On the Dynamics of Planetary Flow Regimes. Part II: Results from a Hierarchy of Orographically Forced Models

FRANCO MOLTENI*

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom and Space and Atmospheric Physics Group, Imperial College, London, United Kingdom

(Manuscript received 7 November 1994, in final form 23 August 1995)

ABSTRACT

The relationship between steady states of the large-scale flow regimes revealed by multimodality in phase space and quasi-resonant axes of a linearized atmospheric model (neutral vectors) is investigated by means of a hierarchy of three hemispheric quasigeostrophic (QG) models, in which the flow is relaxed toward a zonally symmetric equilibrium state and orography provides the only source of asymmetric forcing. Two of these models describe only the interaction between the mean zonal flow and planetary waves of zonal wavenumber 3, including one or two meridional wavenumbers. The third model also has a zonal wavenumber-3 symmetry, but includes waves up to total wavenumber 21, and therefore is able to properly represent the interactions between planetary waves and baroclinically unstable synoptic-scale waves.

Three stationary solutions are found for the highly truncated models, two of which represent states with large wave amplitude but opposite phase. These two steady states produce two well-defined flow regimes in the higher-resolution hemispheric model. This result confirms that flow regimes found in highly truncated models are not merely an artifact of the excessive truncation and can still be found when a sufficiently large number of degrees of freedom is included. The agreement between large-scale steady states and regime centroids is improved if the average effect of high-frequency transients on the planetary-scale flow is parameterized by adjusting the sources of zonal available potential energy and kinetic energy in the highly truncated models.

Using these hemispheric models, it is shown that the leading dynamical and generalized neutral vectors correctly identify the axis linking the centroids of the two regimes as the axis with the smallest linear time derivative. Although anomalies in transient eddies are important in determining the spatial pattern of these centroids, the ultimate source of the two regimes of the hemispheric QG model is the existence of multiple stationary solutions of the large-scale flow. There is no inconsistency between the energetically fundamental role played by high-frequency transients and the capacity of nonlinear large-scale dynamics to determine the multimodal structure of the model's phase space. The neutral vector analysis also confirms that tropical diabatic heating could effectively generate an extratropical response along an axis corresponding to the difference between opposite regime anomalies.

1. Introduction

Thirteen years ago, Hoskins (1983) discussed the relationship between theoretical, observational, and modeling studies in meteorology and concluded that generally an "unhealthy" situation existed, in which aspects of the observed circulation were reproduced in complex numerical models and interpreted in terms of static conceptual models, but there was almost no interaction between theoretical models based on simple dynamical principles and the numerical models whose output could be directly compared with observations.

In an ideal situation, conceptual models used to interpret observations should be the starting point for a hierarchy of dynamical models of increasing complexity, in which the complex numerical models should just represent the other extreme of the spectrum.

Until the mid-1980s, research on planetary flow regimes was suffering from just these sorts of problems. On the theoretical and modeling side, although highly truncated models of the planetary-scale flow with multiple steady states (e.g., Charney and DeVore 1979; Charney and Strauss 1980, CS hereafter) and regimelike behavior (Reindl and Pierrehumbert 1982, RP hereafter) had been developed, there was no proof that these regimes could be reproduced in "intermediate resolution" models that allowed a proper energetic interaction between planetary waves and synoptic-scale, baroclinically unstable eddies (e.g., Cehelsky and Tung 1987, CT hereafter).

On the observational side, evidence of bimodality in the amplitude of planetary waves had emerged (Sutera
midlatitude diabatic heating has been proposed by Mitchell and Derome (1983) and Marshall and So (1990), and one cannot neglect the possibility that the forcing generated by the asymmetric distribution of the main tropical convection areas may also play a catalytic role for the existence of midlatitude regimes (Hoskins and Sardeshmukh 1987; Ferranti et al. 1994a).

In the QG model used in MM93 and Part I, orographic forcing is represented through the inclusion of the term $fh/H_0$ in the potential vorticity (PV) at the lowest level, where $f$ is planetary vorticity, $h$ is the real orographic height, and $H_0$ is the depth of the lowest model layer multiplied by the ratio between the 800-hPa wind speed and the near-surface wind speed (used to compute the vertical velocity at the lower boundary). If this ratio is assumed to be between 2 and 4, then $H_0$ should vary between 5 and 10 km. In MM93, a value of $H_0 = 9$ km was chosen, for reasons which will be discussed in section 3b.

In addition to the orographic term, a further source of forcing for the planetary waves comes from the zonally asymmetric component of the PV forcing terms included in the prognostic equations of the QG model. As described in section 6 of MM93, they have been computed empirically as residual tendencies in order to maintain the model climatology close to the observed one. These terms can be thought of as representative of diabatic heating (in conjunction with the temperature relaxation term), time-averaged PV advection by the tropical divergent circulation, and other physical processes not represented in the QG model. It should be pointed out that they can also compensate for an underestimation or an overestimation of orographic forcing, which adds some uncertainty in the choice of the optimal value for $H_0$.

The disadvantage of computing the PV forcing terms as residual tendencies is that one cannot isolate the contribution of different physical processes; therefore, the role of these processes in the maintenance of the regimes is unclear. This is unfortunately the price one pays in order to obtain a realistic climatology in a relatively simple QG model. However, the confidence in the existence of planetary-scale regimes would increase if one were able to recognize common features between regimes generated by a comprehensive but empirically derived forcing and regimes generated (in a necessarily more idealized environment) by one single physical process.

The second problem that needs further investigation is the relationship between generalized neutral vectors and quasi-stationary solutions of the equations of motions. In MM93, the Lorenz (1963) convection model was used to illustrate the relationship between neutral vectors and stationary states. It was shown how multimodality in probability density functions (PDFs) could reveal the existence of regimes aligned along such vectors; this approach has proved to be effective for neutral vectors computed from the observed and the

1986; Hansen and Sutera 1986), which could be interpreted on the basis of simple theories of multiple equilibria (e.g., Benzi et al. 1986a). However, in the absence of supporting results from more realistic numerical models, the debate on the relevance of the observed bimodality was overshadowed by questions on its statistical significance. [This dispute is by no means settled: see Nitsche et al. (1994).]
QG model climatology. In the case of the generalized neutral vectors defined in Part I, the hierarchical clustering process from which they are computed does not necessarily guarantee the existence of multimodality. Can we prove that, in the case of a simplified system whose quasi-stationary states can be identified independently, the generalized neutral vectors with the longest characteristic times (i.e., those with the smallest residual tendency) do represent the quasi-stationary anomalies?

One can address both these problems by constructing, along the lines advocated by Hoskins (1983), a hierarchy of models that can be seen as progressive simplifications of the global QG model of MM93, but with just one physical process as the source of zonally asymmetric forcing. For historical reasons, in order to shed light on the controversies arisen on the interpretation of the results of CS, RP, and CT (see also comments by Reinhold 1989 and Chelovsky and Tung 1989), we have chosen to study models where the planetary wave forcing is only due to orography.

The three models that we shall consider here have the same vertical discretization and are governed by the same prognostic equations as the global QG model. The only differences are that (a) a spatially uniform drag coefficient has been used in the Ekman dissipation term and (b) the PV forcing terms have only a zonal-mean component and are implicitly defined through an equilibrium streamfunction. However, these models represent only a hemispheric flow with no cross-equatorial wind and with zonal wavenumber-3 symmetry. The first two models just reproduce the interactions between the mean zonal flow and planetary waves of zonal wavenumber 3 (with one or two meridional wavenumbers) and can be seen as the counterparts (in spherical geometry and with three vertical levels) of the beta-channel, 2-level truncated models analyzed by CS. The third model includes all the (meridionally antisymmetric) waves with zonal wavenumber multiple of 3 and total wavenumber \( \leq 21 \). It is an intermediate-truncation model, like the beta-channel model of CT, in which interactions between mean zonal flow, planetary waves, and baroclinically unstable synoptic waves are properly represented.

It should be pointed out that, given the hemispheric domain of these model and their symmetric properties in the zonal direction, the existence of multiple stationary solutions associated with resonant behavior of the planetary waves is obviously favored with respect to the more realistic conditions described by the global QG model. Therefore, the existence of regimes in our intermediate-truncation model cannot be taken as a proof that regimes exist in the real atmosphere. However, in MM93 and Part I we showed that regimes do exist in the global QG model, and we can use the truncated models as interpretative tools that highlight some of the physical mechanisms that are likely to generate regimes in more complex models and in the real atmosphere itself.

The fact that we view the truncated models as idealizations of a more detailed one is, in contrast with the historical development of multiple regime theory, in which the severely truncated models used for the earlier studies were made more “realistic” by the addition of more degrees of freedom in subsequent investigations. The latter approach, however, has led to the paradoxical situation that while results derived from severely truncated models had at least a qualitative agreement with the observational evidence (see Benzi et al. 1986a), results from more complex models appeared to contradict the validity of the previous theoretical results and question the significance of the observational finding about multimodality in the atmosphere (Tung and Rosenthal 1985; CT).

As pointed out by Reinhold (1989), this inconsistency may be attributed to the sensitivity of the behavior of very simple models to the dissipative and forcing parameters. Values of these parameters that are appropriate for severe truncations may not be appropriate when more degrees of freedom are introduced, because of the different energetics that is generated by better-resolved baroclinic waves. It is certainly true that in simplified models a careful tuning of the dissipative and forcing terms is needed to distribute energy between the available degrees of freedom in a realistic way. On the other hand, if the choice of these parameter is arbitrary to a large extent, and a very different dynamical behavior is obtained within the acceptable parameter range, one is still left in doubt about the relevance of such results for the real atmosphere.

Our approach greatly alleviates these problems. In the global QG model of MM93, the dissipative parameters satisfy both requirements of having physically reasonable values and of producing a model climatology comparable to the observed one. These parameters have been fixed in the three truncated models that we are going to analyze. In addition, the spherical geometry of our model allows a direct comparison between the zonal-mean states toward which the models are forced and the real wintertime zonal-mean flow in the Northern Hemisphere. We shall demonstrate that, with a realistic set of dissipative parameters, a remarkable consistency is achieved between two stationary solutions of the severely truncated models and two clearly defined regimes in the intermediate-truncation model. Such consistency is achieved even when the same forcing parameters are used but is enhanced if the average effects of baroclinic eddies are “parameterized” in the truncated planetary wave models by decreasing the sources of available potential energy and increasing the kinetic energy source for the zonal-mean wind (consistently with the structure of the average high-frequency transient forcing discussed in Part I).

Finally, we shall demonstrate that the regimes of the intermediate-resolution model are effectively described
by the leading "dynamical" (i.e., nongeneralized) and
generalized neutral vectors. Therefore, we shall prove
that there is no inconsistency between the ability of
linear theory to reproduce prominent low-frequency
variability patterns as linear responses to anomalous
forcings and the hypothesis that such patterns are ac-
tually axes linking multiple quasi-stationary states gen-
erated by internal nonlinear dynamics.

2. Prognostic equations and truncation of the
simplified quasigeostrophic models

The governing equations for the hierarchy of sim-
plified QG models that we want to develop are the same
as the prognostic equations of the 3-level, global QG
model described in MM93 and Part I, with some mod-
ifications to the dissipative and forcing terms. If we
assume a constant equilibrium state \( (\psi_1^*, \psi_2^*, \psi_3^*) \) for
the streamfunction at the three model levels of 200,
500, and 800 hPa, and define

\[
\begin{align*}
\Delta \psi_1 &= \psi_1 - \psi_2, \quad \Delta \psi_2 = \psi_2 - \psi_3, \\
\Delta \psi_1^* &= \psi_1^* - \psi_2^*, \quad \Delta \psi_2^* = \psi_2^* - \psi_3^*,
\end{align*}
\]

these equations can be written as

\[
\begin{align*}
\frac{\partial q_1}{\partial t} &= -J(\psi_1, q_1) + \tau_{\kappa}^{-1} R_1^{-2} (\Delta \psi_1 - \Delta \psi_1^*) + H D_1, \\
\frac{\partial q_2}{\partial t} &= -J(\psi_2, q_2) - \tau_{\kappa}^{-1} R_1^{-2} (\Delta \psi_1 - \Delta \psi_1^*) \\
&\quad + \tau_{\kappa}^{-1} R_2^{-2} (\Delta \psi_2 - \Delta \psi_2^*) + H D_2, \\
\frac{\partial q_3}{\partial t} &= -J(\psi_3, q_3) - \tau_{\kappa}^{-1} R_2^{-2} (\Delta \psi_2 - \Delta \psi_2^*) \\
&\quad - \tau_{\kappa}^{-1} \nabla^2 (\psi_3 - \psi_3^*) + H D_3,
\end{align*}
\]

where \( J \) and \( \nabla^2 \) are respectively the Jacobian and the
Laplacian of horizontal fields, \( q_i \) is PV at the \( i \)th pres-
sure level, \( R_1 \) and \( R_2 \) are Rossby radii of deformation,
\( \tau_{\kappa} \) and \( \tau_{\kappa} \) are dissipative timescales for temperature
and low-level vorticity, and the \( H D_i \) are horizontal diffusion
terms that act on the time-dependent part of PV (see section 3a for more details). Potential vorticity is defined as

\[
\begin{align*}
q_1 &= \nabla^2 \psi_1 - R_1^{-2} \Delta \psi_1 + f, \\
q_2 &= \nabla^2 \psi_2 + R_1^{-2} \Delta \psi_1 - R_2^{-2} \Delta \psi_2 + f, \\
q_3 &= \nabla^2 \psi_3 + R_2^{-2} \Delta \psi_2 + f(1 + h/H_0),
\end{align*}
\]

where the \( R_1 \) and \( R_2 \) have the same values (700 and
450 km, respectively) as in the global QG model.

Equations (1a–c) differ from Eqs. (A1a–c) of Part I
only by the assumption of a spatially uniform coeffi-
cient \( \tau_{\kappa}^{-1} \) for Ekman dissipation and because the PV
forcing terms are expressed as a function of the equi-
librium state \( \psi^* \). In the following, we shall assume \( \psi^* \)
to be zonally symmetric.

We shall limit the spatial domain of the QG model
to one hemisphere by expanding the streamfunction
and PV field onto spherical harmonics that are antisym-
metric with respect to the equator. As a further simpli-
ication, the eddy part of the flow will be expanded only
on harmonics of zonal wavenumber 3. If a truncation
at total wavenumber 21 is used as in the global QG
model, each single-level field will be described by 77
real degrees of freedom (DOF). In total, this interme-
diate-resolution three-level model will have 231 DOFs,
and we shall refer to it as to the QG231 model. This
model has enough DOFs to properly represent the in-
teractions between zonal flow, planetary waves and
baroclinically unstable, synoptic-scale waves, and to
generate realistic spectra of energy and enstrophy.

The QG231 model will be used here in the same way
as the global QG model was used in MM93. Namely,
we shall perform long (80 winter) integrations of this
model, and we shall look for regimes in the model
phase space using probability density functions (PDFs)
of model states. The three-fold symmetry in the zonal
direction implies that these regimes, if they exist, can
be easily identified by looking at the PDFs of the wave-
number-3 components of the eddy fields in midlati-
dudes. We shall also investigate whether the leading
dynamical and generalized neutral vectors computed
from model data correspond well to the axes linking
the phase-space clusters associated with the regimes.

If regimes with different phase or amplitude of plan-
tary waves exists in the QG231 phase space, the ques-
tion arises of whether they are related in some way to
the existence of multiple stationary solutions of a dy-
namical system that describe only the interactions be-
tween mean zonal flow and planetary waves. To inves-
tigate this problem, we have built a highly truncated
model based on Eq. (1), but in which only the lowest
two meridional wavenumbers of the mean zonal flow
and the zonal wavenumber-3 wave are retained. This
truncation includes a total of six real, orthonormal har-
monic functions:

\[
\begin{align*}
\Phi_1(\mu) &= P_{0,1}(\mu), \\
\Phi_2(\mu) &= P_{0,3}(\mu), \\
\Phi_3(\lambda, \mu) &= P_{3,4}(\mu) \cdot \sqrt{2} \cos(3\lambda), \\
\Phi_4(\lambda, \mu) &= -P_{3,4}(\mu) \cdot \sqrt{2} \sin(3\lambda), \\
\Phi_5(\lambda, \mu) &= P_{5,6}(\mu) \cdot \sqrt{2} \cos(3\lambda), \\
\Phi_6(\lambda, \mu) &= -P_{5,6}(\mu) \cdot \sqrt{2} \sin(3\lambda),
\end{align*}
\]

where \( \lambda \) is longitude, \( \mu \) is the sine of latitude, and \( P_{m,n} \)
is the associated Legendre function of zonal wavenum-
ber \( m \) and total wavenumber \( n \). (Of these functions,
\( P_{0,1} \) and \( P_{3,4} \) do not change sign between the pole and
the equator, while \( P_{0,3} \) and \( P_{5,6} \) change sign once).
The model obtained by projecting the streamfunction field at each model level onto these six functions has a total of 18 DOFs and will be indicated as QG18. Due to the presence of just one zonal wavenumber, only the following three types of nonlinear interactions occur in the QG18 model: interactions between zonal-mean and eddy components, which produce propagation of the wave pattern in the east–west direction; interactions between wave functions with the same meridional wavenumber but different phase in longitude, which affect both the $\Phi_1$ and the $\Phi_2$ zonal-mean components; and interactions between wave functions with different meridional wavenumber and different phase in longitude, which affect only the $\Phi_2$ zonal-mean component.

If Eqs. (1) are linearized around a zonally symmetric basic state that is representative of the wintertime climatology in the Northern Hemisphere, one can find that waves with total wavenumber $n = 6–7$ are close to resonance. Therefore, in QG18 the response to orographic forcing should be dominated by the $\Phi_2/\Phi_6$ waves. The resonant behavior of this wave can be reproduced by considering its interaction with the superrotational component $\Phi_6$ only. Therefore, as a maximum simplification of our 3-level QG model, we shall also consider a 9-DOF model (QG9 hereafter) in which the streamfunction at each level is expressed only as a linear combination of $\Phi_1$, $\Phi_5$, and $\Phi_6$.

As mentioned above, the QG9 and QG18 models can be seen as the counterparts (in spherical geometry and one additional vertical level) of the beta-channel baroclinic models with one or two meridional modes investigated by CS. However, one important difference between the two cases must be pointed out. In beta-channel geometry, it is possible to construct a system based on functions with meridional wavenumber 1, which is an exact subsystem of a model with meridional wavenumber 1 and 2. This means that the stationary solutions of the meridional wavenumber-1 model will also be stationary solutions of the meridional wavenumber-1 and -2 model if the coefficients of the remaining functions are set to zero. Only the stability properties will change when the two different sets of functions are considered. For hemispheric models like QG9 and QG18, no such property exists; consequently, the stationary states of the QG18 model are not equal to those of QG9 and they must be found independently.

It is not our intention here to perform a detailed analysis of the stationary states of the QG9 and QG18 models as a function of the forcing parameters. We shall simply discuss the stationary states of the two models and their stability to large-scale perturbations, for a few realistic combinations of the parameters, and show how simple energetic arguments can be used to understand the differences between the results obtained with the two truncated models and with the intermediate-resolution QG231 model. Our choice for the orography and the equilibrium state of the three models is discussed in the following section.

3. Dissipative and forcing terms
   a. Dissipative processes

In the QG231, QG18, and QG9 models, the temperature relaxation timescale $\tau_T$ has the same value (25 days) used in the global QG model. In the Ekman dissipation term, the increase in the drag coefficient over land surface and orography has been neglected, since this would have acted as a further eddy-generating term in addition to orographic forcing. So, the simplified QG models use a spatially uniform drag coefficient, which is the inverse of the dissipative timescale $\tau_D = 3$ days.

The horizontal diffusion term also has the same formulation as in the global QG model. This term is proportional to $\nabla^2 q'$, where $q'$ is the time-dependent component of PV. In the global QG model, the diffusion coefficient was chosen in such a way that harmonics with $n = 21$ were damped on a 2-day timescale. In the QG231 model, this timescale has been reduced to 1 day by doubling the diffusion coefficient. As far as the QG9 and QG18 models are concerned, the horizontal diffusion terms are extremely small at the scales resolved by these truncated models, and therefore they have been neglected.

In QG231, horizontal diffusion was increased for the following reason. As documented in a number of previous studies with simplified models (e.g., O'Brien and Branscome 1990), models in which the forcing is acting on all the resolved wavelengths have a "redder" energy spectrum than models in which the forcing is only acting on a few large-scale components. Although (as discussed below) the idealized orography of QG231 projects on more than one zonal wavenumber, there is much less forcing on the synoptic scales in this model than in the global QG model. The increase in the diffusion coefficient guarantees that a realistic energy spectrum is obtained in QG231, therefore avoiding the possibility (discussed by CT) that a regimelike behavior is induced in the model by accumulation of energy in the synoptic-scale waves.

   b. Orography

Let us now consider the orographic forcing terms in the simplified QG models. Given the idealized nature of these models, we have assumed a simple spectral representation for $h_\mu$ (the product of the orographic height by the sine of latitude), which has no zonal-mean component and simulates (in QG231) three meridionally oriented ridges with an east–west extension comparable to that of the Rocky Mountains. More precisely, the orographic term in QG231 projects onto waves with zonal wavenumber $m = 3, 6$, and $9$ and meridional wavenumber $n - m = 1$ and 3:
\((h\mu)_{QG231} = H_3 [P_{3,4}(\mu) + P_{3,6}(\mu)] \sqrt{2} \cos(3\lambda) + H_3 [P_{5,3}(\mu) + P_{5,9}(\mu)] \sqrt{2} \cos(6\lambda) + H_3 [P_{9,10}(\mu) + P_{9,12}(\mu)] \sqrt{2} \cos(9\lambda)\), \hspace{1cm} (4)

where again the \(P_{m,n}\) are the associated Legendre functions. Choosing the same amplitude for the two meridional modes of each zonal wavenumber reduces the orography in the tropical regions and enhances it in midlatitudes, while assuming \(H_3 = 2H_0 = 3H_0\) produces a localization of the mountains in the east–west direction. Finally, the value of \(H_3\) is chosen in such a way that the difference between the highest and lowest point of the QG231 orography is about 1000 m. The pattern of \(h\mu\) for the QG231 model is shown in Fig. 1a.

The orographies of QG18 and QG9 are just the projections of the QG231 orography onto the eddy functions represented in the two models. Using the definitions given in Eq. (3), we can write

\[(h\mu)_{QG18} = H_3 [P_{3,4}(\mu) + P_{3,6}(\mu)] \sqrt{2} \cos(3\lambda) = H_3 [\Phi_3(\lambda, \mu) + \Phi_5(\lambda, \mu)] \hspace{1cm} (5)\]

\[(h\mu)_{QG9} = H_3 P_{5,6}(\mu) \sqrt{2} \cos(3\lambda) = H_3 \Phi_9(\lambda, \mu). \hspace{1cm} (6)\]

These orographies are shown in Figs. 1b and 1c, respectively.

In order to be converted into a PV term, the orographies given by Eqs. (4) – (6) are normalized by a height parameter \(H_0\). As discussed in the introduction, the choice of the optimal value for \(H_0\) in the global QG model was complicated by the fact that the time-independent PV sources were able to compensate for some local underestimation or overestimation of orographic forcing. On the basis of preliminary experiments used to select appropriate values of the parameters, it was felt that a value of \(H_0\) of the order of 5 km was appropriate for mountain ranges with maximum altitude (at T21 resolution) of 2 km or less but produced an excessive forcing in regions with higher topography, especially if located at relatively low latitudes (like the Himalayan region). For this reason, a “conservative” value of \(H_0 = 9\) km was chosen in the global model.

In the simplified QG models described here, the orography does not exceed 1 km and is confined to middle and high latitudes. Therefore, a value of \(H_0 = 5\) km has been adopted, which corresponds to the assumption that the near-surface wind used to compute the lower boundary vertical velocity is one-half of the 800-hPa wind. It should be pointed out that, even with this value of \(H_0\), the orographic forcing in our models is smaller than in most idealized models used in previous studies on the influences of large-scale orography.

c. Forcing of the zonal-mean flow

Finally, we shall discuss the definition of the zonally symmetric equilibrium states toward which the zonal-mean flow of the models is relaxed. In order to maintain a realistic mean zonal flow in the lowest model level, a forcing of the 800-hPa zonal wind toward nonzero
values is used here in addition to the baroclinic forcing of the streamfunction thickness $\Delta \psi_1$ and $\Delta \psi_2$. In order to set appropriate values for $\Delta \psi_1^*$, $\Delta \psi_2^*$, and $\psi_3^*$, the corresponding climatological values for January–February have been computed by projecting the Northern Hemisphere time- and zonally averaged streamfunction at the three model levels on the meridionally antisymmetric spherical harmonics. As in MM93, our observed streamfunction “climatologies” is computed as an average of European Centre for Medium-Range Weather Forecasts analyses for January and February 1984–89.

Let us consider the QG9 and QG18 models first, and let $\delta c_{1k}$, $\delta c_{2k}$, and $c_{3k}$ be the projections of the (Northern Hemisphere) climatological $\Delta \psi_1$, $\Delta \psi_2$, and $\psi_3$ on the zonal functions $\Phi_k$, for $k = 1, 2$. In the QG9 model we have set

$$\Delta \psi_1^* = \alpha_1 \delta c_{11} \Phi_1$$  \hspace{1cm} (7a)

$$\Delta \psi_2^* = \alpha_2 \delta c_{21} \Phi_1$$  \hspace{1cm} (7b)

$$\psi_3^* = \alpha_1 c_{31} \Phi_1.$$  \hspace{1cm} (7c)

In QG18 too, the baroclinic forcing depends (for simplicity) on one single coefficient of proportionality:

$$\Delta \psi_1^* = \alpha_r (\delta c_{11} \Phi_1 + \delta c_{12} \Phi_2)$$  \hspace{1cm} (8a)

$$\Delta \psi_2^* = \alpha_r (\delta c_{21} \Phi_1 + \delta c_{22} \Phi_2),$$  \hspace{1cm} (8b)

while for the barotropic component of the equilibrium state, two different coefficients have been assumed for the $\Phi_1$ and $\Phi_2$ components:

$$\psi_3^* = \alpha_1 c_{31} \Phi_1 + \alpha_2 c_{32} \Phi_2.$$  \hspace{1cm} (8c)

The reason for this choice is that, in the real atmosphere, the barotropic “forcing” of $\Phi_1$ and higher-order zonal harmonics is mainly due to momentum transport by transient eddies. By changing the value of $\alpha_2$ while leaving $\alpha_1$ fixed, we shall be able to get some information on the sensitivity of large-scale quasi-stationary states to this momentum transport. A distinction between the coefficients of $\Phi_1$ and $\Phi_2$ is less important for $\Delta \psi_1^*$ and $\Delta \psi_2^*$ since the sources of available potential energy are dominated by the superrotational component anyway; on the contrary, in the 800-hPa climatological flow, the amplitude of the $\Phi_2$ component is 50% larger than the $\Phi_1$ component.

In the QG231 model, the first four zonal harmonics have been used in the definition of the equilibrium state. Again, the baroclinic components $\Delta \psi_1^*$ and $\Delta \psi_2^*$ are given by the climatological values multiplied by one single coefficient $\alpha_r$. For $\psi_3^*$, the first two zonal harmonics are given by Eq. (8c) as in QG18; for the two higher-order harmonics, the climatological values are multiplied by the same proportionality coefficient $\alpha_2$ used for the second zonal harmonic.

4. Stationary states of the severely truncated models

In this section, we shall investigate the stationary solutions of the two severely truncated models that we have previously described. We shall start by studying the steady states of QG9, the simplest of our models, then we shall discuss how the stationary states can be made more realistic by the less severe truncation of the QG18 model.

Once the orography and the values of the dissipation parameters are fixed, the stationary states of QG9 and QG18 depend on the values of the forcing parameters $\alpha_r$, $\alpha_1$, and (for QG18 only) $\alpha_2$, which have been defined in Eqs. (7) and (8). The parameter $\alpha_r$ determines the source of zonal available potential energy (APE) for the two models, while $\alpha_1$ and $\alpha_2$ control the kinetic energy (KE) source for the zonal flow. In a number of previous studies using baroclinic beta-channel models (e.g., CS; RP; O’Brien and Branscome 1990), a KE source was not provided, and the properties of the model were usually studied as a function of parameters controlling the APE source. However, especially for highly truncated models, a KE source avoids the presence of stationary states with midlatitude easterly flow in the bottom level, which cannot find a counterpart in the regimes of the real atmosphere.

If $\alpha_1$ and $\alpha_2$ are fixed to realistic values, the QG9 and QG18 models possess only one stationary state when $\alpha_r$ is below a certain threshold. Above this bifurcation point, three stationary states exist: one of which is close to the equilibrium state, while the other two have a lower mean zonal wind and waves with large amplitude and very different phase. This behavior is qualitatively similar to that of the CS model. In contrast to CS, we shall not investigate a parameter range in which more than three stationary states may exist.

In the QG9 model, we set $\alpha_1 = 2$ and increased $\alpha_r$ in steps of 0.1 until (for $\alpha_r = 1.7$) three stationary solutions were found. Let us examine the results obtained with these values of the parameters. The eddy pattern of the 500-hPa streamfunction for the three stationary solutions are shown in Fig. 2. All the streamfunction fields and coefficients are in units of $10^6$ m$^2$ s$^{-1}$, which corresponds to about 1 dam of geopotential height in midlatitudes. The latitudinal averages between 30$^\circ$ and 60$^\circ$N of the mean zonal wind at the three model levels are listed in Table 1b for the QG9 zonal equilibrium and steady states, and can be compared with the corresponding values for the real climatology in Table 1a. As previously mentioned, the first stationary state is very similar to the equilibrium state, as far as the mean zonal flow is concerned, and has a weak wave pattern with ridges just about 10 degrees upstream of the orographic ridges. In the second stationary state, the zonal-mean wind in the lowest level is reduced by about 30% with respect to the equilibrium state, with progressively smaller reductions in the middle and upper levels. The phase of the wave
pattern is similar to that of the first solution, but its amplitude is more than three times larger. The third steady state has a zonal-mean flow slightly weaker than the second one. This small change is, however, sufficient to bring the flow from a superresonant to a subresonant state: the $\Phi_2$ component (on which the orography projects) has the opposite sign, so that the wave ridges are now almost completely out of phase with the mountain ridges, and the total wave amplitude is 10–20% larger than in the second steady state. For simplicity, following RP, hereafter we shall refer to the three stationary solutions as to the near-zonal, the ridge, and the trough, steady states, with reference to the wave phase above the orographic ridges.

Can the steady states of QG9 be considered as reasonable idealizations of the actual large-scale atmospheric flow? As far as the amplitude of the planetary waves is concerned, the values of the ridge and trough steady states are large when compared with the time-averaged climatological waves but are comparable to planetary wave amplitudes observed in some strongly anomalous situations. When the mean zonal flow is
considered, since in QG9 this is described by just one spectral coefficient, it is impossible to obtain realistic values at different latitudes. Table 1 shows that the midlatitude westerly wind is strongly underestimated at 500 and 800 hPa because of the extreme truncation of the QG9 model.

One can obviously expect that more realistic zonal-mean states are generated by the QG18 model, since the second zonal harmonic allows the presence of tropical easterlies and a midlatitude maximum for the westerlies. In the real atmosphere, the latitudinal position of the jet maximum is strongly affected by the momentum transport by high-frequency transient eddies. As we shall demonstrate in the next section, a model in which baroclinic eddies are properly represented does not need a strong barotropic forcing of the $\Phi_2$ harmonic to generate a realistic pattern of tropical low-level easterlies. If we want to investigate whether multiple steady states of the large-scale circulation exist for the same range of parameters that are appropriate for the QG231 model, the parameter $\alpha_2$ [see Eq. (8c)] should be much smaller than $\alpha_1$. If, on the other hand, we want to simulate in the QG18 model the average momentum transport by baroclinic eddies, then $\alpha_1$ and $\alpha_2$ should have comparable values. We shall present the stationary solutions of QG18 obtained with two very different values of $\alpha_2$, namely 0.2 and 1. In both cases, we have set $\alpha_1 = 2$, as in the QG9 model.

For the $\alpha_2 = 0.2$ case, we shall examine the steady states obtained with $\alpha_T = 1.5$; this is the same combination of forcing parameters for which (as shown in the following section) a clear multiregime behavior can be found in the QG231 model. The midlatitude zonal wind of the equilibrium and the three steady states are listed in Table 1c; the pattern of the 500-hPa eddy streamfunction for the steady states is shown in Fig. 3.

One can immediately see that the three stationary states have the same structure as those of the QG9 model. Interestingly, the model has used the additional degrees of freedom to make the first (near zonal) solution even closer to the equilibrium state than in the QG9 case previously shown. In the other two solutions, the eddy pattern is dominated (as expected) by the $\Phi_3/\Phi_4$ components. The total wave amplitude of the second and third solutions is very similar, as are the zonal-mean components. Again, the ridges in the extratropical wave pattern are just upstream of the mountain ridges in the second (ridge) solution and almost out of phase in the third (trough) steady state.

Table 1c shows that the small barotropic forcing of the $\Phi_2$ component has not succeeded in making the midlatitude zonal flow of the steady states substantially more realistic. There is now a larger difference in the westerly wind between the near-zonal steady state (which has stronger upper- and middle-level westerlies than climate) and the other two; for all solutions, the 500-hPa wind is closer to the climatological value, but the 800-hPa westerlies are still strongly underestimated.

Let us now consider the $\alpha_2 = 1$ case. Now, three steady states can be found for $\alpha_T \geq 1.3$. If, by increasing the value of $\alpha_T$, we want to simulate the effects of the energy conversion due to transient eddies, the increase in the KE source must be accompanied by a decrease in the source of APE. We shall therefore consider the parameter combination $\alpha_T = 1.3, \alpha_2 = 1$, and discuss the effects that this redistribution of energy sources has on the stationary solutions. The midlatitude mean zonal wind of the three steady states are listed in Table 1d, while the 500-hPa eddy streamfunction fields are shown in Fig. 4.

Although the general characteristics of the three solutions are similar to those found with $\alpha_T = 1.5, \alpha_2 = 0.2$, a number of significant differences can be found. First of all, the midlatitude zonal winds are much closer to the observed climatological values than in the previous case. The interaction of the stronger 800-hPa zonal wind with the orography produces a stronger orographic forcing; for this reason, the near-zonal stationary solution is now less close to the equilibrium state, and the amplitude of its wave pattern is considerably larger than in the case with $\alpha_2 = 0.2$. On the other hand, the effect of the lower APE source is to reduce considerably the wave amplitude in the ridge and the trough steady states. In addition, while in the previous case these two states had an almost identical wave ampli-
tude, now the trough steady state possesses a larger wave amplitude than the ridge state, as in QG9.

The difference in 500-hPa streamfunction between the ridge and the trough steady states is plotted in Fig. 5a for the $\alpha_T = 1.5$, $\alpha_2 = 0.2$ case and in Fig. 5b for the $\alpha_T = 1.3$, $\alpha_2 = 1$ case. Apart from the difference in amplitude, in both cases the pattern is dominated by the $\Phi_1$ harmonic, that is, by the component of the resonant wavenumber in phase with orography. The $\Phi_3$ component has only a negligible projection on the difference pattern in the latter case; on the other hand, the stronger interaction between low-level flow and orography due to the increased value of $\alpha_2$ produces a larger difference in the zonal-mean wind between steady states with opposite wave phase.

Despite the differences listed above, one can conclude that many features of the three steady states of the QG18 model have remained the same despite changes in the forcing parameters, which have considerably altered the structure of the zonal-mean flow. These features can also be reproduced in the even more severely truncated QG9 model. Although this can be viewed as an indication of robustness of our results, it does not guarantee that a similar behavior will be found in a more complex model that includes many more degrees of freedom. We shall now turn to the QG231 model for the answer to this crucial question.

5. Regimes in the intermediate-truncation model

We shall now investigate whether the multiple stationary states of the QG9 and QG18 models have any counterpart in flow regimes of QG231. For this purpose, a number of long nonlinear integrations of this model have been performed; the duration of each of them is 8100 days, equivalent to ninety 90-day winters. The first 10 winters have been considered as an adjustment period and have been neglected in the data analysis. All results in this section will therefore refer to 80-winter (7200 day) periods, as we did for the run of the global QG model analyzed in Part I.

Various combinations of the forcing parameters $\alpha_T$, $\alpha_1$, and $\alpha_2$ have been tested. A qualitatively realistic climatology has been obtained setting $\alpha_1 = 2$, $\alpha_2 = 0.2$, and two regimes have been found for $\alpha_T \geq 1.4$. These regimes become well defined for $\alpha_T = 1.5$; therefore, we shall take this particular integration as the reference for the subsequent analyses.

Figure 6a shows the 80-winter average of the modeled mean zonal wind at the three vertical levels. At 200 and 500 hPa, QG231 has produced a midlatitude westerly jet that has a realistic maximum speed but is centered around 45°, instead of 30° as in the real atmosphere. At 800 hPa, a band of easterlies extends up to 30° (observed easterlies only reach 20°N), and again the westerly maximum is located 15° north of its observed counterpart; besides, the maximum wind speed is overestimated by about 30% with respect to the mean zonal wind in the Pacific–North American sector of the Northern Hemisphere (where the real and the modeled orographies are more comparable).

The time-averaged eddy streamfunction field at the three model levels is shown in Fig. 6b–d. The average eddy field shows a modest westward tilt with height: the high-latitude ridges are 35° upstream with respect to the orography maxima at 800 hPa and 45° upstream at 200 h Pa. The amplitude of the mean waves is smaller than the observed one at 500 and 800 hPa but overall quite realistic.

Finally, the standard deviation of the 500-hPa streamfunction in the reference run of QG231 is shown in Fig. 6e. The QG231 model produces variability maxima located around 50°N and about 45° upstream of the mountain top, in good agreement with the position of the Pacific variability maximum (with respect to the Rocky Mountains). The maximum amplitude (corre-
Fig. 6. Statistics from the 80-winter integration of the QG231 model. (a) Average latitudinal profile of mean zonal wind at the three levels of 200 (dashed line), 500 (solid line), and 800 hPa (dotted line); (b)–(d) average streamfunction eddies at (b) 200, (c) 500, and (d) 800 hPa; (e) standard deviation of 500-hPa streamfunction.

responding to a height standard deviation of about 160 m) is also remarkably close to its observed counterpart.

Given the idealized nature of the model, one can conclude that its climatology is (at least qualitatively) a reasonable approximation of the wintertime climatology of the boreal atmosphere. The zonal-mean profile is obviously not as good as that of the global QG model (shown in Fig. 10 of MM93), but one should remember that in QG231 the PV sources are controlled by just three parameters, rather than projecting on all the degrees of freedom of the model.

We shall now turn our attention to the crucial aspect of the QG231 simulations, namely the existence of flow regimes. If regimes exist in the QG231 phase space, and they are related to steady states of the planetary-scale circulation, these regimes should have a strong
signature on the PDF of the variables that characterize
the stationary solutions of the QG18 model. As we have
seen, the difference between such solutions is better
seen in the wavenumber-3 component that is in phase
with orography. Since the QG231 model has its low-
frequency variability maxima around 50°N, we shall
consider here the projections of 500-hPa streamfunc-
tion fields onto the eddy functions $\Phi_3$, $\Phi_4$, $\Phi_5$, and $\Phi_6$
defined by Eq. (3) and use them to reconstruct a me-
ridionally filtered wavenumber-3 profile at 50°:

$$\psi(\lambda, 50^\circ N, 500 \text{ hPa})_{m=3,n=4+6} = c_s \cos(3\lambda) - c_v \sin(3\lambda). \quad (9)$$

The $c_s$ coefficient gives the component of the profile
in phase with the orography, and it is a linear combi-
nation of the projections on $\Phi_3$ and $\Phi_5$; $c_v$ represents
the component one-quarter wavelength upstream of
the orography, and it is a combination of the projections
on $\Phi_4$ and $\Phi_6$. (Note that the amplitude of $\Phi_4/\Phi_6$ is
about twice the amplitude of $\Phi_3/\Phi_5$ at 50°N and that
the projections on these functions are weighted accord-
ingly in the computation of $c_s$ and $c_v$.) As a reference,
Table 2 lists the values of $c_s$ and $c_v$ for the stationary
solutions of the QG18 model discussed in the previous
section.

For the reference integration of QG231, two-dimen-
sional PDFs in the plane $c_s$-$c_v$ have been computed
from running 5- and 10-day mean values of the spectral
coefficients, using the iterative kernel estimation tech-
nique previously applied to the analysis of the global
QG model data [see appendix B of MM93, or Silver-
man (1986) and Kimoto and Ghil (1993) for a more
general discussion]. The regime behavior is better re-
vealed by 10-day mean values; the PDF obtained with
this time filter is shown in Fig. 7. The bimodal nature of
the PDF is unequivocal and much more evident than
in the PDF for the global QG model shown in MM93.
The two density maxima are aligned mainly along the
$c_v$ axis, in qualitative agreement with the structure of
the ridge and trough steady states of the QG18 model.
If the same values of the forcing parameters are con-
sidered, the axis linking the density maxima is almost
parallel to the axis linking the two steady states; be-
sides, each density maximum is roughly aligned in the
direction of the corresponding steady state with respect
to the origin of the $c_s$-$c_v$ plane, which indicates a good
agreement in the phase of the wave patterns. However,
the distance between the QG18 steady states and the
origin (which is a measure of the wave amplitude) is
about twice as large as the distance of the density
maxima.

The bimodal structure of the 80-winter PDF is well
reproduced in the two subsamples, which include the
first 40 winters (after the adjustment period) and the
last 40 winters. The PDFs for the two subsamples are
shown in Fig. 8. One can note, however, that the regime
separation is stronger in the last 40 winters, the relative
amplitude of the two maxima is reversed in the two
periods, and the PDF ridge connecting the two modes
is located in a different position on the $c_v$ axis.

One feature that is consistent between the two sub-
samples is the elongation (toward lower values of both
$c_s$ and $c_v$) of the density maximum with positive $c_s$. If
the 80-winter PDF is recomputed with a smaller value
of the parameter that controls the smoothness of the
density estimate (i.e., the "radius of influence" of each
datum), this maximum is actually split into two, as
shown in Fig. 9. This split is reproduced in each of the
40-winter subsamples (not shown) when less smooth-
ing is applied. Since its position on the $c_v$ axis is very
close to zero, it might be tempting to conclude that the

| Table 2. Wavenumber-3 coefficients $c_s$ and $c_v$ ($10^6 \text{ m}^2 \text{ s}^{-1}$), representing 500-hPa wave components in phase and one wavelength upstream with respect to orography, for the steady states of the QG18 model. The forcing parameters in (a) are the same as in the QG231 model. |
|---------------------------------|----------------|----------------|
| Steady state                    | $c_s$          | $c_v$          |
| (a) $\alpha_T = 1.5$, $\alpha_I = 2.0$, $\alpha_2 = 0.2$ | 2.5            | 0.4            |
| Near zonal                     | 19.9           | 22.9           |
| Ridge                          | -27.6          | 11.9           |
| Trough                         | 9.2            | 4.0            |
| (b) $\alpha_T = 1.3$, $\alpha_I = 2.0$, $\alpha_2 = 1.0$ | 15.1           | 14.8           |
| Near zonal                     | -19.0          | 15.6           |

Fig. 7. Probability density of 10-day mean fields from the 80-
winter integration of the QG231 model, in the plane spanned by
the two components ($c_s$, $c_v$) of zonal wavenumber 3 at 50°N. The $c_s$
coordinate represents the component in phase with orography; $c_v$
represents the component one-quarter wavelength upstream. Units are
$10^6 \text{ m}^2 \text{ s}^{-1}$ for $c_s$ and $c_v$. The black squares indicate the coordinates of the three steady states of the QG18 model with the same forcing
parameters used in the integration of QG231 (cf. Table 2a).
third, weaker maximum corresponds to the near-zonal steady state of the QG18 model.

On the other hand, experiments with slightly different values of the forcing parameters or orographic height do confirm the existence of the two principal maxima and often reveal the presence of a third, weaker maximum, but the position of the third maximum in phase space is highly variable. In general, a regime corresponding to near-zonal state of QG18 cannot be consistently identified from the PDF. A normal-mode stability analysis of the three steady states of QG18 reveals that, in the parameter range considered here, the near-zonal state is already more unstable than the other two states to planetary-scale perturbations (i.e., perturbations of the QG18 variables). Although a stability analysis using all the DOFs of the QG231 model has not been carried out (remember that the QG18 steady states are not stationary solutions of QG231, not even if the same orography is used), the stronger mean zonal wind of the near-zonal state will also make it more baroclinically unstable to synoptic-scale perturbations than the ridge and trough steady states.

We are therefore in a similar situation to that found in RP: only two of the three (or five, depending on the forcing parameters) steady states of the CS model had a counterpart in the regimes of the RP model, and no regime was found corresponding to the zonal Hadley equilibrium. Like the QG231 regimes, the RP regimes had opposite phases of the planetary waves with respect to orography, but the average wave amplitude was smaller than the amplitude of the corresponding CS steady states.

We can try to explain with energetic arguments why the distance in phase space between the density maxima associated with the trough and the ridge regimes is significantly smaller than the distance between the corresponding steady states of QG18. We know from many previous diagnostic studies on observed and model data that baroclinically unstable transient eddies (which are present in QG231) play a major role in the energetic conversion process that eventually transforms the APE of the zonal-mean flow into KE of the zonal flow itself and of the planetary- and synoptic-scale waves. As far as the mean zonal flow is concerned, we can view the conversion process as a sink of APE (which counterbalances the generation of APE by diabatic forcing) and a source of KE, as clearly shown by the structure of the average high-frequency transient forcing in the global QG model discussed in Part I (see Fig. 9 there).

We have seen in the previous section that, if we parameterize this process in the QG18 model by reducing the APE source and increasing the KE source for the zonal flow, we considerably reduce the wave amplitude

---

**Fig. 8.** As in Fig. 7 but for (a) the first 40 winters and (b) the last 40 winters of the integration.

**Fig. 9.** As in Fig. 7 but computed using a smoothing parameter (i.e., the radius of influence of individual data) decreased by 20%.
in the ridge and the trough steady states. In addition, we noticed that this reduction is stronger for the ridge than for the trough solution. This is, once again, in agreement with the behavior of the regimes in QG231: the PDF maximum corresponding to the ridge regime is closer to the origin of the $c_{r-c_{r}}$ plane (that is, to a zonal state) than the PDF maximum of the trough regime, so that the latter has a stronger wave amplitude. We can say that the first-order effect of the transient eddies on the regimes is to move the PDF maxima in phase space from the position of the steady states of the planetary-scale flow (with the same forcing) to a position of quasi-stationarity in the presence of an additional, statistically defined forcing that represents the average effect of high-frequency PV fluxes.

6. Relationship between neutral vectors, regimes, and steady states in the simplified models

a. Definition of neutral vectors for the QG231 model

In MM93 and Part I, we used dynamical and generalized neutral vectors to identify regimes in the real atmosphere and in the global QG model. (Following Part I, we call “dynamical” those neutral vectors that are computed directly from the linearized model as the axes with the smallest linear time derivative.) In both cases, bimodality has been found in the projection of time-filtered fields along dynamical neutral vectors, supporting the hypothesis of the existence of multiple quasi-stationary states. Generalized neutral vectors, in which the linear time derivative is balanced by the anomalous transient eddy forcing, have provided a dynamical interpretation of the multidimensional structure of the model’s phase space as revealed by a hierarchical cluster analysis.

The QG231 model provides an ideal context to test once again the validity of our assumptions. We have seen that, with the parameter values of the “reference” integration, this model possesses two well-defined and nearly equally populated regimes, associated with two stationary states of the large-scale flow. We therefore expect a good correspondence between the dynamical neutral vector with the longest characteristic time and the axis linking the two steady states. We also expect that, if the high-frequency transient (HFT) forcing is able to shift significantly the regime centroids from the stationary states, then the difference between regime centroids will be even better represented by the generalized than by the dynamical neutral vectors.

We have applied the dynamical and generalized neutral vector analysis to the reference run of QG231 in exactly the same way it was applied to the long run of the global QG model. For the sake of clarity, we shall summarize here the main steps of the procedure.

1) The model equations have been linearized around the time-mean state, and a linear time-derivative operator $\hat{L}$ for the streamfunction field has been defined.

2) The dynamical neutral vectors have been computed as the eigenvectors $E_i$ of $\hat{L}^*\hat{L}$ with the smallest eigenvalues (i.e., the longest characteristic times), $\hat{L}^*$ being the adjoint of $\hat{L}$ with respect to a kinetic energy inner product.

3) A sample of 1440 partially overlapping 10-day mean fields of streamfunction have been computed from the 80-winter reference run, together with the transient-forcing fields generated by the deviations from each individual 10-day mean. By subtracting the respective 80-winter mean, the fields in two samples have been converted into “anomalies.”

4) The streamfunction 10-day mean anomalies have been projected onto the full set of 231 eigenvectors $E_i$, while the anomalies of HFT forcing have been projected on the vectors $D_i$, which represent the normalized linear tendencies (i.e., time derivatives) associated with the $E_i$ vectors.

5) From these projections, generalized neutral vectors have been computed as the anomalies for which the closest balance between linear time derivative and anomalous HFT forcing is achieved, using the statistical-dynamical technique described in Part I.

b. The first dynamical neutral vector and its optimal forcing

The three-dimensional structure of the first dynamical neutral vector $E_1$ of QG231 is shown in the left column of Fig. 10. This vector has a characteristic time of 600 days, almost three times longer than that of the second neutral vector (the characteristic time is the ratio between the norm of the vector and the norm of its linear time derivative). As expected, $E_1$ has an equivalent barotropic structure dominated by the resonant wavenumber $m = 3, n = 6$. The wave maxima and minima are located about $10^9$ eastward of the orographic maxima and minima (the sign is of course arbitrary). This means that the wave component in phase with the orography ($\Phi_3$) is stronger than the upstream component ($\Phi_3$), in agreement with the difference between the ridge and the trough steady states shown in Fig. 5.

The neutral vector $E_1$ also possesses a significant zonal-mean component, which produces a clear difference in amplitude between the two extremes of the wave pattern. As is also visible in the difference map in Fig. 5b, the fact that the (positive) extreme located over the mountains has a smaller amplitude than the (negative) extreme half-wavelength upstream means that the mean zonal wind is stronger, and its maximum is shifted northward, when the ridges are located over the mountains (as in the ridge steady state). Therefore, in the first neutral vector the relationship between variations in zonal wind and wave phase is qualitatively the same as in the large-scale steady state, although the zonal wind variations are comparatively stronger.

The right column of Fig. 10 shows the (normalized) time derivative $D_1$ associated with $E_1$. As discussed in
Fig. 10. (Left column) First dynamical neutral vector $E_1$ of the QG231 model, computed using the 80-winter average as a basic state; (top to bottom): 200-, 500-, and 800-hPa streamfunctions. 
(Right column): streamfunction vector $D_1$, defined as the (normalized) linear time derivative of $E_1$, for the same pressure levels; the opposite of $D_1$ is the optimal forcing pattern for $E_1$.

Fig. 11. Streamfunction anomalies at 200, 500, and 800 hPa (top to bottom) representing the generalized neutral vectors $V_1$ (left column) and $V_2$ (right column) of QG231.
Part I, the opposite of $D_1$ gives the forcing pattern that can balance the linear tendency of $E_1$; therefore, $E_1$ can be viewed as the ‘steady linear response’ to a forcing, which is parallel to $-D_1$. Since the first neutral vector has the smallest possible linear tendency, it follows that $D_1$ is the normalized forcing pattern that is able to produce the strongest steady linear response in the linearized QG231 model.

The structure of $D_1$ is strongly baroclinic, with its amplitude concentrated in the middle and the lowest model levels, where the pattern has opposite sign. Interestingly, the maxima of this vector are located in the subtropical area, around $20^\circ N$, south of the areas of maximum amplitude in the middle- and upper-level components of $E_1$. The sign of these subtropical features is such that an anomalous diabatic heating in the 500–800 hPa layer located in the area of the $D_1$ maxima would generate a strong response that is negatively correlated with $E_1$.

The results presented in the previous sections prove unequivocally that the orography is the ultimate source of the two nonlinear regimes that dominate the low-frequency variability in QG231. However, if we tried to explain this variability in terms of steady linear response to an anomalous external forcing, we would find that a pattern very close to the dominant mode of variability of the system can be effectively ‘excited’ by subtropical diabatic heating. Although no external forcing of this kind exists in the reference run of QG231, it certainly exists in the real atmosphere in periods with strong anomalies in tropical sea surface temperature, when the frequency of extratropical regimes can be significantly affected. This result shows that there is no contradiction between the fact that the leading low-frequency modes can be fairly well reproduced as steady linear responses to tropical or subtropical diabatic forcing (e.g., Hoskins and Karoly 1981; Branstator 1985; Navarra and Miyakoda 1988) and the hypothesis (supported by the observed multimodality in large-scale parameters) that such patterns correspond to regime centroids originated by the internal nonlinear dynamics of the extratropical flow.

c. Generalized neutral vectors

We now turn our attention to the generalized neutral vectors of QG231. As for the global QG model analyzed in Part I, the Ward clustering procedure (Anderberg 1973) is used to generate a hierarchy of partitions of the model phase space, from 1440 clusters at the beginning to just 2 clusters at the end. For each new cluster formed in the process, a generalized neutral vector is computed in the neighborhood of the cluster centroid, and its characteristic time (inversely proportional to the norm of residual time derivative) is evaluated. We assume that the ‘best’ partition of the model phase space is the one for which the longest possible characteristic times are obtained for some of the clusters.

For QG231, these maximum values are reached in the four-cluster partition, where two clusters correspond to generalized neutral vectors $V_1$ and $V_2$ with characteristic times of 171 and 169 days (the vectors $V_1$ and $V_2$ associated with the other two clusters have characteristic times of 65 and 53 days). The populations of the four clusters are 24%, 31%, 26%, and 19%, respectively. When the four clusters are merged into two ‘superclusters,’ the clusters associated with $V_1$ and $V_2$ end up in different superclusters; similarly, we have seen in the case of the global QG model that the clusters with longest characteristic times acted as nuclei of aggregation around which bigger clusters were formed at a later stage.

If the difference in characteristic time is indeed an index of the relative persistence in different parts of phase space, then we should expect that $V_1$ and $V_2$ resemble opposite phases of the axis linking the ridge and trough steady states, which are associated with the model regimes (remember that neutral vectors represent anomaly fields, so they cannot reproduce features that are common to both regimes). The 3D structure of these two vectors is shown in Fig. 11. It is evident that the two vectors represent opposite anomalies whose eddy parts are dominated by the resonant wave component in phase with the orography. Their pattern is strongly correlated with the difference pattern between the ridge and the trough steady states of QG18 (see Fig. 5). In the case in which the forcing parameters of QG18 are tuned to simulate the average effect of transient eddies (Fig. 5b), the amplitude of the difference pattern is also comparable with the difference between $V_1$ and $V_2$, which is shown in Fig. 12a for the 500-hPa level.

As in the pattern of the first dynamical neutral vector $E_1$, and to a smaller extent in the steady-state difference in Fig. 5b, the zonal-mean component in $V_1$ and $V_2$ is such that the local maxima (or minima) located over the orographic ridges have a smaller amplitude than the corresponding features located over the valleys of the topography. It is interesting to note that, in the Pacific–North American (PNA) sector, a similar structure can also be found in the difference map between the two observed regimes identified in MM93 and in the composite for the ‘traditional’ PNA pattern as defined by Wallace and Gutzler (1981). The zonal-mean component is stronger in $V_1$ and $V_2$ than in the steady states of QG18 but weaker than in $E_1$; significant differences in the zonal-mean flow were already noticed comparing the generalized and dynamical neutral vectors of the global QG model.

Figure 12b shows the difference between the vectors $V_1$ and $V_2$, which are associated with the two clusters with shorter characteristic times in the four-cluster partition of the QG231 phase space. Again, these two vectors have almost opposite anomalies, and their difference pattern is orthogonal to the difference between $V_1$ and $V_2$. The vector $V_2$ corresponds to an anomaly that
Fig. 12. (a) 500-hPa streamfunction difference between generalized neutral vectors $V_1$ and $V_2$; (b) as in (a) but for $V_3$ minus $V_2$; (c)–(f): 500-hPa streamfunction obtained by superimposing the generalized neutral vectors (c) $V_1$, (d) $V_3$, (e) $V_5$, and (f) $V_4$ to the QG251 climatology (80-winter average).

Fig. 13. Streamfunction tendencies at 200, 500, and 800 hPa (top to bottom) representing the anomalies in high-frequency transient forcing $\tilde{F}_1$ (left column) and $\tilde{F}_2$ (right column) associated with generalized neutral vectors $V_1$ and $V_2$, respectively.
intensifies the ridges located one-quarter wavelength upstream of the mountain maxima and, therefore, increases the total wave amplitude; the opposite is true for $V_4$. The difference between $V_3$ and $V_4$ has a smaller amplitude than the difference between $V_1$ and $V_2$; taking into account the cluster population, one finds that the centroids of the third and fourth clusters explain less than one-third of the variance (i.e., the kinetic energy) explained by the centroids associated with $V_1$ and $V_2$. As shown in the previous section, no local maxima in the PDF can be found corresponding to either $V_1$ or $V_4$; the difference in characteristic time is therefore able to discriminate between clusters corresponding to parts of phase space with different persistence properties.

The total 500-hPa streamfunction fields obtained superimposing $V_1$, $V_2$, $V_3$, and $V_4$ on the 80–winter mean are shown in Figs. 12c–f. In addition to the clear difference in the phase of the planetary waves between $V_1$ and $V_2$, one can see that the ridge regime associated with $V_1$ has a smaller wave amplitude than the trough regime associated with $V_2$, in agreement with the position of the corresponding PDF maxima in the $c_r$–$c_v$ plane. A larger difference in wave amplitude can be seen, as expected, between flows corresponding to $V_3$ and $V_4$.

d. Anomalous forcing by high-frequency transients

We shall now look at the structure of the anomalies in HFT forcing associated with $V_1$ and $V_2$. These forcing anomalies have been converted into streamfunction tendencies over a 10-day period and are shown in Fig. 13. As in the case of the global QG model, the anomalies in HFT forcing tend to reinforce the large-scale flow anomalies. This is particularly evident at the 800-hPa level, where the forcing and flow anomalies have an almost identical longitudinal phase, while at 200-hPa the forcing maxima and minima are located upstream of the maxima and minima in $V_1$ and $V_2$. As discussed in Part I, this phase shift implies that the forcing anomalies are mainly compensated by dissipation at the former level, by advection at the latter.

An interesting feature of the HFT forcing anomalies is that their meridional scale is smaller than the meridional scale of the wave pattern in $V_1$ and $V_2$. In the regions where the large-scale anomalies are positive, the anomalous transient forcing has the shape of a dipole with a high centered around 60°N and a low centered at 30°N. This pattern is clearly reminiscent of atmospheric blocking. As in the global QG model, it appears that in QG231 anomalies the HFT forcing tends to form blocking-type structure in the regions of large-scale ridges, consistently with a number of diagnostic and modeling studies of atmospheric blocking (e.g., Illari and Marshall 1983; Shotts 1986; Haines and Marshall 1987; Vautard and Legras 1988).

At this point, we can ask ourselves the following question: although we know that up-scale energy cascade is very important for the energetics of the planetary-scale flow, is it correct to say that in QG231 the regimes are the response to anomalous transient eddies? We know that for $V_1$ and $V_2$ a very close balance exists between linearized PV advection, dissipative processes, and anomaly in HFT forcing. If we indicate such forcing anomalies as $\tilde{F}_1$ and $\tilde{F}_2$, then we can write

\[
\tilde{LV}_1 + \tilde{F}_1 = 0 \quad (10a)
\]

\[
\tilde{LV}_2 + \tilde{F}_2 = 0. \quad (10b)
\]

From a diagnostic point of view, one is therefore entitled to say that, to a good degree of approximation, $V_1$ and $V_2$ are the steady linear responses to $\tilde{F}_1$ and $\tilde{F}_2$, respectively. This is in agreement with the results of a recent study by Branstator (1992) on the maintenance of low-frequency anomalies in the NCAR Community Climate Model and of a similar study by Ting and Lau (1993) based on the GFDL general circulation model. However, Branstator pointed out that “the maintaining transients are not configured optimally for forcing the low-frequency patterns.” Indeed, despite the similarity between the first dynamical neutral vector $E_1$ and the generalized neutral vectors $V_1$ and $V_2$ (which are actually opposite phases of the same vector), the forcing anomalies $\tilde{F}_1$ and $\tilde{F}_2$ have a rather different structure from $D_1$ (see Fig. 10), which is the forcing pattern that balances $E_1$.

If one admits that the variations in transient eddies are themselves the result of variations in the large-scale flow (e.g., Cai and Mak 1990), then the linear diagnostic approach cannot easily explain why $V_1$ and $V_2$ explain more variance in QG231 than other large-scale patterns. The answer can only be given by nonlinear dynamics: $V_1$ and $V_2$ are the preferred low-frequency anomalies of the QG231 model because they represent regimes corresponding to stationary solutions of the nonlinear equations for the large-scale flow. However, the transient eddies produce a significant shift of the regime centroids from the large-scale steady states, so that the three-dimensional pattern of the centroids is properly described by generalized neutral vectors that are in dynamical balance with anomalies in the HFT forcing.

We can summarize the results presented in this section by saying that the generalized neutral vectors have provided a description of the model low-frequency variability that is fully consistent with the structure of the PDF in the model phase space. In particular, the generalized neutral vectors corresponding to PDF maxima do have longer characteristic times than those associated with clusters that do not represent model regimes, so that we can be confident about the significance of the results obtained in the case of the more complex global QG model.

More importantly, we have provided a clear modeling example of regimes that are generated by extratropical nonlinear, large-scale dynamics, but which are
significantly modified by the effects of transient eddies, and can be easily excited by an external subtropical diabatic heating. Therefore, the positive results obtained in many diagnostic studies of real atmospheric anomalies, which have highlighted the similarity of observed low-frequency patterns with the steady linear response to transient eddy forcing or tropical heating anomalies, do not necessarily contradict an explanation of these patterns based on the existence of regimes generated by multiple quasi-stationary solutions of the nonlinear equations of motion.

7. Summary and conclusions

In our previous studies (MM93 and Part I), we showed how neutral vectors can provide a dynamical understanding of the relationship between steady states of the large-scale flow, regimes revealed by multimodality in phase space and quasi-resonant axes of a linearized atmospheric model. In this paper, the validity of our assumptions has been tested in a simplified, more controlled environment. By progressively reducing the number of degrees of freedom in the global QG model developed in MM93, we built a hierarchy of three hemispheric models, in which the flow was relaxed toward a zonally symmetric equilibrium state and orography provided the only source of asymmetric forcing. Two of these models described only the interactions between the mean zonal flow and planetary waves of zonal wavenumber 3, including one or two meridional wavenumbers [as in the beta-channel model of Charney and Straus (1980)]. The third model also had a zonal wavenumber-3 symmetry but included waves up to total wavenumber 21 and therefore was able to properly represent the interactions between planetary waves and baroclinically unstable synoptic-scale waves.

As in the Charney and Straus study, we found three stationary solutions for the highly truncated models, two of which represent states with large wave amplitude but opposite phase. These two steady states produced two well-defined flow regimes in the higher-resolution hemispheric model. This result is in agreement with the findings of Reinhold and Pierrehumbert (1982) and contradicts the statement by Cehelsky and Tung (1987) that flow regimes found in highly truncated models are an artifact of the excessive truncation and are not found when synoptic-scale transients are adequately represented. It should be stressed, however, that the hemispheric domain and zonal wavenumber-3 symmetry adopted here greatly favor the existence of regimes with respect to a “full” global model. Indeed, the phase-space multimodality found in MM93 using such a model (and a PV forcing acting on all spatial scales) was much less robust than the multimodality simulated here.

Using these hemispheric models, we have shown that the leading dynamical and generalized neutral vectors correctly identify the axis linking the centroids of the two regimes as the axis with the longest characteristic time (i.e., the smallest linear tendency). Although anomalies in transient eddies are important in determining the spatial pattern of these centroids, it is evident that the ultimate source of the two regimes of the hemispheric QG model is the existence of multiple stationary solutions of the large-scale flow. There is no inconsistency between the energetically fundamental role played by high-frequency transients and the capacity of nonlinear large-scale dynamics to determine the multimodal structure of the model’s phase space.

The neutral vector analysis of the hemispheric model also provides a new interpretation for a result that emerged from many dynamical studies of low-frequency variability in the last decade and apparently contradicts the important role of nonlinear dynamics. We refer to the ability of linear models to reproduce some recurrent low-frequency patterns as steady linear responses to anomalous tropical diabatic heating, which provides a simple explanation for the anomalous extratropical circulation occurring during extreme episodes of the El Niño–Southern Oscillation (ENSO) cycle.

Although we have used here neutral vectors to analyze internal midlatitude dynamics (including the effects of flow-dependent diabatic processes in midlatitudes), we can also use their linear resonant properties to investigate which fixed, external forcing would produce the largest response in terms of the dominant low-frequency pattern. In our simplified hemispheric model, this turns out to be a forcing of strong baroclinic structure (like that generated by diabatic heating) located in the subtropical regions. The similarity between the response to such forcing and the spatial pattern of the model regimes is due to the fact that the difference between two steady states becomes an easily excitable neutral vector when the model equations are linearized around an intermediate point in phase space. As discussed in more detail in Molteni et al. (1993), it is therefore possible to interpret the midlatitude response to strong ENSO events as a partial “locking” of the extratropical circulation in one of its regimes, without contradicting the nonlinear origin of the regimes themselves.

When the results of MM93, Part I, and this study are considered altogether, we can conclude that the concept of dynamical and generalized neutral vectors has allowed us to diagnose and interpret the dynamical properties of planetary-scale regimes from the simple conceptual model provided by Lorenz’s three-variable system (Lorenz 1963) to the observed atmospheric data, passing through a hierarchy of dynamical models with consistent formulation but differing in the number of degrees of freedom. In this way, we have hopefully come close to the “optimal” link between theory and observations advocated by Hoskins (1983). The work of Ferranti et al. (1994a) has shown that it is possible to extend our approach to the other extreme of the spec-
trum of atmospheric models, namely the complex general circulation models (GCMs) used for numerical weather prediction or climate studies.

With regard to GCMs, a number of studies have revealed that a large fraction of the perceived systematic errors of these models is due to their partial inability to correctly simulate the occurrence of certain flow types [in particular, blocking and states with large planetary wave amplitude; see Tracton et al. (1989); Tibaldi and Molteni (1990); Miyakoda and Sirutis (1990); Ferranti et al. (1994b)]. This type of flow-dependent error has been partially alleviated in recent years (mainly through improved parametrization schemes for diabatic processes) but still affects in a significant way medium- and extended-range forecasts as well as climate simulations. As demonstrated by the sensitivity study of Ferranti et al. (1994a), neutral vectors can provide modelers with a new tool to investigate the complex relationship between variations in the diabatic forcing of the model atmosphere and the nonlinear response to them (i.e., changes in low-frequency variability and regime frequencies), which is not easily revealed by traditional diagnostic methods.

Acknowledgments. The author is deeply indebted to Prof. J. Marshall for many useful discussions and suggestions during the development of this study. Valuable comments by Prof. B. Hoskins and two anonymous reviewers are also gratefully acknowledged.

REFERENCES


