the southern Indian Ocean and to the south of New Zealand. For mode 2, the second fastest growing mode, the maximum amplitude occurs to the southeast of Australia. Both modes have subsidiary maxima downstream of Australia near New Zealand. These features are also seen in Fig. 7 (b) of Trenberth (1991) which shows the standard deviation of the meridional velocity in the 2-8 day band for July at 300mb. Again, however, the modal maxima tend to occur somewhat further downstream than in the observations. Of course, in the observations, many different modes contribute and, as discussed above, nonlinear effects play a significant role.

The maxima of these modes, in the region of the polar jet, occur close to where the generalised Phillips’ criterion in Fig. 3(b) of Frederiksen and Frederiksen (1993b) has a subsidiary maximum. The primary maximum of the Phillips’ criterion, downstream of Australia near New Zealand, corresponds closely to the secondary maximum in the modal amplitude. We note that, although the generalised Phillips’ criterion may provide some guidance as to where disturbances will tend to occur, it is at best a fairly crude criterion compared with doing the full instability calculations.
(ii) Comma cloud disturbances of South Brazil

In their study of the instability properties of the global three-dimensional climatological flow for January 1979, Frederiksen and Frederiksen (1993a), found a wide variety of disturbances including modes with structures and periods characteristic of the comma cloud disturbances of South Brazil. They again used a two-level primitive equation model but also included a wave-CISK cumulus heating parameterization. Some of their disturbances, such as their moist B mode 23 shown in their Fig. 4, are unusual in that they also have large mean streamfunction amplitudes just downstream of the Andes and are suggestive of lee cyclogenesis. These modes have structures and phase velocities very similar to the inverted comma cloud disturbances of South Brazil (Bonatti and Rao, 1987); with a zonal wavenumber of about 12, the phase frequencies of these modes correspond to about $8 \text{ms}^{-1}$, the same as estimated on the basis of observations by Bonatti and Rao. The e-folding time of these disturbances depend sensitively on the moisture parameterization decreasing with the strength of the cumulus heating.

Bonatti and Rao also performed instability calculations but used the zonally averaged flow over the South American region in a beta-plane multi-level model with Mak’s (1982) moisture parameterization and different heating profiles. They obtained a phase velocity of $3.9 \text{ms}^{-1}$, somewhat shorter than the observed value of $8 \text{ms}^{-1}$, and an e-folding time of just over 1 day, considerably shorter than Frederiksen and Frederiksen’s minimum of 4.2 days for moist B mode 23. Some of these differences in growth rates may be attributed to differences in two- and multi-level models (Frederiksen, 1978b), to basic states taken from different months (April versus the present January case) and the different moisture parameterizations. The results of Frederiksen and Frederiksen (1993a) with a three-dimensional basic state and a primitive equation model with full spherical geometry nevertheless agree with Bonatti and Rao’s qualitative conclusions, based on a quasigeostrophic multi-level beta-plane model with a zonally averaged basic state, that the basic mechanism for the generation of the South Brazil inverted comma cloud disturbances appears to be baroclinic instability modified by latent heat effects.

(iii) Australian northwest cloud band disturbances

Frederiksen and Frederiksen (1993c) studied the instability modes for July basic states taken from general circulation model simulations (by Frederiksen and Balgovićd, 1993) of the atmospheric response to Indian Ocean sea surface temperature anomalies. In cases where the sea surface temperature anomaly produced an increase in the frequency of the so-called Australian northwest cloud band, their two-level primitive equation model with a wave-CISK cumulus heating parameterization also produced normal modes with very similar properties. The northwest cloud band is a broad band of cloud and rainfall extending from the northwest to the southwest corners of Australia and then across New Zealand and into the South Pacific.

Modes with periods between 10 to 14 days emanating from the northwest coast of Australia were found. The northwest cloud band appears to be associated with a fairly large scale baroclinic disturbance intensified, particularly to
the northwest, by the cumulus heating.

4. BLOCKING

Unlike the generally accepted baroclinic instability theory of cyclogenesis (Charney, 1947; Eady, 1949), the importance of the role of instability in the development of other large-scale disturbances has been more controversial.

(a) Northern Hemisphere studies

(i) Climatological basic states and the baroclinic - barotropic dipole instability mechanism.

The study of Frederiksen (1982a) indicated that instability theory with three-dimensional basic state flows could produce a variety of disturbances in addition to the monopole cyclogenesis modes (such as that shown in Figure 1) associated with the storm tracks. It suggested that three-dimensional instability theory could provide a similar basis for understanding both blocking and localized cyclogenesis as zonally averaged instability theory has traditionally provided for cyclone scale disturbances. According to this proposition, the formation of mature anomalies such as blocking events is initiated (at least in some cases) by the generation of onset-of-blocking dipole instability modes upstream of the regions of maximum amplitude of the mature anomalies. These onset modes have westward tilt; baroclinic processes and barotropic processes are relatively important in their formation. They propagate eastward, and as they increase in amplitude, they become quasi-stationary and essentially equivalent barotropic. It was suggested that in many qualitative respects the change from onset-of-blocking dipole modes to mature anomaly modes would occur through nonlinear effects in much the same way as is the case with cyclone scale disturbances and as is found in the life cycle experiments of Gall (1976b), Simmons and Hoskins (1978) and Frederiksen (1981b,c) using zonally averaged basic states. Support for this view also comes from the observational study of Dole (1986).

Here we shall concentrate on instability modes relevant to blocking in the Pacific-North American regions, since this is the case for which there are most detailed observational data for comparison. Figures 5a and 5b show upper level streamfunctions for the third and twenty-second fastest growing modes for case 1 of Frederiksen (1982a, 1983b) (corresponding to a potential temperature difference between the layers of 23°C or a static stability parameter of 11.5°C). We may identify the mode in Figure 5a as a Pacific onset-of-blocking mode which initiates the blocking process and that in Figure 5b as a Pacific-North American mature anomaly mode characteristic of the later stages of block development in this region. The onset-of-blocking mode has an e-folding time for growth of about 2.5 days and a period of 7.4 days. These values are to be compared with an e-folding time of 2.3 days and a period of 3.3 days for the fastest growing monopole cyclogenesis mode in Figure 1 for case 1. Like the monopole cyclogenesis modes, the onset-of-blocking mode has a westward tilt with height, while the mature anomaly mode is equivalent barotropic, has an e-folding time of 5.8 days, and has infinite period. In fact, modes very similar to the mature anomaly mode in Figure 5b have been obtained within nondivergent barotropic models (Simmons et al., 1983; Frederiksen, 1983b; Schubert, 1985), in shallow water equation models (Haarsma and Opsteegh, 1988) and in five-level models (Frederiksen and Bell, 1987).

It is interesting to compare the instability results with
the studies of Dole (1983, 1986), which provided important insight into many aspects of observed anomalies including their time evolution. Figure 6 shows a time sequence of composite analyses of 15 positive anomaly cases at 500 mb leading to establishment of a mature Pacific anomaly pattern. Figure 6a is unfiltered data on day -3 before the appearance (on day 0) of the essentially stationary large-scale positive anomaly in the key region in the north central Pacific, while Figure 6b shows low-pass-filtered data on day 6.

For the period leading up to the appearance of the positive anomaly in the key region, which we refer to as the onset-of-blocking period, Dole notes that the "sequence of development suggests that the initial rapid growth of the main centre is primarily associated with the propagating intensifying disturbance which originates in mid-latitudes near Japan." He also notes that "this disturbance continues to intensify as it becomes quasi-stationary over the key region." The dipole wave train across east Asia and into the Pacific that appears in Figure 6a at day -3 is qualitatively very similar to that in Figure 5a (as well as to that in Figure 6a of Frederiksen, 1982a).

Further evidence that baroclinic onset-of-blocking instability modes are involved in the initial period may be obtained from Figure 20 of Dole (1986), which shows longitude-pressure cross sections at 45°N and 20°N of the unfiltered Pacific composite anomalies at days -3 and 0.

The dipole nature of the developing anomaly is clearly evident, particularly at day -3, and the zonal scale of the anomaly at day -3 is practically the same as shown in Figure 5a. The developing anomaly has a definite westward tilt with height, as do the onset-of-blocking instability modes. Dole notes: "...This feature has pronounced westward tilts with height during this period, suggesting that a substantial baroclinic contribution is involved in its amplification...."

Following day 0 the development of the 500-mb height anomaly follows that shown by Dole (1986), with the anomalies being largely equivalent barotropic during this period. Intensification of the centers occurs with little phase propagation, and by day 4, and especially at day 6, the Pacific-North American pattern is established as shown in Figure 6b.

The observational and theoretical results described above suggest that the development of mature anomalies, such as blocks, may be thought of as consisting basically of two stages. For the Pacific-North American pattern the baroclinic-barotropic dipole instability mechanism may be summarised as follows:

1. A rapidly growing and relatively rapidly eastward propagating onset-of-blocking dipole disturbance mode which tilts westward with height forms in the east Asia-Pacific Ocean region through the combined baroclinic-barotropic instability of the three-dimensional basic state. As the disturbance grows, the regions of largest amplitude propagate into the central Pacific, and the disturbance increases its zonal scale, becoming quasi-stationary and essentially equivalent barotropic through the operation of nonlinear effects. At this stage the mode has a structure reminiscent of the mature Pacific-North American pattern instability mode.
2. The (nonlinear analogue of the) mature anomaly instability mode amplifies without phase propagation and through the operation of largely equivalent barotropic effects to form the large-amplitude mature Pacific-North American anomaly pattern.

The change from the initial onset-of-blocking mode to the formation of the large-amplitude mature anomaly pattern mode is, in fact, a continuous one, and the separation into two distinct stages is to some extent arbitrary, if convenient. The fact that the mode continually changes its structure in a gradual manner to optimize its growth, as nonlinear effects become increasingly important, is also clearly shown in the numerical simulation of Frederiksen and Puri (1985).

Further observational evidence for the existence of onset-of-blocking dipole modes with periods between about 6 and 10 days was presented by Schubert (1986) using filtered data rather than the composite approach of Dole. Figure 21 of Schubert (1986) shows one of Schubert's modes with periods between 6 and 10 days which is quite similar to the onset-of-blocking instability mode in Figure 5a and to Dole's observed onset mode in Figure 6a.

(ii) Barotropic instability

Simmons et al. (1983) proposed an alternative instability theory for the development of mature anomalies which contrasted with the previously described baroclinic-barotropic dipole instability mechanism of Frederiksen (1982a). The Simmons et al. mechanism is direct barotropic instability of the climatological basic state flow. They recognise, however, that the global growth rates for direct barotropic instability may only be about one-third those for the baroclinic-barotropic dipole instability mechanism. In order for their path of development to be competitive, they propose that the local growth rate (i.e., at a particular point) due to barotropic instability may be enhanced (during one-quarter period) over the global growth rate due to phase propagation. However, for this mechanism to work, the mode must be propagating, and Simmons et al. suggest that as a consequence, stationary growing modes may be difficult to excite in reality.

In contrast, the studies of Frederiksen (1989) and Frederiksen and Bell (1990), using instantaneous observed basic states during blocking, found that the fastest growing equivalent barotropic instability modes are in fact non-propagating. Their growth rates are about 5 times those of barotropic modes growing on climatological flows typical for Northern Hemisphere winter. Although the instability calculation used to obtain these modes is a linear one, nonlinear processes have been responsible for distorting the flow to the extent that barotropic instability has become the dominant process at this late stage of block development.

(iii) Observed instantaneous basic states

Frederiksen (1989) and Frederiksen and Bell (1990) studied the instability properties of sequences of daily instantaneous flows during situations leading to blocking over the Gulf of Alaska and over the North Atlantic Ocean. Here we shall concentrate on the North Atlantic blocking during January 1979. Frederiksen and Bell (1990), found that, in addition to rapidly growing cyclogenesis modes, their five level quasigeostrophic model produced rapidly growing onset-of-blocking modes and equivalent barotropic mature blocking modes with e-folding times as short as 1.7 days. The most noteworthy events and most remarkable large-scale localized dipole instability modes were found to occur during the period between 19 and 24 January when the North Atlantic block amplified to a mature disturbance in the region of south Greenland.

Figs 7a and b show the 500mb disturbance streamfunctions for mode 1 on 20 and mode 2 on 22 January, the fastest and second fastest growing modes on the respective days. Mode 1 on 20 January is a propagating westward tilting onset-of-blocking mode which has a smaller scale than the stationary equivalent barotropic mature anomaly mode in Fig. 7b. For example, mode 1 on 20 January has a westward tilt with height between 900 and 100 mb of approximately 30 degrees in the regions of large amplitude. It has largest amplitude at 300 mb and substantial amplitude throughout the troposphere. This onset mode appears to be associated with the rapid deepening of the surface cyclone off the east coast of North America during this period and the subsequent amplification of the 500 mb blocking high over southern Greenland. At 900 mb (not shown) the normal mode has largest amplitude off the east coast of North America near \( \lambda = 300^\circ, \phi = 50^\circ N \) where the surface low is observed on 21 January. In many respects the sequence of events that occur during this period in January 1979 is very similar to that associated with the Atlantic blocking between 6 and 9 January, 1977, analysed in detail by Colucci (1985).

Mode 1 on 21 January (not shown) and mode 2 on 22 January are large-scale equivalent barotropic modes; they are rather remarkable in that they have very little of the wave-train structure usually found with instability modes, but rather are strongly localized in the blocking region. These modes are also quasi-stationary.

As discussed by Branstator (1985) and Frederiksen and Bell (1987, 1990), the adjoint mode (see Eq. (2.10)) provides the largest projection onto a given eigenvector. Thus for modes of a given unit norm it provides the optimal initial condition for asymptotically (as \( t \to \infty \)) exciting large amplitude in the eigenmode.

Frederiksen and Bell (1990) found that the adjoint
Instability theory

The eigenvector for mode 1 on 20 January is a small-scale structure reminiscent of a localized storm off the east coast of North America. It however has largest amplitude at 500 mb rather than at the lowest level where most incipient cyclogenesis modes have their maxima. Thus its structure is more suggestive of a more mature cyclone-scale disturbance. The adjoint eigenmode 1 on 21 January has some of the gross features of the onset-of-blocking mode 1 on 20 January.

The above results suggest that localized and intense cyclogenesis off the east coast of North America that occurred during this period, may possibly be able to trigger the sequence of events which leads to the mature block on 22 January. The situation here appears to be quite similar to the sequence of events leading to mature anomalies found by Frederiksen and Puri (1985) in a numerical simulation study (see their Figs 7, 8 and 12) and by Colucci (1985) on the basis of an observational study. See also Holopainen and Fortelius (1987) and Holopainen (1990).

(iv) The role of topographic instability

Charney and DeVore (1979) originally proposed a theory of block formation and switching between high and low zonal index flow based on the topographic instability mechanism. In their studies with observed instantaneous flows Frederiksen (1989) and Frederiksen and Bell (1990) also examined the role of topographic instability. They found that topographic instability plays at most a very peripheral role during blocking compared with baroclinic - barotropic and barotropic instabilities. This is not to say that topography, as well as land-sea contrasts, is not important in determining the preferred regions of blocking. Frederiksen and Bell (1987) also found that topographic instability does play an important role in the generation of some of the other large scale Northern Hemisphere low frequency anomalies.

Following the work of Charney and DeVore (1979), there have appeared a large number of further studies on the topic of topographic instability (e.g. Hart 1979; Chamey and Straus 1980; Davey 1980, 1981; Deininger 1981; Källen 1981, 1982, 1983; Egger 1981; Egger and Metz 1981; Linden 1983; Benzi et al. 1984; Rambaldi et al. 1985; Vallis 1983; Pyfe and Derome 1986, 1987a, b; Mukougawa and Hirota 1986; Frederiksen and Frederiksen, 1989, 1991; Grotjahn and Wang, 1990; Howel and Nathan, 1990.) As noted by Frederiksen and Frederiksen (1989, 1991) some of the previous studies using the beta-plane approximation contained quantitative or qualitative errors in their formulations.

Pedlosky (1981) presented a weakly nonlinear theory of the interaction of flows over topography while strongly nonlinear studies have been performed by Frederiksen (1985b), Frederiksen and Carnevale (1986), Carnevale and Frederiksen (1987) and McIntyre and Shepherd (1987) for barotropic models and by Frederiksen (1991a,b) for baroclinic models.

(b) Southern Hemisphere studies

(i) Observed basic states

Frederiksen (1984) found rapidly growing normal modes with dipole structures over the south-eastern Australian-Tasman Sea area prior to the formation of observed blocks.

Frederiksen and Frederiksen (1993b) examined the
structures of dipole instability modes obtained using a two-
level primitive equation model with January and July
Southern Hemisphere climatological states. For both
January and July, they found large scale slower propagating
dipole or multipole modes which, at suitable phases, are
consistent with blocking in some or all of the three regions
of blocking described by van Loon (1956). The regional
characteristics of Southern Hemisphere blocking have further
been elaborated on by Taljaard (in van Loon et al. 1972),
Wright (1974), Baines (1983), Coughlan (1983), Trenberth
and Swanson (1983) and Lejenits (1984). The primary
regions of blocking are in the Australian/New Zealand
region, east of the Falklands and to the southeast of Africa.

For the January basic state, Frederiksen and
Frederiksen's eigenmode 32, which has a period of 7.6 days
and an e-folding time of 9.1 days, is typical of blocking
modes. The lower level streamfunction for this mode is
shown in Fig 8 (a). If we take the phase of this mode in
Fig. 8(a) as corresponding to minus the lower level
streamfunction, since the geopotential height and
streamfunction have opposite signs in the Southern
Hemisphere, we note that this mode would correspond to
blocking to the southeast of New Zealand, near the Falkland
Islands and to the southeast of southern Africa. That is, it
is somewhat similar to the triple block shown in Fig. 4 of
van Loon (1956). Although this mode is of considerably
larger scale than the cyclogenesis modes of subsection 3(b)
it has a westward tilt with height in most regions (not
shown), indicating that baroclinic processes are relatively
important in the genesis of this mode.

For their July basic state, a typical blocking mode is
mode 25 which has a period of 6.6 days and an e-folding
time of 8.1 days. Figure 8 (b) shows the structure of the
mean streamfunction for this mode. Depending on its phase
or sign, this mode could be associated with blocking in the
New Zealand region, in the Australian region or to the
southeast of South America and the southeast of southern
Africa. For both the January and July basic states there are
also a number of other modes with similar periods to the
modes described above which have dipole structures in some
of the main regions of blocking in the Southern Hemisphere.

One can only guess at the extent to which nonlinear
effects would increase the scale of blocking modes for both
January and July, decrease their phase speeds and to make
them equivalent barotropic. Such changes would be
expected on the basis of structural changes of baroclinic
waves due to nonlinear processes (Frederiksen, 1981a, b, c
and references therein). As discussed in section 4(a) these
changes are also observed for Northern Hemisphere blocks
(Dole, 1986, 1989; Colucci, 1985; Dole and Black, 1990)
and have been predicted on the basis of three dimensional
instability theory (Frederiksen, 1982a, 1983b). One would
again expect in the Southern Hemisphere that dipole modes
would also become slower moving, more equivalent
barotropic and of larger scale in the nonlinear regime. As
in the case of Northern Hemisphere blocks (Frederiksen and
Bell, 1990; Frederiksen and Frederiksen, 1992), we expect
the instability modes obtained with the current climatological
flows, having smoother potential vorticity gradients than
instantaneous flows, will tend to pick out the average
possible development during the month. In particular, many
of the modes, for both the January and July basic states,
tend to pick up more than one region of blocking around the
hemisphere.
5. LOW-FREQUENCY ANOMALIES AND QUASI-STATIONARY TELECONNECTION PATTERNS

(a) Northern Hemisphere anomalies with climatological basic states

In section 4(a) we discussed the role of instability in the generation of large scale low frequency anomalies such as the Pacific-North American (PNA) teleconnection pattern and large scale blocks over the North Atlantic. Analogs of a number of other teleconnection patterns have been obtained in studies of the instability of horizontally varying observed flows. Simmons et al. (1983) found an analog of Wallace and Gutzler's (1981) East Atlantic (EA) teleconnection pattern within a barotropic model while Frederiksen (1983b) obtained a North Atlantic Oscillation (NAO) instability mode similar to the observations of Walker and Bliss (1982) and van Loon and Rogers (1978) in both two-level baroclinic and barotropic models. Frederiksen and Bell (1987) examined the roles of baroclinic, barotropic and topographic instability in the generation of teleconnection patterns within five level quasigeostrophic and barotropic models. They found analogs of a wide variety of low frequency anomalies including the PNA, EA, NAO, North Soviet Union (or Eurasian), Western Pacific and Western Atlantic quasi-stationary teleconnection patterns.

Frederiksen and Frederiksen (1992) using the Northern Hemisphere January 1979 basic state in a two-level-primitive equation model found a stationary large scale mode which appears to be a combination of the PNA and NAO teleconnection pattern (Fig. 11 of Frederiksen and Frederiksen, 1992). It has a striking similarity to Fig. 9 of Wagner (1979) which shows the 700 mb height change from the first half to the second half of January 1979. The growth of this mode is consistent with a weakening of the Pacific block, the development of the south Greenland block and the transition in flow characteristics between the first and second halves of the month. It was found that the structure of the mode could (at least approximately) be related to the Kuo (1951) barotropic instability criterion which relates the instability of the flow to the gradient of the absolute vorticity.

(b) Southern Hemisphere anomalies with climatological basic states

In their study of Southern Hemisphere disturbances growing on January and July climatological basic states, Frederiksen and Frederiksen (1993b) also found a group of low frequency large scale anomaly modes which resemble some of the primary Southern Hemisphere teleconnection patterns.

For their January basic state, mode 38, with a period of 53 days and an e-folding time of 10 days, and mode 62, with a period of 317 days and an e-folding time of 21 days, are typical low frequency anomaly modes. Both modes have an equivalent barotropic vertical structure. Fig. 9 (a) shows the upper level disturbance streamfunctions for mode 38. These modes have largest amplitudes in the high latitude regions poleward of 60°S where they are essentially equivalent barotropic, as are most of the observed low frequency teleconnection patterns for the Southern Hemisphere (Mo and White, 1985; Mo and Ghil, 1987; Kidson, 1988; Karoly et al., 1989; Karoly, 1989b). Mode
38 resembles the Antarctic, or high latitude, mode corresponding to Kidson's (1988) EOF pattern 1 for summer (his Fig. 5(a)) and winter (his Fig. 5(b)). Our mode clearly has somewhat smaller scale features than the observed teleconnection patterns, although to some extent this is to be expected because of the ensemble averaging and time filtering of the observational data.

As well as large amplitude poleward of 60°S, mode 38 also has significant amplitude between 30° and 60°S, in a band stretching from New Zealand to southern South America. In these respects, it bears some similarity to the seasonal mean summer anomalies of 500mb height for the El-Niño/Southern Oscillation (ENSO) period of 1982/1983 shown in Fig. 9(d) of Karoly (1989a). Mode 38, however, has significantly larger amplitude in the high latitude region and may correspond to a mixture of the high latitude anomaly and the 1982/1983 ENSO anomaly with the former predominating.

Frederiksen and Frederiksen's fastest growing low frequency anomaly mode for July is mode 11 with a period of 177 days and an e-folding time of 6.1 days. Figure 9(b) shows the upper level streamfunction for this mode. It is very similar to Karoly's (1989b) high latitude teleconnection pattern which is shown in his Fig. 1 and to Kidson's (1988) EOF pattern 1 for winter (his Fig. 5(b)). Again this mode is equivalent barotropic with largest amplitude in the upper level.

For both January and July there are also modes which exhibit a definite wavenumber 3 structure similar to the observed wavenumber 3 teleconnection pattern shown in Figs. 4(a) and (c) of Mo and White (1985), Fig. 12 (b) of Mo and Ghil (1987) and Fig. 5(a) of Karoly et al. (1989). Mode 62 for January and modes 35 and 50 for July are such disturbances.

Another interesting low frequency mode is Frederiksen and Frederiksen's (1993b) mode 47 for July which has a period of 21.8 days and an e-folding time of 15.9 days. This mode has a wavetrain structure extending from Australia to the edge of Antarctica and then across to southern South America and is somewhat reminiscent of the wavetrain modes shown in Fig. 5(e) of Karoly et al. (1989) and Figs. 4(b) and (c) of Kidson (1988).

6. MONSOON DISTURBANCES AND INTRASEASONAL OSCILLATIONS

Frederiksen and Frederiksen (1993a) studied the normal mode disturbances for the global monthly averaged three-dimensional flow for January 1979 using a two-level primitive equation model with a wave - CISK cumulus heating parameterization. They examined a wide variety of disturbances but perhaps the most interesting and novel modes they found were quasi-stationary monsoon disturbances and intraseasonal oscillation modes.

(a) Quasi-stationary monsoon disturbances

For the particular specification of the cumulus heating that they referred to as the moist A case, Frederiksen and Frederiksen found that their mode 35 is a stationary mode which is highly localized in the Australian region where it has largest amplitude. Figure 10a shows the lower level streamfunction while Fig. 10b shows the upper level divergence (which is proportional to the 500 mb z-vertical velocity). The vertical structure of the disturbance consists largely of the first internal mode with the upper level streamfunction having essentially the opposite sign (in a given location) but being about 50% larger in magnitude and with some slight tilt in the vertical (see Fig. 5a of Frederiksen and Frederiksen, 1993a).

With the sign of the perturbation as shown in Fig. 10, it would tend to intensify the monsoon that occurred in the north Australian region during January 1979 and move it slightly poleward. With the opposite sign, it would tend to switch off the monsoon on the east coast of Australia.

The structure of this mode changes remarkably little with the different strengths of the cumulus heating, except that without moisture additional centres appear over Antarctica and in the central North Pacific. Thus moisture focuses the mode in the Australian region. The e-folding time of the mode varies from 15.2 days without moisture to 5.5 days for the moist A basic state and depends quite sensitively on the strength of the cumulus heating.

The monsoon disturbance is quite unusual compared with the type of structures previously found in studies of the instability of three dimensional flows, in that it is a stationary localized mode with quite small scale and with a large vertical shear. It appears that both the horizontal and vertical shears (and to some extent moisture) are very important in determining the structures, periods and growth rates of stationary monsoon disturbances.

(b) Intraseasonal oscillations

Frederiksen and Frederiksen (1993a) also found a group of modes with periods between 20 and 60 days, including modes which have properties similar to intraseasonal oscillations of the type originally discovered by Madden and Julian (1971, 1972). In the tropical regions, the observed oscillations have a structure which is dominated by the first internal mode (Knutson and Weickmann, 1987, Fig. 4; Rui and Wang, 1990, Fig. 8; Hendon and Liebmann, 1990, Fig. 5 and references therein).

The fastest growing intraseasonal oscillation mode for
moist A basic state is mode 58 which has a period of 28.1 days and an e-folding time of 8.7 days. Figs. 11a - c show the vertically averaged streamfunction, the vertical shear streamfunction and the upper level velocity potential for this mode at a particular phase (0°). The mean streamfunction in Fig. 11(a) shows wave trains emanating from the Southeast Asian/Indian Ocean region into the Pacific/North American region as well blocking in the North Atlantic and large amplitude in the Eurasian region. The Pacific-North American pattern is very evident at a phase of 90° or about 7 days later (not shown) as is the blocking over Greenland which occurred during late January 1979. The Northern Hemisphere extra-tropical structure of this mode has some of the qualitative features found in the observations of Knutson and Weickmann (1987), Hendon and Liebmann (1990) and Schubert and Park (1991).

The upper level velocity potential in Fig. 11(c) has large negative values, corresponding to rising motion, in particular in the Australian region. A spectral analysis indicates substantial zonal wavenumber 1 and wavenumber 2 components. The velocity potential shows more small scale features than the ensemble averaged observations (Lorenc, 1984; Knutson and Weickmann, 1987; Schubert and Park, 1991) or the slower growing intraseasonal oscillation instability mode discussed below. Short-time averaged or single realizations of the observed intraseasonal oscillation have, however, more small scale features, as is evident in the ten day averaged results in Fig. 4 of Knutson and Weickmann (1987).

Another essential feature of the observed intraseasonal oscillation is the fact that in the equatorial regions it appears to consist, to a large extent, of the first internal mode viz., having a large contribution in the shear streamfunction and shear zonal velocity and in the 500 mb vertical velocity. We note that the shear streamfunction has large magnitudes in the equatorial regions. In the extra-tropical regions the
Fig. 11 As in Fig. 10 for (a) the vertical mean streamfunction, (b) the vertical shear streamfunction and (c) the upper-level velocity potential for mode 58.

mode is more equivalent barotropic with large magnitudes in the winter hemisphere, again in agreement with the observed intraseasonal oscillation (Knutson and Weickmann, 1987; Hendon and Liebmann, 1990; Schubert and Park, 1991). The largest magnitudes of the shear streamfunction in Fig. 11(b) occurs in the Australian/South Pacific region where the shear zonal velocity is also large. At this particular phase the disturbance would strengthen the monsoon circulation, particularly in North-east Australia. Of course, at a phase of 180° (about 14 days later) the mode will have the opposite
sign corresponding to an inactive, or break, period of the Australian monsoon.

Frederiksen and Frederiksen (1993a) find that with no cumulus heating parameterization there is a mode which is very similar to mode 58 for the moist A case as far as the extra-tropical regions are concerned. However, it has virtually no amplitude of the shear streamfunction in the tropical regions. Thus, a cumulus heating parameterization is essential for generating the right tropical response for the intraseasonal oscillation.

Another intraseasonal oscillation mode examined in detail by Frederiksen and Frederiksen (1993a) is their moist A mode 91 with a period of 48.3 days and an e-folding time of 15.2 days. It thus lies within the 40-50 day range originally attributed to the intraseasonal oscillation found by Madden and Julian (1971, 1972). Figures 12a and b show, for this mode, the shear streamfunction and the upper level velocity potential at a particular phase (0°).

The mean streamfunction (Fig. 7a of Frederiksen and Frederiksen 1993a) has clearly defined Pacific-North American and Eurasian patterns. The velocity potential at all phases has a wavenumber one envelope within which are embedded smaller scale features. It has some similarity to that of the velocity potential anomaly associated with the first zonal wind EOF in Fig. 4(a) of Schubert and Park (1991), but with the additional smaller scale features that might be expected in short-time averaged or single realizations of the intraseasonal oscillation. Again, the velocity potential and divergence show eastward propagation with time or increasing phase.

At a phase of 90°, this intraseasonal mode interacts strongly with the Australian monsoon (Fig. 7g of Frederiksen and Frederiksen, 1993a). Like moist A mode 58, it can tend to switch off and on the Australian monsoon during its time evolution.

7. STRATOSPHERIC DISTURBANCES

The instability characteristics of stratospheric flows with
zonal variations have been examined by Frederiksen (1982) using a multi-level quasigeostrophic model and by Manney et al. (1989) and Manney and Nathan (1990) using a barotropic model. Frederiksen used basic-state distorted polar night jets, typical of the Northern Hemisphere winter stratosphere and mesosphere, obtained from Koermer et al. (1983) numerical simulations of sudden warmings. Their instability was studied just prior to the onset of the model sudden warming. The zonally averaged polar vortex was found to be stable. However the distorted vortices studied were unstable to three dimensional disturbances dominated by zonal wavenumber 2 and with doubling times of between about 5 and 10 days. The results indicated that wave instability may be an important contributing factor in the sudden warming. Frederiksen and Bell (1987) in their study of the instability of the three-dimensional Northern Hemisphere climatological flow for January 1978 also found a mode (shown in their Fig. 11) which would initiate a stratospheric sudden warming, as well as block formation in the troposphere, should it reach sufficient amplitude in a given synoptic situation.

Manney et al. (1989) and Manney and Nathan (1990) examined the barotropic instability of idealized stratospheric flows consisting of a zonal jet and a single planetary wave. They related some of their disturbance eigenmodes to the two-day zonal wavenumber 3 wave in the upper stratosphere and mesosphere.

8. NON-NORMAL MODES, ADJOINT MODES AND OPTIMAL PERTURBATIONS

As discussed in section 4a(iii), within the linear context with constant basic state, the adjoint mode provides the most efficient way of obtaining largest amplitude of the growing eigenmode as \( t \to \infty \). However as pointed out by a number of authors (e.g., Lorenz, 1965; Leith, 1979; Farrell, 1984; Rheinboldt, 1986; Lacarra and Talagrand, 1988; Frederiksen and Bell, 1990; Borges and Hartmann, 1992) depending on the norm chosen, there are other (single) non-normal modes which may give the largest amplification initially or for a short time before nonlinear effects become important. For example, in the absence of \( f \), the tendency of the \( L_2 \) norm squared can be written from Eq. (2.6) in the form

\[
\frac{\partial}{\partial t} \| \Psi \|^2 = \langle \Psi, (D^* + [D]^*) \Psi \rangle \tag{8.1}
\]

where \([D] = -i[A]^{-1}[B]\), superscript + denotes adjoint and the inner product is defined by Eq. (2.11a). Thus the perturbation which maximises \( (\partial/\partial t) \| \Psi \|^2 \) at \( t = 0 \) is the eigenvector of \([D] + [D]^*\) corresponding to the largest eigenvalue \( \lambda \) of this matrix. In fact \( \lambda \) may be considerably larger than \( 2\phi \) for the fastest-growing eigenmode. For example, with their basic state for 21 January 1979, Frederiksen and Bell (1990) find the optimal perturbation has \( \lambda = 5.5 \) \( \text{d}^{-1} \) compared with \( \lambda = 2\phi = 1.22 \) \( \text{d}^{-1} \) for the eigenmode within a five-level quasigeostrophic model. The 500 mb structure of the optimal perturbation is shown in Fig. 11 of Frederiksen and Bell (1990). It is also strongly localized in the blocking region over southern Greenland.

Frederiksen and Bell (1990) examined the evolution of the fastest-growing normal mode on 21 January using the 500mb basic state within a linearized barotropic model and compared it with the corresponding evolution of the fastest-growing adjoint mode and with the optimal perturbation for which the amplitude growth at \( t = 0 \), as measured by the \( L_2 \) norm, is the largest. In all cases the structures were normalized to having the same \( L_2 \) norm initially. During the 10-day integration, the normal mode of course does not change its structure (since \( \partial \phi / \partial t = 0 \)) but increases its amplitude with a growth rate of \( \alpha = 0.69 \) \( \text{d}^{-1} \). Its initial norm squared growth rate is just \( \lambda = 2\phi \). The adjoint mode however undergoes very rapid change in structure so that by day 1 it is very similar to the normal mode but with 2.05 times the amplitude, which increases to 5.6 times on day 10.

The optimal perturbation, with largest initial norm squared growth factor, has \( \lambda = 4.3 \) \( \text{d}^{-1} \) which is 3 times that of the eigenmode. However, this optimal perturbation does not maintain its initial rapid growth, but undergoes complex changes in its structure and it is not until about day 8 that it is very similar to the eigenmode. By day 10 it only has one half of the amplitude of the eigenmode. Table 1 summarizes the changes in the \( L_2 \) norm and kinetic energy during the integration. It is clear that during this period the adjoint mode provides the best method (of the three considered) of getting a large amplification of the eigenmode.

As well as the optimal perturbations as \( t \to \infty \) (the fastest growing adjoint mode) and at \( t = 0 \), it is possible to define optimal perturbations over a given finite time interval (Lacarra and Talagrand, 1988; Farrell, 1989; Borges and Hartmann, 1992; Farrell and Moore, 1992; Moore and Farrell, 1993). This then leads to whole families of disturbances depending on the optimizing time interval, some of which may exhibit features similar to those found

<table>
<thead>
<tr>
<th>Day</th>
<th>( A(A) )</th>
<th>( A(K) )</th>
<th>( A(D) )</th>
<th>( A(K) )</th>
<th>( A(V) )</th>
<th>( A(\lambda) )</th>
<th>( A(\lambda) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.01</td>
<td>4.04</td>
<td>4.11</td>
<td>6.34</td>
<td>3.67</td>
<td>5.55</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.03</td>
<td>16.27</td>
<td>12.87</td>
<td>44.51</td>
<td>6.47</td>
<td>14.33</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>( 1.07 \times 10^5 )</td>
<td>( 1.14 \times 10^6 )</td>
<td>( 6.02 \times 10^6 )</td>
<td>( 8.78 \times 10^7 )</td>
<td>( 5.52 \times 10^5 )</td>
<td>( 1.32 \times 10^6 )</td>
<td></td>
</tr>
</tbody>
</table>
in particular synoptic situations. Unlike normal modes which are uniquely determined by the basic state flow and the dissipation parameterizations, the optimal perturbations depend both on the choice of norm (e.g., $L_2$ norm, kinetic energy norm, enstrophy norm etc.) and on the optimizing time interval. As well as this arbitrariness, there is the problem of a mechanism of selecting the disturbance.

The fastest growing eigenmode will always be selected asymptotically within a linear model irrespective of the initial conditions (including from random noise). In contrast, optimal perturbations must be prepared (by some other instability process or nonlinear effect etc.) to have their initial specific structures. In the case of adjoint modes, we noted in section 4a(iii) that the adjoint mode 1 on 20 January 1979 is very similar to a mature cyclone scale mode and the adjoint mode 1 on 21 January has some of the gross features of the onset-of-blocking mode 1 on 20 January. These are then examples of a previous nonlinear baroclinic instability process and a previous baroclinic-barotropic instability process approximately preparing adjoint modes.

In general, to convincingly establish the relevance of a particular optimal perturbation in a given synoptic situation, it would seem to be essential to elucidate the mechanisms for selecting the disturbance. The probability of it appearing at random, particularly given the above arbitrariness with respect to norm and optimizing time interval, must be exceedingly small.

**SUMMARY AND CONCLUSIONS**

This review has considered the progress over the last decade in the application of normal mode instability theory with three-dimensionally varying basic states to understanding the mechanisms of formation of large-scale atmospheric disturbances. Our survey has included regional and local cyclogenesis, blocking in both hemispheres, low stratospheric disturbances of South Brazil and Australian northwest cloud band disturbances. It has been argued that normal mode instability theory, generalized to observed three-dimensionally varying basic states and possibly including a cumulus heating parameterization, is a natural starting point for understanding the dynamical generation mechanisms of the above disturbances. Both physically and mathematically it provides a natural generalization and union of the zonally averaged instability theory of Charney (1947) and Eady (1949) and the Rossby wave dispersion theory of Rossby (1949) and Yeh (1949), as discussed in detail by Frederiksen and Webster (1988).

One would of course seldomly expect to find an atmospheric disturbance of pure normal mode form. Indeed we have surveyed examples of the relevance of non-normal mode disturbances (consisting of superpositions of normal modes) and discussed studies emphasizing the roles of nonlinear effects in the developments of extra-tropical cyclones and low frequency fluctuations. Nevertheless, normal mode instability theory does provide a very useful guide to the zoo of atmospheric disturbances with presumably many more exhibits yet to be classified.

**REFERENCES**
