Q. J. R. Meteorol. Soc. (1998), **124**, pp. 2149–2173

Internal interannual variability of the extratropical stratospheric circulation: The low-latitude flywheel

By R. K. SCOTT* and P. H. HAYNES†
University of Cambridge, UK

(Received 14 May 1997; revised 21 January1998)

SUMMARY

A new mechanism for internal interannual variability of the extratropical stratospheric circulation is described. The variability is internal in the sense that it arises in a stratosphere-only model without any interannual variability being imposed externally. In particular the wave forcing in the model, representing the effect of the tropospheric circulation, is kept constant from year-to-year and there is no imposed quasi-biennial variability in the equatorial winds. It is argued that the internal variability arises because of the longer 'memory' of the stratospheric flow at low latitudes. The smaller Coriolis parameter at low latitudes means that a given wind signal has associated with it a smaller temperature signal and is therefore less affected by radiative damping than a corresponding signal in the extratropics. The circulation at low latitudes can therefore act as a large flywheel, retaining memory of zonal mean quantities on an annual time-scale or longer.

The mechanism for the variability is studied using a 'mechanistic' primitive-equation model of the stratosphere. Waves are forced by a constant wavenumber one perturbation to the geopotential height field at the lower boundary and a seasonal cycle is prescribed through Newtonian cooling towards a time-dependent temperature field. Interannual variability is found within a certain range of forcing amplitude. The variability typically takes the form of a biennial oscillation, but more complex behaviour is also found under some circumstances. Diagnostics suggest that zonal flow anomalies in the subtropics persisting from the end of one winter to the beginning of the next are responsible for the interannual variability. Further experiments, in which the zonal flow at low latitudes is constrained in a particular configuration, provide further evidence that the role of the subtropical flow is crucial since constraining the subtropical flow reduces or eliminates the forcing range for which interannual variability is obtained.

The importance of the contrast between high latitudes and low latitudes for such internal modes of variability is consistent with their absence in stratospheric models such as the Holton–Mass model with a single degree of freedom for representing latitudinal structure. To illustrate the mechanism for the interannual variability further, an extended version of the Holton–Mass model is formulated that differentiates between high- and low-latitude regions. As in the case with full latitudinal resolution, interannual variability is obtained when the amplitude of the wave forcing lies within a certain range.

KEYWORDS: Dynamics Numerical modelling Stratosphere

1. INTRODUCTION

Quantifying and understanding interannual variability is a vital part of assessing the evidence for long-term trends. Furthermore, the simulation of variability that is similar in magnitude and spatial and temporal structure to that observed is an important requirement on general circulation models. The issue of interannual variability has, of course, long been central to tropospheric general circulation studies. More recently, stratospheric interannual variability has received attention as the number of years for which stratospheric observational datasets are available have gradually increased and as multi-year runs of troposphere–stratosphere general circulation models have become computationally feasible.

At least two distinct mechanisms for interannual variability in the extratropical stratosphere have been suggested. The first is associated with the interannual variability of the tropospheric circulation. A whole string of observational and modelling studies have provided evidence that the tropospheric circulation is important in providing a forcing for stratospheric planetary waves that may (in part) control the occurrence of sudden stratospheric warmings etc. (e.g. Labitzke 1982; Randel 1987; Mechoso *et al.* 1988; Fairlie

* Current affiliation: College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA.
† Corresponding author: Centre for Atmospheric Science, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK.

et al. 1990; for observational studies and e.g. O'Neill and Pope 1988; Farrara *et al.* 1992; Scinocca and Haynes 1998; for modelling studies). It follows that interannual variability in the tropospheric circulation will almost inevitably result in a corresponding variability in the stratospheric circulation. Some of this tropospheric variability may be associated with well defined features such as the El-Nino Southern Oscillation (ENSO) signal. (e.g. Wallace and Chang 1982; Labitzke 1982; Hamilton 1993).

A second important mechanism for interannual variability in the extratropical stratosphere arises from the equatorial quasi-biennial oscillation (hereafter QBO). The QBO is, of course, the dominant manifestation of interannual variability in the tropical stratosphere and it is believed to exert an important influence on the extratropics. Various observational and modelling studies provide increasing evidence that the low-latitude circulation associated with a particular phase of the QBO affects the propagation of stratospheric planetary waves in the extratropics (e.g. Holton and Tan 1982; Wallace and Chang 1982; Dunkerton and Baldwin 1991; for observational studies and (e.g. Dameris and Ebel 1990; Holton and Austin 1991; O'Sullivan and Dunkerton 1994; for modelling studies).

Both these mechanisms for interannual variability in the extratropical stratosphere might be regarded as externally forced. In this paper we investigate another distinct mechanism for interannual variability in the extratropical stratosphere that, in particular, does not require variability in the tropospheric forcing. This mechanism is therefore most straightforwardly investigated in a 'mechanistic' stratosphere-only model of the type used for many previous investigations, in which the effect of the tropospheric circulation on the stratosphere is represented by forcing at an artificial lower boundary at 100 mb or so.

The hypothesis to be investigated is that the system comprising the stratospheric circulation, plus a representation of the tropospheric forcing that is constant from year to year, and without an equatorial QBO, has its own internal modes of interannual variability. In the context of the real atmosphere, such internal modes of variability would be expected to interact with the existing external variability of the QBO and the troposphere. For example, on the one hand weak external variability could excite the internal modes whilst, on the other hand, the internal modes could modulate strong variability due to external effects.

The internal interannual variability of the stratospheric circulation has already been investigated in a simple model, the Holton–Mass model (Holton and Mass 1976), by Yoden (1990). A seasonal cycle was imposed by applying radiative relaxation towards a seasonally varying temperature field. Yoden found no evidence for interannual variability and argued that the lack of it was due to the effects of radiative relaxation. It seemed that such relaxation was strong enough to dissipate during the summer any anomaly in the circulation remaining at the end of the previous winter. Thus any anomaly in the circulation arising during a particular winter was 'forgotten' by the system before the beginning of the following winter and there was no possibility of anomalies in one winter affecting the next to give any interesting mode of variability.

Is it possible that some mode of variability relevant to the atmosphere has been missed in Yoden's model? It might be the case, for example, that the effect of the radiative relaxation has been in some way overestimated in his experiments. The relaxation rate that he used is not unreasonable. It is not the case that if one simply reduces the relaxation rate in such experiments then interannual variability appears, since one complicating effect of the reduction is that the circulation is no longer so effectively constrained to a quasi-realistic seasonal cycle. However, one important aspect of the Holton–Mass model is that it is based on a mid-latitude β-plane channel. This, of course, will have implications for a number of different aspects of the model circulation. One key aspect that seems potentially important here is the effect of radiative relaxation on zonally symmetric dynamical quantities. There

is an important difference between the effect on a latitudinally truncated β-plane channel and the effect on a sphere. In particular, the smallness of the Coriolis parameter at low latitudes on the sphere implies that a given radiative relaxation rate will damp dynamical anomalies of a given vertical scale less at lower latitudes than in mid-latitudes. In this sense the tropical stratosphere resembles a large, zonal flywheel, relatively unconstrained by the effects of radiative relaxation, and with a longer 'memory' of zonal mean quantities. This effect is crucial in determining the time-scale required to reach a steady state in problems involving forced zonally symmetric circulations. It is therefore implicit in the results of investigations of these problems such as Haynes *et al.* (1991), who solve using an expansion in Hough functions and Holton *et al.* (1995), who give relevant scaling arguments. The implications for the latitudinal structure of the QBO are discussed in Haynes (1998). A simple illustration of the latitudinal variation in the effective damping rate is given in the appendix.

Note the term 'flywheel' has been used in the stratospheric context before, particularly in Snieder and Fels (1988). In that paper the emphasis is on the fact that the mechanical inertia of the circulation is important in the thermal response to radiative heating (and it is therefore as if the mechanical inertia modifies the thermal inertia).

Since the latitudinal variation was not represented in the β-plane geometry of Yoden's experiments, it is natural to determine whether internal interannual variability exists in a model that includes full latitudinal variation. This is the main topic to be investigated in this paper. The structure of the paper is as follows. In section 2 experiments are described using a spherical model with latitudinal structure, and one zonal wave component. In section 3 similar experiments, but this time with low-latitude velocities constrained to a particular configuration, are described. This is seen to have an important impact on the existence of internal modes of variability. In section 4 the Holton–Mass model used in Yoden's experiments is extended, *ad hoc*, to include a crude representation of the low-latitude effect. Variability is observed in such an extended model. Finally, conclusions are given in section 5.

2. Numerical experiments in a mechanistic model

(*a*) *Model description*

The model used in this study is a mechanistic primitive equation model written by Dr R. Saravanan, with pressure as the vertical coordinate and a spectral representation in the horizontal. The time-stepping uses the semi-implicit scheme, treating terms associated with gravity waves implicitly, and uses a Robert time-filter to damp the computational mode associated with the semi-implicit scheme (e.g. Haltiner and Williams 1980). There are 21 pressure levels, equally spaced in log-pressure, $z = -H \ln(p/p_s)$ where $p_s = 10^5$ Pa, spanning $z_B = 11.4$ km (2×10^4 Pa) to $z_T = 68.6$ km (3.4 Pa), at intervals $\Delta z = 2.86$ km, and 21 latitudinal Legendre modes. The reference temperature profile is that of an isothermal atmosphere with $T_{ref} = 240$ K, corresponding to a scale height, H, of 7 km. In the zonal direction only the zonal and wave-1 components are retained. The severe zonal truncation has often been used historically in studies of stratospheric dynamics, but has fallen out of favour as computational power has increased. We reinstate it in this work to allow the possibility of large sets of multi-year runs needed to investigate interannual variability. Why such severe truncation can capture some aspects of the nonlinear evolution is discussed in e.g. Haynes and McIntyre (1987) and references therein. For the results presented in sections 3 and 4 the model was run with hemispheric symmetry.

The lower boundary wave forcing is through a perturbation in the geopotential height field near the lowest model level. This is of the form

$$\Phi' = h_0 E(t) G(\phi) \tag{1}$$

where $E(t)$ is grown smoothly from 0 at $t = 0$ to 1 at $t > t_0$, with $t_0 = 10$ days. The latitudinal structure is given by $G(\phi) = 4\hat{\mu}^2(1 - \hat{\mu}^2)$, where $\hat{\mu}(\mu_0 < \mu < 1) = (\mu - \mu_0)/(1 - \mu_0)$, $\hat{\mu}(0 < \mu < \mu_0) = 0$, and where $\mu = \sin\phi$. Choosing $\mu_0 = \sin(\pi/6)$ gives a maximum $G(\phi) = 1$ located at $\phi = 60°$, $G(\phi < 30°) = 0$. Here, h_0 is the constant forcing amplitude, varied as an external parameter between integrations.

The seasonal cycle is included by radiative relaxation to a seasonally varying potential temperature field, $\theta_R(\phi, z, t)$, which is a sinusoidal superposition of summer and winter potential temperature fields, $\theta_{RS}(\phi, z)$ and $\theta_{RW}(\phi, z)$, given by $\theta_R = r\theta_{RS} + (1 - r)\,\theta_{RW}$ with $r(t) = (1 + \cos 2\pi t/T)/2$. The potential temperature fields, θ_{RS} and θ_{RW}, are specified by first specifying velocity profiles, U_{RS} and U_{RW}, and calculating the geopotential and hence potential temperature fields that are in balance with these velocity profiles. The velocity fields are given by $U_{RS} = U_R(-\phi \geqslant 0)$ and $U_{RW} = U_R(\phi \geqslant 0)$, with

$$U_R(\phi, z) = \cos\phi \cos\left(\frac{\pi}{2}\frac{(z - z_B)}{z_U - z_B}\right)[u_0 \tanh(b_0(\phi - \phi_0)) + J_1 + J_2 + J_3] \tag{2}$$

for constants u_0, b_0, ϕ_0, with $z_U = 90$ km, and where the $J_i(\phi, z)$, are defined by

$$J_i = u_i \operatorname{sech}(b_i(\phi - \phi_i))\operatorname{sech}(a_i(z - z_i)), \quad i = 1, 2, 3 \tag{3}$$

for constants u_i, b_i, ϕ_i, a_i, z_i. The values of the constants for the velocity profile used in this investigation are:

$$\begin{aligned}
&u_0 = 20\ \text{m s}^{-1}, \quad b_0 = .1, \quad \phi_0 = 20°, \\
&u_1 = 340\ \text{m s}^{-1}, \quad b_1 = .04, \phi_1 = 60°, \quad a_1 = .05, z_1 = 65\ \text{km}, \\
&u_2 = -20\ \text{m s}^{-1}, \quad b_2 = .1, \quad \phi_2 = 15°, \quad a_2 = .3, \quad z_2 = 30\ \text{km}, \\
&u_3 = -220\ \text{m s}^{-1}, b_3 = .03, \phi_3 = -55°, a_3 = .08, z_3 = 70\ \text{km}.
\end{aligned} \tag{4}$$

The summer and winter velocity profiles U_{RS} and U_{RW} are shown in Fig. 1. When balancing, the zonal mean geopotential at the lower boundary, as well as the interior temperature field, is determined by requiring that $\mathbf{T}^T \cdot \boldsymbol{\beta} \equiv 0$, where $\mathbf{T}$ is a vector of temperatures at each pressure level, and $\boldsymbol{\beta}$ is a binomial vector of alternating sign (e.g. for 5 levels $\boldsymbol{\beta} = (1, -5, 10, -10, 5)$). This is one way of ensuring that a two-grid wave in the temperature field is minimised and is a simple modification (for pressure coordinates) of the approach set out in Hoskins and Simmons (1975).

The radiative relaxation rate is given as a function of height by $\alpha(z) = [1.5 + \tanh((z - 35\ \text{km})/7\ \text{km})] \times 10^{-6}\ \text{s}^{-1}$, as in Holton (1976). Rayleigh friction is applied above $z = 50$ km, in a sponge layer, relaxing to a zero velocity field at rate $\kappa(z) = [1.02 - \exp((50\ \text{km} - z)/40\ \text{km})] \times 5 \times 10^{-6}\ \text{s}^{-1}$. The graphs of these damping rates are shown in Fig. 2 together with the latitudinal profile of the wave forcing. Finally, horizontal scale selective damping is applied using a ∇^8 hyperdiffusion with a damping rate of 4/day at the smallest scales (highest wavenumbers).

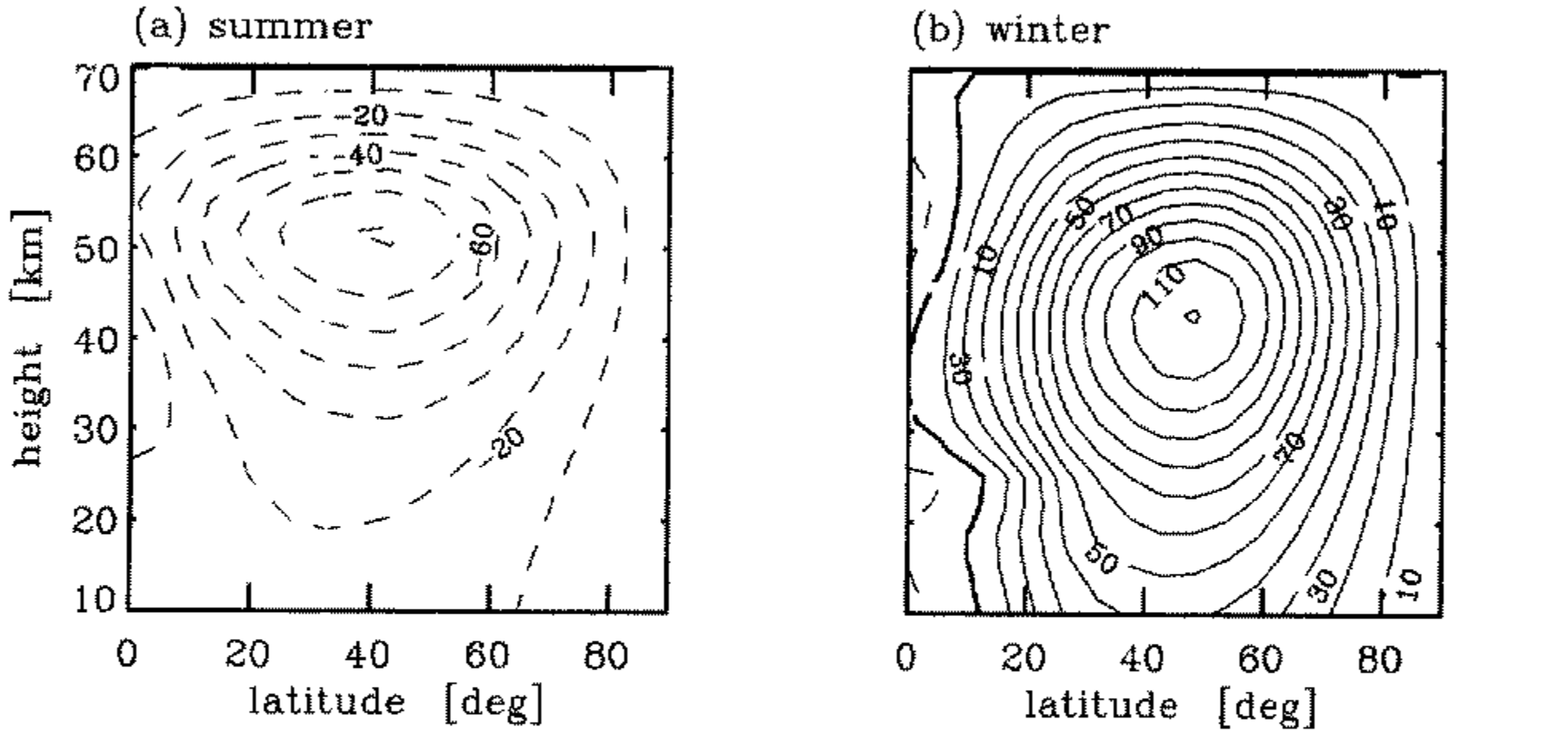

Figure 1. (a) Summer and (b) winter velocity profiles, U_{RS} and U_{RW}, used to calculate the summer and winter radiative basic states.

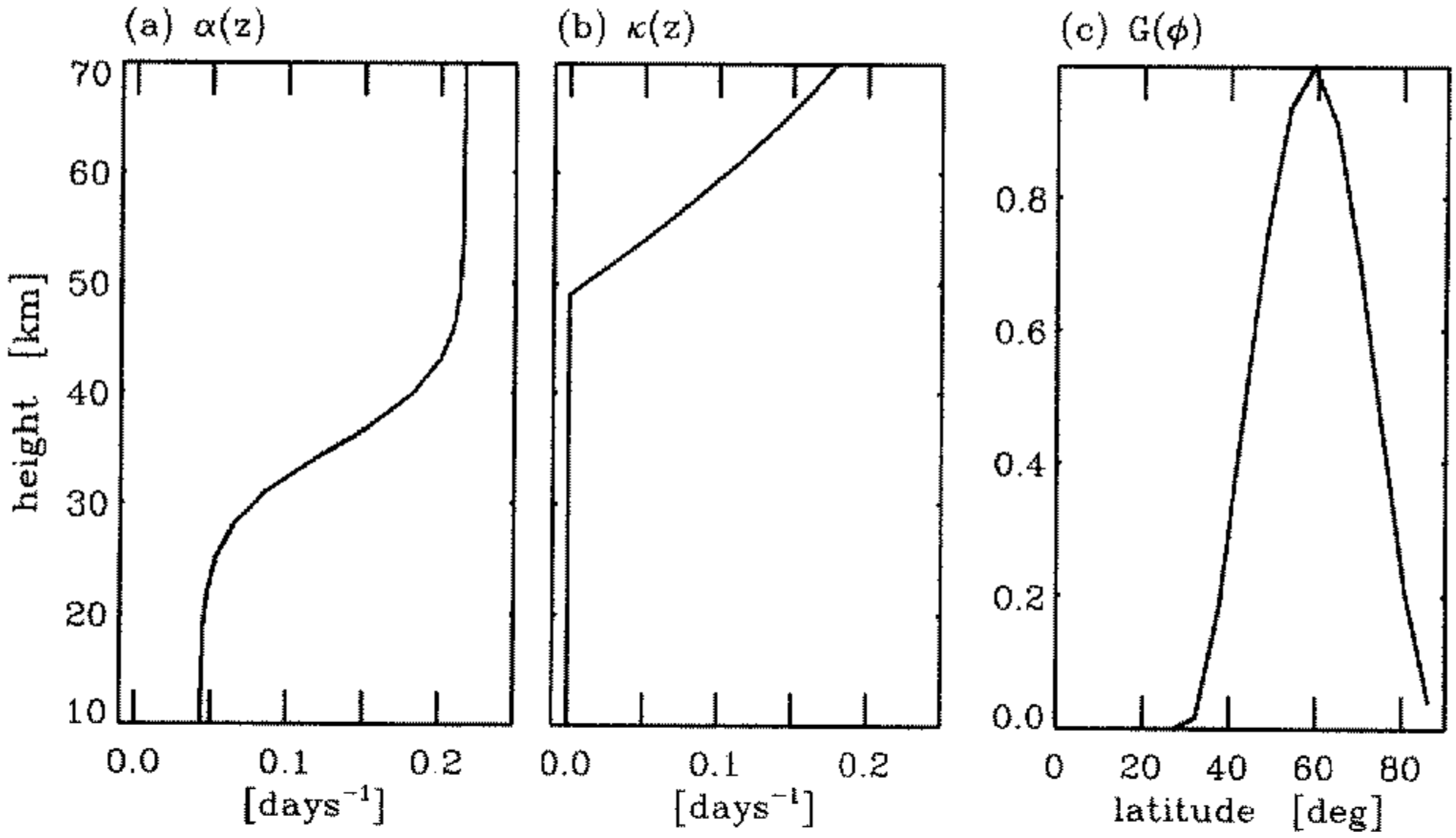

Figure 2. (a) The height dependent radiative relaxation rate, α, in units day^{-1}. (b) The height dependent Rayleigh friction rate, κ, in units day^{-1}. (c) The latitudinal structure, $G(\phi)$, of the lower boundary wave forcing (dimensionless).

(*b*) *Results*

The time evolution of zonal mean velocity at 60° latitude and 40 km altitude is shown in Fig. 3 (solid curves), for ten year runs with various values of the wave forcing amplitude, h_0, between 220 m and 270 m. For $h_0 = 220$ m the response is, to a good approximation, annually periodic, with the zonal velocity in each winter close to the velocity of the radiatively determined basic state. As h_0 is increased, there is a transition to a biennial response. Here the winter responses are alternately dynamically quiescent, characterized by the zonal velocity close to the radiatively determined state, and dynamically disturbed, characterized by the zonal velocity far from the radiatively determined state. There is a large range of values of h_0 for which the response is of this biennially periodic form; for

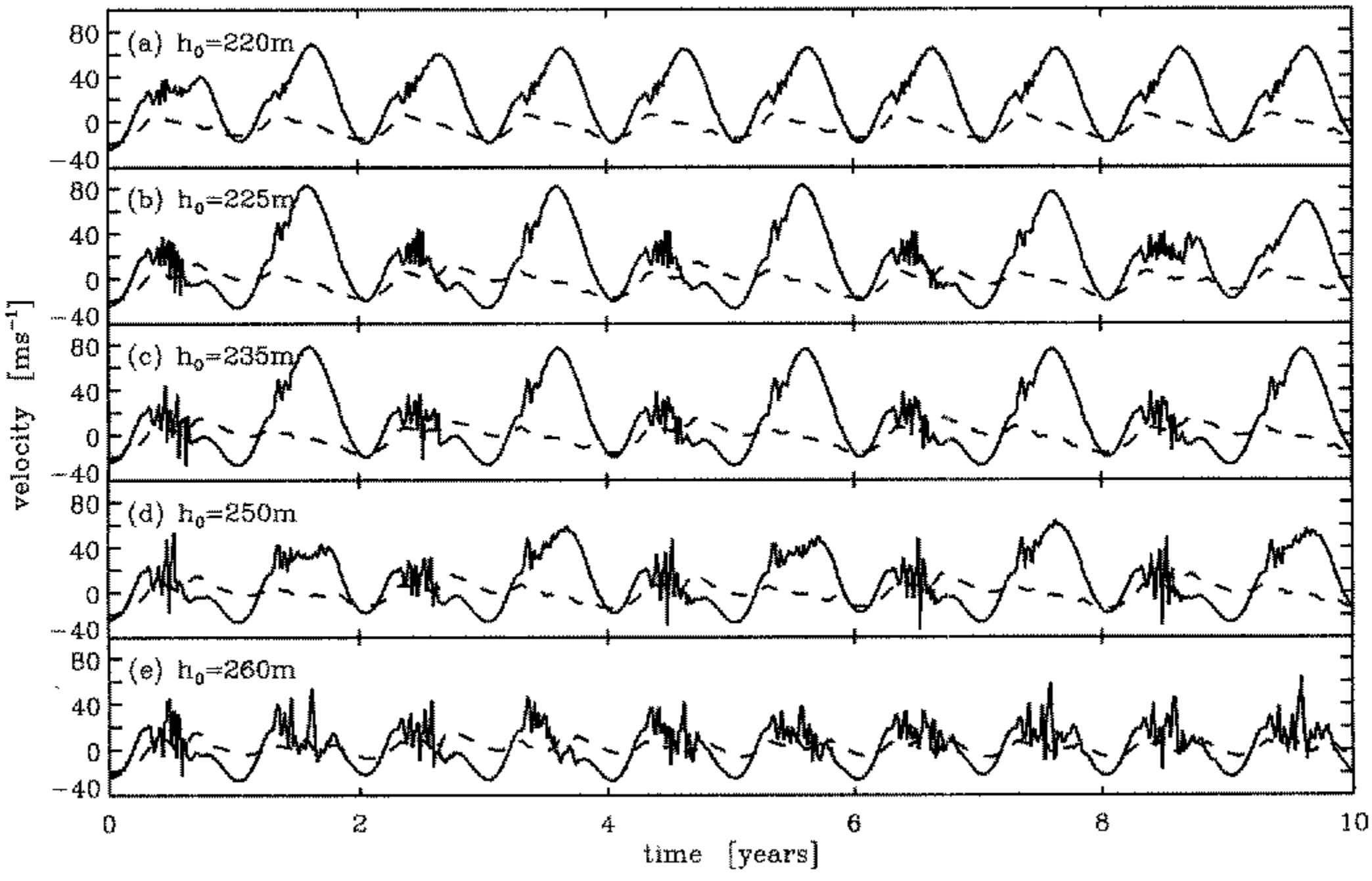

Figure 3. Evolution of the zonal mean velocity, at $\phi = 60°$, $z = 40$ km (solid curves) and at $\phi = 20°$, $z = 25$ km (dashed curves), for ten year integrations with different wave forcing amplitudes: (a) $h_0 = 220$, (b) $h_0 = 225$, (c) $h_0 = 235$, (d) $h_0 = 250$, (e) $h_0 = 260$. Time $t = 0$ corresponds to midsummer.

different h_0 in this range there are only slight differences in the degree to which the flow is disturbed or quiescent. At the other end of this range, $h_0 \simeq 250$ m, there appears a more gradual transition from the biennial response to a response consisting entirely of disturbed winters. Although all the winters are disturbed, there is still a degree of apparently random interannual variability here. We shall focus on the biennial variability of intermediate forcing values, rather than the random variability of stronger forcing values.

Consideration of the time evolution of the zonal mean velocity at 20° latitude and 25 km altitude, highlights differences between the low-altitude flow response in the disturbed and quiescent winters for the various forcing amplitudes, and is also shown in Fig. 3 (dashed curves). Generally the low-latitude flow is more easterly* (negative) during those winters with quiescent high-latitude flow than during those winters with disturbed high-latitude flow. Thus, after a quiescent winter the low-latitude flow is also more easterly than after a disturbed winter and, given the longer time-scale for decay of anomalies at these low latitudes, is also more easterly at the beginning of the next winter.

It may therefore be seen that, when the response is biennial, there is an easterly low-latitude anomaly before a disturbed winter and a westerly low-latitude anomaly before a quiescent winter. It will be argued below that these low-latitude anomalies play an important role in determining the subsequent high-latitude flow evolution. For h_0 outside this range, the forcing is either too weak or too strong for the low-latitude structure to have a strong influence on the high-latitude evolution. Specifically, if h_0 is too strong then the high latitudes are disturbed whatever the low-latitude flow, whereas if h_0 is too weak then the low-latitude signature is too weak (at this forcing amplitude) to influence the high-latitude evolution.

* Or, equivalently, less westerly (positive); what is important is the relative strength between years.

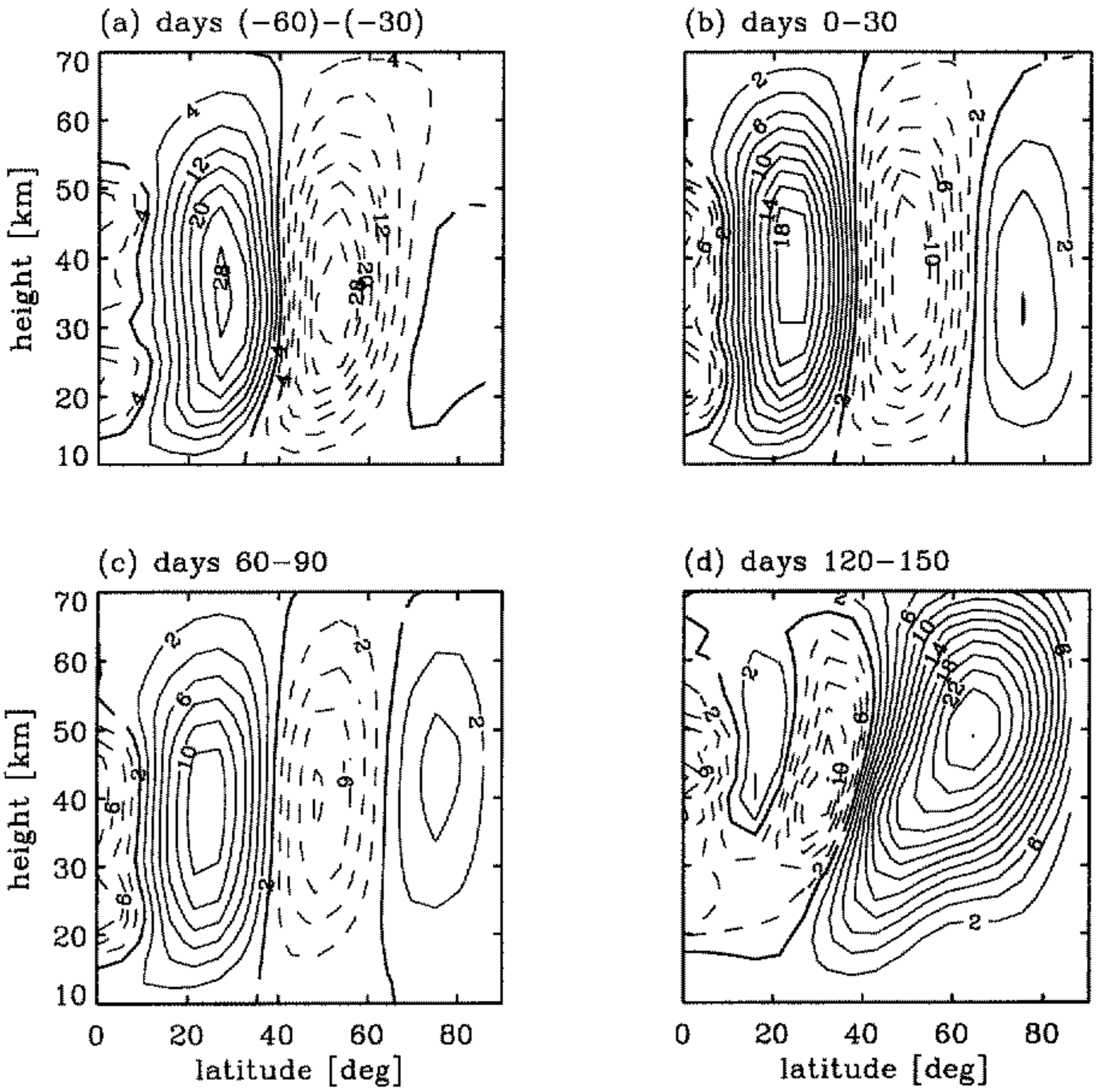

Figure 4. Monthly-mean zonal velocity anomalies, i.e. the difference of quiescent year minus disturbed year, for $h_0 = 235$ m. Monthly means are averaged over days (a) (−60)–(−30), (b) 0–30, (c) 60–90, (d) 120–150, and day 0 corresponds to midsummer. Here, and in the following, solid contours are positive, dashed are negative and heavy is zero. The contour interval is 4 m s^{-1} in (a) and 2 m s^{-1} in (b) to (d).

To validate this hypothesis concerning the role of the low-latitude anomalies it is necessary to examine carefully the latitudinal structure of the velocity field and other dynamical quantities leading up to and following disturbed or quiescent winters. In particular, a method of determining the influences on the high-latitude evolution, and the effect it has on the low latitudes of the following year, will be described in section 3. This method, which constrains the low-latitude winds in a given configuration, will help to indicate the latitudinal extent of the low-latitude wind anomaly that determines the selection of a particular high-latitude winter response. Firstly though, it is instructive to note the quantities such as EP flux divergence and potential vorticity gradient. These will give an indication of the effect that disturbed or quiescent winters have on the wave propagation characteristics of the mean flow. Consideration will be given to which features persist through the summer evolution, from the end of one winter to the beginning of the next.

Figure 4 shows the meridional structure of the monthly-mean zonal-mean zonal velocity difference between the dynamically quiescent and dynamically disturbed years eight and nine of the $h_0 = 235$ m simulation. Day 0 is the summer solstice and 'quiescent year' or 'disturbed year' refers to a year, summer to summer, containing respectively a quiescent winter or a disturbed winter. Figure 4(a,b) thus represents the *relative* (quiescent year

minus disturbed year) velocity anomaly in the summer (days (−60)–(+30), where day 0 is the solstice) before a quiescent winter, or equivalently, in the summer after a disturbed winter. The positive (westerly) anomaly between ∼10° and ∼40° corresponds to the stronger easterlies following a quiescent winter, mentioned above. From Fig. 4(c) (days 60–90) it can be seen that this subtropical anomaly is long-lived, persisting until the onset of the wave events at the beginning of the winter. The extratropical easterly anomaly between ∼40° and ∼65°, of similar strength to the subtropical anomaly in late winter (Fig. 4(a)), decays at a faster rate during the summer and is correspondingly weaker at the beginning of the following winter.

Two natural questions now arise in connection with this low-latitude anomaly. Firstly, which features of the disturbed or quiescent winters lead to the anomaly? Secondly, how does the anomaly effect the subsequent evolution, in particular, the selection of a disturbed or quiescent winter response?

To address these questions we first compare cross-sections of the Eliassen–Palm flux for the quiescent and disturbed winters. The clearest picture of the dynamical cause of the anomalies is obtained by analysing the difference in the flux divergence per unit mass,

$$\mathscr{F} = \frac{1}{\rho_0 \cos\phi} \nabla \cdot \mathbf{F} \tag{5}$$

between quiescent winters and disturbed winters, since this may be regarded as the forcing that leads to the observed wind anomalies. (Here $\mathbf{F} = \{F^{(\phi)}, F^{(z)}\}$ is the Eliassen–Palm flux as defined in, e.g., Andrews *et al.* 1987). This is shown in Fig. 5. In early winter (days 90–120, Fig. 5(b)) it is seen that there is a region of anomalously negative $\mathscr{F}$ (EP flux convergence) at low latitudes, which is responsible for the anomalous easterlies seen in Fig. 4(d) (days 120–150). Additionally, in late winter (days 275–335, Fig. 5(e,f)) there is a broader region of anomalously negative $\mathscr{F}$, extending in the vertical throughout the whole of the stratosphere, which reinforces and deepens the anomalous easterlies.

These early and late winter flux convergence anomalies are a consequence of early and late winter warming-like events in the quiescent winters. These quiescent winters thus resemble a typical southern hemisphere winter evolution of the type reported by, e.g., Randel (1988) and studied by, e.g., Plumb (1989), characterized by increased wave amplitudes in the early and late winter, and a 'midwinter minimum' of wave amplitude just after midwinter (∼day 230 in this case). A possible explanation of this behaviour in simple models is that the mean flow passes through a state for which the wave forcing is resonant e.g., Dunkerton *et al.* 1981; McIntyre 1982; Clark 1992; Scott 1996). In this case (and those of Clark and Scott) the slow variation of the mean flow through possible resonant states is allowed by the seasonal cycle. Figure 6 (dot-dashed line) shows the latitudinally averaged vertical EP flux through the lower boundary, defined by

$$\langle F^{(z)}|_B \rangle = \int_0^{\frac{\pi}{2}} F^{(z)} \cos\phi \, \mathrm{d}\phi. \tag{6}$$

The sharp increase around days 70–90 and days 280–310 is certainly consistent with resonance, i.e. with the existence of a free mode of the same phase speed as the lower boundary forcing, i.e. zero phase speed, at those times.

The strong time variation of the EP flux through the lower boundary is a reminder that this aspect of the wave forcing is determined by the model and, if the lower boundary geopotential is imposed, cannot be externally imposed. As discussed by McIntyre (1982), it is not at all clear that imposing the geopotential at an artificial lower boundary correctly represents the effective tropospheric forcing of the stratosphere, particularly in view of

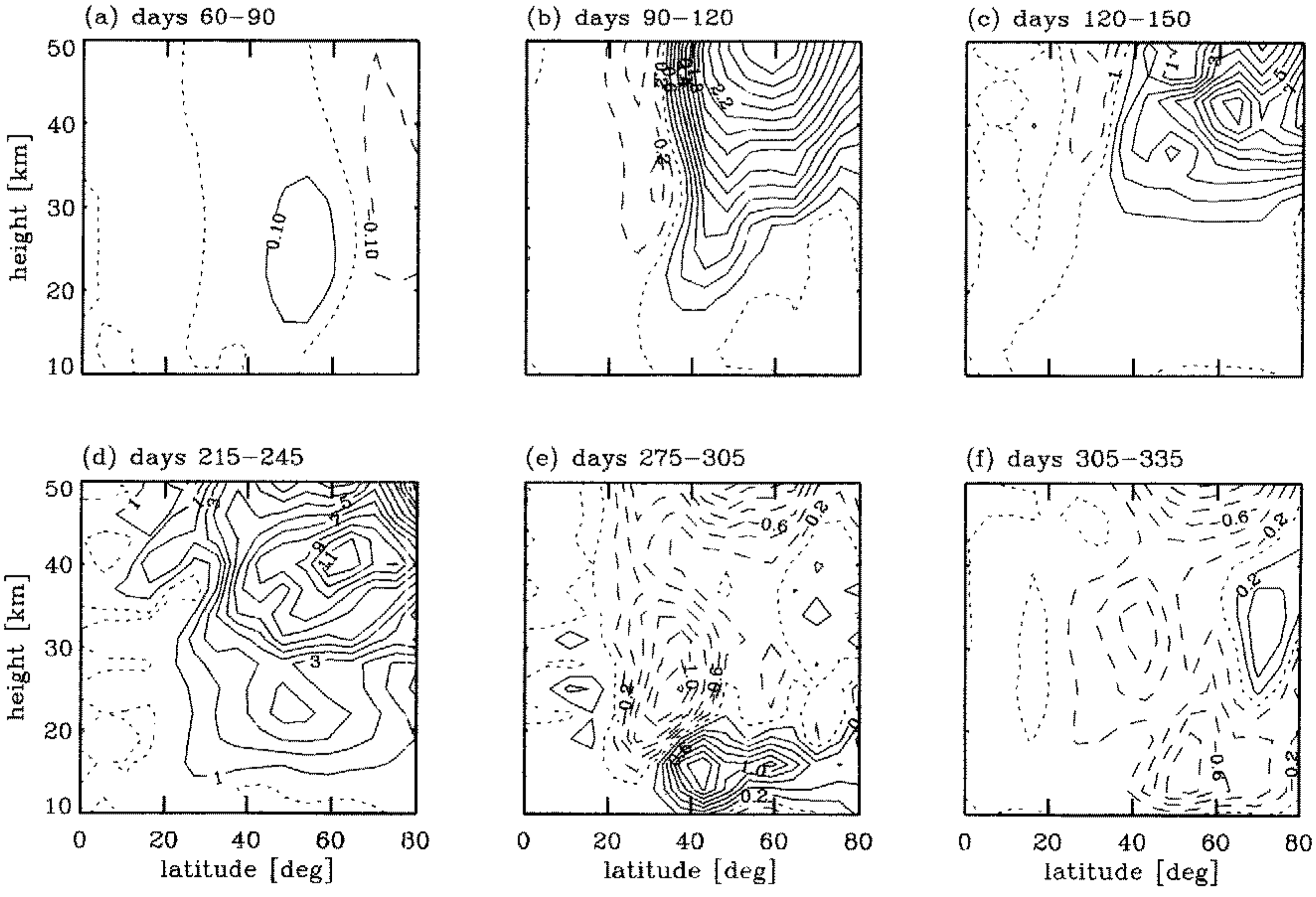

Figure 5. Monthly-mean wave-force per unit mass, $\mathcal{F}$, anomalies, quiescent year minus disturbed year, for $h_0 = 235$ m. Monthly means are averaged over days (a) 60–90, (b) 90–120, (c) 120–150, (d) 215–245, (e) 275–305, (f) 305–335. Thus (a,c) corresponds to Fig. 4(c,d). The contour interval is 0.1, 0.2, 1.0, 1.0, 0.2, 0.2 m s^{-1} day^{-1} in (a)–(f) respectively, and the zero contour is dotted (as before solid contours are positive and dashed are negative).

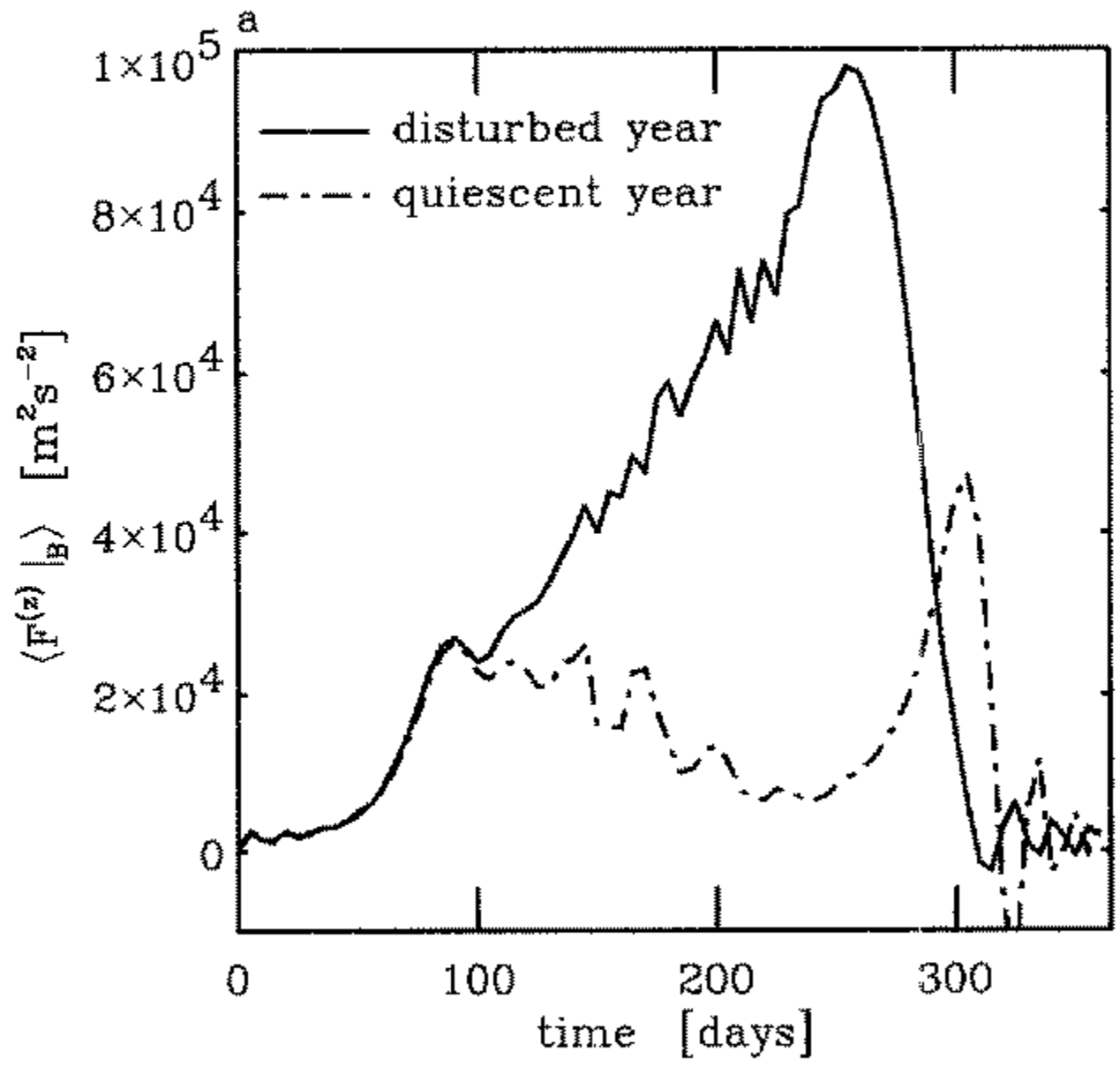

Figure 6. Latitudinally averaged, vertical EP flux through the lower boundary, $\langle F^{(z)}|_B\rangle$, for $h_0 = 235$ m. The solid curve depicts a disturbed year (from $t = 8$ to $t = 9$ years in Fig. 3(c)) and the dot-dashed curve depicts a quiescent year (from $t = 9$ to $t = 10$ years in Fig. 3(c)).

the strong resonance effects noted above. Nonetheless, wishing to address the question of interannual variability without first having to solve the difficult problem of the optimal representation of the tropospheric forcing, we here followed the precedent of much modelling work over the last fifteen years or so, including all of that mentioned in section 1 in connection with investigation of the extratropical QBO, in using the geopotential lower boundary condition.

Another important feature of Fig. 5(c,d) (and the intervening winter months, not shown) is the large region of anomalously positive $\mathcal{F}$ (EP flux divergence), persistent throughout much of the winter at most latitudes in the stratosphere. This is naturally associated with larger EP-flux convergence in the disturbed winters, and suggests that there might be greater EP flux into the domain interior in the disturbed winters than in the quiescent winters. This may be confirmed from Fig. 6, by comparing the latitudinally averaged vertical EP flux through the lower boundary, $\langle F^{(z)}|_B\rangle$, for the two years. If we again associate large upward EP fluxes through the lower boundary with near-resonant conditions, the differences highlighted in Fig. 6 suggest that the flow during a disturbed winter is generally closer to resonance conditions than the flow during a quiescent winter.

To focus more closely on the second question above—which features of the anomaly lead to different wave propagation characteristics in different winters—two further diagnostic fields are considered: cross-sections of the difference in latitudinal gradient of potential vorticity, $\bar{Q}_\phi$ between quiescent and disturbed winters, shown in Fig. 7, and cross-sections of zonal velocity for disturbed and quiescent winters taken separately, shown in Fig. 8. The former is rescaled in the vertical following Lait (1994). The greater EP flux through the lower boundary into the domain in disturbed years than in quiescent years, suggests that there will generally be more mean flow deceleration then, to some extent regardless of the meridional structure of the $\bar{u}$ or $\bar{Q}_\phi$ fields, and to which latitudes the waves propagate. However such fields do allow some explanation of the detailed pattern

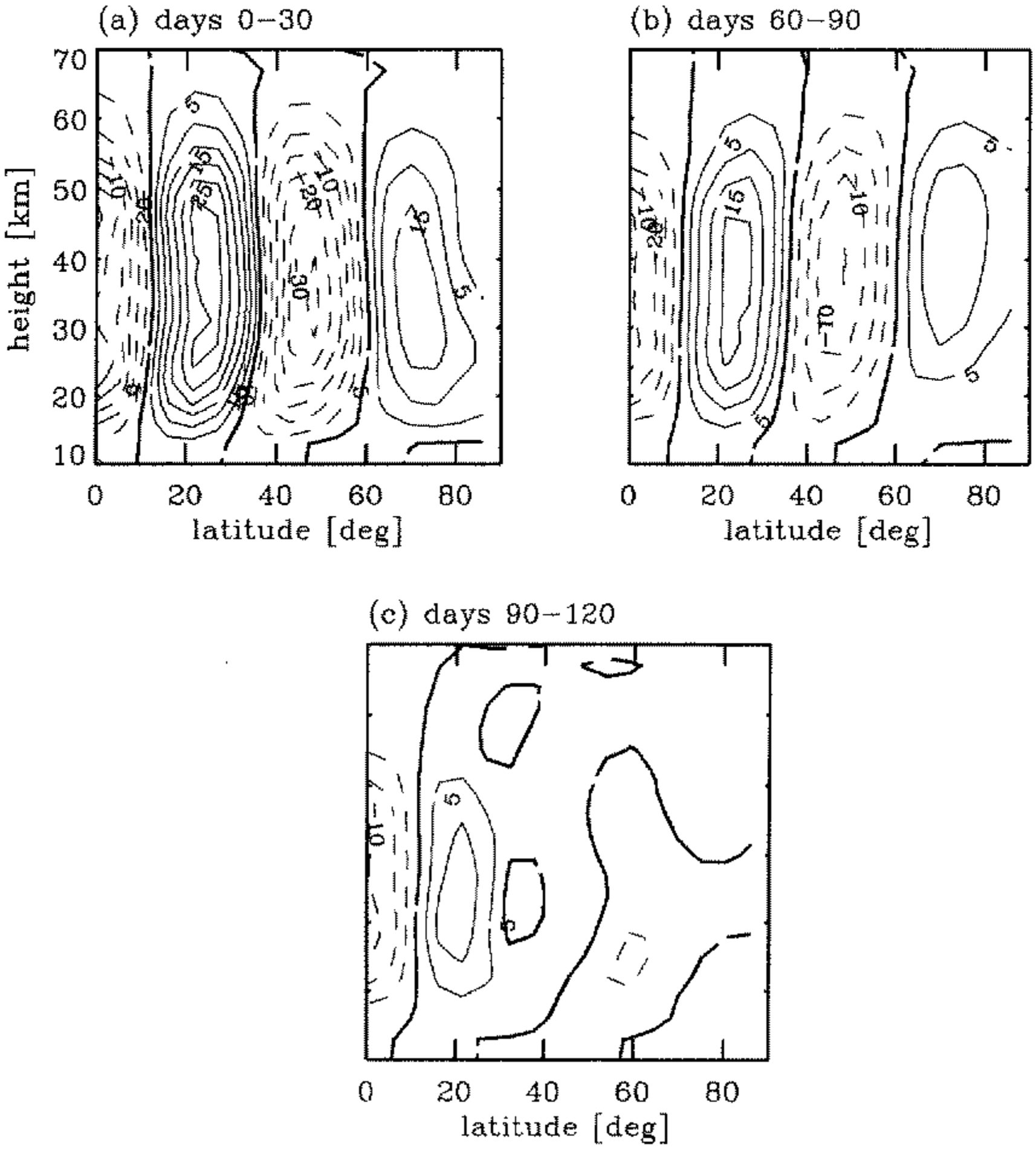

Figure 7. Monthly-mean latitudinal gradient of PV at constant potential temperature, $\bar{Q}_\phi$ anomalies, quiescent year minus disturbed year. Monthly means are averaged over days (a) 0–30, (b) 60–90, (c) 90–120. Thus (a,b) corresponds to Fig. 4(b,c). The contour interval is 5×10^{-6} K kg^{-1} m^{-2} s^{-1}.

of EP flux divergence anomalies observed in Fig. 5, especially those associated with the early and late winters of the quiescent years.

Insight into the likely differences between wave propagation in quiescent versus disturbed winters is best given by examining the $\bar{Q}_\phi$ anomalies, shown in Fig. 7. These show a markedly increased potential vorticity gradient at low latitudes (10°–30°) in the quiescent winters, which persists until day 120. Thus, in the quiescent winters there is likely to be greater wave propagation at these latitudes: the larger PV gradients at low latitudes will allow waves to enter that region and be absorbed there. At higher altitudes the wave amplitudes become larger, and the thermal and momentum damping stronger (Rayleigh friction above 50 km) and the waves are dissipated, causing the region of low-latitude flux convergence shown in Fig. 5(b,c).

In late winter the situation is different, and the wind fields themselves provide insight into the EP flux convergence patterns of Fig. 5(e,f). Following a disturbed winter, from around day 242, easterlies develop at middle latitudes and descend until they are deep and span the vertical, Fig. 8(c,d). Since this latitude range is also that at which the lower

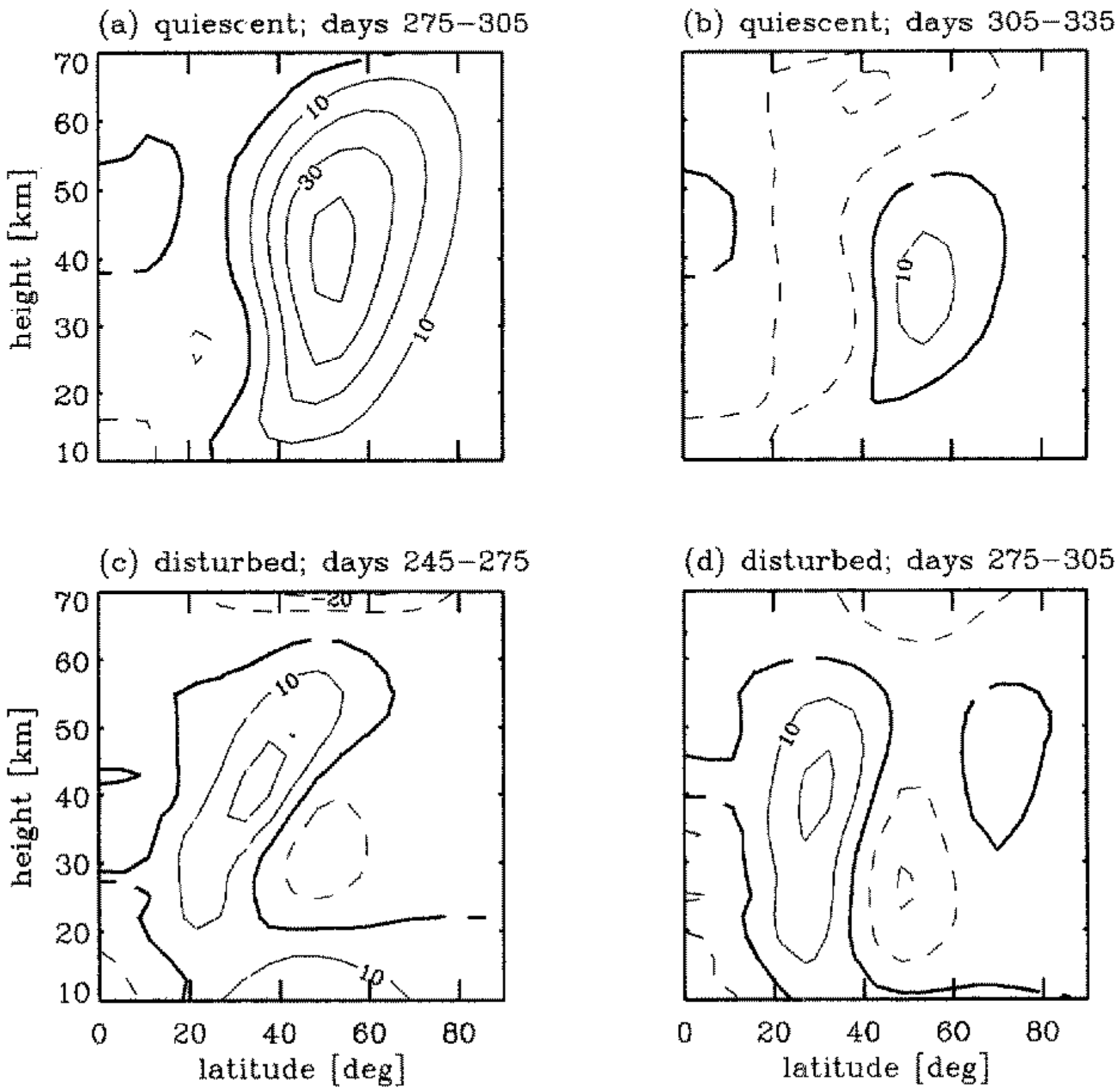

Figure 8. Monthly-mean zonal velocity fields in late winters of the quiescent and disturbed years, for $h_0 = 235$ m (from $t = 9$ to $t = 10$ years and from $t = 8$ and $t = 9$ years respectively in Fig. 3(c)): (a,b) quiescent, averaged over days 275–305 and days 305–335; (c,d) disturbed, averaged over days 245–275 and days 275–305. The contour interval is 10 m s^{-1}.

boundary forcing is applied, vertical wave propagation is prevented and the waves do not penetrate into altitudes above 30 km, this height descending as the zero wind line descends from day 240 to day 300. Following a quiescent winter, on the other hand, westerlies persist at most latitudes until day 300 and later, Fig. 8(a,b). Thus wave propagation is possible and takes two separate courses: one to low latitudes, following great circles, and dissipating in these regions where there are weak westerlies; the other upwards to high altitudes, and dissipating in these regions where there are increased damping rates. As mentioned above, it is the forcing of the mean flow associated with low-latitude dissipation of the waves that is important in setting the anomalous low-altitude easterly winds at the end of the quiescent winter.

3. Sensitivity to constraints on the low-latitude velocities

To confirm the hypothesis that it is the low-latitude wind anomalies that are responsible for the selection of either a disturbed or a quiescent winter evolution, simulations have been carried out in which the low-latitude zonal wind field is constrained to follow a particular temporal evolution. Both the latitudinal extent of the constraint and the associated

wind field itself have been varied in these simulations, to try to ascertain exactly which features, and at which latitudes, lead to a particular winter response.

The constraint is applied by including Rayleigh friction in the zonal mean momentum equation, relaxing the zonal-mean zonal velocity, $\bar{u}$, to the required velocity field, on a short latitude-dependent time-scale, $1/\kappa_L$, where

$$\kappa_L(\phi) = \frac{1}{5 \text{ days}} \begin{cases} \cos^2 \frac{\pi}{2} \frac{\phi}{\phi_L} & 0 \leqslant \phi \leqslant \phi_L \\ 0 & \phi > \phi_L \end{cases} . \tag{7}$$

The parameter ϕ_L therefore determines the width of the constrained region. There is no such additional friction applied to the wave-1 components of the velocity.

Two different time varying constraining velocity fields have been used, both taken from an unconstrained simulation that exhibits interannual variability (that shown in Fig. 3(c) with $h_0 = 235$ m). In one the constraining velocity field, U_Q, is one that leads into a quiescent winter; in the other the constraining velocity field, U_D, is one that leads into a disturbed winter. Both of these time evolving velocity fields are taken from the midwinter *before* the quiescent or disturbed winter through to the midwinter of that year itself. In terms of the evolution shown in Fig. 3(c), $U_Q(\phi, z, t)$ is the velocity field from time $t = 8.5$ to time $t = 9.5$ and $U_D(\phi, z, t)$ is the velocity field from time $t = 7.5$ to time $t = 8.5$. Each of these time series is then repeated with annual periodicity.

If ϕ_L were large enough the constraint would operate at almost all latitudes and the response would follow U_Q or U_D nearly everywhere, i.e. all winters would be either disturbed or quiescent, with no variability. On the other hand, if ϕ_L were very small the constraint should make very little difference to the flow response: if other parameters are as before the response should still be interannually variable with alternately disturbed and quiescent winters. If it is the case that the low latitudes in early winter control whether the subsequent evolution is disturbed or quiescent, then it should be the case that increasing ϕ_L past some critical, and relatively small, value should eliminate the variability.

Figure 9 (solid curves) shows the high-latitude (60°N, 40 km) evolution of six simulations using the two different constraining velocities, U_Q and U_D, and three different values of ϕ_L: 15°, 20° and 25°. It is seen that for $\phi_L = 15°$ (Fig. 9(a,d)) the constraint makes little difference and the response is still biennial as before. On the other hand, for $\phi_L = 25°$ (Fig. 9(c,f)) the constraint has sufficient influence over the high latitude evolution to produce either all disturbed or all quiescent winters, depending on which constraining velocity is used, U_D or U_Q. For the intermediate value, $\phi_L = 20°$, (Fig. 9(b,e)) an intermediate response is obtained: with the quiescent constraining velocity, U_Q, (Fig. 9(b)) all the winters are still quiescent, but less so than before; with the disturbed constraining velocity, U_D, (Fig. 9(e)) there is a clear biennial response, although each winter is still relatively disturbed.

Similar behaviour is also observed in the low-latitude (20°N, 25 km) evolution of the same six cases Fig. 9 (dashed curves). For $\phi_L = 15°$ (Fig. 9(a,d)) there is little difference between constraining to U_D or to U_Q. For $\phi_L = 25°$ (Fig. 9(c,f)) the 'response' is entirely governed by U_D and U_Q at 20°, which is to be expected since the velocities are constrained at this latitude. For $\phi_L = 20°$ (Fig. 9(b,e)) again the intermediate nature of the response is apparent: if the constraining velocity is U_Q, the velocities at the beginning of each year are between those of U_Q and U_D; if the constraining velocity is U_D, the velocities at the beginning of each year are variable, but within a much smaller range than the unconstrained case of Fig. 3(c).

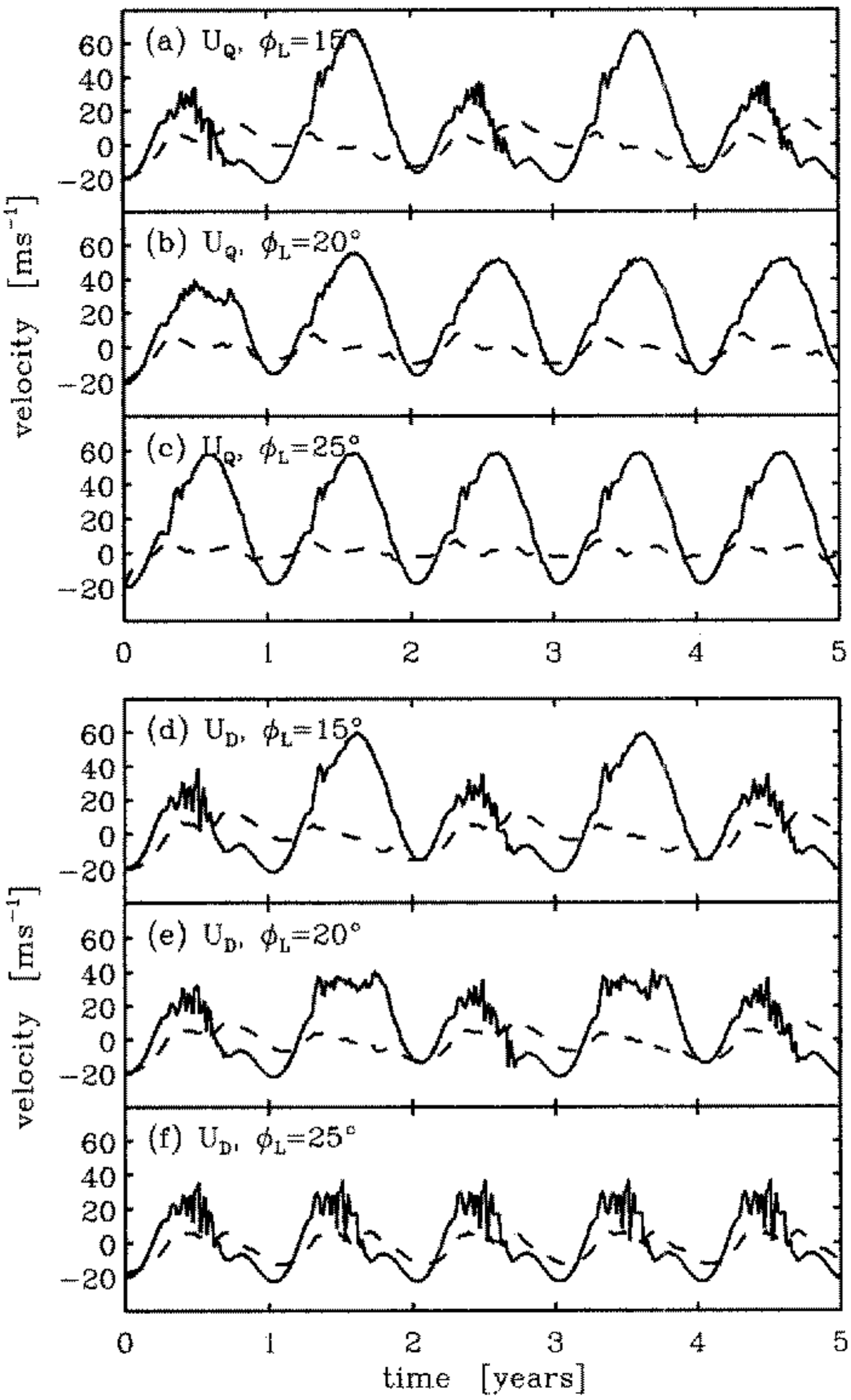

Figure 9. Evolution of the zonal mean velocity at 60°, 40 km (solid curves) and at 20°, 25 km (dashed curves), for $h_0 = 235$ km. These integrations are similar to Fig. 3(c) but now include low-latitude relaxation to the constraining velocities U_Q or U_D: (a) to (c) using U_Q, with $\phi_L = 15°, 20°, 25°$; (d) to (f) using U_D, with $\phi_L = 15°, 20°, 25°$.

Since, as seen from Eq.(7), the time-scale of the constraining relaxation is not constant over the latitude range $[0°, \phi_L]$, it is necessary to check the actual latitude range at which the velocity evolution is close to U_Q and U_D. Differences between quiescent and disturbed years were used as in Fig. 4, for the constrained cases above that produced variability, Fig. 9(a,d,e). It was found that for $\phi_L = 15°$ the velocities are constrained to within ± 2 m s^{-1} between 0° and 8°, and for $\phi_L = 20°$ the velocities are constrained to within ± 2 m s^{-1} of U_D between 0° and 13°. For $\phi_L = 25°$, there is no interannual variability so it is meaningless to take differences between quiescent and disturbed years. However, the difference between the velocity evolution $\bar{u}$, and the constraining velocities, U_Q and U_H shows that the velocities are constrained to within ± 2 m s^{-1} between 0° and 16°. Thus, for example, in the case

of $\phi_L = 25°$ with U_Q in Fig. 9(f), the tendency of the evolution to produce anomalous easterlies at low latitudes following the quiescent winters is being significantly reduced within [0°, 16°]. The velocities in this example are forced to follow a structure that would, in the unconstrained case, be present after a disturbed winter.

To be certain that imposing the low-latitude constraint for a particular ϕ_L is really altering or eliminating the interannual variability, rather than merely changing the forcing range for which variability occurs, four cases, with $\phi_L = 20°$ and $\phi_L = 30°$ for each of the constraining velocities U_Q and U_D, were repeated with different values of the lower boundary forcing. In these cases, constraining with U_Q always led to relatively quiescent winters and constraining with U_D always led to relatively disturbed winters. Therefore the cases with U_Q were repeated with stronger forcing until disturbed winters occurred and, conversely, the cases with U_D were repeated with weaker forcing until quiescent winters occurred.

The results are shown in Fig. 10 for $\phi_L = 20°$, and Fig. 11 for $\phi_L = 30°$. The general picture is that for $\phi_L = 20°$, stronger or weaker forcing values do exist that give rise to interannual variability in both the case constrained with U_Q and the case constrained with U_D. On the other hand for $\phi_L = 30°$, such forcing values do not exist in either the case U_Q or the case U_D.

A few details should be noted. In the case $\phi_L = 20°$, U_Q (Fig. 10(a–c)) increasing the forcing to $h_0 = 240$ m still results in quiescent winters though less quiescent than those resulting from $h_0 = 235$ m. Further increase to $h_0 = 245$ m leads to variability*. Despite the stronger forcing, the quiescent winters are more quiescent than with $h_0 = 240$ m. This is consistent with the idea that a disturbed winter in one year results in conditions more favourable to a quiescent winter in the following year. A similar pattern occurs in the case $\phi_L = 20°$, U_D (Fig. 10(d–f)). First, as the forcing is reduced to $h_0 = 230$ m, the magnitude of the variability increases—the quiescent winters are more quiescent. Then, as the forcing is reduced to $h_0 = 225$ m, the variability ceases, and all winters are quiescent. This time, despite the weaker forcing, these winters are less quiescent than in the variable case with $h_0 = 230$ m. On the other hand, in the case of $\phi_L = 30°$ (Fig. 11), the two sequences constrained with U_Q and U_D both show a gradual transition from either the quiescent or the disturbed regime to the other as the forcing is increased or decreased. Here variability is not observed for any of the intermediate forcing values considered. (The variability in the case U_U with $h_0 = 250$ m is slight and confined to the second half of each winter; each is essentially disturbed.)

It is interesting to note how the latitudinally averaged vertical EP flux through the lower boundary, $\langle F^{(z)}|_B \rangle$, is affected as the strength of the forcing is varied as above. An example is shown in Fig. 12(a)–(c), which corresponds to the three cases shown in Fig. 11(d)–(f), constraining to U_D with $\phi_L = 30°$, and $h_0 = 230, 225, 220$ m. The changes in $\langle F^{(z)}|_B \rangle$ more or less reflect the state of the high latitude winds: large $\langle F^{(z)}|_B \rangle$ corresponds to disturbed winters and small $\langle F^{(z)}|_B \rangle$ corresponds to quiescent winters. In this sense, it is possible to use $\langle F^{(z)}|_B \rangle$ as an index of the flow regime of a particular winter. This was done for a further set of experiments on the sphere, described briefly in the conclusions below. Further, large increases in $\langle F^{(z)}|_B \rangle$ can be associated with the flow being close to conditions for resonance, as investigated by Dunkerton *et al.* (1981), Clark (1992) and Hsu (unpublished, described in Clark 1992, Ch. 2). From this point of view, both the $\phi_L = 20°$ and $\phi_L = 30°$ constraints are insufficient to determine the resonant properties of the flow: for strong or weak enough forcing resonant or non-resonant like responses occur for both

* The collapse to zero winds in Fig. 10(c), and in Fig. 11(c) below, in the third model year is a result of numerical instability.

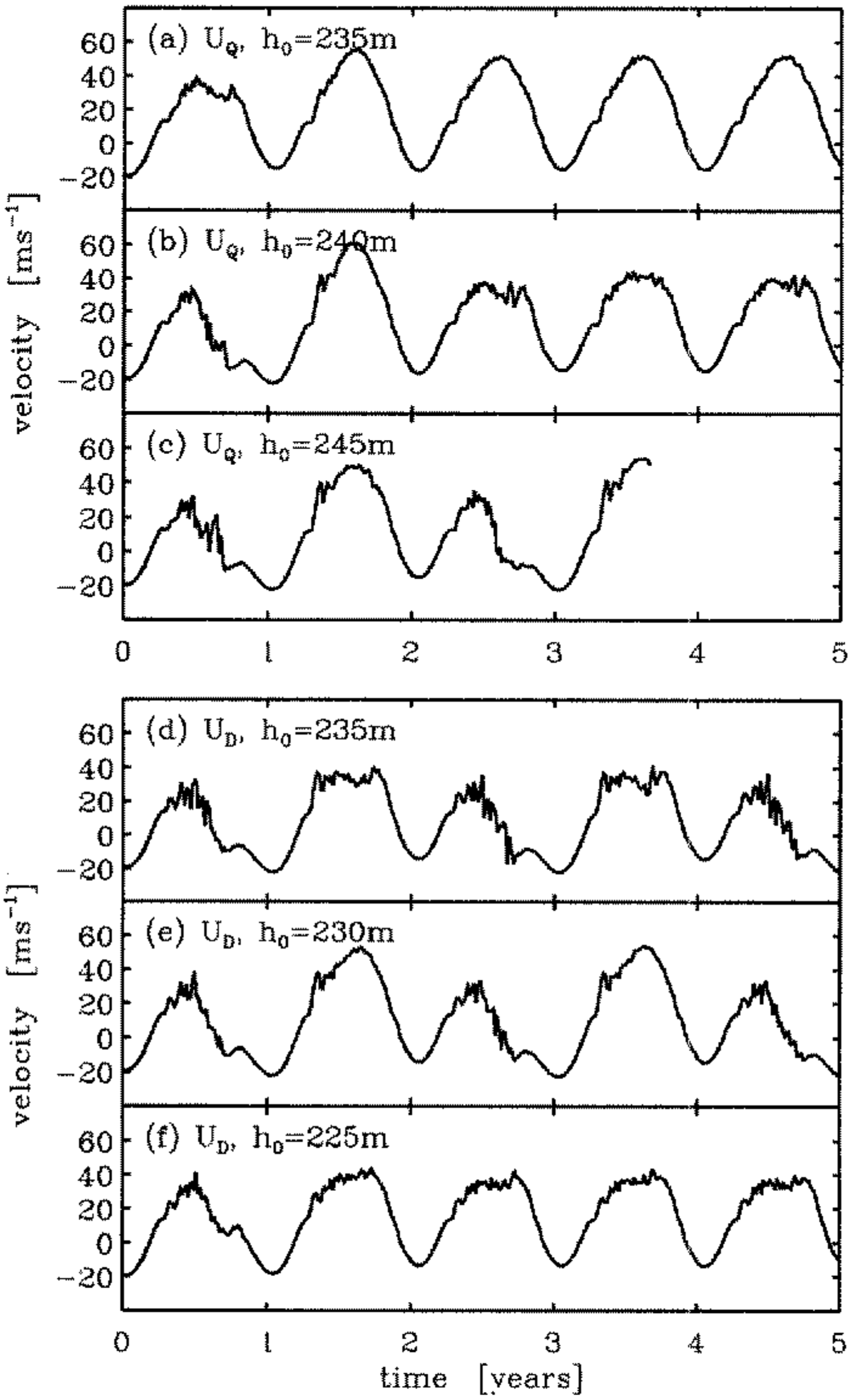

Figure 10. Evolution of the zonal mean velocity at 60°, 40 km constraining to U_Q or U_D with $\phi_L = 20°$. These are similar to Fig. 9(b,e) but with different wave-forcing amplitudes, h_0: (a) to (c) constraining to U_U, with $h_0 = 235$, 240, 245 m; (d) to (f) constraining to U_D with $h_0 = 235$, 230, 225 m.

U_Q and U_D. On the other hand, the $\phi_L = 30°$ constraint is sufficient to eliminate interannual variability whereas the $\phi_L = 20°$ constraint is not.

4. Biennial variability in a 'two-stream' model

(a) *Model formulation*

In this section a modification is introduced to the Holton–Mass model (Holton and Mass 1976, hereafter HM) to allow for the different time-scales for the persistence of velocity anomalies at different latitudes. The modification appeals to the properties of

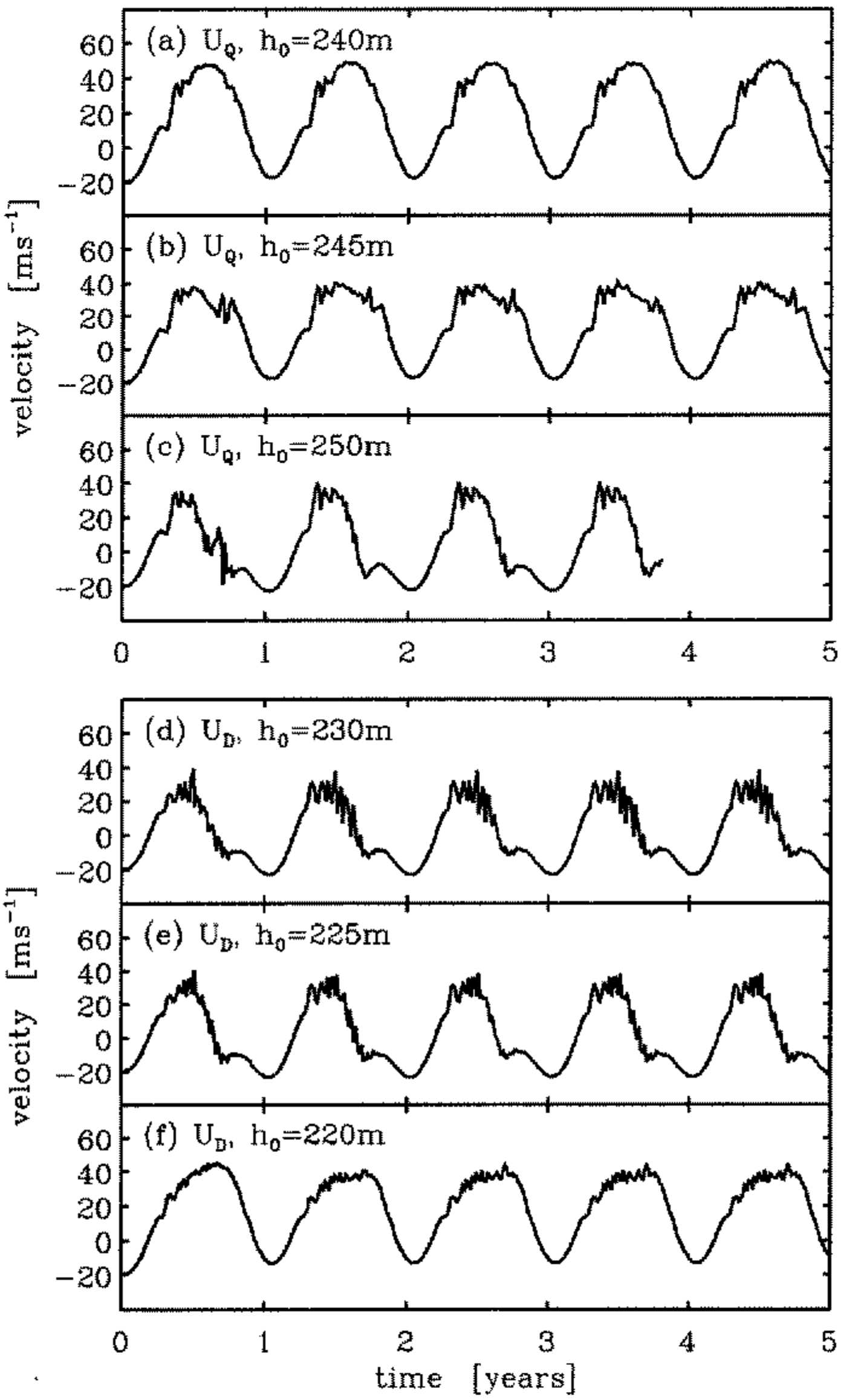

Figure 11. Evolution of the zonal mean velocity at 60°, 40 km constraining to U_Q or U_D with $\phi_L = 30°$. Again the wave-forcing amplitude is different between cases: (a) to (c) constraining to U_Q, with $h_0 = 240$, 245, 250 m; (d) to (f) constraining to U_D with $h_0 = 230$, 225, 220 m.

linear wave propagation on a sphere. It is well known (e.g. Karoly and Hoskins 1982) that Rossby waves generated in mid or high latitudes will tend to propagate equatorwards, following a great circle. Thus, planetary scale Rossby waves, propagating from, say, 100 mb in midlatitudes, will favour low latitudes in the absence of other factors, such as the effect of the mean flow structure. (Of course, if the low-latitude winds are in a state unfavourable to wave propagation, e.g. easterly, the waves will be confined to high latitudes.) Differences between high and low latitudes are not represented in the Holton–Mass model, which is based on a mid-latitude channel and assumes that zonal mean quantities and waves have a pre-specified latitudinal structure, usually the greatest Fourier mode across the

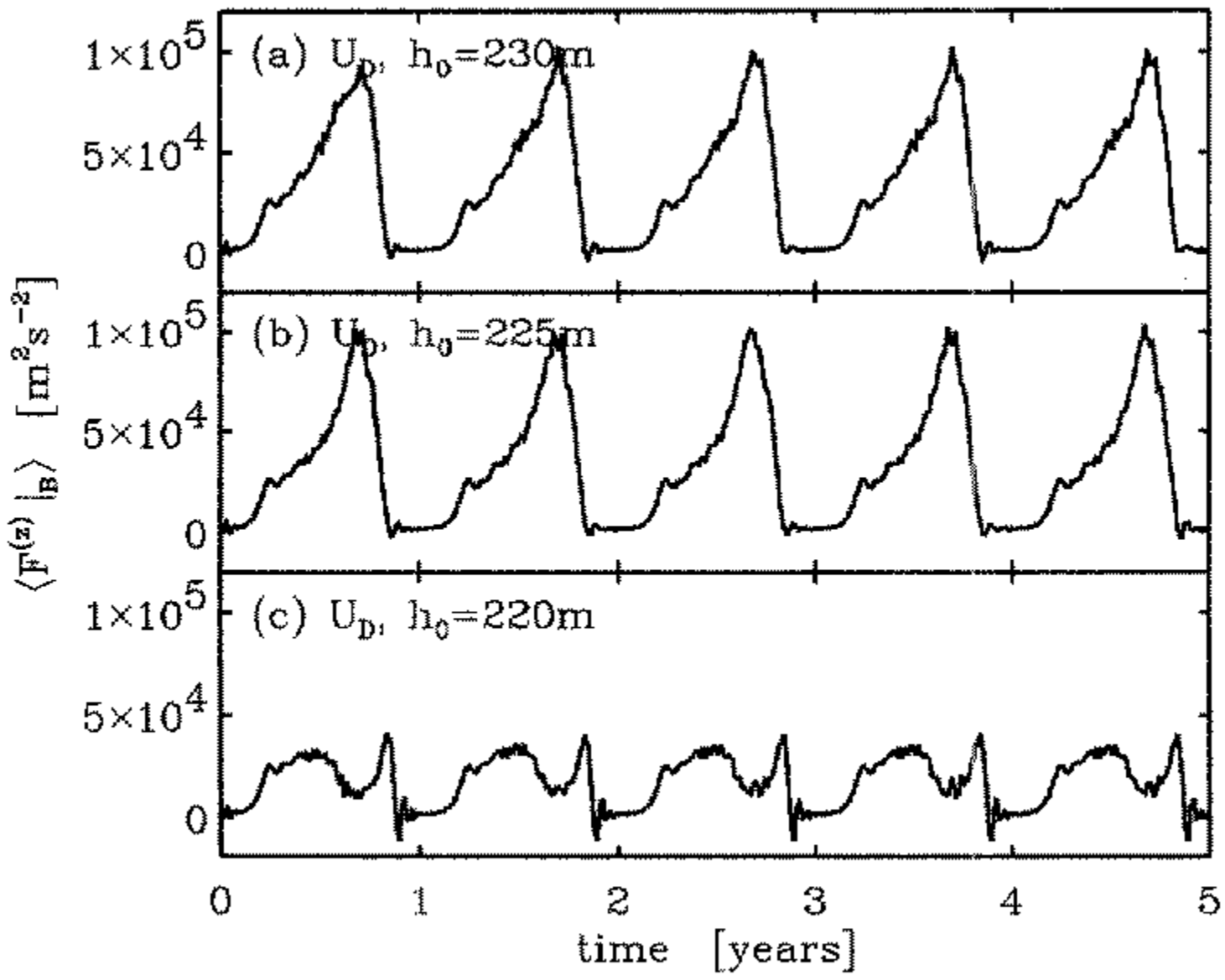

Figure 12. Latitudinally averaged, vertical EP flux through the lower boundary, $\langle F^{(z)}|_{\mathrm{B}}\rangle$, for the three cases shown in Fig. 11(d) to (f): constraining to U_{D} with $\phi_{\mathrm{L}} = 30$ and $h_0 = 230$, 225, 220 m respectively in (a) to (c).

channel. The simplest relevant modification to the Holton–Mass model, therefore, is to allow two zonal velocities, one low-latitude and one high-latitude, which receive more or less of the wave amplitude according to their strength. This allows the possibility that the relaxation rate for the low-latitude velocity is less than that for the high-latitude velocity. We emphasize that the model thus formed is not intended to be a realistic representation of the atmosphere. Rather, the intention is to illustrate how the addition of latitudinal-like structure in a very simple model can lead to internal interannual variability of the form discussed above.

The Holton–Mass model evolution equations derive from the zonally and latitudinally truncated quasi-geostrophic potential vorticity equation (HM, Eq.(1)). This can be written in the form of evolution equations for zonal velocity, $U(z)$, and perturbation streamfunction, $\Psi(z)$ (HM, Eqs.(9), (10)). Neglecting damping terms, these evolution equations take the form

$$\mathscr{L}_\Psi \Psi = \mathscr{F}_\Psi[U, \Psi] \tag{8}$$

$$\mathscr{L}_U U = \mathscr{F}_U[\Psi] \tag{9}$$

where $\mathscr{L}_\Psi$ and $\mathscr{L}_U$ are linear operators, which are independent of, respectively, U and Ψ, and $\mathscr{F}_\Psi$ and $\mathscr{F}_U$ are functionals of U and Ψ. Equation (8) describes the time evolution of the waves and Eq. (9) the time evolution of the zonal flow. The form of these operators and functionals depends on the zonal and latitudinal wavenumbers.

For the modified model we now assume that there are two separate regions into which the waves can propagate, low-latitude and high-latitude, including separate zonal velocities U_{L} and U_{H}, and that the wave evolution in turn depends on some combination of these velocities. Our modified equations take the form

$$\mathscr{L}_\Psi \Psi = \mathscr{F}_\Psi[a_{\mathrm{L}} U_{\mathrm{L}} + a_{\mathrm{H}} U_{\mathrm{H}}, \Psi] \tag{10}$$

$$\mathscr{L}_U U_{\mathrm{L}} = b_{\mathrm{L}} \mathscr{F}_U[\Psi] \tag{11}$$

$$\mathcal{L}_U U_{\mathrm{H}} = b_{\mathrm{H}} \mathcal{F}_U[\Psi] \tag{12}$$

for some choice of a_{L}, a_{H}, b_{L}, b_{H}. Equation (10) is the modified analogue of Eq. (8), describing the evolution of the waves. Since it is envisaged that the waves (in general) extend over both high-latitude and low-latitude regions, it is expected that Ψ depends both on U_{L} and U_{H}. The original velocity, U has therefore been replaced in the wave equation (Eq. (10)) by a linear combination of the new low- and high-latitude velocities, $a_{\mathrm{L}}U_{\mathrm{L}} + a_{\mathrm{H}}U_{\mathrm{H}}$, to represent this dependence in the simplest possible way. Equation (11) and Eq. (12) are the modified analogues of Eq. (9), but separately describe the evolution of the low- and high-latitude velocities. The coefficients b_{L} and b_{H} indicate the effect of the waves on the mean flow in each of the two regions. The further assumption that the total effect of the waves on this combination of high- and low-latitude velocities is the same as the effect of the waves on the single original velocity in Eq. (9), i.e.,

$$a_{\mathrm{L}}\mathcal{L}_U U_{\mathrm{L}} + a_{\mathrm{H}}\mathcal{L}_U U_{\mathrm{H}} = \mathcal{F}_{\mathrm{U}}(\Psi) \tag{13}$$

implies the condition:

$$a_{\mathrm{L}}b_{\mathrm{L}} + a_{\mathrm{H}}b_{\mathrm{H}} = 1. \tag{14}$$

We require that the model be consistent with two important limiting cases. When the waves are entirely confined to low latitudes they should be independent of the high-latitude velocity, so $a_{\mathrm{H}} = 0$. Similarly the high-latitude velocity should be independent of the waves, so $b_{\mathrm{H}} = 0$. The model should then be equivalent to a 'low-latitude' Holton–Mass model, hence $a_{\mathrm{L}} = b_{\mathrm{L}} = 1$. Correspondingly, when the waves are entirely confined to high latitudes $a_{\mathrm{H}} = b_{\mathrm{H}} = 1$ and $a_{\mathrm{L}} = b_{\mathrm{L}} = 0$. We further require that $0 \leqslant a_{\mathrm{L}}, a_{\mathrm{H}} \leqslant 1$.

To allow the model to represent the atmosphere in the sense that wave propagation favours either the high-latitude region or the low-latitude region we make a_{L}, a_{H}, b_{L}, b_{H} functions of the low-latitude flow U_{L} at $z = 20$ km. We define

$$\gamma = \begin{cases} 1 - \dfrac{U_{\mathrm{L}}}{U_{\mathrm{C}}} & 0 < U_{\mathrm{L}_{20}} < U_{\mathrm{C}} \\ 1 & U_{\mathrm{L}_{20}} \leqslant 0 \\ 0 & U_{\mathrm{L}_{20}} \geqslant U_{\mathrm{C}} \end{cases} \tag{15}$$

for some U_{C}, where $U_{\mathrm{L}_{20}} = U_{\mathrm{L}}$ ($z = 20$ km), and set

$$a_{\mathrm{L}} = b_{\mathrm{L}} = \cos\frac{\gamma\pi}{2}, \quad a_{\mathrm{H}} = b_{\mathrm{H}} = \sin\frac{\gamma\pi}{2} \tag{16}$$

which preserves the condition of Eq.(14). Thus when $U_{\mathrm{L}_{20}} \leqslant 0$, $a_{\mathrm{L}} = b_{\mathrm{L}} = 0$, ($\gamma = 1$), waves do not propagate into the low-latitude region and the zonal velocity evolves only at high latitudes; when $U_{\mathrm{L}_{20}} \geqslant U_{\mathrm{C}}, a_{\mathrm{H}} = b_{\mathrm{H}} = 0, \gamma = 0$), waves do not propagate into the high-latitude region and the zonal velocity evolves only at low latitude; and when $0 < U_{\mathrm{L}_{20}} < U_{\mathrm{C}}$ there is a smooth transition between these two extremes with wave propagation and zonal velocity evolution in both regions. This choice of a_{L}, a_{H}, b_{L}, b_{H} is not the only possible one, but serves adequately and requires the choice of only one extra external parameter, U_{C}.

The damping is reintroduced to Eqs.(10)–(12) using separate relaxation time-scales α_{L} and α_{H} for relaxation to separate 'radiative' equilibrium velocities U_{RL} and U_{RH} in low- and high-latitude regions respectively. Consistent with the arguments presented in

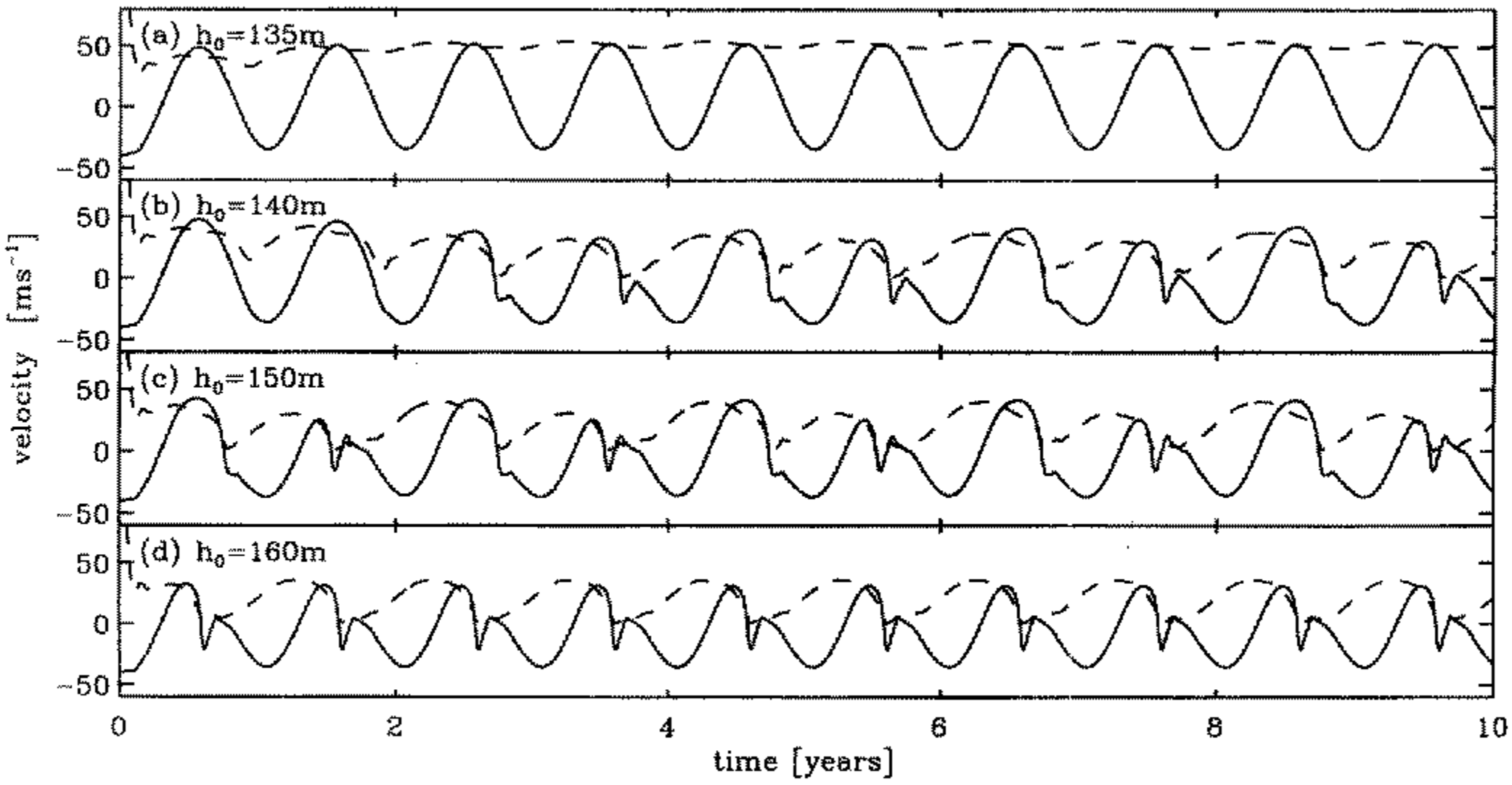

Figure 13. Evolution of the high-latitude velocity, U_H, at $z = 35$ km (solid curves) and low-latitude velocity, U_L, at $z = 20$ km (dashed curves, exaggerated five-fold) for ten year integrations of the two-stream Holton–Mass model with different forcing amplitudes: (a) $h_0 = 135$ m, (b) $h_0 = 140$ m, (c) $h_0 = 150$ m, (d) $h_0 = 160$ m. Time $t = 0$ corresponds to midsummer.

the appendix, the low latitude time-scale is generally chosen to be longer than the high latitude time-scale. For simplicity the thermal relaxation applied to the waves is set equal to α_H. This is justified on the basis that the latter represents not just thermal dissipation, but a number of different processes that lead to wave dissipation, including wave breaking.

(*b*) *Results*

Consecutive years of U_H and U_L (exaggerated by a factor of five) for various forcing values are shown in Fig. 13, for the case $U_C = 12\,\mathrm{m\,s^{-1}}$. The radiative equilibrium velocities used here are given by $U_{RL} = 22\,\mathrm{m\,s^{-1}}$ and $U_{RH}(t) = [10 - 2z' \cos \frac{2\pi t}{365}]\,\mathrm{m\,s^{-1}}$, with z' in km above the lower boundary at 10 km, and t in days after midsummer. The radiative relaxation rates, α_H and α_L, are such that $\alpha_L = \frac{1}{10}\alpha_H$, with α_H as in HM Eq.(12). The results show a similar pattern to that obtained using the primitive equation model on the hemisphere: for weak forcing, $h_0 = 135$, each year is quiescent; for strong forcing, $h_0 = 160$, each year is disturbed; for intermediate forcing, $136 \leqslant h_0 \leqslant 155$, disturbed and quiescent years alternate. Interannual variability was also found for a range of values of the parameters U_C and U_{RL}, with larger U_C requiring larger U_{RL} and larger h_0 for similar behaviour.

Figure 13 shows that for cases that exhibit disturbed winters, i.e. for $h_0 > 135$, U_L lies in the range [0, 8]. Equivalently, γ lies in the range [0.3, 1]. Further, γ is typically smaller (U_L is larger) in the early winters. Since small γ corresponds to wave propagation into the low-latitude region, it follows that if U_L is large enough in early winter there is less wave propagation into high latitudes and a quiescent winter follows. By mid-late winter the low-latitude waves decrease U_L so that waves are focused again into high latitudes and a final warming type event occurs. Because of the increase in EP flux convergence associated with this event, U_L also decreases sharply at this time, focusing more waves into high latitudes.

On the other hand, if U_L is smaller in early winter, there is greater wave propagation into high latitudes. This results in an earlier sudden warming type event, occurring this time in midwinter. As before, this also forces a sudden decrease in U_L to nearly zero,

at which point waves are again directed only into the high-latitude regions. Since this deceleration occurs earlier in the disturbed winter, U_L also begins its recovery to U_{RL} earlier in the disturbed winter. This results in a larger U_L after the disturbed winter than after the quiescent winter.

Thus, the winters are characterized by the timing of the sudden warming-type event: a midwinter—major warming—occurrence corresponding to a disturbed winter, and a late winter—final warming—occurrence corresponding to a quiescent winter. Since the low latitude velocity, U_L, exhibits similar decreases to U_H, the picture obtained is no longer the simplified one in which waves either affect the low latitudes or the high latitudes, one or the other. The variability is characterized by the timing of the warming, which in turn leaves a signal that persists until the following winter. The difference in low- and high-latitude damping rates implies that the signal remains strongest in the low-latitude region.

The crucial role of the relative weakness of the low-latitude relaxation in allowing the interannual variability was confirmed by varying α_L. Recall that in the simulations reported above the relaxation rates satisfy $\alpha_L = 0.1\alpha_H$, giving $\alpha_L \simeq \frac{1}{100\text{ days}}$ in the lower stratosphere. Focusing on the case $U_C = 12$, $U_{RL} = 22$ above and increasing α_L, shows that no variability occurs for $\alpha_L > 0.12\alpha_H$, or $\alpha_L \simeq \frac{1}{180\text{ days}}$. Varying h_0 for this value of α_L, obtains instead a gradual transition from quiescent to disturbed winters with increasing h_0. This is characterized by the appearance of a sudden warming that occurs in late winter for small h_0—quiescent winters—and occurs in earlier winter for larger h_0—disturbed winters. Again we conclude that interannual variability occurs only when the 'memory' of the low-latitude flow lasts from year to year.

5. Conclusions

The results presented in this paper have demonstrated the possibility of internal interannual variability in stratosphere-only models, arising from the relatively long memory of low-latitude velocity anomalies. The longer memory is due to the fact that low-latitude velocity anomalies have, due to the latitudinal variation of the Coriolis parameter, relatively weaker associated temperature anomalies than those at high latitudes. The interannual variability has been demonstrated both in a hemispherical model and in an *ad hoc* extended, 'two-stream' β-plane model.

The interannual variability was manifested by a biennial signal in the winter evolution, with alternately dynamically disturbed and dynamically quiescent winters. (However one might imagine that less coherent modes of interannual variability might also be allowed by the memory from one year to the next.) The biennial variability was found for a range of forcing amplitudes. Forcing amplitudes that lie above or below this range produce, respectively, a winter response that is always disturbed or always quiescent.

Both the models considered have been severely truncated in the zonal direction to make computationally feasible the possibility of many sets of multi-year runs. We argue that such models contain the essential ingredients of two-way interactions between waves and mean flow that would be present in models with isotropic horizontal resolution, but clearly extending this investigation to such models is a priority for future work.

For the hemispherical model discussed in section 2 it was noted that a deep westerly anomaly centred around 20° latitude, persistent from the end of one winter through to the beginning of the next, was the likely cause of the selection of a quiescent winter. The anomaly itself was seen to result from the previous winter evolution, or, more precisely, from the difference between the previous disturbed and the quiescent winter evolutions.

Various pieces of evidence have been presented to strengthen the claim that this subtropical anomaly, rather than an equatorial anomaly also observed, is a vital component

of the variability. In particular, a low-latitude relaxation on the zonal velocity was used to damp out anomalies within a specified latitude range about the equator. Even when the constraining velocity eliminated the equatorial anomaly between 0° and 12° interannual variability was still observed. It was necessary to eliminate the subtropical anomalies observed in the unconstrained simulations (westerly before a quiescent winter) to completely eliminate the variability.

All results presented here were for models on a hemispheric domain. Similar sequences of experiments have been performed on a full spherical domain, with wave forcing only in one hemisphere and show a similar biennial variability. In addition, less regular variability, in which a quiescent winter occurred approximately every three years was also obtained for a certain forcing range. In this regime half of the disturbed years were followed by a quiescent year and half by another disturbed year. Analysis revealed that the only difference between these cases was the presence of a subtropical anomaly before the quiescent year. A second approach indexed each year with the annually averaged vertical EP flux through the lower boundary, the higher the index, the more disturbed the year. There was a strong correlation between the lowest values of the index of a particular year and the strength of the subtropical westerly anomaly at the beginning of that year.

What is the relevance of this work to processes in the real atmosphere? Variability in the real stratosphere arise at least partly through variability in the external conditions, e.g. the magnitude of disturbances in the troposphere, or the low latitude winds influenced by the phase of the equatorial QBO (regarded as external to the extratropical stratosphere). One natural question, therefore, is this: if the internal modes illustrated here are also present in the real atmosphere, to what extent will the variability associated with these dominate or be dominated by the effects of external variability? Even if external variability dominates, the mechanism identified here may be important in determining the response of that variability.

Regarding the tropospheric variability, some indication of the relative importance of the mechanisms might be gained by comparing the magnitude of the interannual variability of geopotential height disturbances at, say, the 200 mb height level with the range of model forcing amplitudes that lead to model variability. In the studies in this chapter, the width of the forcing range that produces variability is approximately 15%. Thus it could be argued that, if this range is of the same order as that observed in the upper troposphere, variability arising from the longer memory of low latitudes might not be entirely swamped by the variability of the upper troposphere.

Certainly a better understanding of the mechanisms for internal variability will allow better statistical interpretation of long term variability. The longer memory of the subtropics has been clearly demonstrated here and may well have to be taken into account in such interpretation. Current statistical analyses of long term interannual variability typically treat each year independently of all other years, i.e. the flow is assumed to have no memory between years. The results of this investigation suggest that this may not always be a valid assumption.

Acknowledgements

We are grateful to Michael McIntyre for helpful conversations on this work, and to two anonymous referees and to Ted Shepherd and John Hampson for useful comments on the paper. Financial support was provided by the UK Natural Environment Research Council and the UK Meteorological Office through a CASE studentship. The work of the Centre for Atmospheric Science is supported by the UK Natural Environment Research Council through the UK Universities Global Atmospheric Modelling Programme and by the Isaac

Newton Trust. RKS is currently supported through the US Office of Naval Research grants N00041-93-1-1301 and N00041-95-1-0047.

Appendix

Effective relaxation at low-latitudes

We illustrate here how the rate at which momentum anomalies are damped by thermal or momentum damping depends upon the Coriolis parameter. These ideas have been illustrated in, e.g., Garcia (1987); Holton *et al.* (1995); Haynes (1998), for spherical geometry; they are here described on the f-plane for simplicity.

Under the quasi-geostrophic and Boussinesq approximations, the linearized TEM primitive equations for zonally symmetric flow on the f-plane are

$$\frac{\partial u}{\partial t} - fv = -\kappa(u - u_R) \tag{A.1}$$

$$\frac{\partial \theta}{\partial t} + w\theta_{0z} = -\alpha(\theta - \theta_R) \tag{A.2}$$

$$fu_z + \frac{g}{\theta_0}\theta_y = 0 \tag{A.3}$$

$$v_y + w_z = 0 \tag{A.4}$$

using standard notation; κ is the momentum damping rate, α is the heating/cooling rate. We consider perturbations to a steady, balanced, system with $u = u_R, \theta = \theta_R$. Then Eqs.(A.1)–(A.4) with $u_R = \theta_R = 0$ govern the evolution of the perturbation. These can be combined to form a single equation for u:

$$\left(\alpha + \frac{\partial}{\partial t}\right) f^2 u_{zz} + \left(\kappa + \frac{\partial}{\partial t}\right) N^2 u_{yy} = 0 \tag{A.5}$$

where $N^2 = g\theta_{0z}/\theta_0$. Substituting $u \propto \exp\{ily + imz + i\omega t\}$ into Eq.(A.5), leads to the dispersion relation

$$\omega = -i\alpha\left(1 + \frac{N^2 l^2}{f^2 m^2}\right)^{-1} - i\kappa\left(1 + \frac{f^2 m^2}{N^2 l^2}\right)^{-1}. \tag{A.6}$$

Since $\mathrm{Im}(\omega) < 0$ this is an exponentially decaying response. The rate of decay due to thermal damping ($\kappa = 0$) tends to zero as f tends to zero, for fixed l and m. On the other hand, the rate of decay due to momentum damping ($\alpha = 0$) tends to κ as f tends to zero.

Since the seasonal cycle in the real atmosphere arises as a result of seasonally varying radiative heating or cooling rather than frictional momentum damping, any anomaly in the momentum or temperature fields will decay at the latitudinally dependent rate given by Eq.(A.6) with $\kappa = 0$. As can be seen, this decay rate also depends on the meridional structure of the anomaly: vertically deep structures with small m will decay more slowly than shallow structures with larger m.

References

Andrews, D. G., Holton, J. R. and Leovy, C. B. 1987 *Middle atmosphere dynamics*. Academic Press, Orlando, USA

Clark, P. D. 1992 'Internal variability in the middle atmosphere.' *Ph.D. thesis*, University of Cambridge, UK

Dameris, M. and Ebel, A. 1990 The quasi-biennial oscillation and major stratospheric warmings: a three dimensional model study. *Annales Geophys.*, **8,** 79–86

Dunkerton, T. J. and Baldwin, M. P. 1991 Quasi-biennial modulation of planetary-wave fluxes in the Northern Hemisphere winter. *J. Atmos. Sci.*, **48,** 1043–1061

Dunkerton, T. J., Hsu, C.-P. F. and McIntyre, M. E. 1981 Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, **38,** 819–843

Fairlie, T. D. A., O'Neill, A. and Pope, V. D. 1990 The sudden breakdown of an unusually strong cyclone in the stratosphere during winter 1988/89. *Q. J. R. Meteorol. Soc.*, **38,** 819–843

Farrara, J. D., Fisher, M., Mechoso, C. R. and O'Neill, A. 1992 Planetary-scale disturbances in the southern stratosphere during early winter. *J. Atmos. Sci.*, **49,** 1757–1775

Garcia, R. R. 1987 On the mean meridional circulation of the middle atmosphere. *J. Atmos. Sci.*, **44,** 3599–3609

Haltiner, G. J. and Williams, R. T. 1980 *Numerical prediction and dynamic meteorology.* Wiley, New York.

Hamilton, K. 1995 Interannual variability in the northern hemisphere winter middle atmosphere in control and perturbed experiments with the GFDL SKYHI general circulation model. *J. Atmos. Sci.*, **52,** 44–66

Haynes, P. H. 1998 The latitudinal structure of the quasi-biennial oscillation. *Q. J. R. Meteorol. Soc.* (in press)

Haynes, P. H. and McIntyre, M. E. 1987 On the representation of translating Rossby-wave critical layers and wave breaking in zonally-truncated model. *J. Atmos. Sci.*, **44,** 2359–2382

Haynes, P. H., Marks, C. J., McIntyre, M. E., Shepherd, T. G. and Shine, K. P. 1991 On the 'downward control' of extratropical diabatic circulations by eddy-induced mean zonal-forces. *J. Atmos. Sci.*, **48,** 651–678

Holton, J. R. 1976 A semi-spectral numerical model for wave, mean flow interactions in the stratosphere: application to sudden stratospheric warmings. *J. Atmos. Sci.*, **33,** 1639–1649

Holton, J. R. and Austin, J. 1991 The influence of the equatorial QBO on sudden stratospheric warmings. *J. Atmos. Sci.*, **48,** 607–618

Holton, J. R. and Mass, C. 1976 Stratospheric vacillation cycles. *J. Atmos. Sci.*, **33,** 2218–2225

Holton, J. R. and Tan, H. C. 1982 The quasi-biennial oscillation in the Northern Hemisphere lower stratosphere. *J. Meteorol. Soc. Japan*, **60,** 140–148

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B. and Pfister, L. 1995 Stratosphere–troposphere exchange. *Revs. Geophys.*, **33,** 403–439

Hoskins, B. J. and Simmons, A. J. 1975 A multi-layer spectral model and the semi-implicit method. *Q. J. R. Meteorol. Soc.*, **101,** 637–655

Karoly, D. and Hoskins, B. J. 1982 Three-dimensional propagation of planetary waves. *J. Meteorol. Soc. Japan*, **60,** 109–123

Labitzke, K. 1982 On the interannual variability of the middle stratosphere during the northern winters. *J. Meteorol. Soc. Japan,*, **60,** 124–139

Lait, L. R. 1994 An alternative form for potential vorticity. *J. Atmos. Sci.*, **51,** 1754–1759

McIntyre, M. E. 1982 How well do we understand the dynamics of stratospheric warmings? *J. Meteorol. Soc. Japan*, **60,** 37–65

Mechoso, C. R., O'Neill, A., Pope, V. D. and Farrara, J. D. 1988 A study of the stratospheric final warming of 1982 in the Southern Hemisphere. *Q. J. R. Meteorol. Soc.*, **114,** 1365–1384

O'Neill, A. and Pope, V. D. 1988 Simulations of linear and nonlinear disturbances in the stratosphere. *Q. J. R. Meteorol. Soc.*, **114,** 1063–1110

O'Sullivan, D. and Dunkerton, T. J. 1994 Seasonal development of the extratropical QBO in a numerical model of the middle atmosphere. *J. Atmos. Sci.*, **51,** 3706–3721

Plumb, R. A. 1989 On the seasonal cycle of stratospheric planetary waves. *Pure Appl. Geophys.*, **130,** 223–242

Randel, W. J. 1987 A study of planetary waves in the southern winter troposphere and stratosphere. Part I: Wave structure and vertical propagation. *J. Atmos. Sci.*, **44,** 917–935

1988 The seasonal evolution of planetary waves in the Southern Hemisphere stratosphere and troposphere. *Q. J. R. Meteorol. Soc.*, **114,** 1385–1409

Scinocca, J. F. and Haynes, P. H. 1998 Dynamical forcing of stratospheric waves by the tropospheric circulation. *J. Atmos. Sci.*, **55,** 2361–2392

Scott, R. K. 1996 'Seasonal and interannual variations of the stratospheric circulation.' *Ph.D. thesis*, University of Cambridge, UK

Snieder, R. K. and Fels, S. B. 1988 The flywheel effect in the middle atmosphere. *J. Atmos. Sci.*, **45,** 3996–4004

Wallace, J. M. and Chang, F. C. 1982 Interannual variability of the wintertime polar vortex in the northern hemisphere middle stratosphere. *J. Meteorol. Soc. Japan*, **60,** 149–155

Yoden, S. 1990 An illustrative model of seasonal and interannual variations of the stratospheric circulation. *J. Atmos. Sci.*, **47,** 1845–1853