

THESIS

LOW-ORDER MODELS OF SUDDEN STRATOSPHERIC WARMINGS

Submitted by

Jeremiah P. Sjoberg

Department of Atmospheric Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2010

Master's Committee:

Department Head: Richard Johnson

Advisor: Thomas Birner

Richard Eykholt

Wayne Schubert

ABSTRACT

LOW-ORDER MODELS OF SUDDEN STRATOSPHERIC WARMINGS

Sudden stratospheric warmings (SSWs) are large scale events that are of first-order importance to understanding the wintertime dynamical variability in the polar stratosphere. These events are dynamically driven by wave-mean flow interaction of tropospherically-generated, vertically propagating planetary waves. These waves break at critical levels and subsequently force reversal of the zonal wind and a warming of the polar cap. Understanding the interaction between these planetary waves and the stratospheric zonal flow is of first-order importance in forcing SSWs. Characteristics of these planetary waves are investigated in simple models of stratospheric wave-mean flow interaction.

A linear stability analysis of the equilibrium states of these systems as functions of the incoming wave forcing amplitude is performed. Bifurcations of the zonal wind exist within the models, such that for sufficiently large quasi-stationary wave amplitudes, the zonal wind is forced into a SSW-like state. Transient momentum forcing due to quasi-random gravity wave activity is also taken into account by means of an additive noise term in the zonal momentum equation. This small-scale forcing needs to be parameterized in climate models and is usually considered to be of little importance in driving SSWs. It is shown that this noisy small-scale forcing can strongly affect the nature of stratospheric wave-mean flow interaction and the occurrence of SSWs in the truncated models.

This study also examines the zonal wind response to transient forcing of tropospheric planetary waves, motivated by the fact that as the forcing amplitude in the real atmosphere is observed to undergo strong temporal fluctuations. It is shown that for wave forcing periods less than the internal time scales of the model, modeled sudden warmings are highly

sensitive to the time scale of the prescribed tropospheric forcing in both systems. Forcing of SSWs in this transient case can only occur for wave amplitudes that are large relative to the quasi-stationary wave amplitudes necessary to force SSWs. Simple statistics derived from the ERA-40 and ERA-Interim datasets are shown to largely support this result, obtained with simple models, although more work is required to confirm this observational support.

ACKNOWLEDGEMENTS

I would like to thank Thomas Birner, whose tremendous guidance has taught me so much and led me on the path to this thesis. As well, I would like to thank Wayne Schubert and Richard Eykholt for being committee members fore my Master's defense. Thanks go out to Paul Williams who provided the Fokker-Planck code utilized in this study. Thanks also to my friends and family members who supported me and gave me the will to complete this task.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
1 INTRODUCTION	1
2 THE MODELS	8
2.1 THE HOLTON-MASS MODEL	8
2.2 THE LOW-ORDER MODEL	14
2.3 MULTIPLE WAVENUMBERS	17
3 STABILITY	20
3.1 SINGLE WAVENUMBER EQUILIBRIUM STATES	20
3.2 MULTIPLE WAVENUMBER EQUILIBRIUM STATES	26
3.3 LINEAR STABILITY ANALYSIS	31
3.4 HOLTON-MASS MODEL AND GCM STABILITY	38
4 SMALL-SCALE VARIABILITY	42
4.1 NOISY PROCESSES	42
4.2 THE FOKKER-PLANCK EQUATION	45
5 TRANSIENT FORCING OF SUDDEN WARMINGS	54
5.1 TRANSIENT WAVE AMPLITUDES	54

5.2	LOW-ORDER MODELING OF TRANSIENT AMPLITUDES	57
5.3	HOLTON-MASS MODELING OF TRANSIENT AMPLITUDES	69
5.4	OBSERVATIONAL AMPLITUDE TRANSIENCE	77
6	SUMMARY	84
6.1	CONCLUSIONS	84
6.2	OUTLOOK	85
	REFERENCES	90

1 INTRODUCTION

Sudden stratospheric warmings (SSWs) are large scale events that are of first-order importance to understanding the wintertime dynamical variability in the polar stratosphere. These dramatic events, first observed by Scherhag (1952), are marked by a polar vortex reversal on the order of days with a temperature increase that can be 50 K on the same timescale. The wind reversal is a downward propagating feature that begins near the stratopause — located at approximately 50 km altitude — and that descends to the lowest levels of the stratosphere — approximately 15 km. Observational evidence has also been presented that suggests significant tropospheric circulation modification can result from these events (Baldwin and Dunkerton 2001, Limpasuvan et al. 2004, Thompson et al. 2002).

The World Meteorological Organization (WMO) definition of a SSW requires that the 10 hPa zonal mean temperature gradient between the pole and 60°N reverse to become positive and that the 10 hPa zonal mean zonal wind at 60°N reverse to become easterly. If both these criteria are satisfied simultaneously, then the SSW event is labeled as a major warming event. If only the temperature gradient reversal criterion is satisfied, then the SSW event is labeled as a minor warming event. In this study, we focus solely on major warming events.

One way in which SSWs may be observationally identified is through potential vorticity (PV) contours on surfaces of constant potential temperature (isentropic surfaces). When viewed this way, one can observe that as solar forcing of the stratosphere wanes with the onset of winter, a large vortex of high potential vorticity builds over the polar region — 60°N-90°N. These high potential vorticity values are indicative of both the strong westerly

flow and the cold temperatures that denote the polar vortex. Figure 1 is potential vorticity on the 550 K isentropes for two dates in the winter of December 2001 – February 2002. The 550 K isentropes are chosen because they represent the lower stratosphere over the polar region. In Figure 1, the left plot shows a polar vortex that is slightly displaced off the pole's center, but otherwise is stable. This vortex of high PV will remain isolated from the lower PV values surrounding it so long as the vortex is stable.

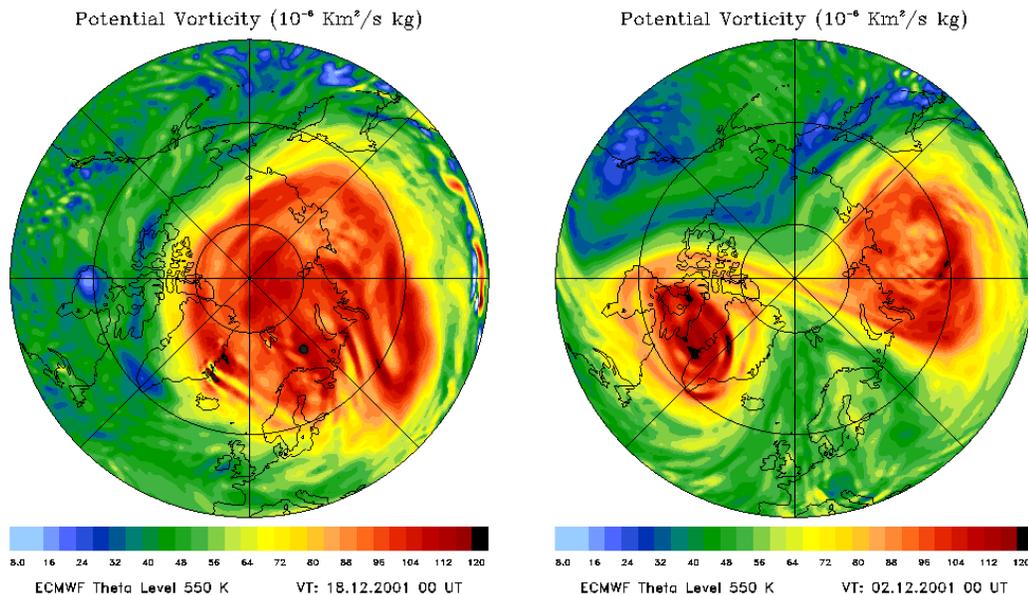


FIG. 1. Plots of potential vorticity on the 550 K isentropes. The 550 K isentropes are located in the lower stratosphere over the polar region. The left plot is of a stable polar vortex (noted in the shades of yellow to red). The right plot is of a SSW event that, following Charlton and Polvani (2007), is described as a polar vortex split event.

When a SSW occurs, one observes that the high PV of the stable vortex rapidly mixes with the surrounding environment, indicative of the zonal wind reversal and large temperature increase associated with the warming event. Charlton and Polvani (2007) noted that one way of classifying a SSW event is through observation of the PV (or perhaps geopotential) structure during an event. They note that SSWs fall into one of two modes of variability: a vortex split or a vortex displacement. A vortex split is characterized by the single vortex of high PV splitting into two largely independent vortices of high PV. An example of such a vortex split is plotted on the right side of Figure 1. The characteristic

feature of a vortex displacement, on the other hand, is that a majority of the high PV vortex has moved outside of the polar region (latitudes less than 60°N). If the sudden warming occurs in early to midwinter (early December to late January), the vortex can often be observed to restore either by the displaced vortex or by one of the split vortices moving back into the high latitudes (Schoeberl 1978).

The first physical insight into the dynamics of SSWs was proposed by Matsuno (1971). Matsuno suggested a spectral model of wave-mean flow interaction wherein quasi-stationary planetary waves of large amplitude generated in the troposphere propagate vertically into the stratosphere and break at critical levels on the polar vortex jet. These critical levels are identified by Charney and Drazin (1961) as levels where the zonal wind speed is westerly but less than some critical value (see Chapter 3). As the waves break, they deposit momentum and heat at the critical level. From this, the breaking waves induce an equatorward circulation and polar stratospheric heating. This southward flow is turned westward by the Coriolis torque, forcing a deceleration of the zonal flow. If this zonal flow deceleration continues, the westerly jet will be reversed and the level of wave breaking and zonal wind deceleration propagate downwards.

The dynamics of SSWs may also be seen through use of the transformed Eulerian mean (TEM) equations presented by Andrews and McIntyre (1976, see also Holton 2004, Chapter 10.2.2). We use TEM because it allows us to cast the large-scale eddy forcing terms in the stratosphere as a single forcing term affecting the zonal wind equation. To do this, Andrews and McIntyre begin by defining both a residual meridional circulation \bar{v}^* and a residual vertical circulation \bar{w}^* and combining with the quasi-geostrophic equations (discussed in Chapter 2), they arrive at the TEM zonal wind and continuity equations

$$\partial_t \bar{u} - f_0 \bar{v}^* = \frac{1}{\rho} \vec{\nabla} \cdot \vec{F} + \bar{X} \quad (1)$$

$$\partial_y \bar{v}^* + \frac{1}{\rho} \partial_z (\rho \bar{w}^*) = 0 \quad (2)$$

where $\vec{\nabla} \cdot \vec{F}$ is the Eliassen-Palm flux (EP flux) divergence. This term is the only term in the TEM equation set that depends explicitly on eddy fields and represents the single large-scale forcing term of the equations. The $\vec{\nabla} \cdot \vec{F}$ term is related to the meridional PV flux by

$$\frac{1}{\rho} \vec{\nabla} \cdot \vec{F} = \overline{v'q'}. \quad (3)$$

In a steady state with no friction ($\overline{X} = 0$), the zonal flow does not accelerate and the TEM zonal wind equation (1) becomes

$$f_0 \overline{v^*} = -\overline{v'q'}. \quad (4)$$

Then by taking the quasi-geostrophic potential vorticity equation (see equation (11), Chapter 2) and linearizing about a basic state, we may formulate the perturbation potential vorticity equation

$$\partial_t q' + \overline{u} \partial_x q' + v' \partial_y \overline{q} = S' \quad (5)$$

where S represents sources and sinks of PV. Multiplying 5 by the the perturbation PV and zonally averaging, we arrive at the meridional PV flux equation (see Holton and Dunkerton 1978)

$$\overline{v'q'} = \left[-\partial_t \left(\frac{1}{2} \overline{q'^2} \right) + \overline{v'S'} \right] / \partial_y \overline{q}. \quad (6)$$

The first term on the right-hand side represents wave transience and the second term represents wave dissipation.

The implication from (6) is that for vertically propagating waves ($\partial_t \left(\frac{1}{2} \overline{q'^2} \right) > 0$) that break at some critical level ($\overline{v'S'} < 0$) in the stratosphere, there is an equatorward flux of potential vorticity. Then from (3), there is a convergence of EP flux within the stratosphere. From (4), this potential vorticity flux induces a residual mean meridional circulation that is directed poleward. From the TEM continuity equation (2), this subsequently induces a

high latitude residual mean downward circulation. This induced circulation is closed by a lower level equatorward \bar{v}^* and associated lower latitude upward \bar{w}^* .

In a SSW event, this induced residual circulation is not enough to balance the EP flux convergence in (1), thus a zonal flow deceleration is induced. The residual downward circulation meanwhile warms the polar region through adiabatic compression. As in Matsuno (1971), the induced zonal flow deceleration lowers the wave breaking critical level and the region of induced residual circulation propagates downward until the EP flux convergence is weak relative to the residual circulation. Those regions above the critical level are then no longer affected by the wave forcing and gradually restore to the stable polar vortex state through radiative damping.

Matsuno (1971) — who did not have the TEM framework to aid in his interpretation — developed a quasi-geostrophic numerical model through which he was able to capture many of the fundamental characteristics of a SSW when forced with an abrupt amplification of tropospheric planetary waves. Geisler (1974) followed Matsuno’s work by considering a quasi-geostrophic beta plane channel model of wave-mean flow interaction. He showed that the essential dynamics of Matsuno (1971) can be well reproduced in this simpler, one-dimensional model.

Holton and Mass (1976) developed a model similar to Geisler’s quasi-geostrophic β -plane channel model, but with a zonal wind that varies meridionally. As Holton and Mass state, this assumption more closely models the real atmosphere because it allows greater vertical propagation of planetary waves. In this model, Holton and Mass demonstrate that waves with constant amplitude produce one of two mean zonal wind states: a state where the winds remain close to a prescribed radiative equilibrium speeds, and a state where the winds vacillate about 0 m/s. These vacillations are reminiscent of a SSW, though they are highly simplified. Holton and Mass show that the final state for a given bottom boundary wave amplitude (which we denote in this study by the parameter h) is dependent on the magnitude of that amplitude.

The value of this model (Holton-Mass model hereafter) is in its ability to produce primitive SSWs while only resolving the vertical dimension of a wave-mean flow interaction process with steady forcing. Many of the fundamental dynamics of SSWs are contained even within this simple system and may more easily be determined because the model does not contain the dimensional nor the parameter space complexities of global climate models (GCMs). We may extract further value from the Holton-Mass model by showing that with proper assumptions introduced by Ruzmaikin et al. (2003), the model may be further simplified in the vertical such that the one-dimensional Holton-Mass model may be expressed as a three layer model where all resolved fields exist only in the middle layer.

With this vertically-truncated form of the Holton-Mass model, hereafter referred to as our low-order model, one is able to model the most primitive version of a sudden stratospheric warming. The simplicity of the low-order model allows us to retain only the most fundamental physics associated with the stratospheric wave-mean flow interaction of a SSW. As Held (2005) states, there is great importance in developing a model hierarchy wherein the complexities of global climate models are first understood through more simple models capturing the underlying interactions that the higher order models seek to resolve. Following Held's reasoning, this low-order system allows us to model SSWs in a way which excludes complex parameterizations or chaotic behavior brought about through the full inclusion of nonlinearity.

Stemming from this simplicity, we can show that it is straightforward in the low-order model to increase the number of resolved wavenumbers. Whereas the models considered by Matsuno (1971) and Holton and Mass (1976) resolve only a single wavenumber, the low-order model may easily be modified to include multiple wavenumbers, allowing for deeper analysis of this wave-mean flow interaction. Furthermore, quantification of the stationary solutions and stability points of the model can be done analytically (or at least semi-analytically) rather than numerically. Another advantage of such a simple model is that the primary inputs into the system are likewise simplified to a set of terms: the bottom

boundary wave forcing and its associated wavenumber, the bottom boundary wind speed, and the vertical shear.

In this study, we shall prescribe the bottom boundary wind speed and vertical shear terms, thus allowing the primary input into the model to be the bottom boundary wave forcing. We reduce the main input in the model to only the bottom boundary wave amplitude term h for the following reason: it is the troposphericly generated waves that propagate vertically into the stratosphere that Matsuno (1971) showed were the principle cause of SSWs. Then if one were to conceptualize the most primitive form of a model for determining sudden stratospheric warming dynamics, this bottom boundary forcing should be most important input into that conceptual model.

These troposphericly generated planetary waves that are at the heart of this study. We shall firstly show how the models (Holton-Mass model and low-order model) are derived as a function of the bottom boundary forcing in Chapter 2. In Chapter 3, we explore the equilibrium states and stabilities of the models with respect to a stationary or quasi-stationary bottom boundary forcing. Chapter 4 describes a simple parameterization of small-scale variability that we include in the low-order model. Transient wave amplitudes and their effects in the Holton-Mass and low-order models are shown in Chapter 5.

2 THE MODELS

2.1 THE HOLTON-MASS MODEL

Our goal here is to derive a low-order model of stratospheric wave-mean flow interaction that captures the essential dynamics of sudden stratospheric warmings while only retaining a small parameter-space. To accomplish this, we consider the Holton-Mass model with appropriate boundary conditions, and apply vertical truncation following Ruzmaikin et al. (2003). The so-called Holton-Mass model we shall first consider is presented in Holton and Mass (1976).

Holton and Mass' model is derived from the quasi-geostrophic (QG) primitive equations of dry atmospheric motion on a β -plane. The form of these are presented in Andrews et al. (1987):

$$D_g u_g - f_0 v_a - \beta y v_g = X, \tag{7}$$

$$D_g v_g + f_0 u_a + \beta y u_g = Y, \tag{8}$$

$$\partial_x u_a + \partial_y v_a + \frac{1}{\rho} \partial_z (\rho w_a) = 0, \tag{9}$$

$$D_g \theta_e + w_a \partial_z \theta_0 = Q, \tag{10}$$

where

$$D_g \equiv \partial_t + u_g \partial_x + v_g \partial_y,$$

subscripts “g” represent geostrophic wind components, and subscripts “a” represent ageostrophic wind components. θ_e represents a small departure from a reference potential temperature $\theta_0(z)$. X and Y here represent friction and other nonconservative processes; for this analysis, these will be assumed to be zero. Q is a radiative heating term which restores the

temperature to T_0 , the radiative equilibrium temperature. Q is assumed to be

$$Q = -\alpha c_p (T - T_0),$$

while T_0 is assumed to satisfy thermal balance,

$$\partial_y T_0 = -f_0 \frac{H}{R} \partial_z u_R,$$

for $\partial_z u_R = \Lambda(t)$ such that $u_R(z, t) = u_R(0, t) + \Lambda(t)z$ is the mean radiative zonal wind.

Taking $\partial_x(8) - \partial_y(7)$ and combining with the thermodynamic equation (10) gives the QG potential vorticity equation

$$D_g q + \frac{f_0^2}{\rho} \partial_z \left(\frac{\alpha \rho}{N^2} \partial_z \psi \right) = 0 \quad (11)$$

where

$$q = f_0 + \beta y + \nabla^2 \psi + \frac{1}{\rho} \partial_z \left(\rho \frac{f_0^2}{N^2} \partial_z \psi \right)$$

and ψ is the geostrophic streamfunction. Equation (11) describes the conservation of QG potential vorticity with the inclusion of a diabatic term that is a function of the radiative heating.

Following Dickinson (1969) we take the zonal mean of the fields in (7-10) and combine to make the zonal wind equation. With the linearized form of the QG potential vorticity equation, our set of equations become

$$(\partial_t + \bar{u} \partial_x) q' + \partial_y \bar{q} \partial_x \psi' + \frac{f_0^2}{\rho} \partial_z \left(\frac{\alpha \rho}{N^2} \partial_z \psi' \right) = 0, \quad (12)$$

$$\begin{aligned} \partial_t \left[\partial_{yy} \bar{u} + \frac{f_0^2}{\rho N^2} \partial_z (\rho \partial_z \bar{u}) \right] &= -\frac{f_0^2}{\rho N^2} \partial_z [\alpha \rho \partial_z (\bar{u} - u_R)] \\ &+ \frac{f_0^2}{N^2} \partial_{yy} \left[\frac{1}{\rho} \partial_z (\rho \overline{\partial_x \psi' \partial_z \psi'}) \right], \end{aligned} \quad (13)$$

where the perturbation QG potential vorticity is

$$q' = \nabla^2 \psi' + \frac{f_0^2}{\rho^2} \partial_z \left(\frac{\rho}{N^2} \partial_z \psi' \right) ,$$

and the basic state potential vorticity gradient is

$$\partial_y \bar{q} = \beta - \partial_{yy} \bar{u} - \frac{f_0^2}{\rho} \partial_z \left(\frac{\rho}{N^2} \partial_z \bar{u} \right) .$$

This model assumes Newtonian damping to a specified radiative state, where the damping rate is given by

$$\alpha(z) = [\alpha_{mid} + \tanh\{(z - z_{mid})/H\}] * 10^{-6} \text{s}^{-1} . \quad (14)$$

Here $\alpha_{mid} = 1.5 \cdot 10^{-6} \text{ s}^{-1}$ is the midlevel radiative damping, $z_{mid} = 35 \text{ km}$ is the midlevel altitude, and $H = 7 \text{ km}$ is the scale height. Fig. 2 plots this radiative damping with respect to z for the Holton-Mass model (solid curve) and for the low-order model (dash-dot curve, discussed in Section 2.2). Expressing the damping in terms of a time scale, the lowest levels of the model damp on a time scale of approximately 23 days, the mid levels on a scale of approximately 8 days, and the upper levels on a scale of approximately 5 days.

Holton and Mass confine this to a β -plane channel centered at 60°N with a notional channel width of 60° . The channel height used here is 16 km (z_B) to 106 km (z_T), which differ from Holton and Mass (1976) and are amended from Holton (see `sudden_warming_model.m` file, Chapter 12, 2004). The model assumes zero normal flow across the latitudinal boundaries of the model such that the geostrophic streamfunction ψ' and the meridional wind $\partial_x \psi'$ are zero at these boundaries. It is likewise assumed that both the energy density of the wave perturbations and the thermal wind vanish at the upper boundary. At the lower boundary, Holton and Mass specify a winter zonal wind and wave perturbation.

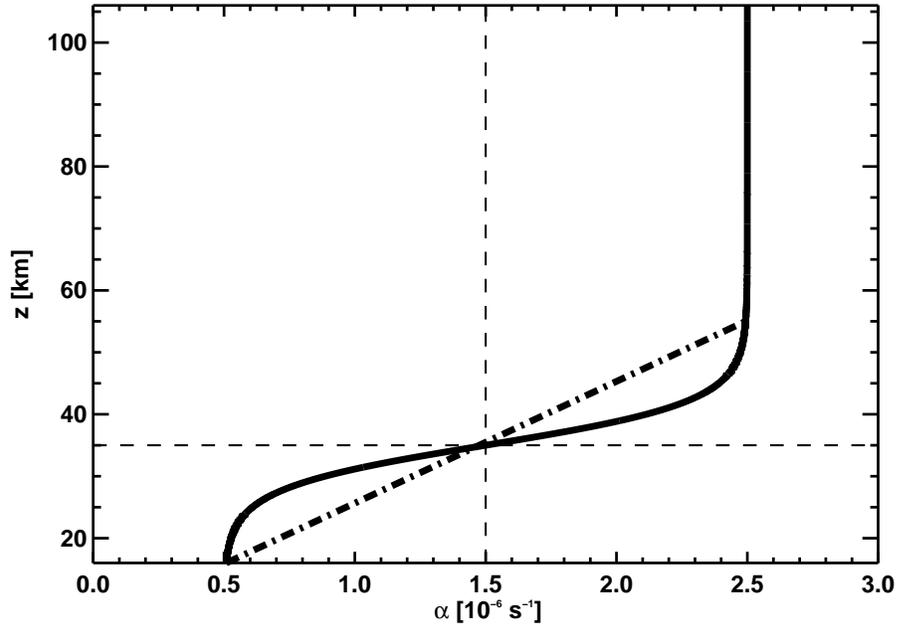


FIG. 2. Newtonian damping profile α as a function of height. The solid curve plots the profile for the Holton-Mass model, while the dash-dot curve plots the profile for the low-order model (discussed in Section 2.2). The dashed lines are plotted to show the value of α at the model midlevel.

The wavelike solution assumption is made, given by

$$\psi'(x, y, z, t) = \text{Re} [\Psi(z, t) e^{ikx}] e^{z/2H} \sin ly \quad (15)$$

$$\bar{u}(y, z, t) = U(z, t) \sin ly, \quad (16)$$

where $k = s/(a \cos \pi/3)$ is the zonal wavenumber, with $s = 1, 2, \dots$, the zonal planetary wavenumber. The meridional wavenumber $l = 3/a$ is specified such that $\sin ly$ is maximized at the channel center and 0 at the boundaries. A schematic of this is given as the solid curve in Fig. 3.

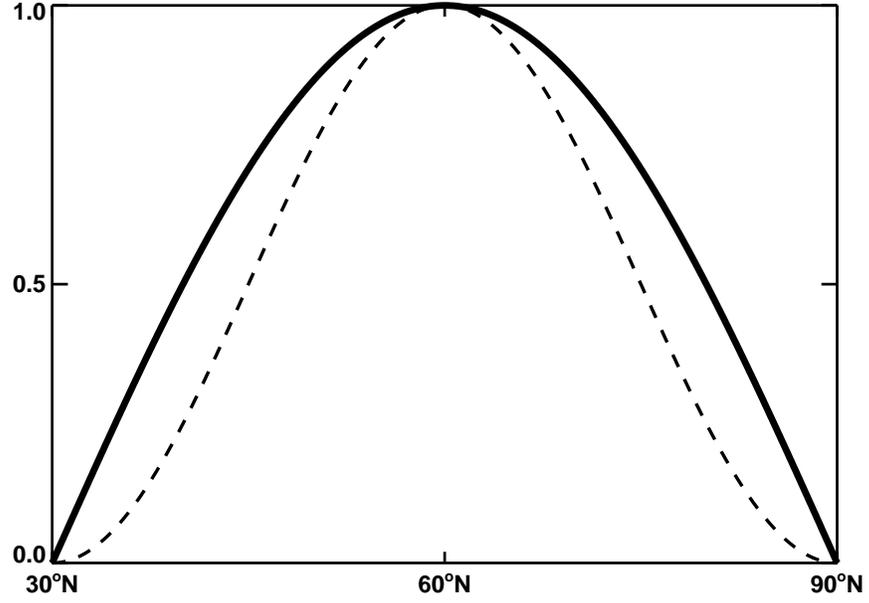


FIG. 3. Magnitudes of $\sin ly$ (solid) and $\sin^2 ly$ (dashed) as functions of latitude.

Using these wavelike solutions result in the following boundary conditions (described above):

$$\begin{aligned} \Psi(z_B, t) &= \frac{g}{f_0} h(t), & U(z_B, t) &= U_R(0, t), \\ \Psi(z_T, t) &= 0, & \partial_z U|_{z_T} &= \partial_z U_R|_{z_T} = \Lambda(t), \end{aligned}$$

where $h(t)$ is the wave forcing amplitude at the bottom of the model, the radiative equilibrium wind U_R is specified by the vertical shear $\partial_z U_R = \Lambda(t)$, and $\Lambda(t)$ is typically taken to be between 1 and 3 $\text{m s}^{-1} \text{ km}^{-1}$. Also from the inclusion of the wavelike solutions will arise terms with meridional dependencies defined by $\sin^2 ly$. Holton and Mass expand $\sin^2 ly$ into a Fourier sine series approximation

$$\sin^2 ly \approx \epsilon \sin ly$$

where higher order terms have been ignored and Holton and Mass assume $\epsilon \equiv 8/(3\pi)$. For

our experiments, we shall use the approximation $\epsilon \equiv 1$. We assume this approximation because the resolved fields for the model are at channel center, where $\sin^2 ly = \sin ly$. This may be seen in Figure 3, where the magnitudes of $\sin ly$ (solid curve) and $\sin^2 ly$ (dashed curve) are plotted within the nominal channel.

With the wavelike assumptions for the linearized quasi-geostrophic potential vorticity equation and the zonal mean zonal wind equation, together with the associated boundary conditions, we may solve these equations to the form given by Holton and Mass (1976):

$$\begin{aligned}
0 = & - \left[k^2 + l^2 + \frac{f_0^2}{N^2} \frac{1}{4H^2} \right] \partial_t \Psi \\
& + \frac{f_0^2}{N^2} [\partial_t + \bar{u}ik + \alpha] \partial_{zz} \Psi + \frac{f_0^2}{N^2} \partial_z \alpha \partial_z \Psi \\
& + \left[-\bar{u}ik (k^2 + l^2) + \partial_y \bar{q}ik + \frac{f_0^2}{N^2} \left(-\frac{ik}{4H^2} \bar{u} + \frac{1}{2H} \partial_z \alpha - \frac{\alpha}{4H^2} \right) \right] \Psi,
\end{aligned} \tag{17}$$

$$\begin{aligned}
0 = & -l^2 \partial_t U + \frac{f_0^2}{N^2} \left[-\frac{1}{H} \partial_t + \partial_z \alpha - \frac{\alpha}{H} \right] \partial_z U \\
& + \frac{f_0^2}{N^2} [\partial_t + \alpha] \partial_{zz} U + \frac{f_0^2}{N^2} \left[\frac{\alpha}{H} - \partial_z \alpha \right] \partial_z U_R \\
& - \frac{f_0^2}{N^2} \alpha \partial_{zz} U_R - \frac{f_0^2}{N^2} k l^2 \epsilon e^{z/2H} \text{Im} [\Psi \partial_{zz} \Psi^*],
\end{aligned} \tag{18}$$

where Im is the imaginary part and Ψ^* is the complex conjugate of Ψ . Here, N^2 ($= 4 \cdot 10^{-4} \text{ s}^{-2}$) is the buoyancy frequency squared; f_0 ($= 1.26 \cdot 10^{-4} \text{ s}^{-1}$) is the Coriolis parameter at 60°N ; β ($= 1.14 \cdot 10^{-11} \text{ s}^{-1} \text{ m}^{-1}$) is the meridional derivative of the Coriolis parameter at 60°N ; and ρ ($= \rho_0 \cdot \exp(-z/H) \text{ kg m}^{-3}$) is the standard atmospheric density. The chosen parameter set for integrating the Holton-Mass equations are $U_R(z_B, t)$ ($= 15\text{m/s}$), $\Lambda(t)$ ($= 1\text{m/s/km}$), and Δt ($= 3600\text{s}$). The model includes 46 vertical layers.

The model is integrated using third-order Adams-Bashforth time differencing (Durran 1999), while the vertical derivatives, for any variable M , are defined as

$$\begin{aligned}\frac{\partial M}{\partial z} &= \frac{M_{j+1} - M_{j-1}}{2\Delta z}, \\ \frac{\partial^2 M}{\partial z^2} &= \frac{M_{j+1} - 2M_j + M_{j-1}}{\Delta z^2},\end{aligned}$$

where $j = 0, 1, \dots, J$, for $J+1$ vertical levels and J corresponds to $z_T = J\Delta z$.

2.2 THE LOW-ORDER MODEL

Following Ruzmaikin et al. (2003), we may further simplify the system by truncating the vertical resolution of the Holton-Mass model. This is done by assuming a three layer discretization of the vertical derivatives with the levels included being the upper and lower boundaries, and the midlevel height. Taking the boundary conditions into account with this simplification, we get the discretized derivatives

$$\begin{aligned}\partial_z \Psi &= \frac{\Psi_2 - \Psi_0}{2\Delta z} = -\frac{gh}{2f_0\Delta z}, \\ \partial_{zz} \Psi &= \frac{\Psi_2 - 2\Psi_1 + \Psi_0}{(\Delta z)^2} = -\frac{2\Psi}{(\Delta z)^2} + \frac{gh}{f_0(\Delta z)^2}, \\ \partial_z U &= \frac{U_2 - U_0}{2\Delta z} = \frac{U - U_B + \Lambda\delta z}{2\Delta z}, \\ \partial_{zz} U &= \frac{U_2 - 2U_1 + U_0}{(\Delta z)^2} = \frac{-U + U_B + \Lambda\Delta z}{(\Delta z)^2}.\end{aligned}$$

Now splitting the streamfunction into its imaginary and real parts as $\Psi = X + iY$, and utilizing the above discretizations with the Holton-Mass model, we arrive at a low-order model of polar stratospheric wave-mean flow interaction. To best express the coefficients in terms of atmospheric parameters, we present the system in canonical form as found in

Ruzmaikin et al. (2003):

$$\begin{aligned}
a_1 \dot{X} &= a_2 X + [a_3 + a_4 \Lambda + a_5 U_B] Y + a_6 U Y + a_7 h + a_8 \dot{h}, \\
a_1 \dot{Y} &= a_2 Y - [a_3 + a_4 \Lambda + a_5 U_B] X - a_6 U X + a_9 h U, \\
b_1 \dot{U} &= b_2 (U - U_R) - b_3 h Y - b_5 \dot{U}_B + b_4 \dot{\Lambda},
\end{aligned}$$

where overdots represent derivatives in time and where

$$\begin{aligned}
a_1 &= k^2 + l^2 + \frac{f_0^2}{N^2} \left(\frac{1}{4H^2} + \frac{2}{(\Delta z)^2} \right), & a_8 &= \frac{f_0 g}{N^2 (\Delta z)^2}, \\
a_2 &= \frac{f_0^2}{N^2} \left(\frac{\partial_z \alpha}{2H} - \alpha \left[\frac{1}{4H^2} + \frac{2}{(\Delta z)^2} \right] \right), & a_9 &= \frac{\epsilon k f_0 g}{N^2 (\Delta z)^2}, \\
a_3 &= -k\beta, & b_1 &= l^2 + \frac{f_0^2}{N^2} \left(\frac{1}{(\Delta z)^2} + \frac{1}{2H\Delta z} \right), \\
a_4 &= \epsilon k \frac{f_0^2}{N^2} \left(\frac{1}{\Delta z} - \frac{1}{2H} \right), & b_2 &= \frac{f_0^2}{N^2} \left(\frac{\partial_z \alpha}{2\Delta z} - \alpha \left[\frac{1}{(\Delta z)^2} + \frac{1}{2H\Delta z} \right] \right), \\
a_5 &= \epsilon k \frac{f_0^2}{N^2} \left(\frac{1}{(\Delta z)^2} + \frac{1}{2H\Delta z} \right), & b_3 &= -\frac{\epsilon k l^2 g f_0}{N^2 (\Delta z)^2} \exp(\Delta z/H), \\
a_6 &= \epsilon k \left(k^2 + \frac{f_0^2}{N^2} \left[\frac{1}{4H^2} + \frac{1}{(\Delta z)^2} - \frac{1}{2H\Delta z} \right] \right), & b_4 &= -\frac{f_0^2}{N^2} \left(\frac{1}{2H} - \frac{1}{\Delta z} \right), \\
a_7 &= \frac{g}{2\Delta z} \frac{f_0}{N^2} \left(\frac{2\alpha}{\Delta z} - \partial_z \alpha \right), & b_5 &= -\frac{f_0^2}{N^2} \left(\frac{1}{2H\Delta z} + \frac{1}{(\Delta z)^2} \right).
\end{aligned}$$

The terms b_5 and \dot{U}_B are new terms included to allow time variations in the bottom boundary zonal wind U_B . While we shall include only constant U_B in this study, we have included this term for posterity.

We note here that the Newtonian damping term α is the same as for the Holton-Mass model (Figure 2 solid curve), but that we have discretized its form into a linear function with height (Figure 2 dash dot curve). This allows us to be consistent with the Newtonian damping values of the Holton-Mass model while also not including values of its vertical derivative $\partial_z \alpha$ that can not exist in our three layer model. One observes in Figure 2 that the vertical derivative of α for the Holton-Mass model at the midlevel is large relative to

the lower and upper levels. While choosing the true value of this vertical derivative at the midlevel is correct for a model with many layers, we opt for the discretized value of α because it is more correct physically with respect to the three layer model truncation.

To then simplify the notation of the system, we define parameters as in Birner and Williams (2008) where

$$\begin{aligned}\alpha_1 &= -\frac{a_2}{a_1}, \quad \alpha_2 = -\frac{b_2}{b_1}, \quad r = -\frac{a_3 + a_4\Lambda + a_5U_B}{a_1}, \\ s &= \frac{a_6}{a_1}, \quad \xi = -\frac{a_7}{a_1}, \quad \zeta = \frac{a_9}{a_1}, \\ \eta &= -\frac{b_3}{b_1}, \quad \delta_h = \frac{a_8}{a_1}, \quad \delta_{U_B} = \frac{b_5}{b_1}, \quad \delta_\Lambda = \frac{b_4}{b_1}.\end{aligned}$$

It is noted that δ_{U_B} is new for this study. The inverse of the terms α_1 and α_2 represent the model relaxation time scales of the waves and of the radiative damping, respectively. While the model wave relaxation time scale depends on the wavenumber, we note here that the model radiative damping time scale, $1/\alpha_2$, is constant and approximately 24 days.

The system, hereafter referred to as the low-order system, may thus be represented in the simple form

$$\dot{X} = -\alpha_1 X - rY + sUY - \xi h + \delta_h \dot{h}, \quad (19)$$

$$\dot{Y} = -\alpha_1 Y + rX - sUX + \zeta hU, \quad (20)$$

$$\dot{U} = -\alpha_2(U - U_R) - \eta hY - \delta_{U_B} \dot{U}_B + \delta_\Lambda \dot{\Lambda}. \quad (21)$$

The parameters for the above low-order system are identical to those in the Holton-Mass model, except for the vertical depth and a few additions. The lower boundary is now taken to be 15 km while the top is 55 km; this implies that $z_{mid} = 35$ km while $\Delta z = 20$ km. For all runs of the simple model, the bottom boundary wind $U_B = 10 \text{ m s}^{-1}$, the vertical shear $\Lambda = 1 \text{ m s}^{-1} \text{ km}^{-1}$, and the time step $\Delta t = 0.1$ days.

2.3 MULTIPLE WAVENUMBERS

Whereas the previous dynamical systems are wave-mean flow interaction models for a single wave mode, we now consider forms of the models which resolve more than one wavenumber. The primary consideration here is that since we have assumed linear wave propagation, the zonal wind will feel the effects of each wavenumber independently of one another. Thus in the case of the low-order model, if one computes the streamfunction of each wavenumber individually, the only coupling that shall arise is through modification of the zonal wind by the vertical EP flux divergence of the individual wavenumbers.

To more explicitly demonstrate this, consider the wave assumption made for the derivation of the Holton-Mass model:

$$\psi'(x, y, z, t) = \text{Re} [\Psi(z, t) e^{ikx}] e^{z/2H} \sin ly.$$

This assumption was made such that only one zonal wave mode was included in the model. This implies that the geostrophic streamfunction ψ' should be expressed as the spectral geostrophic streamfunction ψ'_m , where the subscript m is the zonal wavenumber. Thus the full geostrophic streamfunction wave assumption should be written

$$\psi'(x, y, z, t) = \sum_m (\text{Re} [\Psi_m(z, t) e^{imx/(a \cos \pi/3)})] e^{z/2H} \sin ly).$$

Since the Holton-Mass model — as derived by Holton and Mass (1976) — only includes one wavenumber, it excludes wave-wave interactions. We note that by assuming orthogonality in the x dimension and the same y -dependence for all wavenumbers, this full wave assumption retains the exclusion of wave-wave interactions.

Written as such, it is straightforward to see that should one wish to retain the single wavenumber system (m_{sing}), all one must do is set $\Psi_m(z, t) = 0$ for all wavenumbers m_{zeros} , where $m_{\text{zeros}} \notin m_{\text{sing}}$. Including multiple wavenumbers in both the Holton-Mass

model and the low-order model is then a matter of setting all $\Psi_{m_{\text{zeros}}} = 0$ for all wavenumbers m_{zeros} .

Recall that the single wavenumber Holton-Mass model is comprised of both an evolution equation for the streamfunction and an evolution equation for the zonal wind. The streamfunction interacts with the zonal wind through the EP flux divergence of that streamfunction. For n resolved wavenumbers, since our assumptions are that Ψ is wavenumber separable and that these wavenumbers do not interact, there should exist n evolution equations for the streamfunction which interact with the one evolution equation for the zonal wind. This is likewise true for the low-order model, except that Ψ has been separated into its real (X) and imaginary (Y) parts. Thus we should have n evolution equations for X and n evolution equations for Y for a total of $2n + 1$ equations.

The Holton-Mass model EP flux divergence is denoted by the final term in equation (18). Then for the set of resolved wavenumbers m , the Holton-Mass equations take the form

$$\begin{aligned}
0 = & - \left[k_m^2 + l^2 + \frac{f_0^2}{N^2} \frac{1}{4H^2} \right] \partial_t \Psi_m & (22) \\
& + \frac{f_0^2}{N^2} [\partial_t + \bar{u} i k_m + \alpha] \partial_{zz} \Psi_m + \frac{f_0^2}{N^2} \partial_z \alpha \partial_z \Psi_m \\
& + \left[-\bar{u} i k_m (k_m^2 + l^2) + \partial_y \bar{q} i k_m + \frac{f_0^2}{N^2} \left(-\frac{i k_m}{4H^2} \bar{u} + \frac{1}{2H} \partial_z \alpha - \frac{\alpha}{4H^2} \right) \right] \Psi,
\end{aligned}$$

$$\begin{aligned}
0 = & -l^2 \partial_t U + \frac{f_0^2}{N^2} \left[-\frac{1}{H} \partial_t + \partial_z \alpha - \frac{\alpha}{H} \right] \partial_z U & (23) \\
& + \frac{f_0^2}{N^2} [\partial_t + \alpha] \partial_{zz} U + \frac{f_0^2}{N^2} \left[\frac{\alpha}{H} - \partial_z \alpha \right] \partial_z U_R \\
& - \frac{f_0^2}{N^2} \alpha \partial_{zz} U_R - \sum_m \frac{f_0^2}{N^2} k_m l^2 \epsilon e^{z/2H} \text{Im} [\Psi_m \partial_{zz} \Psi_m] .
\end{aligned}$$

Written in the context of the model presented in (19)-(21), the vertical EP flux divergence for the low-order system is given by $\eta h Y$. Then for the set of resolved wavenumbers

m , the model takes the form

$$\dot{X}_m = -\alpha_{1,m}X_m - r_m Y_m + s_m U Y_m - \xi_m h_m + \delta_{h_m} \dot{h}_m , \quad (24)$$

$$\dot{Y}_m = -\alpha_{1,m}Y_m + r_m X_m - s_m U X_m + \zeta_m h_m U , \quad (25)$$

$$\dot{U} = -\alpha_2(U - U_R) - \delta_{U_B} \dot{U}_B + \delta_\Lambda \dot{\Lambda} - \sum_m (\eta_m h_m Y_m) . \quad (26)$$

One observes that (24) and (25) for each wavenumber are decoupled from the other wavenumbers, and the last term of (26) represents the EP flux divergence due to all wavenumbers.

3 STABILITY

3.1 SINGLE WAVENUMBER EQUILIBRIUM STATES

With any dynamical system, it is imperative to determine both the equilibrium states and the stability regions of the system. For high-order systems, this process is primarily determined through numerical processes. Given our low-order model of wave-mean flow interaction, the method by which these equilibrium states and stability regions are determined is semi-analytical. As will be shown, the stability analysis requires information from the equilibrium states of the system. Thus we begin by determining these equilibrium states of the low-order model.

The method of identifying equilibrium states relies on finding the roots of the stationary system. To find the roots of the stationary system, we set all time-dependent terms in the low-order system to zero and combine the equations into a cubic form in U . We set all other parameters to constants such that the coefficients of this function are dependent on the wave forcing amplitude h . This allows us to piecewise determine the roots of U given a specified h . For the standard parameter set, the resulting real roots for wavenumber 2 are plotted in Figure 4.

The two primary solutions that arise from this analysis are the topmost and bottommost solutions, plotted in blue and red respectively. We shall demonstrate in the stability analysis that the middle solution, plotted in green, is an unstable solution that defines a bistable region. Of note for the two primary solutions is that the top solution only exists for the lower to mid bounds of forcing amplitude and is the only solution at the lowest forcing amplitudes. One observes that the top solution falls on or near the prescribed radiative equilibrium wind speed of our system. This solution is termed the radiative solution.

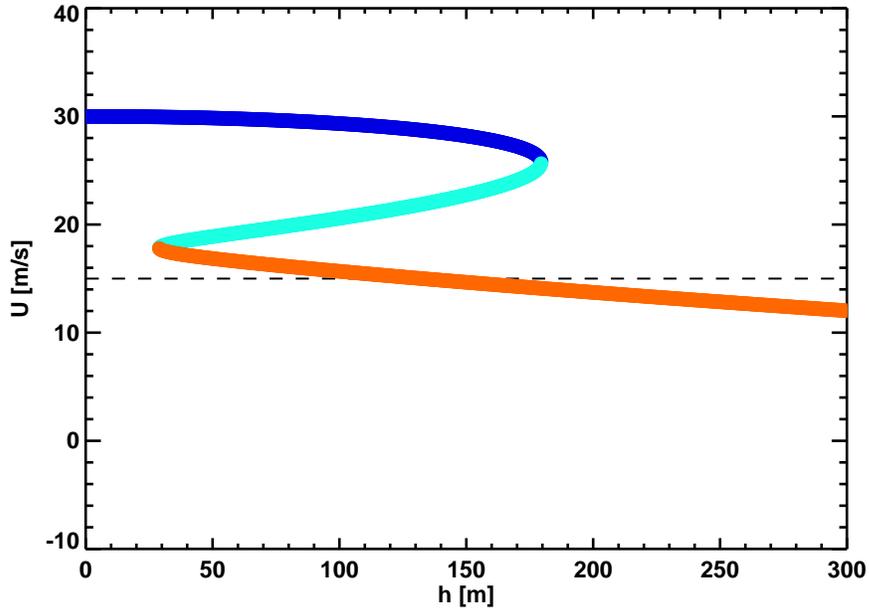


FIG. 4. Equilibrium states diagram for the low-order model with wavenumber 2 forcing. The blue curve is the radiative solution, the red curve is the sudden warming solution, and the green curve is the unstable middle solution. The black dashed line represents the arbitrary separation between radiative solution above and sudden warming solution below (see discussion of wavenumber 1 equilibrium states).

Conversely, the bottom solution is the only solution past some critical forcing amplitude. In the simplest of interpretations, large wave amplitudes in the atmosphere lead to large zonal wind reductions and ultimately to SSWs. Thus we term this bottom solution as the sudden warming solution. We note that the model does not fully capture wind reversal but rather a wind reduction.

This figure shows that for a stationary bottom boundary forcing, any wave forcing amplitude larger than 180 m does not have a radiative solution and must be in the sudden warming solution. Forcing amplitude less than 29 m does not permit a sudden warming solution to exist and thus the system must be in the radiative solution. We now define these critical points by the state to which the model transitions: 180 m is the sudden warming critical point and 29 m is the radiative critical point. For stationary waves with amplitudes h that fall between the two critical points, the zonal wind either stably resides in the sudden warming solution or in the radiative solution depending on the initial conditions.

Existence of the bifurcation in our low-order system is crucial as it demonstrates the possibility of producing sudden warmings within the model. Though the figure is for equilibrium states of the system, the behavior of this figure can be captured qualitatively by numerically integrating the model with a slowly increasing bottom boundary forcing as in Birner and Williams (2008). Slowly increasing shall be defined here as when the amplitude is increasing on a time scale much larger than radiative damping such that the amplitude is considered quasi-stationary. The results of this integration are plotted in Figure 5.

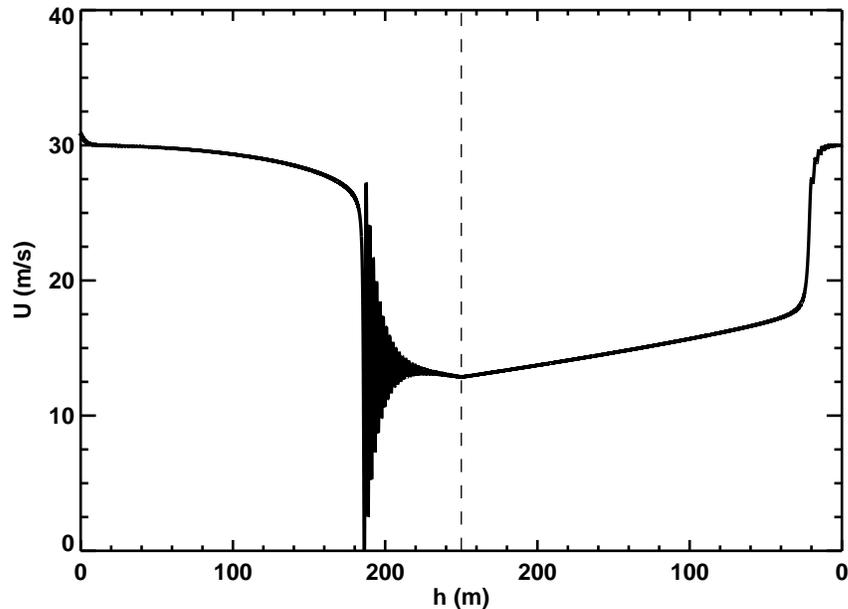


FIG. 5. As in Birner and Williams (2008), zonal wind from a 4000 day numerical integration of the model with wavenumber 2 forcing. The abscissa plots the wave amplitude which is monotonically increased to 250 m over the first 2000 days of the integration, and is monotonically decreased to 0 m over the next 2000 days. The vertical dashed line indicates the point at which the forcing stops ramping up to 250 m and begins ramping down to 0 m.

The forcing here is ramped up from 0 m to 250 m over 2000 days, and is subsequently ramped down to 0 m over the next 2000 days. A ramping period of 2000 days is chosen because this is much longer than the radiative time scale of the model. One should note that the rapid transition between radiative solution and sudden warming solution occurs near 180 m and thus well models the behavior of the equilibrium states. The restoration

from sudden warming solution to radiative solution is also well modeled since the zonal wind quickly restores to radiative solution speeds at approximately 30 m of forcing.

Figure 6 represents the stationary solutions diagram for wavenumber 1 forcing. What one should note from this figure is that there is no bifurcation between the radiative solution and the sudden warming solution. This figure instead suggests that if the forcing wave increases in amplitude on time scales much longer than radiative damping time scales, then there is no sudden warming strictly from wavenumber 1 forcing in the low-order model. Rather, the system evolves as a gradual polar vortex deceleration for strong forcing.

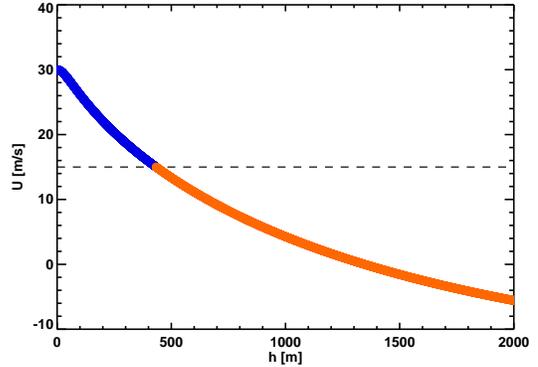


FIG. 6. As in Fig. 4, but for wavenumber 1 forcing.

To separate these two solutions of the model more appropriately, wind speeds less than 15 m/s are arbitrarily considered to always be within the sudden warming solution for the given parameter set. This separation is plotted in Figures 4, 6, and 8 as the horizontal dashed line. This definition is only to provide a sudden warming solution state in the low-order model. As such, it does not redefine the region where the sudden warming solution is larger than 15 m/s; so long as there exists bifurcation within the equilibrium states, the bottommost solution will always be termed as the sudden warming solution. Then to force the model with wavenumber 1 forcing into this definition of sudden warming solution, the magnitude of the forcing would need to be approximately 450 m.

It is important to note that this behavior of a slowly decelerating vortex is sensitive to the choice of parameters. Once the vertical shear term Λ is larger than 2 m/s/km, a bifurcation does appear in the equilibrium states for wavenumber 1, though a significant bifurcation structure does not appear until Λ is 3 m/s/km or larger. While this change

allows for wavenumber 1 sudden warmings, this also produces unphysical mean winter zonal wind speeds aloft (approximately 100 m/s at 10 hPa and 60°N).

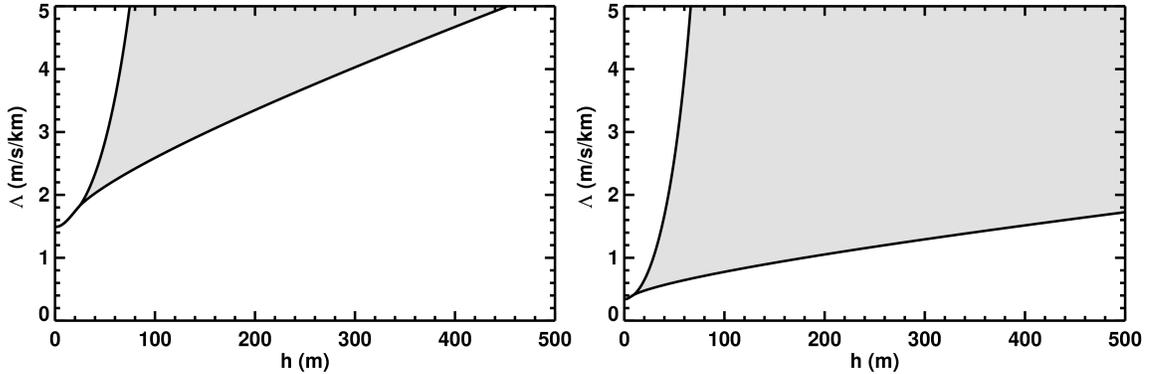


FIG. 7. Reproduction of Figure 1 of Ruzmaikin et al. (2003) for wavenumber 1 forcing (left) and wavenumber 2 forcing (right). The shaded region denotes where in the parameter space of Λ and h the equilibrium states contain multiple solutions. The existence of multiple solutions implies that a bifurcation exists.

This dependence on the vertical shear for multiple equilibrium states is explored in Figure 1 of Ruzmaikin et al. (2003) and reproduced here in Figure 7 for wavenumber 1 (left) and wavenumber 2 (right). The shaded region denotes where multiple equilibrium states exist in the parameter space of h and Λ . Multiple equilibrium states for a single choice of parameters implies the existence of a bifurcation in this model setup. One observes that for our choice of $\Lambda = 1$ m/s/km, there exists multiple equilibria for wavenumber 2, but not for wavenumber 1. This numerically demonstrates that a finite region of multiple equilibrium states exists for wavenumber 1-only forcing and for vertical shear greater than 2 m/s/km.

For completeness, we include the wavenumber 3 and wavenumber 4 equilibrium states diagrams in Figure 8. As with wavenumber 2, the wavenumber 3 equilibrium states for the low-order model contain a bifurcation, though here the sudden warming critical point does not occur until a wave amplitude of approximately 1600 m. While there is not a one-to-one relationship between the model and the observations, this wave amplitude is near to an order of magnitude larger than is observed at any wavenumber in the lower stratosphere. For wavenumber 4 forcing, there exists a radiative solution critical point at an amplitude

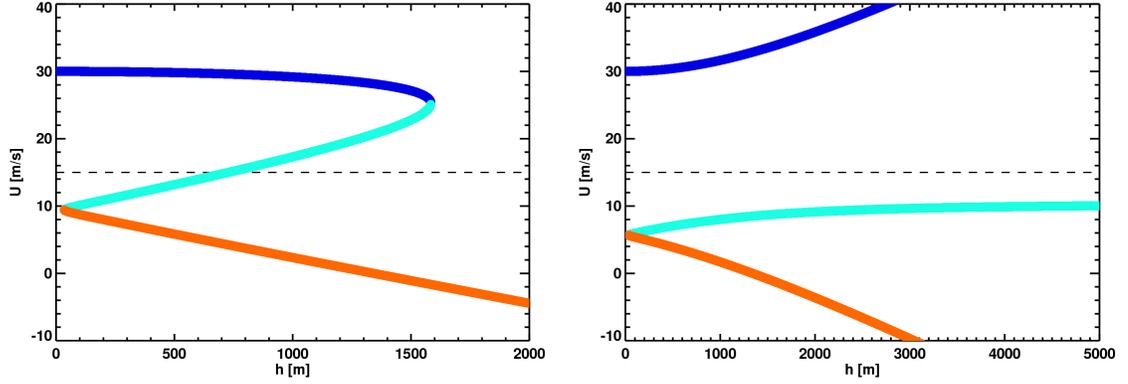


FIG. 8. As in Fig. 4, but for wavenumber 3 forcing (left) and wavenumber 4 forcing (right).

of 36.5 m but there does not exist a sudden warming critical point. One observes that the radiative solution zonal wind increases with increasing wave forcing. Since the necessary amplitude for a purely single wavenumber forced transition to the sudden warming solution is large for wavenumber 3 and nonexistent for wavenumber 4, we should not expect sudden warmings to be driven only by these higher wavenumbers in the low-order model.

Likewise, Charney and Drazin (1961) present criteria for vertical propagation of stationary waves: zonal mean velocity must be greater than 0 m/s (and thus, westerly) but less than the critical velocity

$$U_c = \beta / \left[k^2 + l^2 + \frac{f_0^2}{4H^2N^2} \right].$$

For wavenumbers 1 and 2, the critical velocities are approximately 22 m/s and 14 m/s, respectively. Since our bottom boundary zonal wind speed is chosen to be 10 m/s, this implies that vertical energy propagation occurs for these wavenumbers in the model. The critical velocities for wavenumbers 3 and 4 are approximately 8.7 m/s and 5.7 m/s, respectively. This implies that vertical energy propagation does not occur for these wavenumbers. Then we should not expect wavenumbers 3 and 4 to produce sudden warming solutions in the model not only from the equilibrium states diagrams in Figure 8, but also from the Charney and Drazin (1961) criteria.

Though the above equilibrium states diagrams present the evolution of the system for stationary or quasi-stationary forcing amplitudes, they should give one a general sense of the evolution of sudden warmings in the real atmosphere. We have shown that high wavenumber forcing amplitudes are insufficient to drive the low-order system to produce a sudden warming with physically reasonable forcing values. The model does capture the rapid transitions from radiative state to sudden warming state for wavenumber 2. Also, this wavenumber 2 sudden warming critical point occurs at forcing amplitudes on the order of observed amplitudes. For the chosen parameter set, wavenumber 1 forcing can not produce sudden warmings, but rather gradual polar vortex deceleration. Sudden warmings may be forced by wavenumber 1 forcing in the low-order model only for significantly higher vertical wind shear.

Stratospheric sudden warmings in the real atmosphere exhibit both wavenumbers 1 and 2 breakdowns, with vortex displacements being wavenumber 1 events and vortex splits being wavenumber 2 events. The low-order, single wavenumber model analyzed thus far is too simplistic to include both of these types of events. One simple method of including both events is to increase the number of resolved wavenumbers such that both wavenumbers are simultaneously forcing the zonal wind.

3.2 MULTIPLE WAVENUMBER EQUILIBRIUM STATES

Analyzing equilibrium states with the multiple wavenumber, low-order system is slightly more complex than in the case of a single wavenumber. Whereas the zonal wind in the previous diagrams may be solved as a function of forcing amplitude, there are two independent variables in the multiple wavenumber system. Likewise, semi-analytical solutions involve solving the roots of a quintic equation in U for a two wavenumber system. Thus our approach must be different to effectively yet succinctly describe the stationary solutions with multiple wavenumbers.

We firstly present Figure 9 which plots the equilibrium states diagram as a function of wavenumber 1 forcing for wavenumbers 1 and 2 resolved in the model. For this equilibrium states diagram, we hold wavenumber 2 forcing constant at 20 m. This amplitude of wavenumber 2 forcing is less than the radiative solution critical point, implying that if the model only resolved wavenumber 2 forcing, then it would only ever be in the radiative solution. One observes from Figure 9 that even with such small wavenumber 2 forcing amplitude, the wavenumber 1 equilibrium states now contain a bifurcation. The significance of such a result is that we may now force primitive sudden warming events in the low order model with wavenumber 1 forcing, rather than simply forcing gradual polar vortex deceleration.

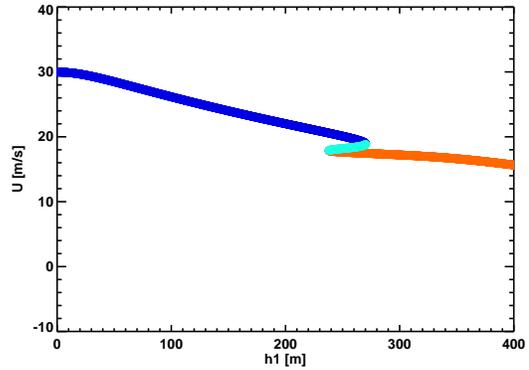


FIG. 9. Equilibrium states diagram for the multiple wavenumber model with wavenumbers 1 and 2 resolved as a function of wavenumber 1 forcing where wavenumber 2 forcing is held constant at 20 m. Even with small wavenumber 2 forcing, the wavenumber 1 equilibrium states contain a bifurcation.

It can be shown that the equilibrium states of this multiple wavenumber low-order model with any combination of two of the three lowest wavenumbers contains a bifurcation. Since these bifurcations exist at nearly all values of the forcing amplitudes, we may further this analysis by determining only the amplitudes at both the sudden warming critical point and the radiative critical point. These two critical points can readily be determined for specified amplitudes of each wavenumber and qualitatively describe the bifurcation. Figure 10 plots the sudden warming critical point (solid curve) and the radiative critical point (dashed curve) as functions of both wavenumber 1 and wavenumber 2 forcing amplitude. Inserts are provided to give a graphical representation of the zonal wind transition each line signifies.

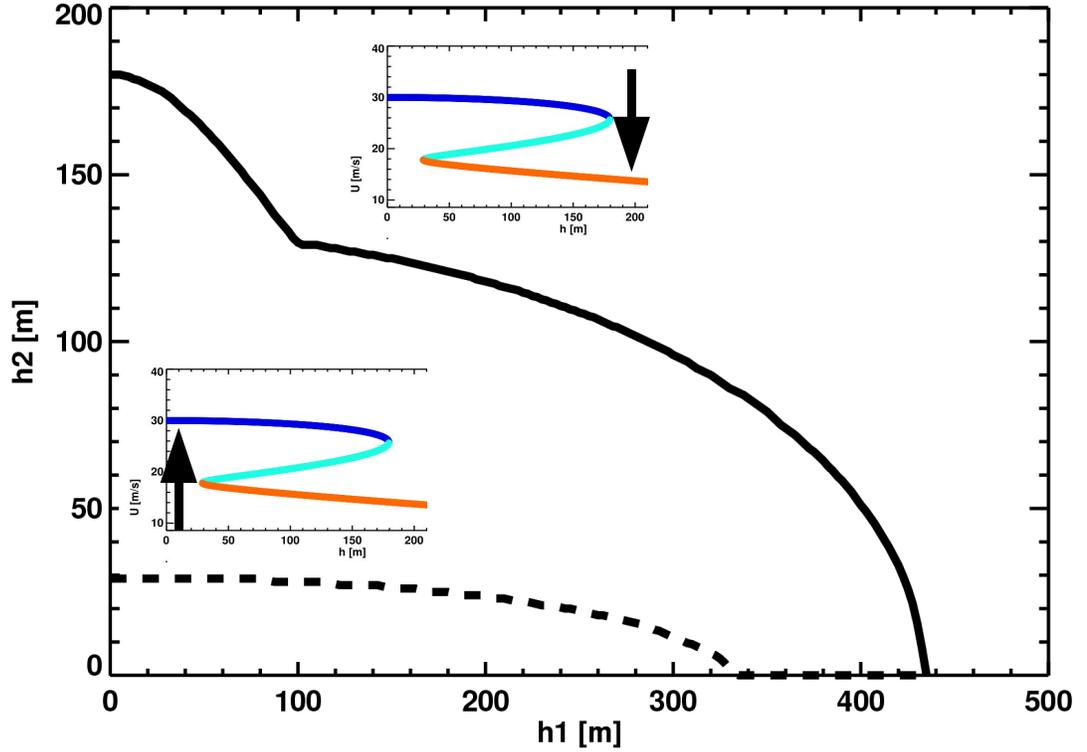


FIG. 10. Curves of sudden warming critical points (solid) and the radiative critical points (dashed) as functions of both wavenumber 1 and wavenumber 2 forcing. Inserts are provided to give graphical representation of the zonal wind transition that each critical point curve signifies.

One may note that at the limits as each wavenumber's amplitude goes to 0 m, the critical points approach the values for the opposing wavenumber's equilibrium states in the single wavenumber model. As wavenumber 1 forcing goes to 0 m, the critical points approach the values for the wavenumber 2-only equilibrium states. Likewise as wavenumber 2 forcing goes to 0 m, the sudden warming critical point value approaches the critical point value for wavenumber 1.

We note that the critical point plots for resolved wavenumbers 1 and 2 forcing have interesting behaviors near the bounds of the wavenumber 1 forcing domain. For the sudden warming critical point curve (solid) near 100 m of wavenumber 1 forcing, there exists a nondifferentiable point in the curve. At wavenumber 1 amplitudes less than 100 m, the

critical point curves rapidly go to the wavenumber 2-only values. This is attributable to the dominance of wavenumber 2 forcing at wavenumber 1 forcing amplitudes less than 100 m. Once wavenumber 1 forcing is larger than 100 m, the two wavenumbers interact until wavenumber 1 forcing is larger than 330 m. The radiative solution critical point curve beyond this amplitude of wavenumber 1 forcing goes to 0 m, implying that the critical point ceases to exist. Thus for wavenumber 1 forcing larger than 330 m, wavenumber 1 forcing is dominant and the critical points rapidly go to the wavenumber 1-only values.

Of interest to note in Figure 10 and the following multiple wavenumber critical points plots is that the curve plotting the radiative critical point varies little with respect to the curve plotting the sudden warming critical point. This behavior holds between all multiple wavenumber runs and indicates that there is a minimum forcing at which the radiative damping is always stronger than the wave forcing. This may be seen through a scale analysis of the zonal wind evolution equation.

Presented in nondimensional units for wavenumbers 1 and 2, the zonal wind equation for the model initialized in the sudden warming solution scales as

$$\dot{U} \approx 10^{-2} - 10^{-4} \cdot h$$

where h is presented in dimensional values. This analysis demonstrates that the total amplitude of the forcing must be at least order 10^2 m to decelerate the zonal wind. Since the radiative damping remains nearly constant so long as the zonal wind is in the sudden warming solution state, the radiative critical point should likewise remain nearly constant.

Figure 11 presents the results of the bifurcation analysis with wavenumbers 1 and 3 resolved and with wavenumbers 2 and 3 resolved. In contrast to the wavenumbers 1 and 2 case, the curve representing the sudden warming critical point simply connects the single wavenumber solutions while the curve representing the radiative critical point varies little. The interactions of wavenumber 3 with wavenumbers 1 and 2 manifests much weaker than

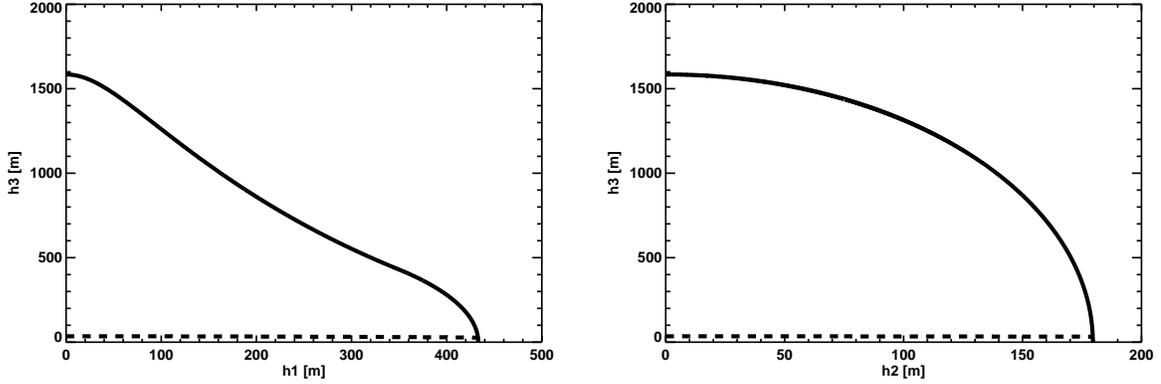


FIG. 11. As in Fig. 10, but for wavenumber 1 and wavenumber 3 forcing (left), and for wavenumber 2 and wavenumber 3 forcing (right).

the interaction between wavenumbers 1 and 2 as evidenced by the lack of any sudden discontinuities in the critical point curves.

While the diagrams presented for the multiple wavenumber system do not contain a large degree of complexity, that does not imply that they are not important. These figures demonstrate that there is significant interaction between wavenumbers 1 and 2 through the mean flow when these wavenumbers are resolved together. It is also shown that wavenumber 3 weakly interacts with the lowest wavenumbers. The figures also help to verify that the multiple wavenumber system is behaving properly. That is, the equilibrium states are identical to those for the single wavenumber system when one wavenumber is omitted in the multiple wavenumber system.

It was previously noted that bifurcations exist throughout almost all of the domain of forcing amplitudes in the multiple wavenumber system. That these bifurcations occur at finite, physically reasonable forcing amplitudes demonstrates that the multiple wavenumber system is modeling a primitive sudden stratospheric warming.

We note that this multiple wavenumber system is not intended to replace the single wavenumber system in this study. Rather, the single wavenumber system is a special case of the multiple wavenumber system. For the following results, we will typically present

results from both the single wavenumber low-order model and the multiple wavenumber low-order model.

We remind the reader that the previous results only hold for stationary or quasi-stationary wave amplitudes. Thus, this analysis has not considered a system in which the forcing is fluctuating on the order of or shorter than the time scale of the radiative damping in the model. We reasonably expect that the qualitative results hold even for time varying forcing: that bifurcations exist and the model simulates a primitive sudden warming. Yet we should expect that the numerical behavior of the model is not represented here when the forcing amplitudes are allowed to vary on short time scales. This topic of forcing amplitude transience will be further explored in Chapter 5.

3.3 LINEAR STABILITY ANALYSIS

We have thus far identified the equilibrium states of the low-order model. Qualification of the stability of these equilibrium states follows from this analysis. Here we shall utilize linear stability analysis techniques to determine the stabilities and instabilities of the system. The theory of this work is best described through the classical description of a ball resting in a potential well.

The ball — here representing the state of the system — is at rest at each equilibrium state. To determine whether the state is stable or unstable, we lightly perturb the ball about the equilibrium state. If the ball returns to the equilibrium state then the solution must be stable. Should the ball roll away from the initial state — i.e. the perturbed system evolves to a different equilibrium state — then the initial equilibrium state is unstable. A schematic of stable and unstable states are plotted in Figure 12.

To perform this stability analysis for the low-order system, we find the Jacobian of the stationary solutions of the system. This Jacobian for the single wavenumber model is given



FIG. 12. Schematic of stability where the solid ball represents the system resting in an equilibrium state. The ball, or system, is perturbed in the direction of the arrow and the dashed ball represents the perturbed position of the system. In a stable equilibrium state (left), the ball will roll back down the slope and return to the original resting position. In an unstable equilibrium state (right), the ball will roll down the slope and not return to the original resting position, but rather to some new equilibrium state.

by

$$\frac{\partial (\dot{X}, \dot{Y}, \dot{U})}{\partial (X, Y, U)} = \begin{bmatrix} -\alpha_1 & -r + sU_{STA} & sY_{STA} \\ r - sU_{STA} & -\alpha_1 & -sX_{STA} + \zeta h \\ 0 & -\eta h & -\alpha_2 \end{bmatrix}$$

where the subscript STA denotes a stationary solution value.

The signs of the real part of the eigenvalues of this matrix give the stability or instability of a solution. If all real parts of the eigenvalues corresponding to a particular solution are negative, the solution is stable. If any one of the real parts is positive, then the solution is unstable. The eigenvalues λ for a given solution are given by solving

$$0 = \begin{vmatrix} -\alpha_1 - \lambda & -r + sU_{STA} & sY_{STA} \\ r - sU_{STA} & -\alpha_1 - \lambda & -sX_{STA} + \zeta h \\ 0 & -\eta h & -\alpha_2 - \lambda \end{vmatrix}$$

Figure 13 plots the real parts of the eigenvalues for wavenumber 1 as a function of forcing amplitude. For this and subsequent eigenvalue plots, the color coding of the eigenvalues is the same as the equilibrium states plots — blue is the radiative solution, red is the sudden warming solution, and green is the middle solution. As the figure shows, the equilibrium states for wavenumber 1 are stable at all strengths of forcing. This should not be unexpected as only one real solution exists for the given parameters. Should more so-

lutions exist — such as for wavenumbers 2 and 3 — then there may exist regions where an equilibrium state is unstable. Under such situations, the system will preferentially move towards a stable state.

For wavenumbers 2 and 3, Figure 14 shows that the middle solution of the equilibrium states is unstable while the upper and lower solutions are stable. For wavenumber 2, both the radiative solution and the sudden warming solution have only negative eigenvalues, and are thus stable solutions. One may infer that if the forcing amplitude is quasi-stationary, or slowly evolving relative to radiative damping time scales, then the system will stably remain in either solution until the amplitude reaches a bifurcation point. Note that this behavior is what was modeled in Figure 5.

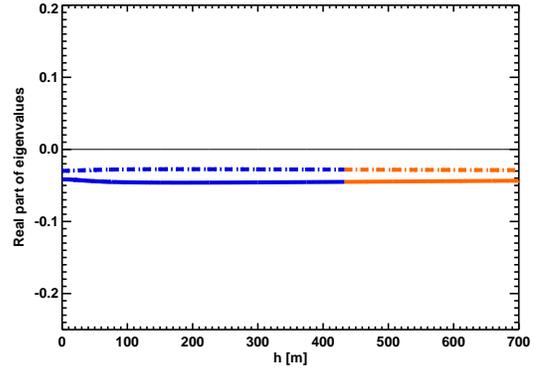


FIG. 13. Real parts of the eigenvalues for the Jacobian of the wavenumber 1 stationary solutions.

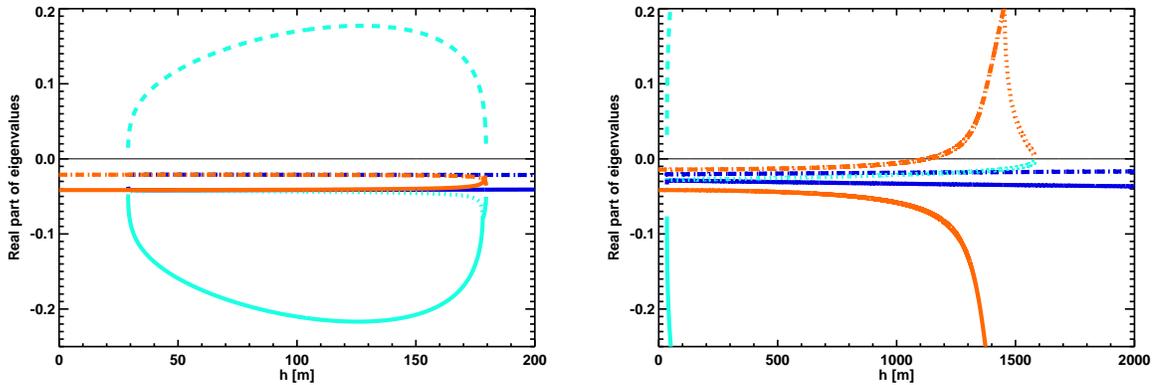


FIG. 14. As in Fig. 13, but for wavenumber 2 forcing (left), and for wavenumber 3 forcing (right).

The wavenumber 3 eigenvalues exhibit an instability distinct from the lower wavenumbers. As with wavenumber 2, the middle stationary solution is entirely unstable and the

sudden warming solution is entirely stable. However, the radiative solution becomes unstable if the forcing amplitude is larger than approximately 1150 m.

The existence of this instability is not obvious as the wavenumber 3 equilibrium states are qualitatively equivalent to wavenumber 2, but wavenumber 2 does not show this instability of the radiative solution. We elicit this instability by numerically integrating the model twice in which quasi-stationary wavenumber 3 forcing was ramped up to both a maximum amplitude of 1100 m and of 1175 m. For both runs, the maximum amplitude was maintained for 1000 days. The resulting zonal winds are plotted in Figure 15.

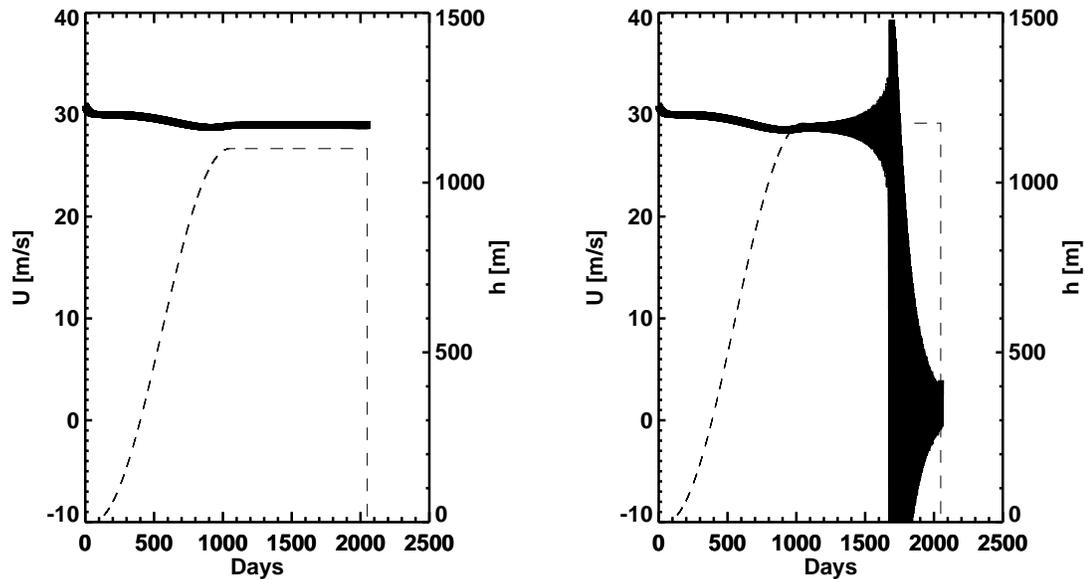


FIG. 15. Zonal wind (solid curve) from the model forced with wavenumber 3 forcing (dashed curve) where the maximum forcing amplitude was 1100 m (left) and 1175 m (right). Note the instability of the radiative solution for 1175 m but not for 1100 m of maximum amplitude forcing.

The run with 1100 m maximum forcing amplitude stabilized about the radiative solution after the ramping was completed and remained in this stable state until the end of run. But for the run with 1175 m of maximum forcing amplitude, after the model stabilized to the radiative solution, it experienced a growing instability that triggered a warming after approximately 700 days of constant forcing. Thus this radiative solution instability does

exist in the low-order model for wavenumber 3. We note here that the time scale required for this instability to produce a sudden warming is greater than an order of magnitude longer than the radiative damping time scale.

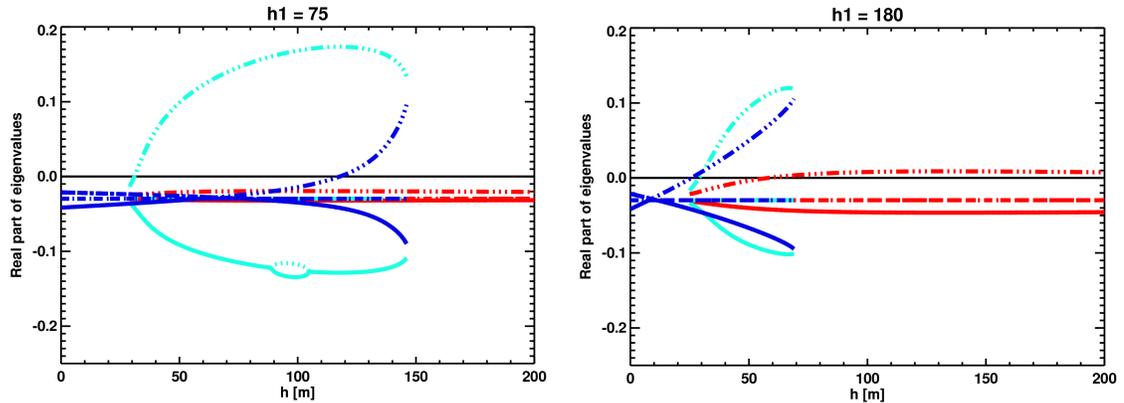


FIG. 16. Real parts of the eigenvalues for the model with wavenumbers 1 and 2 resolved. The left plot is for a constant wavenumber 1 forcing of 75 m while the right plot is for forcing of 180 m.

Linear stability analysis may also be utilized for the coupled system. Though this system adds a degree of complexity to the analysis, the methods remain semi-analytical. We again find the Jacobian of the stationary solutions of the multiple wavenumber system and subsequently find the eigenvalues of this Jacobian. Whereas the single wavenumber system has only three possible eigenvalues for each stationary solution, the two wavenumber system has five.

The real parts of the eigenvalues for the system with wavenumbers 1 and 2 resolved are plotted in Figure 16 as a function of wavenumber 2 amplitude. The left plot is for a constant wavenumber 1 forcing of 75 m while the right plot is for 180 m of wavenumber 1 forcing. We note here that the plot for which wavenumber 1 forcing is held constant at 0 m (not shown) is identical to the left plot in Figure 14.

What one observes is that as wavenumber 1 forcing increases, the eigenvalues of the radiative solution at high amplitudes become more positive. The range of this instability increases as the constant wavenumber 1 forcing is also increased, such that once

wavenumber 1 forcing is 180 m, the instability of the radiative solution covers more than half the domain of that solution. Similar to the wavenumber 3-only stability, this implies that the radiative solution of the system becomes unstable at amplitudes less than where the radiative solution ceases to exist.

Similar to what was done for the wavenumber 3-only stability, the model is run with a constant wavenumber 1 forcing of 75 m and with a slowly increasing wavenumber 2 forcing. Once wavenumber 2 reached 130 m of amplitude, all the forcing was held constant for the remaining 3000 days. The right plot of Figure 17 plots the zonal wind and wavenumber 2 forcing for this model run; the left plot is for a maximum wavenumber 2 amplitude of 120 m. After 2000 days of constant forcing, the instability of the radiative solution manifested and the model transitioned to the sudden warming solution for the 130 m forcing while the radiative solution was stable for the forcing of 120 m.

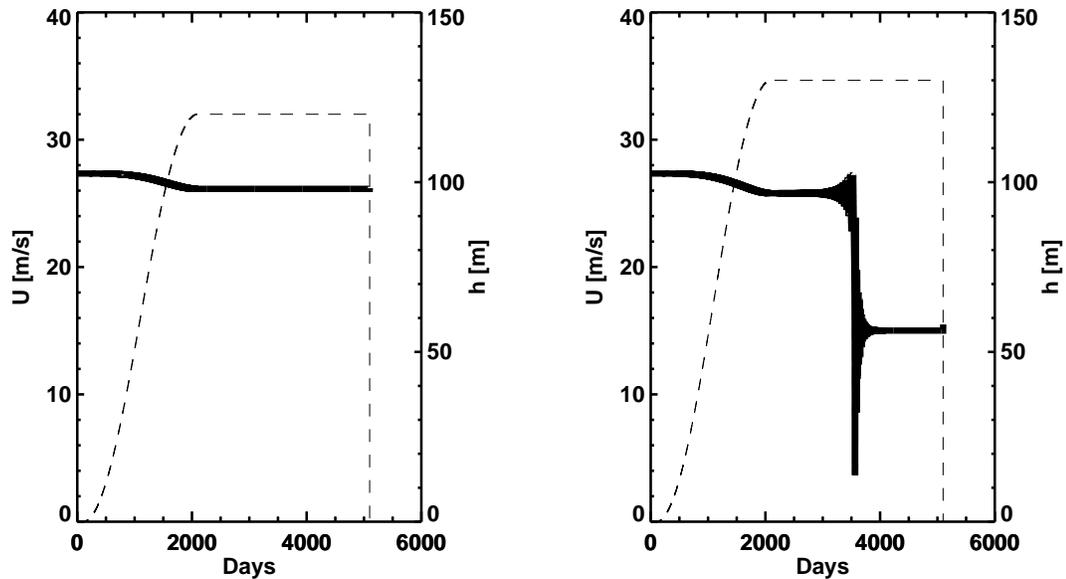


FIG. 17. Zonal wind (solid line) from the model for constant wavenumber 1 forcing of 75 m and for ramping wavenumber 2 forcing (dashed line) where the wavenumber 2 maxima are 120 m (left) and 130 m (right).

Again this instability of the radiative solution is shown to be existent in the model. As with the wavenumber 3-only integration, the time scale for this instability to produce a

warming is much longer than the radiative damping time scale. Only for amplitudes very near the sudden warming critical point (on the order of 1 m difference), does this instability manifest on time scales on the order of the radiative damping.

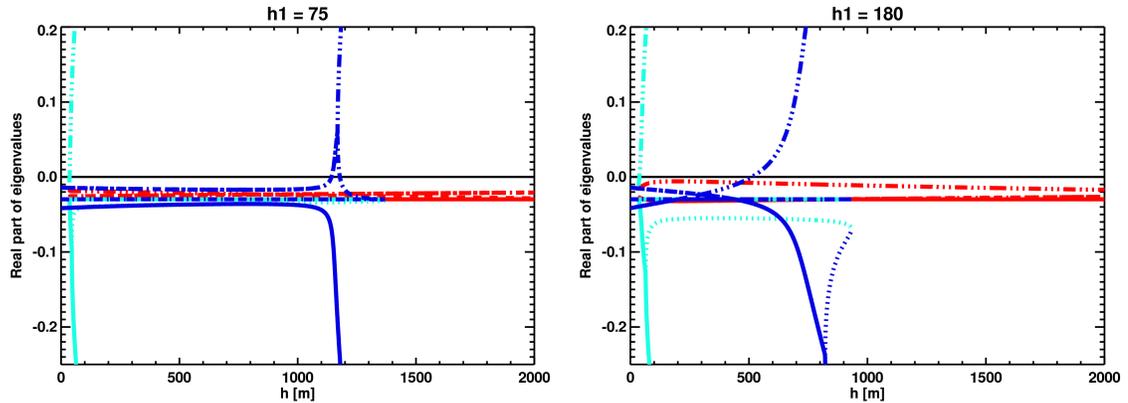


FIG. 18. As in Figure 16, but for wavenumbers 1 and 3.

Figure 18 plots the real parts of the eigenvalues for resolved wavenumbers 1 and 3 as functions of wavenumber 3 forcing amplitude. The left plot is for constant wavenumber 1 forcing of 75 m while the right is for 180 m. This figure has been kept at the same plotting scale as the previous eigenvalue plots and thus does not show the eigenvalue behavior above 0.2 units. In doing so, the figure focuses on the eigenvalues close to zero but excludes most of the rapid increase in positive eigenvalues for the radiative solution. The start of this increase can be observed in Figure 18, though the continuation of the large values up to the sudden warming critical point can not be observed. Thus it is incorrect to conclude from Figure 18 that the instability does not persist until the sudden warming critical point.

Here again one observes that as the constant, low wavenumber forcing is increased, the radiative solution's domain of stability becomes smaller relative to the domain of the solution. Whereas the instability only covers the final 200 m of the radiative solution for 75 m of wavenumber 1 forcing, the instability covers 450 m for 180 m of wavenumber 1 forcing. Compare these ranges to the domain of the radiative solution of 1350 m and of 950 m, respectively. Approximately half the radiative solution is unstable once the wavenumber 1 forcing is 180 m.

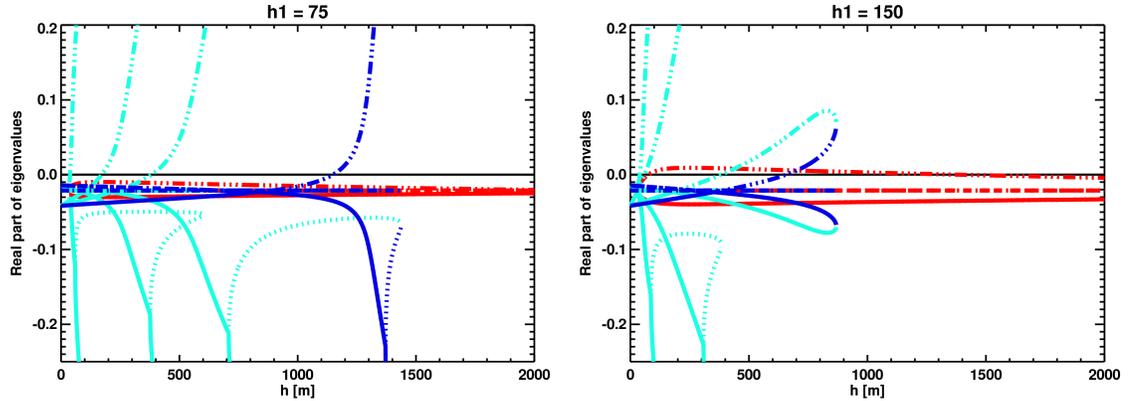


FIG. 19. As in Figure 16, but for wavenumbers 2 and 3.

The previous conclusions from these plots hold for when the model resolves wavenumbers 2 and 3. The primary difference is that the sudden warming critical point relative to wavenumber 2 forcing occurs at a lower amplitude than for wavenumber 1. This is why the right plot of Figure 19 is not for 180 m of wavenumber 2 forcing: only the sudden warming solution exists there.

3.4 HOLTON-MASS MODEL AND GCM STABILITY

We have thus far presented the equilibrium states to the low-order model, both in the single wavenumber and in the multiple wavenumber setups, and the stability regimes for these equilibrium states. It is shown that for our parameter set and for a slowly increasing (decreasing) wave amplitude forcing, the system undergoes a rapid transition between stable states, here between radiative solution and sudden warming solution (between sudden warming solution and radiative solution). As this low-order system is derived from the Holton-Mass model, we seek to demonstrate this behavior in that 1-dimensional system also.

Holton and Mass (1976) observed that their model of stratospheric vacillations would undergo a transition from a radiative state to a continual sudden warming state under only slightly altered conditions. The primary condition affecting this transition was identified to be the bottom boundary forcing.

Chao (1985) and Yoden (1987b) further investigated this hysteretic behavior for the Holton-Mass model. Chao first noted that the Holton-Mass model produced behavior well described by catastrophe theory: for a very slowly increasing wave forcing, there exists a critical amplitude beyond which the model produces a rapid transition to a continual sudden warming state. Chao likewise showed that for the model initialized in the sudden warming state, slowly decreasing the forcing until a different critical value rapidly transitioned the model back into a radiative state. The behaviors of these critical points matches well with the behavior noted in the low-order system.

Yoden furthered these results by examining the bifurcation properties of the Holton-Mass model. The main result relevant to our discussion is from Yoden (1987b), Figure 3, shown in Figure 20. Here he presents the bifurcation diagram at 25 km as a function of bottom boundary wave forcing given on the x axis. This figure is qualitatively the same as what was produced from the low-order model. There exists a stable state radiative solution of the model that persists for slowly varying wave forcing amplitude, until a critical point at which the model transitions to a sudden warming state. And as shown in Chao (1985) and in the low-order model, if the system is in the sudden warming state, there is a minimum wave amplitude at which the model transits rapidly back to the radiative state. Finally, there exists an unstable solution in the region between the two critical points. For further discussion of the stabilities of the Holton-Mass model, see Christiansen (2000).

It is clear then that the stability results of the Holton-Mass model hold for the low-order model. The low-order model is simply a severely truncated version of the Holton-Mass model, thus one should not expect otherwise. What is missing is the ties between these stability states in the Holton-Mass model and in a global climate model (GCM).

It is Boville (1986) who perhaps provides this evidence of distinct stability states through use of the NCAR CCM to model wave-mean flow interaction. Boville demonstrated that there existed two clear mean states separated by not only differences in zonal mean temperature and in zonal mean wind, but also by a well-modeled sudden warming. The first

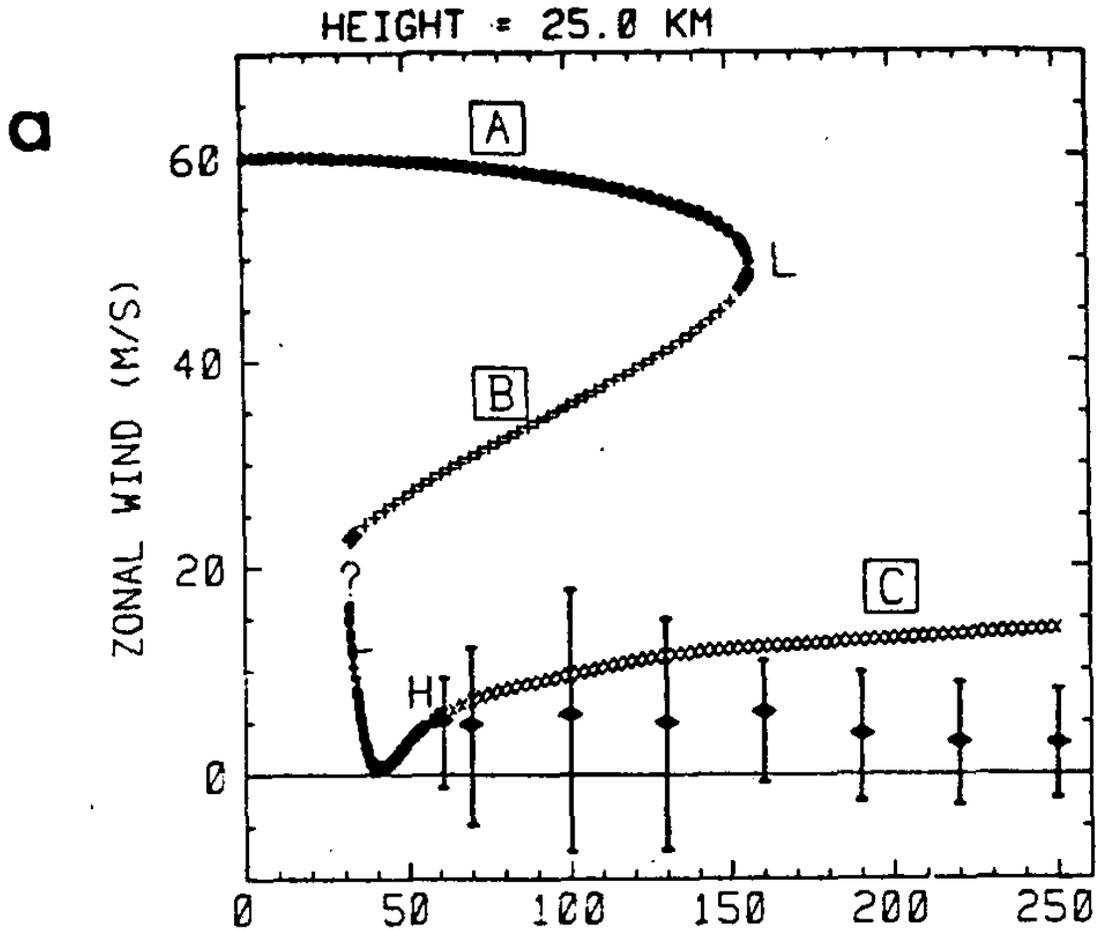


FIG. 20. Figure 3 (a) from Yoden 1987b. This is the bifurcation diagram plotting zonal wind as a function of h at 25 km in the Holton-Mass model and for wavenumber 2 forcing. Stable solutions are represented by solid circles while unstable solutions are represented by X's. This bifurcation diagram is nearly identical to the low-order model bifurcation diagram.

of these states most directly resembles the radiative state modeled in both the Holton-Mass and low-order systems. Similarly, the second state resembles that of the sudden warming state, though this state is only quasi-stable in that it gradually restores to the radiative state. Though there do not exist direct parallels between Boville's results and the low-order model results, one should note that the number of similarities are greater than that of the differences. We might therefore adopt the terminology of Yoden (1987b) and state that the Holton-Mass model is a "conceptual model of the multiple planetary flow regimes in Boville's GCM."

Finally, we reference Monahan et al. (2003) as a source for possible identification of two independent mean states in the real atmosphere. They use nonlinear principal component analysis applied to observations to identify distinct climate regimes in the stratosphere. What they observed was three circulation regimes: two displaying characteristics similar to the radiative solution state of the low-order and Holton-Mass models and the third “associated with stratospheric sudden warmings” (Monahan et al. 2003). That there exists observational evidence of two distinct flow regimes in the real atmosphere suggests that the low-order model is capturing the fundamental stratospheric dynamics. We therefore describe the low-order model as the simplest prototype of a sudden stratospheric warming.

4 SMALL-SCALE VARIABILITY

4.1 NOISY PROCESSES

Scale separation in a numerical model of atmospheric dynamics was notably utilized by Charney and DeVore (1979) in a truncated spectral model of atmospheric flow, not unlike the low-order system we use in this study (see also Egger 1981 and Sura 2002). In their model, Charney and DeVore showed that there existed several distinct flow regimes in the troposphere. However, Tung and Rosenthal (1985) note that certain results of the Charney and DeVore model did not persist once full nonlinearity was considered because scale separation is not applicable to the troposphere. Once full nonlinearity was included in the Charney and DeVore model, the regime behavior that was observed went away. One principal difference between the the Charney and DeVore model and the models we consider in this study is that our low-order model (and the Holton-Mass model) describes stratospheric flow while Charney and DeVore describes tropospheric flow. Therefore Tung and Rosenthal's critique does not apply to our models because Charney and Drazin (1961) demonstrate that scale separation is applicable in the stratosphere. It is from this work that we are able to justify the use of a spectral model containing the gravest wavenumbers (1 and 2) for describing stratospheric flow.

By only including the gravest wavenumbers in both the low-order and Holton-Mass models, we eliminate the effects of small-scale variability from the high wavenumbers. This variability may arise from a number of natural events: internal small-scale nonlinear interactions, gravity wave breaking, or possibly synoptic-scale eddy perturbations. One should not anticipate these small-scale forcings to produce strong dynamical effects relative to the low wavenumber forcings. But the small forcings can perturb the system between

the radiative solution and the sudden warming solution if the low wavenumber forcing is strong, yet insufficient enough to drive the transition alone.

This high-wavenumber noise is represented in the low-order system as a stochastic term added to the zonal wind. As in Birner and Williams (2008), this noise term is modeled as Gaussian white noise. This term is represented by χ_a in the zonal wind equation:

$$\dot{U} = -\alpha_2(U - U_R) - \eta h Y - \delta_{U_B} \dot{U}_B + \chi_a \quad (27)$$

where

$$\overline{\chi_a(t) \chi_a(t')} = \sigma_a^2 \delta(t - t'). \quad (28)$$

As written here, σ_a represents the strength of the noise in units m/s/day. δ is the Dirac delta distribution. This term is everywhere zero except at the origin ($t - t' = 0$) where its value is infinite; the Dirac delta distribution must also satisfy the condition

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

For a discussion on how χ_a is handled numerically in the model, see Birner and Williams (2008).

The form of the zonal wind noise is additive, otherwise described as independent of the state of the zonal wind which it modifies. One may physically interpret this kind of noise as decoupled, high-wavenumber forcing of the zonal wind since χ_a only arises within the zonal wind equation. This is important since our wave-mean flow interaction models are such that the wave forcing on the mean flow is the sum of each wavenumber's EP flux divergence. Thus this additive noise simulates the summation of EP flux divergences for wavenumbers much higher than resolved wavenumbers 1 and 2. We may also consider this

to be gravity wave-like forcing as in Fritts and Alexander (2003):

$$\partial_t \bar{u} \approx \frac{1}{\rho} \vec{\nabla} \cdot \vec{F} - \partial_z \overline{u'w'}$$

where the first term on the right-hand side describes the EP flux divergence and the final term describes the gravity wave forcing. Written like this, we assume the EP flux divergence term to only include the large-scale eddies. This assumption implies that there is scale separation between the large-scale and small-scale eddies, thus allowing us to write the gravity wave forcing as a separate, explicit term.

The fallacy of this argument is the mean flow should affect the wave forcing in our models, but this is not done for the noisy term. We acknowledge that there may be other means of properly parameterizing this short wavelength variability, but this is certainly the easiest step for including small-scale variability. Also, one could argue that in such simple models, accurate resolution of small-scale variability is not necessary to capture the large-scale dynamical features. This is especially true if the forcing at these high wavenumbers is small relative to the low wavenumber forcing. Thus for our experiments, the strength of the noise will be set to small values so as to minimize the error of our white noise parameterization. The range of values considered here falls between 0.1 to 1.0 m/s/day, which are on the order of the observed values for gravity wave drag in the polar stratosphere (Hamilton 1997, Pulido and Thuburn 2006), with 0.1 m/s/day to be used for most experiments in this work. For 0.1 m/s/day noise strength, the acceleration from the noise is much less than the acceleration resulting from the planetary wave breaking, which is often on the order of 5 m/s/day (Randel et al., 1987).

Besides modeling small-scale variability, this noise term also allows us to test for soundness of our results. The addition of this small, random perturbation to the zonal wind allows for the possibility of producing dramatically differing solutions for the same initialization. While observation of many differing solutions is interesting, the true aim of

running these models over many realizations is to achieve statistical confidence. Thus for the remainder of the paper, we choose to run the model 100 times when including noise.

As an example of how we test the confidence, if we seek to test the behavior of the low-order model at some specified constant wave forcing amplitude, then we run the model 100 times with the noise and observe the characteristics of the solution set. If the model produces mainly sudden warming (radiative) solutions, then we infer that the specified forcing amplitude will likely always produce a sudden warming (radiative) solution. If the model produces a nearly equal number of sudden warming and radiative solutions, then this would have other implications, perhaps that the model with that wave forcing is in an unstable state.

4.2 THE FOKKER-PLANCK EQUATION

To more exactly understand the characteristics of the model with respect to the strength of the noise, to the wave forcing amplitude, and to other parameter changes without performing many iterations of the model, we may employ the Fokker-Planck equation to eke out these characteristics. A Fokker-Planck equation in its simplest form allows one to determine the probability density function of the phase space for a system of stochastic differential equations describing a Markov process (Gardiner 1985). With our prior assumption of parameterizing the small-scale variability as white noise, a Fokker-Planck equation exists for our low-order system. The advantage of using Fokker-Planck analyses is that a stochastic differential system's evolution is solved through use of a deterministic partial differential equation (Birner and Williams 2008, Sura 2002).

If we write our system in the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(\mathbf{x}) + \mathbf{B}(\mathbf{x}) \chi_m + \chi_a, \quad (29)$$

where $\mathbf{A}(\mathbf{x})$ is the vector of deterministic components for the system, \mathbf{B} is the square

matrix of coefficients modifying the multiplicative noise term χ_m , and χ_a are the additive noise components, then the Fokker-Planck equation takes the general form (Gardiner 1985; Horsthemke and Lefever 1984)

$$\begin{aligned}
\frac{\partial p(\mathbf{x}, \mathbf{t})}{\partial t} = & - \sum_i \frac{\partial}{\partial x_i} [A_i p(\mathbf{x}, \mathbf{t})] \\
& - \alpha \sum_i \frac{\partial}{\partial x_i} \sum_{j,k} (\sigma_{mi})^2 p(\mathbf{x}, \mathbf{t}) B_{jk} \frac{\partial}{\partial x_j} B_{ik} \\
& + \frac{1}{2} \sum_{i,j} (\sigma_{mi})^2 \frac{\partial^2}{\partial x_i \partial x_j} (\mathbf{B}\mathbf{B}^T)_{ij} p(\mathbf{x}, \mathbf{t}) \\
& + \frac{1}{2} \sum_i (\sigma_{ai})^2 \frac{\partial^2}{\partial x_i^2} p(\mathbf{x}, \mathbf{t}) .
\end{aligned} \tag{30}$$

The first term on the right-hand side is called the deterministic drift and it describes the dynamics of the deterministic system. The next term is called the noise-induced drift and it arises from state-dependent noise χ_m in the stochastic differential equation (29). The third and fourth terms on the right-hand side are diffusive terms and arise from the multiplicative noise and from the additive noise, respectively.

One important note here is that the noise-induced drift is modified by an α term (unrelated to the damping α terms). This α is dependent on the interpretation of stochastic calculus involved in the system (Sura 2002). For Itô calculus, discrete fluctuations are assumed to be uncorrelated and thusly best modeled as white noise. Stratonovich calculus assumes the fluctuating quantities have a finite correlation time but that we model these fluctuations as white noise where the noise strength is small relative to the deterministic quantities. Under the Itô assumption, $\alpha = 0$, while under the Stratonovich assumption, $\alpha = 1/2$. Our modeling of the noise falls under the Stratonovich calculus and thus includes a nonzero α . But since the current setup of our system does not include any state-dependent noise, the Fokker-Planck equation for our system does not include any noise-induced drift and thus the integration of the Fokker-Planck equation is independent of the interpretation of stochastic calculus.

We take advantage of the existence of only two stable solutions in the low-order system. For a dynamical system with many stable solutions, the probability density function (PDF) in phase space would perhaps be more convoluted than that for the low-order system. Instead, integration of the Fokker-Planck equation will only fluctuate the values of the probability density function p between the two solutions. This straightforward evolution helps to present simple solutions from an otherwise complex analysis.

The evolution of the dependent variable p provides not only information on the preferred states of the phase space for a given noise strength, but also provides information on the time evolution of the system to this preferred state. The more quickly one observes the solution converge to the equilibrium in phase space, the more quickly one should expect the system to likewise converge to its equilibrium state for a given set of parameters. For the following results, the parameter which we shall vary is the strength of the noise.

We mention here that it is possible in principle to obtain the final solution of the PDF simply by repeatedly integrating the system. By counting the number of times the system rests in either the sudden warming state or the radiative state, we could numerically calculate the quantity p . The downside to this is it would require many integrations (perhaps on the order of 10^4) to well-approximate the final values of p . While this is one possibility that was considered, integration of the Fokker-Planck equation was found to be equally viable for our purposes.

To solve the Fokker-Planck equation, we adopt the same techniques as described in Birner and Williams (2008). The equation is integrated using a second-order centered-in-time and centered-in-space finite difference scheme. A nondimensional integration domain is defined by $X \in (-0.06, 0.04)$, $Y \in (-0.05, 0.05)$, and $U \in (0.00, 0.80)$, where the respective dimensions are $20 \times 20 \times 40$. The integration is performed to 200 days with a nondimensional time step of 10^{-4} . The solution for p is initialized as a Gaussian distribution that is of width one-twentieth of the domain size and is centered on the radiative solution.

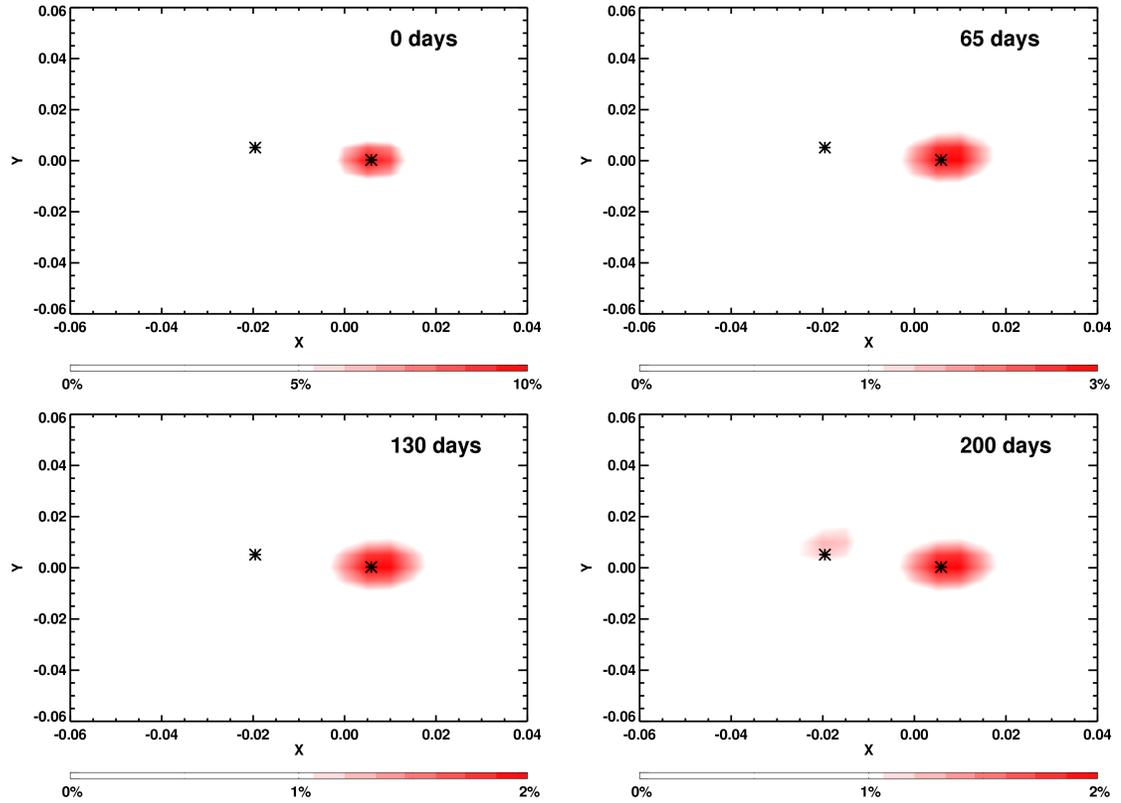


FIG. 21. The numerically integrated probability density function from the Fokker-Planck equation for $h = 80$ m of wavenumber 2 forcing and 0.1 m/s/day additive noise. The contours are the probability density function integrated over the U dimension and plotted as functions of X and Y for 0 days (upper-left), 65 days (upper-right), 130 days (lower-left), and 200 days (lower-right). The asterisks represent the stable solutions of the low-order model, where the equation has been initialized about the radiative solution.

Figure 21 plots four points in the integration of the Fokker-Planck equation for $h = 80$ m of wavenumber 2 forcing and an additive noise strength of 0.1 m/s/day. The plots are the probability density function at the specified time integrated over the U dimension, and presented as functions of the nondimensional X and Y terms. $h = 80$ m was chosen as a forcing strength because it falls well within the bistable regime for steady forcing of wavenumber 2. This allows a situation where if the noise is of sufficient strength, it may perturb the solutions towards the sudden warming solution. The asterisks denote the two stable solutions; as a reminder, the Fokker-Planck equation is initialized about the radiative solution.

What can be seen from Figure 21 is that the probability density function does not evolve much over the period of integration until 200 days. One may note that the scales of contouring are decreasing with time, indicative of a spreading of the probabilities. As well, a local maximum appears over the sudden warming solution. However, the global maximum of the probability density function remains stably over the radiative solution. Thus 0.1 m/s/day noise for 80 m of wavenumber 2 forcing is sufficient to produce sudden warmings in the model, but the probability that it does at this low forcing amplitude is small relative to stability about the radiative solution. The primary effect of the noise in this integration is to broaden the maximum of the probability density function about the radiative solution.

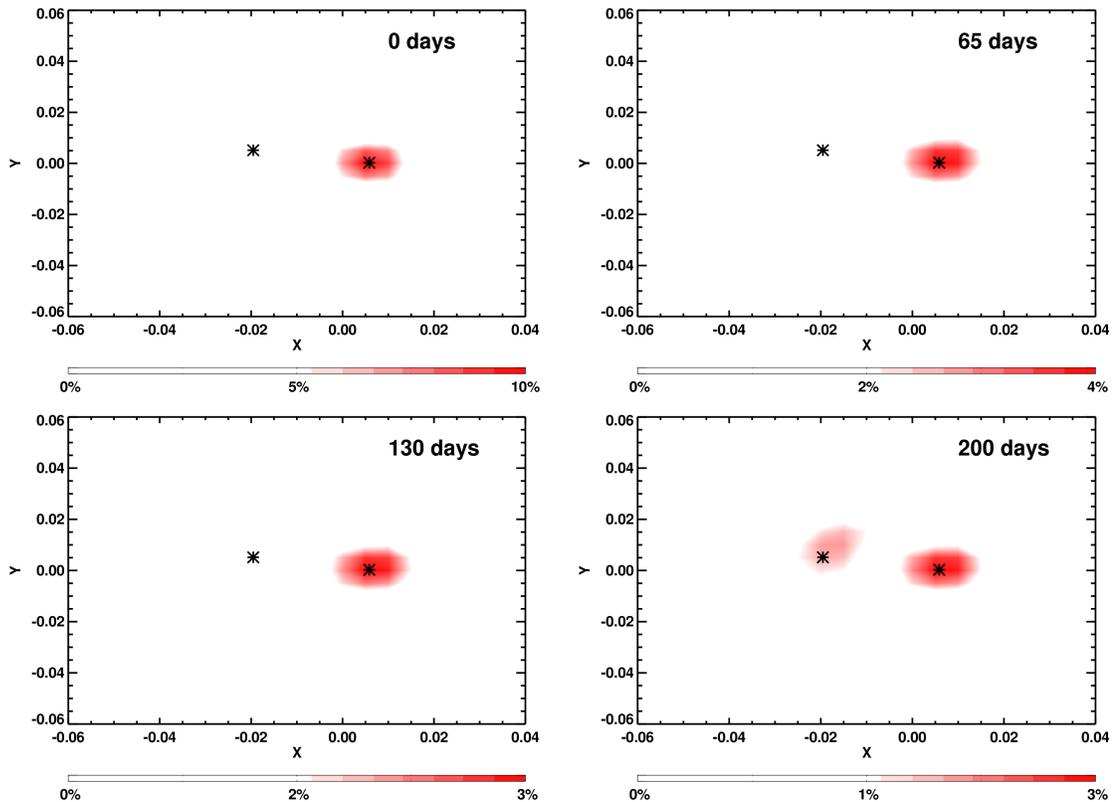


FIG. 22. As in Figure 21, but for 1.0 m/s/day noise strength.

One observes that even for an order of magnitude increase in this noise strength the system is not so perturbed so as to always stabilize to the sudden warming solution, as can be seen in Figure 22 which is for 80 m of wavenumber 2 forcing and 1.0 m/s/day noise

strength. The evolution of the Fokker-Planck equation nearly mirrors that for 0.1 m/s/day noise strength in that the probability density function broadens around the radiative solution until near the end. At 200 days, one observes that a local maximum has developed over the sudden warming solution.

One should note that the shading scales for Figures 21 and 22 are not equal. This should inform the reader that while the evolution qualitatively matches between the two noise strengths, the probability that sudden warming solutions appear for a noise of 1.0 m/s/day is greater than for noise of 0.1 m/s/day. Summing the values of the PDF over the range of the plotted maximums, approximately 22% more sudden warming solutions occur for 1.0 m/s/day noise strength than for 0.1 m/s/day noise strength.

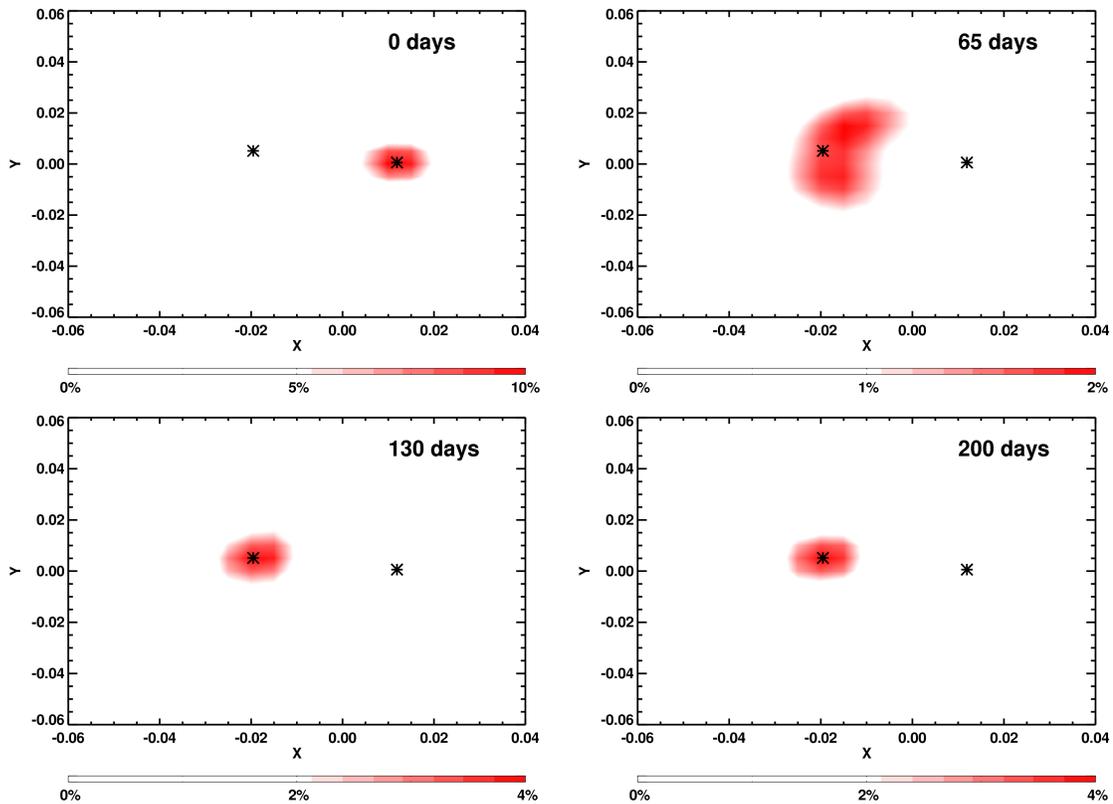


FIG. 23. As in Figure 21, but for a forcing amplitude of 150 m.

To further highlight the effect of the noise on the low-order system, we present Figure 23 wherein the Fokker-Planck equation is integrated with 150 m of wavenumber 2 forcing

and 0.1 m/s/day noise strength. $h = 150$ m was chosen because it is 30 m less than the sudden warming critical point, placing it well within the bistable regime but still close to sudden warming critical point. The probability density function evolves rapidly towards the sudden warming solution such that this solution is the most likely state of the model for these parameters at only 130 days into the integration.

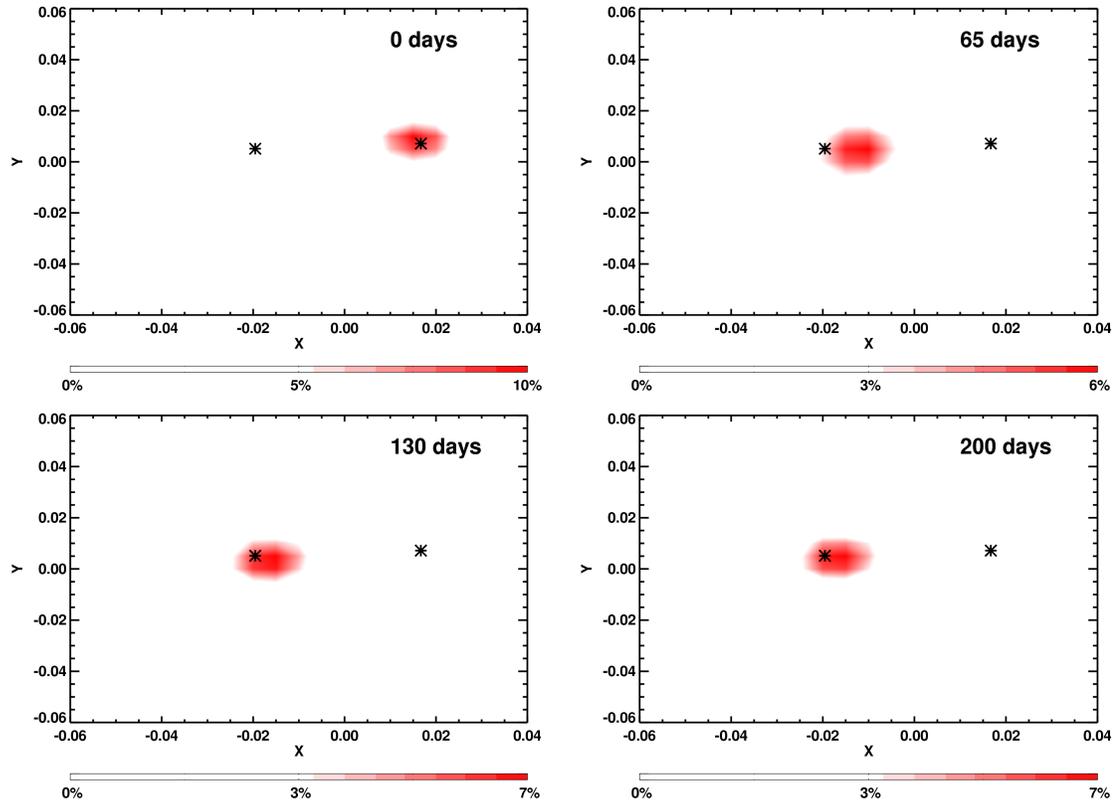


FIG. 24. As in Figure 21, but for wavenumber 1 forcing amplitude of 200 m and 0.1 m/s/day noise strength.

Here we contrast these results with the Fokker-Planck results from Birner and Williams (2008), noting that they considered only a wave amplitude of 100 m. What we show is that the sensitivity of the system to the strength of the noise is itself sensitive to the magnitude of the wave amplitude. For wave amplitudes near to the stationary wave sudden warming critical point ($h = 180$ m for wavenumber 2), only small noise strengths are required to perturb to system into the sudden warming state. For wave amplitudes nearer to the

radiative critical point ($h = 29$ m for wavenumber 2), large noise strengths are required to consistently perturb the system to the sudden warming solution.

We continue by integrating the Fokker-Planck equation with wavenumber 1 forcing. The basis of our understanding of the noise is that it will perturb the zonal wind in the low-order system between stability states, primarily from radiative solution to sudden warming solution. For wavenumber 2 this can be conceptualized as jumping the stationary solution from the upper branch of the bifurcation diagram to the lower branch. For wavenumber 1, there does not exist such a bifurcation. Then the noise for this case can not force a rapid modulation of the system.

Figure 24 plots the evolution of the probability density function as in Figure 21 but for 200 m of wavenumber 1 forcing and 0.1 m/s/day noise strength. $h = 200$ m was chosen because this amplitude of the forcing places the stationary solution for wavenumber 1 within the radiative solution but still near to the smooth transition between radiative solution and sudden warming solution. One observes that a maximum of probability forms over the sudden warming solution by 65 days. The rapid development of this maximum continues throughout the remainder of the integration until the probability density is stably integrated to the sudden warming solution.

The significance of this result relates to the lack of bifurcation for wavenumber 1. From Chapter 3, we observed that the behavior of the zonal wind with respect to constant wavenumber 1 forcing amplitudes required an amplitude of approximately 450 m to force the system into the sudden warming solution (here arbitrarily defined as zonal wind < 15 m/s). What this Fokker-Planck experiment evidences is that even for constant forcing amplitudes much less than this value, the inclusion of small-scale variability through a noisy process can produce a (non-sudden) warming solution. One of the primary problems we previously noted for our choice of parameters was that the low-order model with only wavenumber 1 forcing required unphysical values of incoming wave amplitude. What

we show here however is that such unphysical amplitudes are not necessary so long as the model parameterizes small-scale variability.

What these Fokker-Planck integrations demonstrate is the ability of the noise parameter to perturb the model — which is initialized in the bistable regime — from the stable radiative solution to the stable sudden warming solution. When this perturbation is included for small forcing amplitudes (relative to the range of the bistable region), the probability that the model transits to the sudden warming solution is small. For forcing amplitudes near to the critical point of the bistable regime, the probability that the model transits to the sudden warming solution is greatly increased. This noise also helps to dramatically decrease the necessary strength of the forcing for wavenumber 1 to force the model to the sudden warming solution.

5 TRANSIENT FORCING OF SUDDEN WARMINGS

5.1 TRANSIENT WAVE AMPLITUDES

The basis of Holton and Mass (1976) was that stratospheric vacillations are possible even under steady tropospheric forcing. That is, for stationary tropospheric longwaves of sufficient amplitude, one may observe the modeled system to undergo a change from a stable vortex state to a state of repeated SSWs. While this has important implications for sudden warmings, much of the troposphere to stratosphere wave flux in the real atmosphere is transient in nature.

Physically, the picture one should have in mind for this discussion of transient forcing is of a topographically-forced planetary wave in the troposphere. As the speed of the zonal wind blowing over the topography varies, so too does the amplitude of the resulting planetary waves. These waves with transient amplitudes propagate into the stratosphere and subsequently form the bottom boundary to the Holton-Mass model. It is the goal of this section to explore this transience of the incoming tropospheric waves and how this amplitude transience affects the behavior of sudden warmings.

Two problems arise in our consideration of these topographic planetary waves with transient amplitudes. Firstly, Yoden (1987a) discusses the use of topographically-forced planetary waves within a β -plane channel model framework. He notes that so long as the stratosphere does not effect the tropospheric flow, then the bottom boundary condition of a prescribed zonal wind and of a prescribed geopotential perturbation is satisfactory. Yet sudden warmings in the real atmosphere do contribute to variations in tropospheric flow (Baldwin and Dunkerton 2001). Then our prescribed bottom boundary conditions require that the zonal wind over the topography be modified by some other means and not be

influenced by the stratospheric circulation. While these conditions may be unphysical, they are correct to first order especially considering that the model framework we use in this study does not allow for stratosphere-troposphere coupling.

Secondly, we can not resolve waves of varying phase speed in a model that has assumed stationary waves. The Holton-Mass model contains the fundamental physics and simplicity which we desire to study, but not the capacity for transient forcing. Nonetheless, we feel that with careful explanation of the physical assumptions that have been made do allow for us to still utilize the Holton-Mass model (and derived low-order model) with transient wave amplitudes.

In the derivation of the Holton-Mass model, the wave assumption

$$\psi'(x, y, z, t) = \text{Re} [\Psi(z, t) e^{ikx}] e^{z/2H} \sin ly$$

was made to describe the geostrophic streamfunction. We note that $\Psi(z, t)$ is related to the incoming wave amplitude through the bottom boundary condition

$$\Psi(z_B, t) = \frac{g}{f_0} h(t)$$

where h is prescribed as a function of time. This allows us to mathematically include a wave forcing amplitude that varies with time despite this violating the assumption within the model that the phase speeds of the waves are constant.

Randel (1987) shows the vertical propagation speeds for transient waves of the first three wavenumbers (see also Alexander and Shepherd, 2010). Through the use of cross-correlation techniques he demonstrates that wavenumber 1 vertical propagation time scales from troposphere to middle stratosphere are on the order of four days while wavenumbers 2 and 3 propagate on the order of one to two days. Since observations show that transient wave vertical propagation time scales are short relative to radiative damping, we

can reasonably include these transient waves within both the Holton-Mass model and the low-order model.

One previous study of incoming wave amplitude transience is Harnik (2009), wherein this transient behavior of incoming waves is explored in the context of wave reflection. Through use of the Holton-Mass model, it is shown that for wave pulses of equal maximum magnitude, shorter wave pulses produce a fundamentally different response than longer wave pulses in the model. Short waves generate a weaker, more short-lived deceleration that was confined to upper levels of the model as compared to waves of long duration which generate a strong deceleration that penetrates the depth of the stratosphere.

Comparing runs of equal maximum magnitude and varying duration, one observes that the incoming integrated wave activity is much larger for longer duration forcing. Thus the model was likewise run with wave forcing of varying duration but equal integrated wave activity. This maintained a constant total incoming wave activity between runs which further allowed for analysis of the effects of wave duration. As before, those waves with longer duration generate a much deeper deceleration despite the decrease to the amplitude of the forcing. Figure 25 is Figure 6 from Harnik (2009), which summarizes these results. In Figure 25 (b-d), the thick curve plots the wave pulse, the thin curves plot the zonal wind speeds, and the dashed curve is the minimum wind line.

Through these results, Harnik demonstrated how transient wave amplitudes effect the Holton-Mass model. However these experiments only consider the model forced with an idealized pulse of wave activity. We wish to expand on these results by forcing the models with both a similar idealized forcing and with a larger ensemble of transient wave forcings.

It is our hypothesis that both the low-order model and the Holton-Mass model display a high sensitivity to the duration of the incoming forcing. We expect that wave pulses with long duration will force the model to the SSW stability state at a much lower amplitude than wave pulses with short duration.

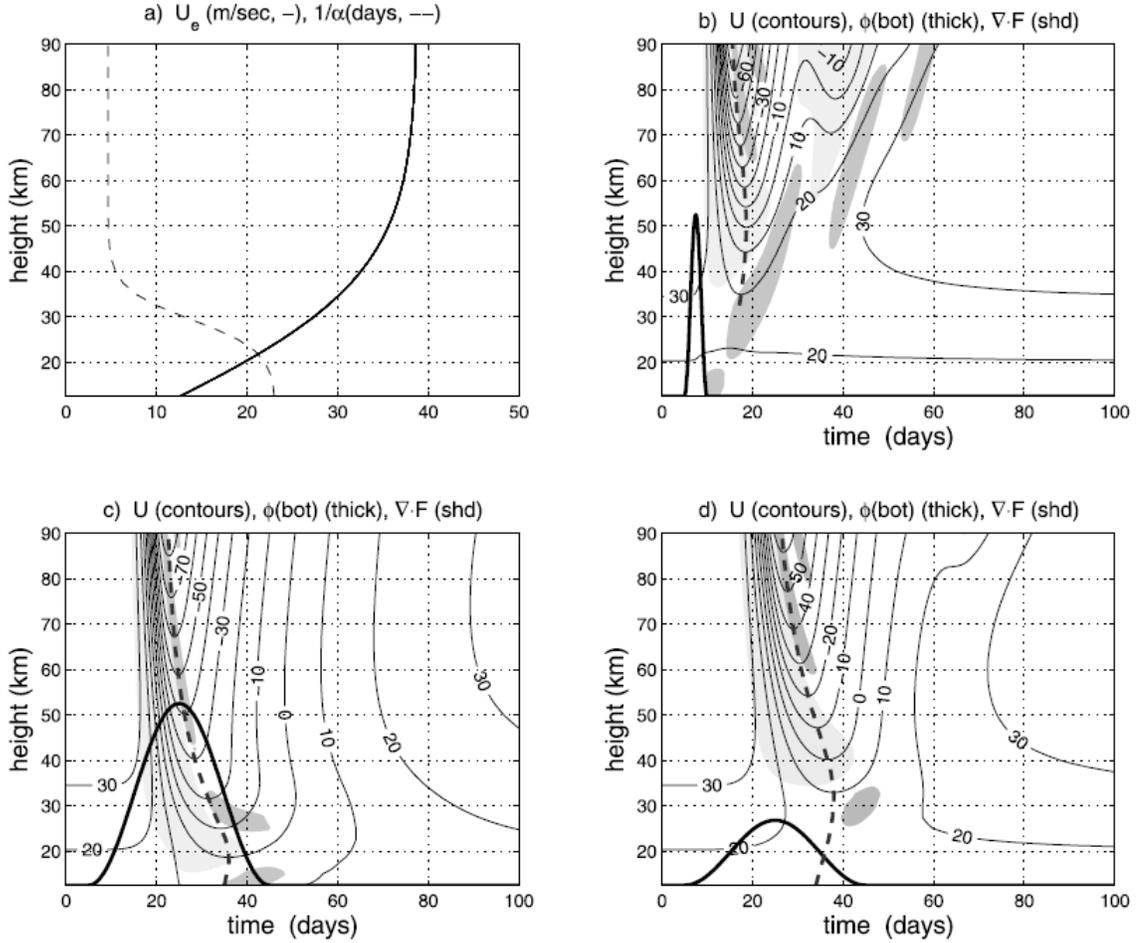


FIG. 25. Figure 6 from Harnik (2009). (a) Profiles of the radiative equilibrium zonal wind profile (solid) and the radiative damping time scale (dashed) Harnik used in the Holton-Mass model runs. (b)-(d) Time-height plots of the zonal wind (contours) for the plotted wave pulses (thick solid line). The dashed lines represent the line of minimum winds. Figure (c) is for a pulse of longer duration than (b), but with equal maximum amplitude. Figure (d) is for a pulse with equal integrated wave activity as in (b).

5.2 LOW-ORDER MODELING OF TRANSIENT AMPLITUDES

We begin here with the low-order model presented by Ruzmaikin (2003) and derived previously. By forcing this simple model with both highly-idealized wave amplitude forcing and quasi-random forcing, we can eke out this reduction in warming occurrence from transient wave behavior. To further these model results, we shall carry out the following tests in the Holton-Mass model in the next subsection.

The form of the idealized forcing is a monotonically ramping forcing amplitude followed by a constant forcing at the maximum amplitude (h_{\max}). By controlling the time of the ramping (τ_{ramp}) and the time of the constant forcing (τ_{const}), we can vary the model forcing between that of a stationary amplitude and of a highly-transient amplitude. A schematic of this idealized forcing is plotted in Figure 26. Long ramping time scales or long constant forcing time scales mimic stationary amplitudes, while short time scales mimic transient amplitudes.

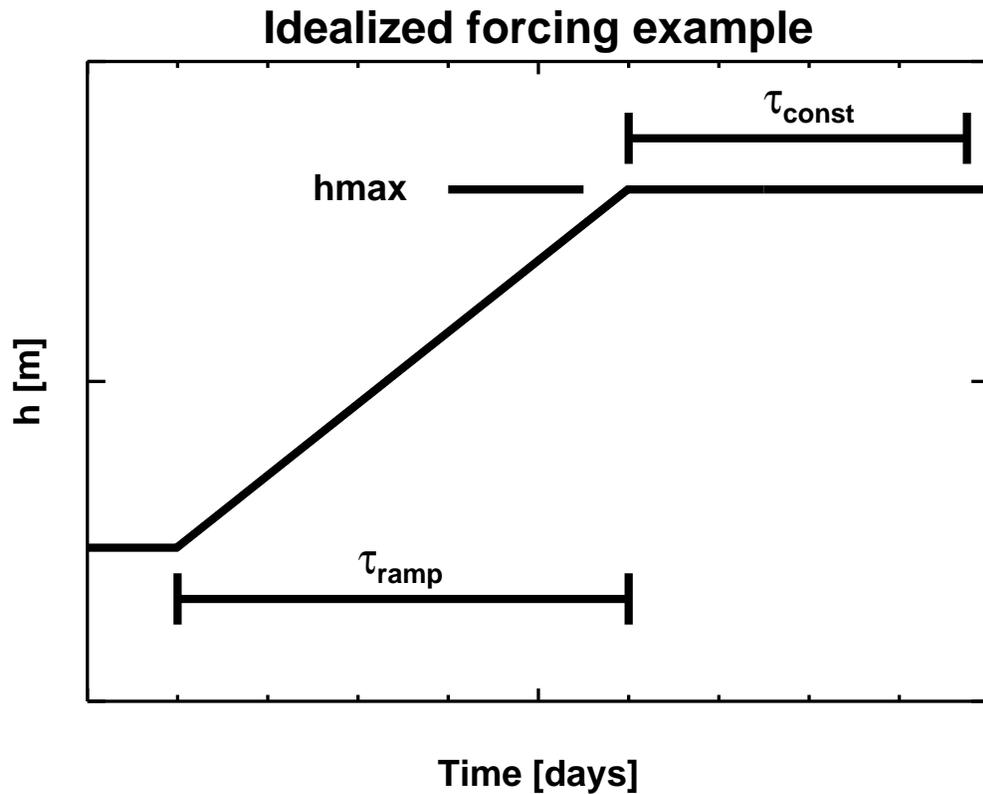


FIG. 26. Schematic of the idealized forcing used for testing the low-order model with transient wave amplitudes. τ_{ramp} represents the duration during which the forcing amplitude is ramped up to the maximum amplitude, represented by h_{\max} . τ_{const} represents the duration during which the forcing amplitude at h_{\max} is held constant.

The purpose of using this idealized forcing is to identify the smallest forcing h_{\max} that produces a warming solution in the low-order model. With the goal being to show that the forcing of sudden warmings decreases as the waves become more transient, our expectation

was to observe that the variable h_{\max} increases as the wave becomes more transient. This would imply that short wave pulses must be of much larger amplitude than the amplitude for stationary wave pulses to force the zonal wind into the sudden warming solution.

These runs were initialized in the radiative solution and with a forcing amplitude of 80 m. The model was allowed to stabilize for 50 days, after which time noise was introduced at a constant forcing of 0.1 m/s/day. To ensure statistical confidence of these results, the model was run 100 times for each experiment.

A contour plot of h_{\max} with respect to τ_{const} and to τ_{ramp} is shown in Figure 27, where wavenumber 1 forcing is on the left and wavenumber 2 forcing is on the right. From the wavenumber 1 plot, one can observe that for τ_{const} longer than 10 days and τ_{ramp} shorter than 20 days, the necessary magnitude to force the model to a warming state is less than is required to force the sudden warming with a steady amplitude. This reduction appears to only occur for decreasing τ_{ramp} and stabilizes to approximately 230 m for h_{\max} so long as τ_{const} is longer than 10 days.

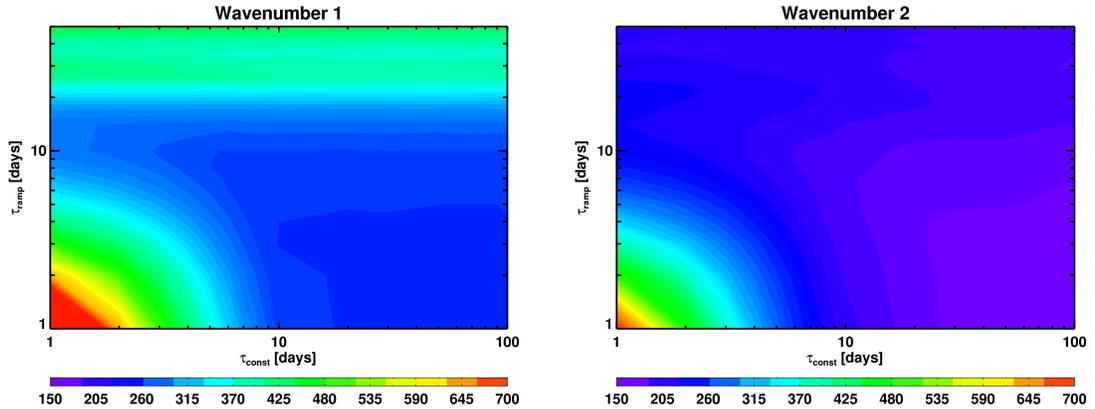


FIG. 27. Contours of h_{\max} as a function of the ramping time τ_{ramp} and the constant forcing time τ_{const} for wavenumber 1 (left) and wavenumber 2 (right).

The more relevant feature which also appears is a dramatic increase in h_{\max} for τ_{ramp} and τ_{const} time both less than 10 days. These time scales of forcing simulate a highly transient wave amplitude and the increase in h_{\max} with the increase in transience is in agreement with our hypothesis.

From the wavenumber 2 plot, one may observe that for τ_{ramp} longer than approximately 10 days and a τ_{const} longer than approximately 10 days, the model behaves as if forced by a steady wave amplitude: the required forcing magnitude for a sudden warming is at most 10% larger than the stationary solution amplitude. Yet for τ_{ramp} less than 10 days and τ_{const} below 10 days, we capture the transient wave amplitude behavior we expected. As the τ_{ramp} decreased (i.e., as transience increased), the wave amplitude needed to force the model to the warming solution increased dramatically.

This behavior is more easily observed in Figure 28 where the curves plot h_{max} as a function of τ_{ramp} for the given τ_{const} . In this figure, where the left plot is for wavenumber 1 forcing and the right plot is for wavenumber 2 forcing, the solid curve plots h_{max} for τ_{const} of 100 days while the dashed curve plots h_{max} for τ_{const} of 1 day. Again, as τ_{ramp} and τ_{const} both decrease, the transience of the wave increases.

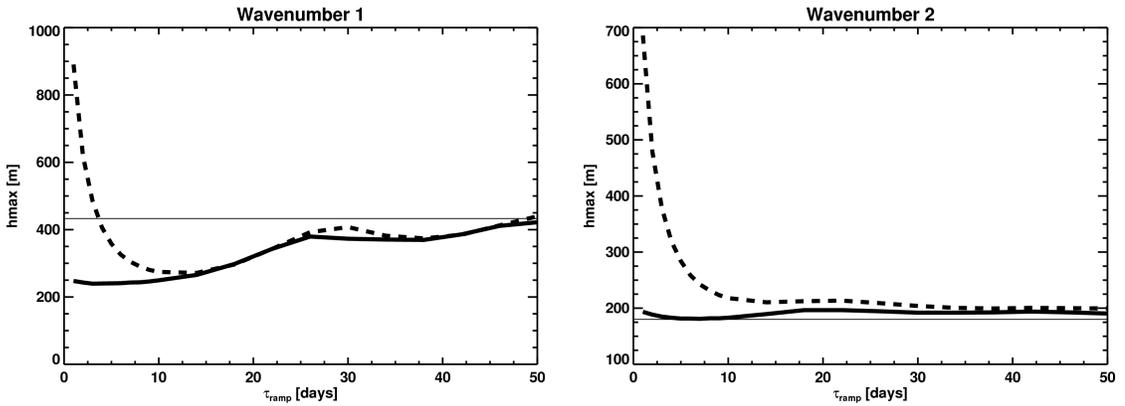


FIG. 28. h_{max} as a function of τ_{ramp} for wavenumber 1 forcing (left) and wavenumber 2 forcing (right). The thick solid curve plots h_{max} for τ_{const} of 100 days (quasi-stationary wave amplitude) while the dashed curve plots h_{max} for τ_{const} of 1 day (transient wave amplitude). The thin solid line plots the stationary wave amplitude solution for each wavenumber.

In Figure 28, one observes that wavenumber 2 forcing behaves as was expected. For long τ_{const} , the necessary magnitude h_{max} to force a sudden warming solution is only 10% larger than determined from the steady forcing solutions. For short τ_{const} , h_{max} rapidly increases once the ramping time is short relative to radiative damping time scales.

Wavenumber 1 mimics these results, though one observes the decrease in h_{\max} with decreasing ramping time down to approximately 10 days noted in Figure 27.

These idealized forcing tests present transient wave amplitude behavior for increasing transience with variable maximum amplitude. We follow these results by testing for the case of increasing transience with constant maximum forcing. A schematic of this experiment is given in Figure 29. Our idea here was to force the model with a transient forcing

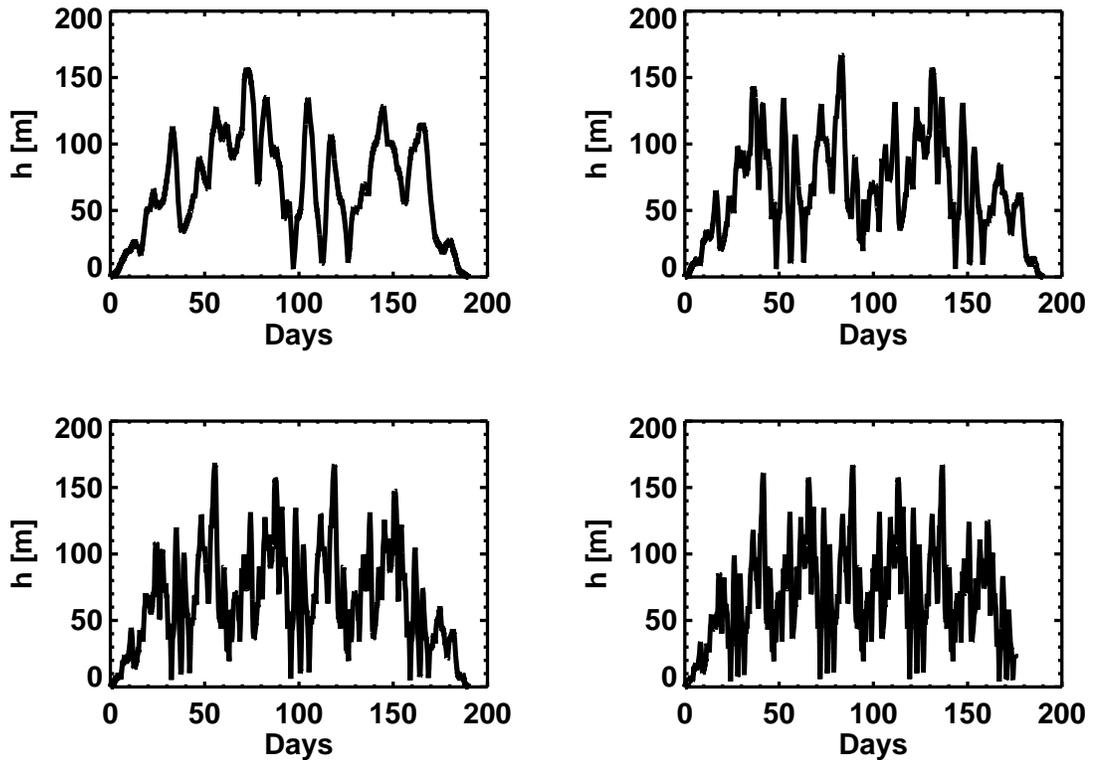


FIG. 29. Four forcing profiles utilized in the low-order model for the replication experiments. The upper left plot gives an initial forcing profile. The upper right plot is the single replication case where the initial forcing profile has been repeated but interpolated onto the same duration. Two replications (bottom left) and three replications (bottom right) are also shown.

profile (see upper left plot of Figure 29) and observe if this profile produced a warming solution. We then replicated the forcing profile by interpolating two instances of the same profile into the same time period. The upper right plot of Figure 29 gives the one replication case where the initial input has been repeated after itself, but interpolated to the same 190

day period as the initial profile. Thus the magnitude of the forcing was invariant while the transience of the forcing was increased. This process was continued out to 20 replications (3 of which are shown in Figure 29). Our expectation from these runs was that as the profile was replicated on to itself more times, the number of warmings resulting from these runs would decrease.

Important to note here is that the forcing was ramped up for 50 days from 0 m to the full magnitude of the forcing, and ramped down for 50 days from the full magnitude of the forcing to 0 m. Then for these experiments, only the central 90 days were considered so that the results presented have been limited to winter time scales.

To perform this test, we utilized an ensemble of quasi-random forcing extracted from the ERA-40 dataset geopotential at 100 hPa and 60°N latitude (to be discussed later with the observational data). Rather than producing our own set of randomized forcing profiles, these extracted data produced a physically-reasonable set of profiles with highly variable time scales and amplitudes. We ran the model 100 times for each year with an additive noise value of 0.1 m/s/day, which gave us an ensemble of 4500 runs for each number of replications. We note here that wavenumber 1 amplitudes were doubled over extracted values in the single wavenumber experiments since these unaltered amplitudes were too small to cause enough warming instances to demonstrate the behavior of the model. The plots of sudden warming solutions with respect to replication numbers for wavenumbers 1 and 2 are plotted in Figure 30.

Of the 4500 runs for wavenumber 1 forcing with no replications, 3806 instances of the model produced a sudden warming solution. As the forcing is replicated, this number of produced sudden warmings decreases rapidly. The total warming solutions goes to approximately 2391 out of 4500 for one replication; this represents nearly 40% fewer instances than for no replications. By seven replications, the total number of sudden warming solutions has gone to zero.

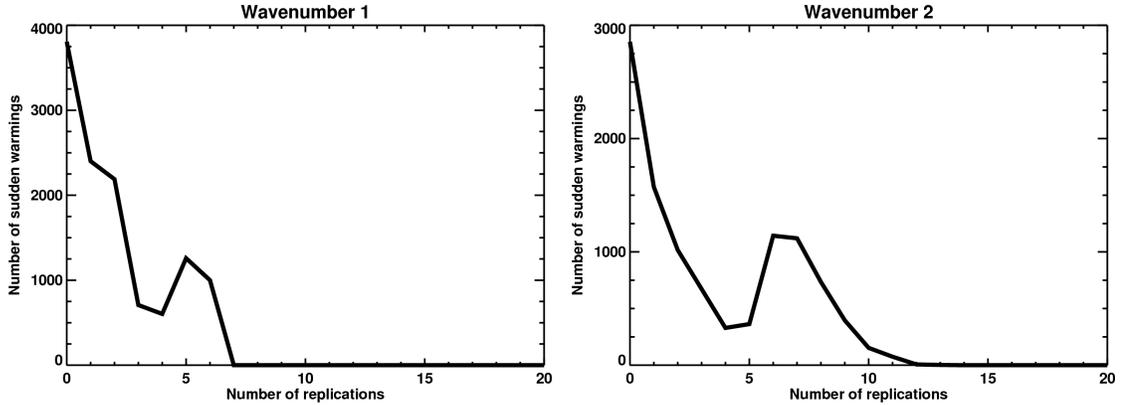


FIG. 30. Number of sudden warmings plotted as a function of the replications of the initial forcing profile. 0 replications corresponds to the unaltered initial forcing profile. There were 45 initial forcing profiles run 100 times at each number of replications for a total of 4500 runs at each number of replications. Wavenumber 1 forcing is plotted on the left while wavenumber 2 forcing is plotted on the right. Transience of the forcing increases with the number of replications such that one observes that the forcing of sudden warming solutions in the low-order model decrease with an increase in transience.

The right plot of Figure 30 shows the sudden warming solutions as a function of replication for wavenumber 2. The initial forcing profiles for this wavenumber forcing produces 2877 out of 4500 possible sudden warming solutions. After one replication, this number is reduced by nearly 45% to 1575 warmings. As with wavenumber 1 forcing, this trend continues until between five and six replications, where another abrupt increase in sudden warming solutions is observed. After this feature, the warming solutions steadily decline until the twelfth replication when the replicated forcing profiles produces no warming solutions.

One may observe that in Figure 30, the reduction in warming solutions is not a monotonic decrease as a function of replications for both wavenumbers. For example, there is a marked increase in warming solutions between five and six replications. While this feature is consistent between the wavenumbers, there exists the possibility that this behavior arises from the stochastic nature of the model runs. The noise strength was chosen such that this noise would neither always perturb the model towards the sudden warming solution at these amplitudes nor be of such small influence that the model behaves deterministically.

To test whether it is this random element within these experiments that produces the increases around five replications, we run the experiment again without noise. As shown in Figure 31, these sudden increases exist even in deterministic runs of the model. Thus it can not be the noisy forcing that creates this enhancement around five to six replications.

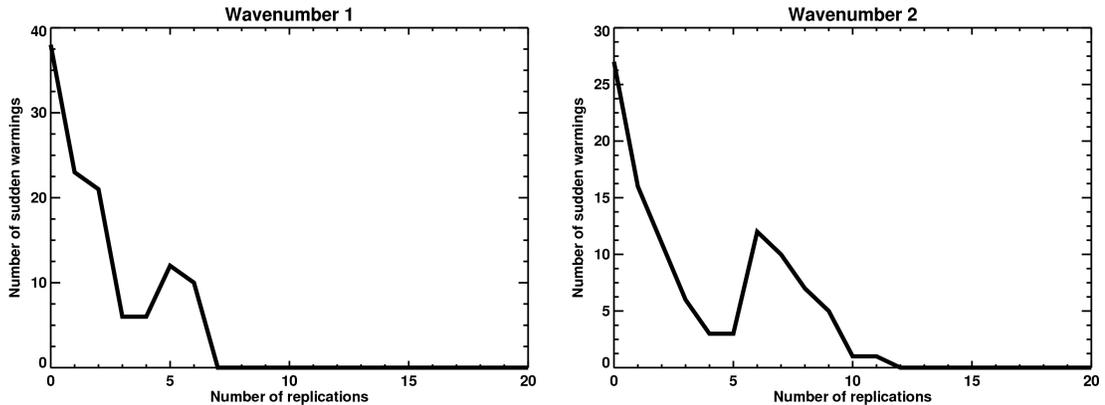


FIG. 31. As in Figure 30, but for no noise. Here we run the model only once for each replication for a total of 45 runs at each number of replications. This demonstrates that the noisy processes in Figure 30 are not causing the sudden increase in sudden warmings near five replications.

To explain these sudden increases we first note that the ensemble of forcing profiles which we have created is derived from reanalysis data and centered on January 15. If this ensemble is based off of observations, it should therefore contain the subseasonal cycle found in those observations. The subseasonal cycle of forcing can be qualitatively observed in Figure 32 which is Figure 4.6 in SPARC CCMVal (2010).

Figure 32 plots the observed and model-derived maximum amplitude of the 10 hPa climatological stationary wave for the Northern Hemisphere (left) and for the Southern Hemisphere (right). One may note that there exists a climatological maximum in stationary wave amplitudes near the middle of January, corresponding to the central point of our forcing profiles.

What this means for our ensemble is that each profile of spectral forcing amplitude inherently contains a single low-frequency variation in that amplitude. This wave in our ensemble has a 180 day period and is minimized at the bounds and maximized at the central

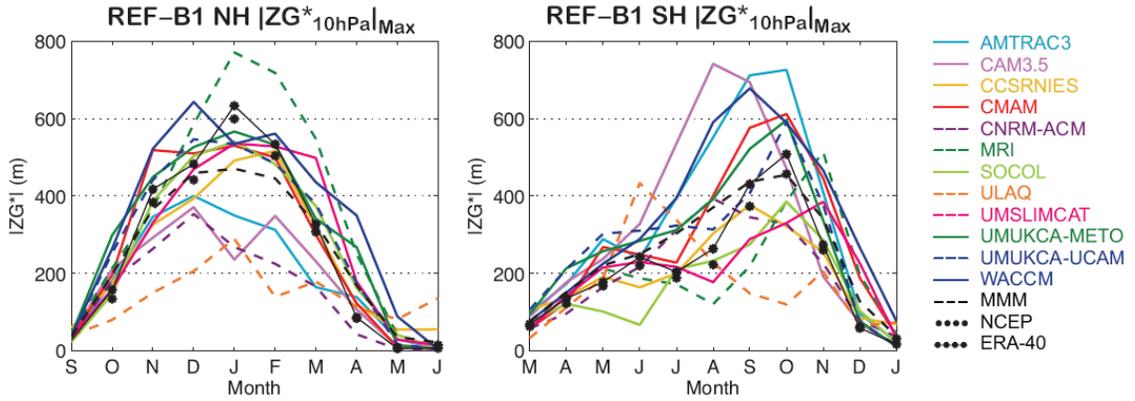


FIG. 32. Figure 4.6 of SPARC CCMVal (2010). Seasonal variation of the maximum amplitude of the Northern Hemisphere (left) and of the Southern Hemisphere (right) 10 hPa climatological stationary wave.

point. In the context of the replication experiment this implies that as the forcing profile is replicated, the period of this wave will decrease. At five replications, this means that the low-frequency variation in maximum amplitude for the initial forcing profile has a 36 day period. Noting also that the sudden increase observed in the sudden warming solutions extends to approximately seven replications, this means that the smallest period in the sudden increase is approximately 26 days.

The significance of forcing the model with a large amplitude inherent wave with a period between 36 and 26 days is that this period range falls close to the radiative damping time scale. Through the course of this study, we have observed that forcing the low-order model with a wave of forcing amplitude that has a period close to the radiative damping time scale produces a zonal wind behavior that is reminiscent of resonance.

Even for forcing amplitudes much less than what are required to force the model to the sudden warming solution, sudden warming solutions were found to exist if these resonant-like modes were forced. However, this aspect of the model has not yet been explored in great detail so we are unable to provide solid evidence for or against this argument. Thus we state without demonstration that it is likely that the sudden increase in sudden warming solutions in the five to seven replications range is due to forcing the model with a frequency that elicits a quasi-resonant response in the zonal wind.

The sudden increase in warming solutions aside, we have shown that by increasing transience of the forcing amplitude while holding the maximum amplitude constant, the low-order system produces fewer sudden warmings. One shortcoming of these replication runs is that as the transience is increased in this fashion, the vertically-propagating integrated wave activity is variable with each replication. Thus we ran the model with a similar set of experiments wherein the forcing was replicated, but then reduced by a factor to maintain constant total wave activity at the bottom boundary. Note that this technique is analogous to the wave activity bounded experiment done by Harnik (2009) and discussed in Figure 25.

This is performed by summing the forcing amplitude over the time domain to obtain the time-integrated wave activity for the forcing. The value of this integrated wave activity for no replications represents the initial total wave activity to which all replications will be set. This normalization to the initial wave activity is a straightforward process as the integrated values merely differ by a constant factor. As with the constant amplitude forcing ensemble, we use these constant total wave activity, replicated wave amplitude profiles to force the model so as to count the warmings with respect to the number of replications.

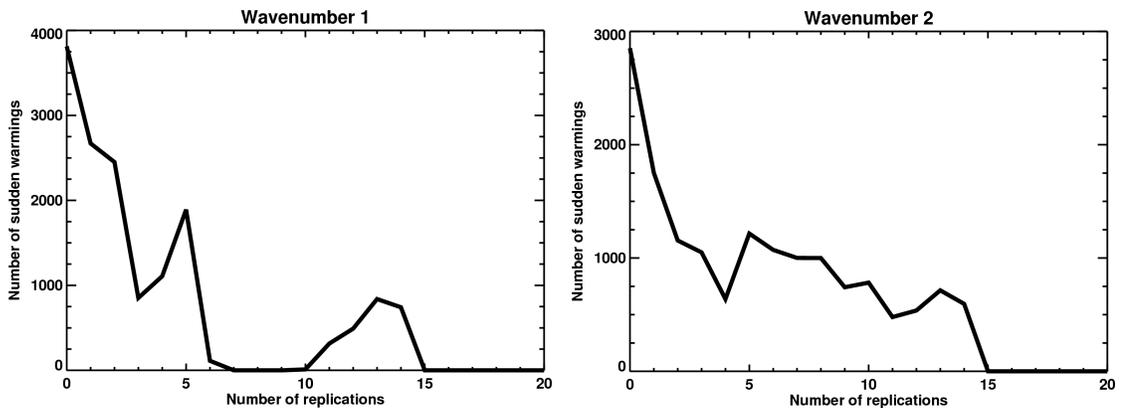


FIG. 33. As in Figure 30, but the amplitudes of the forcing profiles for each replication are multiplied by a factor such that the integrated wave activity of each forcing profile for each replication is held constant.

The results of this experiment are plotted in Figure 33. The response of the model to this wave activity bounded forcing is qualitatively the same as in the bounded amplitude experiments. There exists nearly the same sudden increases in sudden warmings as the transience increases, but overall the number of sudden warming solutions goes to zero with increasing transience.

Thus far we have presented transience tests only for the single wavenumber low-order model. It was previously shown that it is straightforward and in some ways beneficial to resolve multiple wavenumbers in the model. We now present results from the model for which wavenumbers 1 and 2 are resolved simultaneously within the model.

For the replication experiments with one wavenumber, the single wavenumber was replicated a certain number of times and then used to force the model. With two wavenumbers, we shall adopt four different replication tests. The first is to replicate both wavenumber forcing profiles; the second and third are to replicate each wavenumber individually while keeping the other at its initial profile. The final method is to replicate the wavenumber 1 forcing where its magnitude has been doubled while wavenumber 2 is kept at its initial profile. As in the single wavenumber model replication tests, the unaltered amplitude of wavenumber 1 forcing replications produces too few sudden warmings to provide conclusive evidence. By doubling the amplitude, we increase the number of sudden warmings over those forced by the unaltered wavenumber 1 amplitudes.

Figure 34 plots the results of these four replication methods: replication of both wavenumbers at top left, replication of unaltered wavenumber 1 at top right, replication of wavenumber 2 at bottom left, and replication of doubled wavenumber 1 at bottom right. Here again, the model produces fewer sudden warmings as transience increases in the forcing. We also note that the inclusion of both wavenumbers has not eliminated the occurrence of the sudden increases in warming solutions at approximately five replications.

The exception to the above result is for replication of unaltered wavenumber 1 (Figure 34 — top right). The understanding here is that the unaltered wavenumber 1 forcing

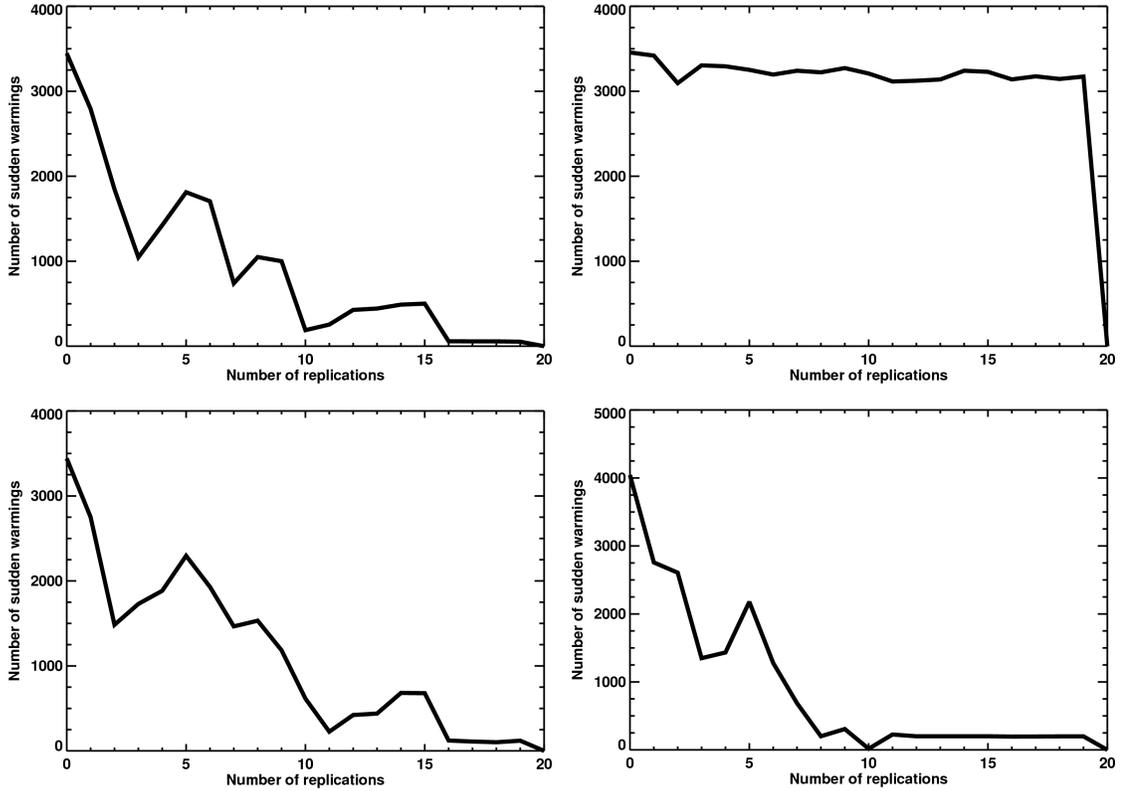


FIG. 34. Sudden warming solutions as a function of the number of replications performed on the initial forcing profiles but for wavenumbers 1 and 2 resolved in the low-order model. The top left plot is for both wavenumbers 1 and 2 replicated simultaneously, the top right plot is for wavenumber 1 replicated alone, the bottom left plot is for wavenumber 2 replicated alone, and the bottom right is for wavenumber 1 replicated alone but with wavenumber 1 amplitudes doubled over the initial values.

amplitudes are insufficient to drive the zonal wind to the sudden warming state and thus increasing transience in this forcing would not manifest. Thus the reason for the nearly constant sudden warming solutions is the model is primarily driven by wavenumber 2 forcing, which is unaltered here.

Integrated wave activity experiments were likewise performed with the two resolved wavenumbers. As in the single wavenumber case, holding the integrated wave activity constant created little variation from the results where the maximum amplitude is held constant (not shown).

We have thusly demonstrated in the low-order model that transient wave amplitude forcing decreases the number of sudden warmings. The picture one should take from these results is that short time scale wave pulses can only slightly perturb the polar vortex due to the brevity of its forcing. This small disturbance is easily restored with ample restoration time between wave pulses. Though if the wave amplitude is of large enough magnitude, the forcing on the zonal wind will be strong enough to transit the model to a sudden warming solution regardless of the short time scale.

To further these low-order model results, we now expand the study to the Holton-Mass model such that we may observe these effects in a one-dimensional system.

5.3 HOLTON-MASS MODELING OF TRANSIENT AMPLITUDES

By moving to the Holton-Mass model, we have increased the vertical resolution of the experiments from a three layer model to a 46 layer model. We repeat the previous transient forcing experiments within this less truncated framework to verify that the results are not due to the extreme vertical simplification utilized to form the low-order model.

To perform these Holton-Mass experiments, we alter a few model parameters. The bottom boundary zonal wind is prescribed to be 15 m/s. The low-order model bottom boundary condition of 10 m/s within the Holton-Mass model allows for very small amplitudes of the forcing wave to produce sudden warming solutions (see Holton and Dunkerton 1978, Yoden 1987b). Also, noise has been eliminated from these runs. Parameterization of the small-scale variability within the Holton-Mass model is not as straightforward as in the low-order model because some of this variability, such as gravity wave drag, has high sensitivity to altitude (Pulido and Thuburn 2006). Thus we exclude noise and shall present deterministic runs only.

Another alteration made for experiments with the Holton-Mass model is of the definition of sudden warming solution. In the case of the low-order model, this sudden warming solution condition was identified through use of the stationary solutions of the model but

arbitrarily defined to be any zonal wind less than 15 m/s. For the 46-layer Holton-Mass model, we alter this sudden warming solution condition to be zonal wind less than or equal to 0 m/s. While this does not follow the WMO definition of a sudden warming — zonal wind and temperature gradient reversal at 10 hPa — we take up this definition because we are not always concerned with the wind reversal only at 10 hPa (approximately the 36 km layer in the Holton-Mass model). Instead, this allows us to identify a “sudden warming solution” at any level within the model.

We begin as for the low-order model by exploring the effects of an idealized forcing presented schematically in Figure 26. Again, the purpose of forcing the model with such an idealized incoming wave is to identify the smallest maximum forcing amplitude necessary to drive the model to a sudden warming solution.

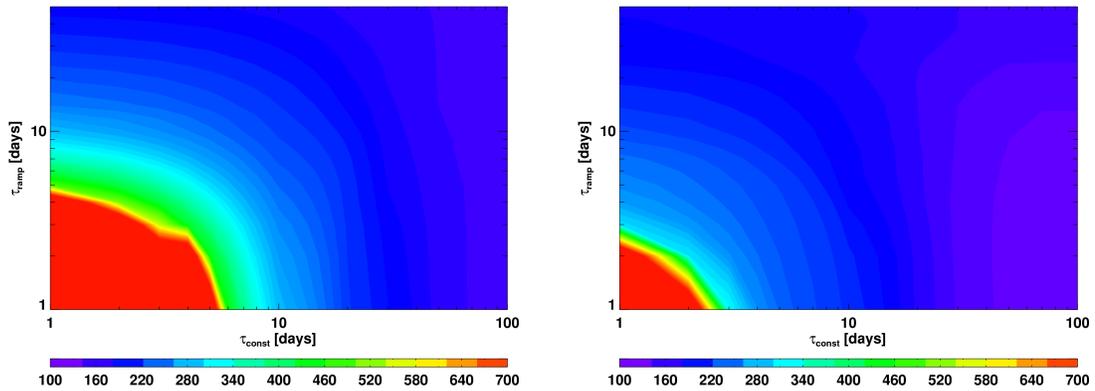


FIG. 35. As in Figure 27 but derived from the Holton-Mass model at the 36 km level and for wind reversal rather than zonal wind less than 15 m/s.

As was expected, the idealized forcing results of the low-order model hold for the Holton-Mass model. Figure 35 is a contour plot as in Figure 27 for wavenumber 1 (left) and wavenumber 2 (right) but for the 36 km level. Long τ_{const} (τ_{ramp}) allow the model to effectively feel a steady wave perturbation regardless of the length of τ_{ramp} (τ_{const}). For τ_{const} (τ_{ramp}) under 10 days, shorter τ_{ramp} (τ_{const}) — or higher transience — required a much higher maximum forcing amplitude to perturb the model to a warming solution.

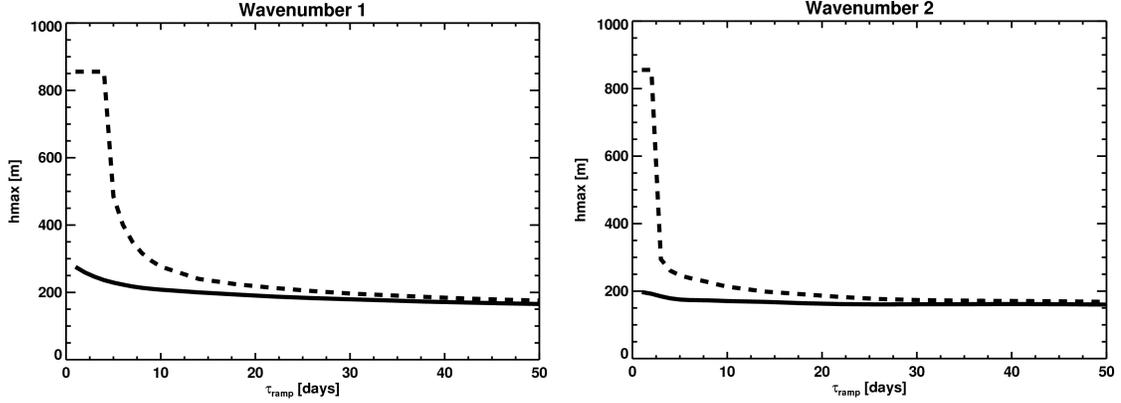


FIG. 36. As in Figure 28 but derived from the Holton-Mass model at the 36 km level.

It is also important to note that similar to the low-order model results, the variable h_{\max} varies only slightly with changes to the forcing period and to the τ_{ramp} if either the τ_{const} or the τ_{ramp} is longer than 10 days. This nearly invariant behavior, and the largely variant behavior for short τ_{ramp} and short τ_{const} , is evident in Figure 36.

We continue by repeating the replication experiments of the low-order model for the Holton-Mass model. We first force the model with wavenumber 1 forcing and count the sudden warming solutions at each vertical level with respect to the number of replications. To properly display the results from these experiments, Figure 37 plots the sudden warming solutions as a function of number of replications for four levels: 26 km (top left), 36 km (top right), 46 km (bottom left), and 56 km (bottom right). One observes that though the number of sudden warming solutions varies with altitude, the pattern of decreasing sudden warming solutions with increasing transience holds at all altitudes. Also from Figure 37, one may observe why it is that we consider the wind reversals at all levels rather than simply at 36 km: there are few instances of wind reversal at this level and thus this alone does not give the full result.

Figure 38 plots the number of sudden warming solutions as a function of number of replications as in Figure 37, but for wavenumber 2 forcing. Except that the number of sudden warming solutions at each level are increased, the behavior is nearly identical to

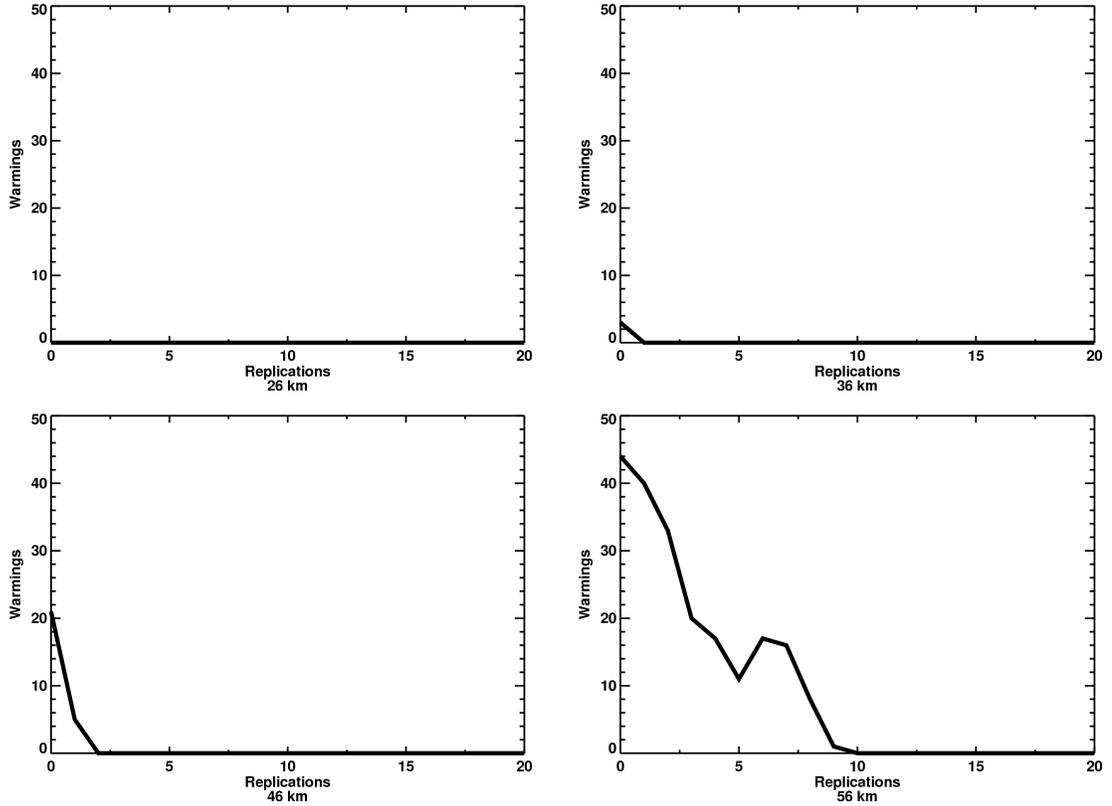


FIG. 37. The Holton-Mass model derived number of sudden warming solutions as a function of replications of the initial forcing profile of wavenumber 1 forcing. This figure includes four levels where the sudden warming solutions are most variable with the number of replications. These levels are 26 km (top left), 36 km (top right), 46 km (bottom left) and 56 km (bottom right).

that for wavenumber 1. As the number of replications increase, transience increases while the number of sudden warming solutions decrease.

As in the low-order model we also perform the same replication experiments but with invariant integrated wave activity from the initial forcing profiles rather than invariant wave amplitudes from the initial forcing profiles. Figure 39 presents the sudden warming solutions as a function of number of replications for this constant integrated wave activity at the same levels as in Figure 37. One observes that between Figures 37 and 39, the number of sudden warming solutions at each level are nearly unchanged despite the former holding wave amplitudes constant and the latter holding integrated wave activity invariant. The same results for wavenumber 2 forcing are presented in Figure 40. Comparing these results

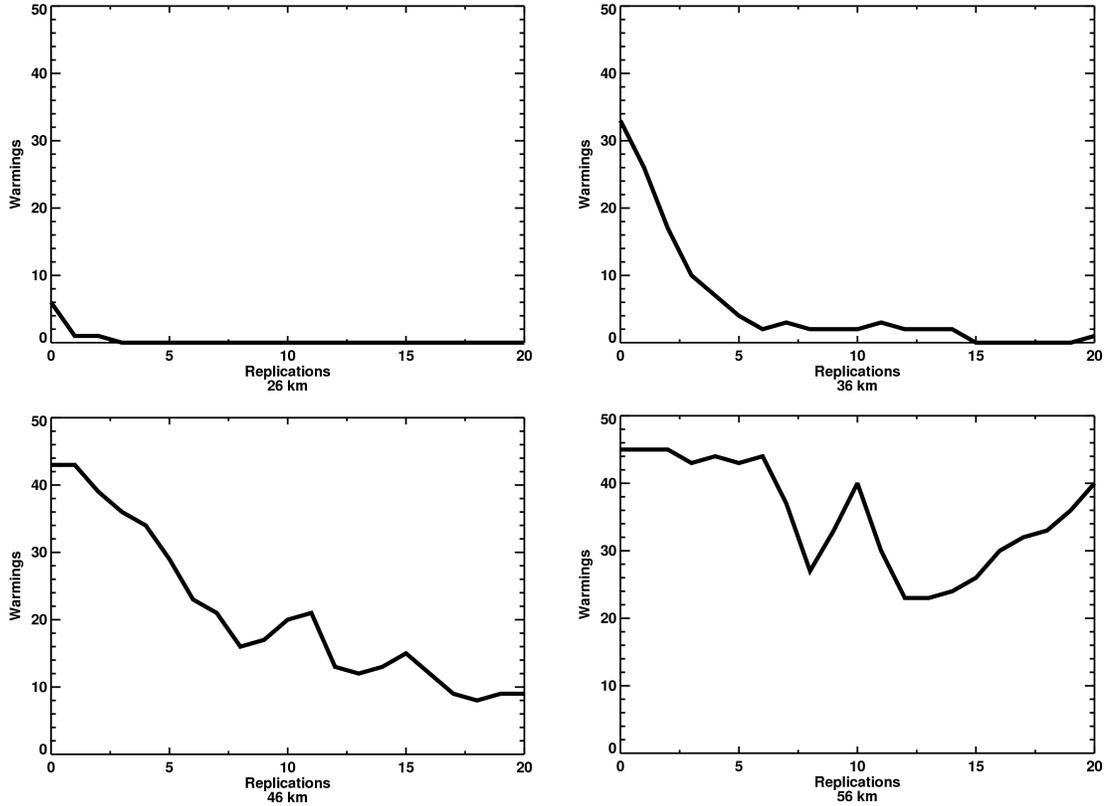


FIG. 38. As in Figure 37, but for replications of wavenumber 2 forcing.

for wavenumber 2 with Figure 38, one observes that there is no qualitative difference between not varying the wave amplitudes and not varying the integrated wave activity, though the number of sudden warming solutions at higher numbers of replications are higher in the case of invariant integrated wave activity.

Proceeding as with the low-order model experiments, we now perform these replication experiments for the Holton-Mass model with two resolved wavenumbers. Before presenting these results, we first note that the principal variability in the number of sudden warming solutions in Figures 37 to 40 occurs between the 26 km level and the 46 km level. Then to more succinctly present the results from these experiments, we may sum the number of sudden warming solutions between and including these layers as a function of the number of replications. We shall hereafter term results of this nature as “integrated sudden

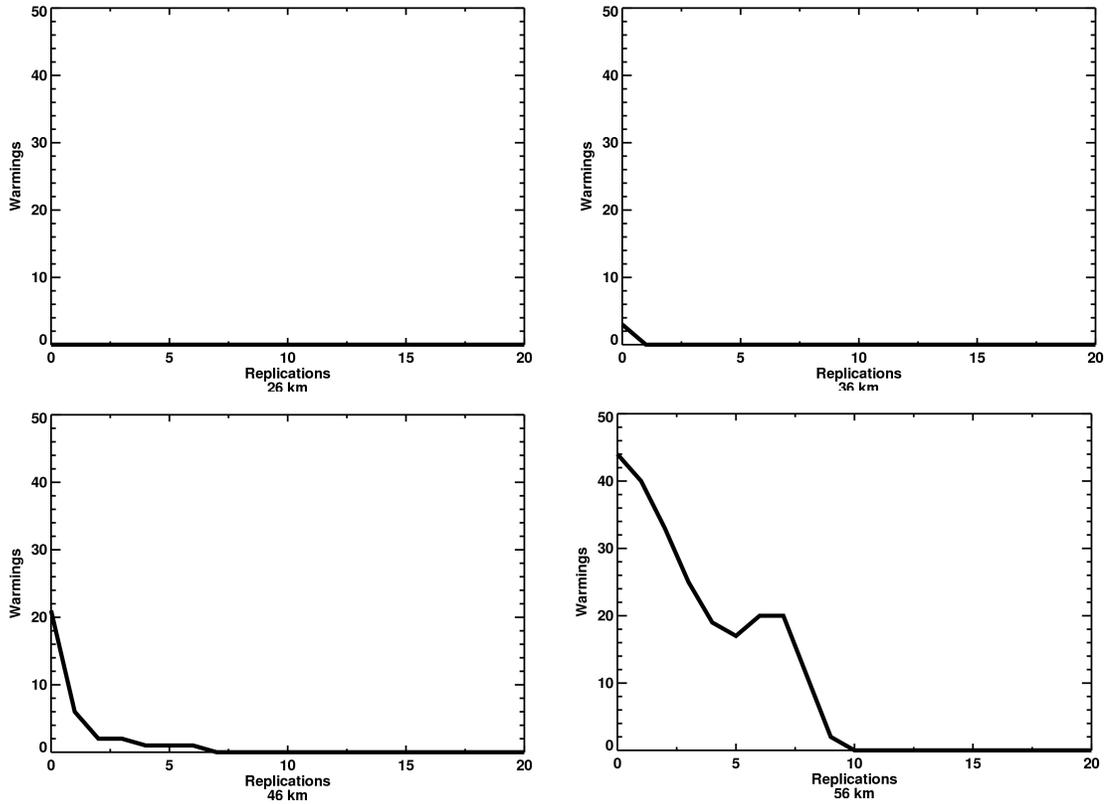


FIG. 39. As in Figure 37, but for invariant integrated wave activity from the initial forcing profiles rather than invariant wave amplitudes.

warming solutions.” To clarify this, we have replotted the results of Figures 37 and 38 as the integrated sudden warming solutions in Figure 41.

As in the low-order model, we perform four experiments: replications of both wavenumbers 1 and 2, replications of wavenumber 1, replications of wavenumber 2, and replications of wavenumber 1 with wavenumber 1 amplitudes doubled. Figure 42 plots the integrated sudden warming solutions for all four experiments — wavenumbers 1 and 2 in the top left, wavenumber 1 in the top right, wavenumber 2 in the bottom left, and wavenumber 1 doubled in the bottom right.

Foremost in Figure 42 is the significant reduction in integrated sudden warming solutions when wavenumbers 1 and 2 are both replicated. By five replications of both wavenumbers there has been an almost 50% reduction in wind reversals between 26 km and 46

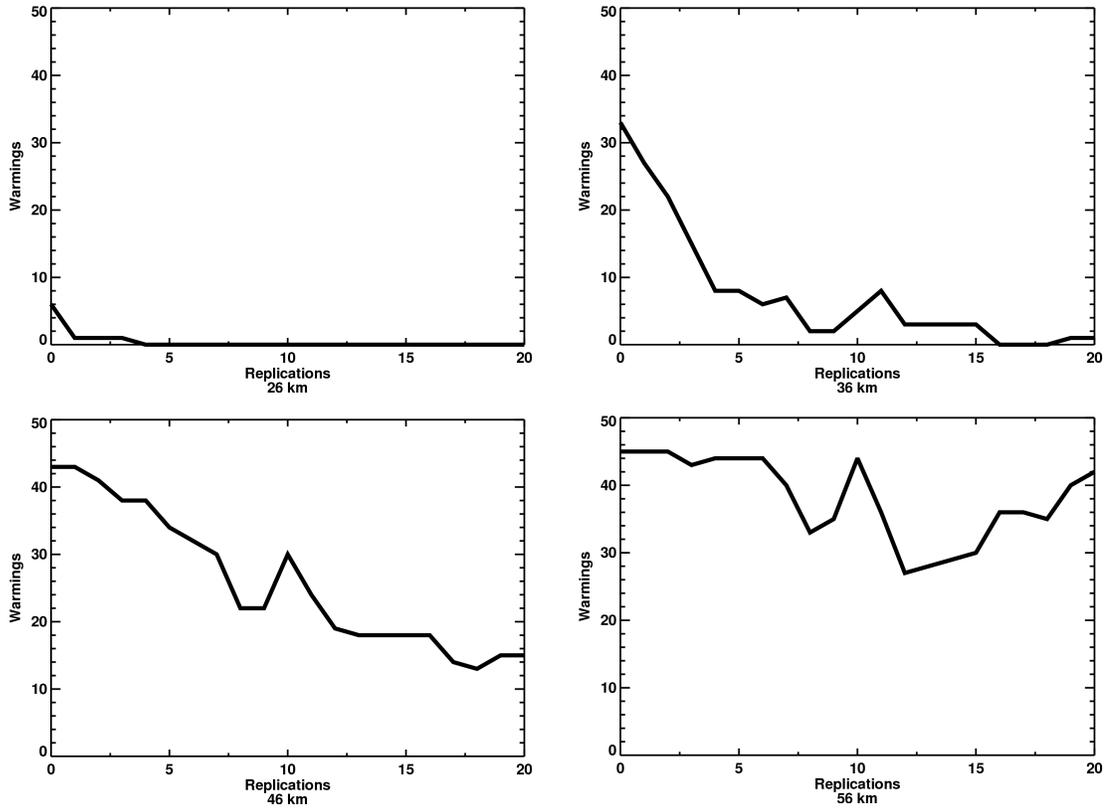


FIG. 40. As in Figure 38, but for invariant integrated wave activity from the initial forcing profiles rather than invariant wave amplitudes.

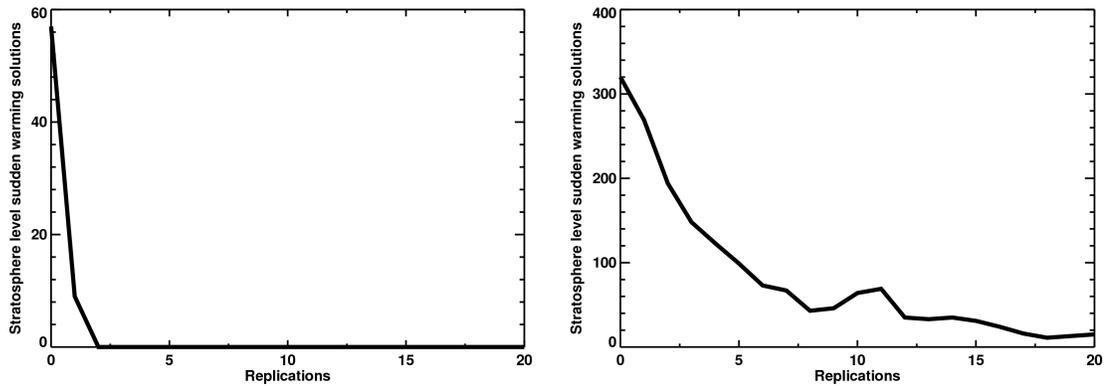


FIG. 41. The integrated sudden warming solutions as a function of replications for wavenumber 1 (left) and wavenumber 2 (right). These solutions are defined by summing the number of sudden warming solutions from the 26 km level to the 46 km level.

km, while at 20 replications that reduction is approximately 67%. Similarly when only wavenumber 2 is replicated but the Holton-Mass model is also forced with the initial

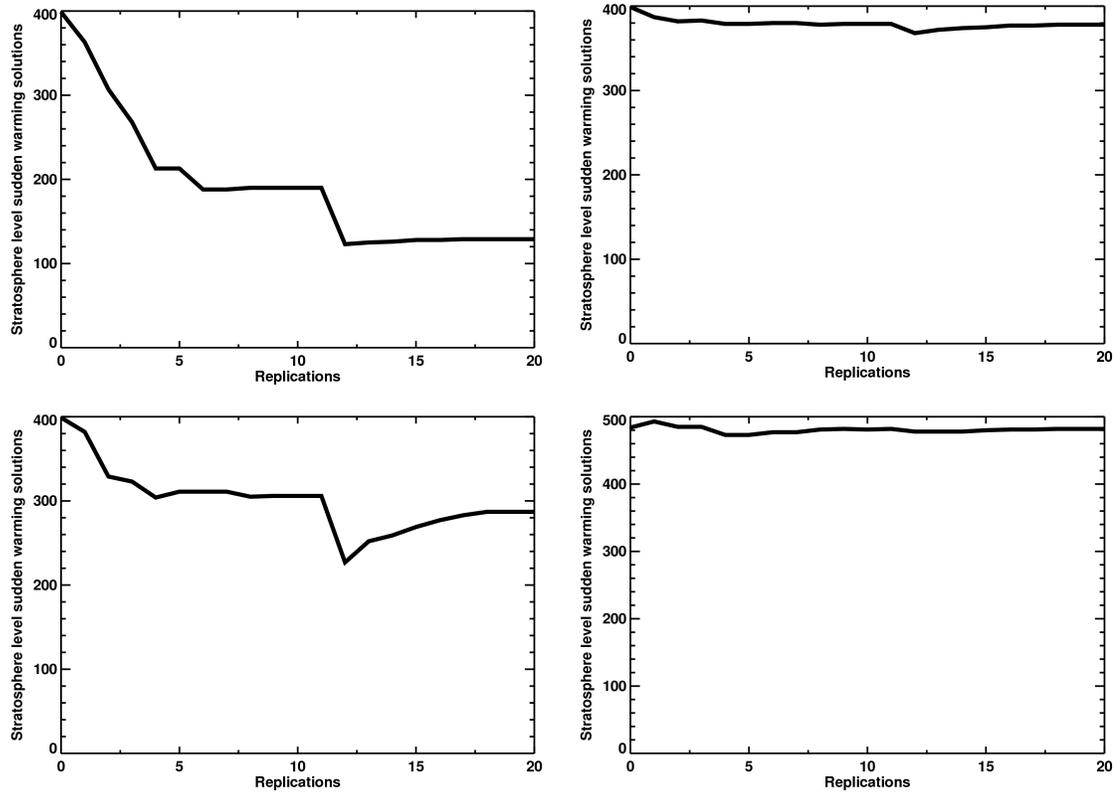


FIG. 42. The integrated sudden warming solutions as a function of replications for replications of both wavenumbers 1 and 2 (top left), of wavenumber 1 (top right), of wavenumber 2 (bottom left), and of wavenumber 1 with wavenumber 1 amplitudes doubled (bottom right).

wavenumber 1 forcing profile, the integrated sudden warming solutions decrease with increasing transience of the forcing. There is a 25% reduction in sudden warming solutions by five replications and this reduction percent remains nearly constant out to 20 replications.

In contrast, one may note that the replication runs with only wavenumber 1 replicated do not show any significant reduction in integrated sudden warming solutions. As in the low-order model, this is likely due to wavenumber 2 forcing being dominant within the model. The initial forcing profiles of wavenumber 2 were able to trigger approximately 320 warming solutions in the single wavenumber case as compared to 55 for wavenumber 1. This shows that the initial forcing profiles for wavenumber 2 can produce warmings more

likely than can wavenumber 1. Thus if wavenumber 2 forcing is the predominant forcing within the multiple wavenumber model, then increasing the transience of wavenumber 1 and thereby decreasing the forced sudden warming solutions by wavenumber 1 will cause a small reduction relative to the total sudden warming solutions forced by the multiple wavenumbers.

As in the low-order model, we exclude results of the replication experiments with multiple wavenumbers and constant integrated wave activity because there is minimal differences between them and the results from holding the maximum wave amplitude constant.

We have thus performed the transience experiments with the Holton-Mass model that were undertaken for the low-order model. From the idealized forcing experiments we noted that there is a significant increase in the necessary wave forcing amplitude to drive the system to a sudden warming state when the transience of that wave increases. The replication experiments confirmed results from the low-order model that forced sudden warmings are decreased as transience of the forcing is increased.

To complete this analysis of transient wave forcing of SSWs, we also seek to identify this transient behavior in observations in the next section. A simple analysis of wave amplitude transience was performed in this study to demonstrate that there exists observational evidence of our hypothesis that more transient forcing reduces SSW occurrence.

5.4 OBSERVATIONAL AMPLITUDE TRANSCIENCE

Our goal with the observational data is to study the climatology of wave amplitude transience in hopes that we could observe that there exists a strongly differing wave amplitude transience between periods where a sudden warming occurred and periods where a sudden warming was absent. Furthermore, we had in mind that the decade of the 1990s had markedly fewer warmings than the first decade of the 2000s. A clear variance in the transience of the wave amplitudes between these two periods would lend credibility to our hypothesis.

For the low-order system, the levels modeled are the lower stratosphere, the stratopause, and approximately 10 hPa, while the principle variables of the system are geopotential and zonal wind. From the boundary conditions, the atmospheric parameters about which we are most concerned are the incoming wave amplitudes and the zonal wind in the lower stratosphere, the geopotential and the zonal wind at 10 hPa, and the vertical shear throughout the depth of the stratosphere. 100 hPa was chosen as the bottom boundary because this level is situated within the lower stratosphere and is also well removed from the tropopause.

The datasets used in this analysis are the ECMWF Re-Analysis (ERA) datasets ERA-40 and ERA-Interim. These reanalysis datasets are derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) modeling of atmospheric conditions using observations from a large pool of sources. The principle of these analyses is to formulate a standardized set of atmospheric conditions with the same time steps, horizontal and vertical resolutions, and resolved atmospheric fields. The ERA-40 dataset covers the time period from September 1957 through August 2002 with a time resolution of six hours, horizontal resolution of 2.5° , and 23 vertical levels. For our purposes, there are 11 middle atmospheric levels which we define as 100 hPa and above.

The ERA-Interim dataset is the continuation of the ERA-40 reanalysis, though with a few notable changes. This reanalysis includes 4D-var data assimilation which improves fluidity of the reanalysis dataset by incorporating spatial data from other times into each computation. This dataset covers the time period from January 1989 to present, though this study limited the range to December 2009. The time resolution is six hours, the horizontal resolution is increased to 1.5° , and the number of vertical levels are increased to 37. The 11 middle atmospheric levels have not changed from the ERA-40 dataset, though we should expect increased accuracy at these levels due to better tuning of the reanalysis model and to better data assimilation.

To perform the transience analysis, we first spectrally analyzed the 60°N latitude zonal mean, 100 hPa geopotential for the period between January 1989 and December 2009.

These data represent the incoming wave amplitudes in the lower stratosphere and along the edge of the polar vortex. This spectral analysis produced zonal mean wave amplitudes at 60°N for each wavenumber as a function of time. We note that these data were time averaged into daily values to smooth both the diurnal cycle and any errors brought about by the smallest time resolution of the reanalysis models.

To get what we call the transience of the wave amplitudes, we chose to take a ± 10 day root mean square (RMS) of the time rate of change of the wave amplitude about a given day to isolate the magnitude of the time rate of change. This was done since the time rate of change of the wave amplitudes does not by itself give as accurate a representation of the transience due to its inclusion of both positive and negative values. The results that follow from this dataset of RMS transience values are not highly dependent on the choice of 10 days. We also exclude results from wavenumbers higher than 2 as the main forcing of SSWs is from wavenumbers 1 and 2.

What was observed was a significant difference between the mean of the wave amplitude transience for years with no sudden warming and for years with a sudden warming in the ERA-Interim dataset. The 95% confidence interval (in units m/day) for population mean RMS transience is presented in Table 1.

Table 1.

	Nonwarming years	Warming years
Wavenumber 1	$15.7 < \mu < 16.2$	$13.6 < \mu < 13.8$
Wavenumber 2	$14.8 < \mu < 15.3$	$13.4 < \mu < 13.9$

95 % confidence intervals (in units m/day) for population mean RMS transience for the ERA-Interim dataset.

While warming years (years with at least one SSW) are well-separated from nonwarming years in this analysis, these results include the diminished incoming wave activity that follows a sudden warming. The downward propagation of the decelerating winds produces a lower breaking level for incoming waves, until the waves break below the 100 hPa level. Thus it is more proper that we analyze the period leading up to the warming, and exclude the period including and following the warming event. It is also not proper that days far

in advance of a sudden warming be included. The primary reasoning here is that if there was no warming 20 days in advance of an actual warming event, then these days leading up to the warming may exhibit characteristics of nonwarming years. Thus we choose 10 days prior to a warming event as a period for consideration because it is shorter than the radiative damping time scale, but not so short as to exclude significant information prior to the event. We note here that other ranges of days prior to warming were considered, but the results are most clear for 10 days prior to an event.

The 95% confidence interval for the wavenumber 1 population mean of the RMS transience for the period 10 days prior to the warming is $14.7 < \mu < 16.7$ m/day, while that for wavenumber 2 is $13.2 < \mu < 14.8$ m/day. The population mean confidence intervals now overlap slightly, which does not lend credibility to observationally proving our hypothesis. However this statistic may not be the best tool for analyzing the differences of transience between nonwarming and warming years as it considers the two populations independently and does not account for differences in sample size.

Thus we instead utilize the difference of means statistic for further analysis (Wilks 2006). Since the hypothesis is that more transient wave amplitudes do not force the system to sudden warmings, we adopt the convention that positive difference of means interval values imply that the transience of nonwarming years is larger than for warming years (or for the periods leading up to warming events). As an example, the 95% confidence difference of means between nonwarming years and warming years for wavenumber 1 is $2.25 < \mu < 3.00$ m/day, while that for wavenumber 2 is $1.05 < \mu < 1.81$ m/day. These results imply that nonwarming years have greater transience than warming years.

Then returning to the comparison of transience for nonwarming years to the transience for period 10 days prior to a warming event, the difference of means for wavenumber 1 is $-0.607 < \mu < 1.06$ m/day while for wavenumber 2 is 0.253 to 1.91 m/day. This more concisely suggests that there exists higher transience in environments where no SSW

occurs compared to environments that directly precede a SSW, though that difference may be miniscule.

One further consideration, that these statistics have not considered, is that we may consider two types of warmings — split and displacement warmings — each with differing dynamical interactions. Thus we separate the analysis of transience prior to a SSW between split events and displacement events. Charlton and Polvani (2007) present a table of sudden warming events through 2002, and the warming type associated with each event. Likewise, as shown for both a displacement and a split event in Figure 43, we plotted 10 hPa geopotential for all events and visually identified the events to complete the analysis through 2009. This visual analysis was compared to and found to be identical to that of Charlton and Polvani (2007) for the overlapping dates.

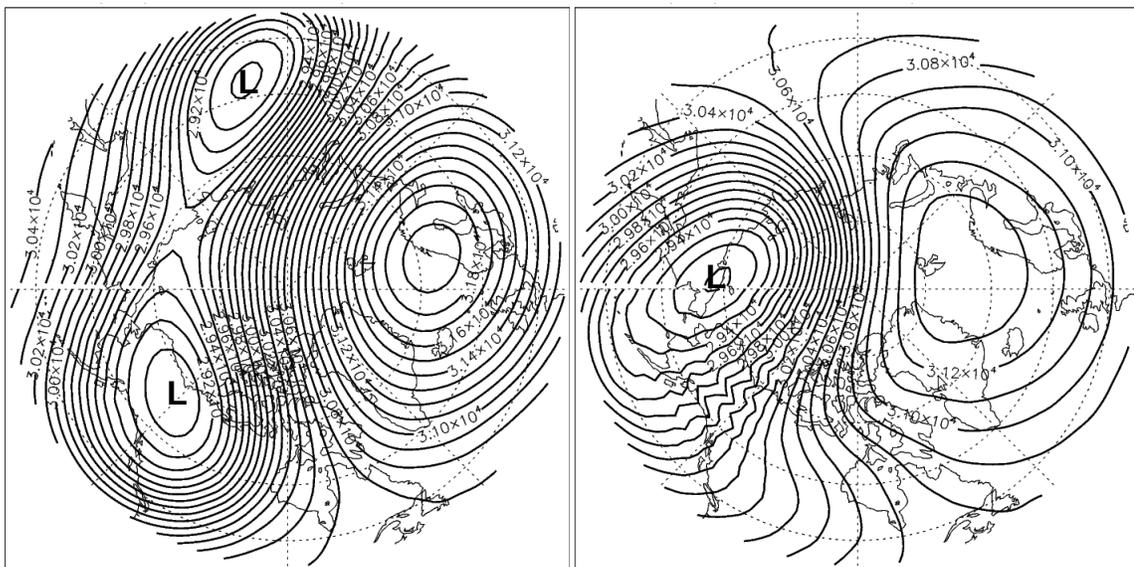


FIG. 43. Geopotential contour plots of two sudden stratospheric warming events. On the left is the February 22, 2008, SSW and represents a vortex displacement type event. The right plot is the January 18, 2003, SSW and represents a vortex split type event. The centers of low geopotential are marked with “L”.

Over the data range, we identified four split events and seven displacement events. The 95% confidence difference of means between the transience of nonwarming years and of the period 10 days prior to a warming event (represented as μ) is summarized in Table 2.

Splitting the analysis shows that for our period, nonwarming years have higher transience

Table 2.

	Displacement events	Split events
Wavenumber 1	$1.03 < \mu < 3.05$	$-4.92 < \mu < -0.967$
Wavenumber 2	$1.15 < \mu < 3.13$	$-3.07 < \mu < 1.52$

95% confidence difference of means between the transience of nonwarming years and of the period 10 days prior to a warming event (represented as μ) for the ERA-interim dataset.

than the 10 days prior to displacement events but have lower transience than 10 days prior to split events.

It is important to note that the sample size for both the split events and for the displacement events with the ERA-Interim dataset are small and may not fully represent the true population. Thus, we also analyzed the ERA-40 dataset in similar fashion as the ERA-Interim dataset. The principle advantage of doing so is that the total number of sudden warming events for the analysis is increased to 21, with 10 split events and 11 displacement events. The difference of means for the above statistics are presented in Table 3.

Table 3.

	All warmings	Split events	Displacement events
Wavenumber 1	$-0.182 < \mu < 0.916$	$0.267 < \mu < 1.98$	$-1.05 < \mu < 0.408$
Wavenumber 2	$0.107 < \mu < 1.31$	$-2.08 < \mu < -0.278$	$1.60 < \mu < 3.24$

95 % confidence difference of means between the transience of nonwarming years and of the period 10 days prior to a warming event (represented as μ) for the ERA-40 dataset.

The ERA-40 analysis for all warming events is nearly identical to the ERA-Interim dataset analysis, primarily that the period 10 days prior to a sudden warming is less transient than during nonwarming years, with the exception being possibly wavenumber 1 transience. The analysis considering each warming type independently differs from the ERA-Interim analysis, however. What the difference of means statistics imply is that for split displacement events — a primarily wavenumber 1 event — the transience of wavenumber 1 prior to the warming is likely greater than of wavenumber 1 during nonwarming years.

Likewise, it is implied that wavenumber 2 transience prior to a split event is greater than wavenumber 2 transience during a nonwarming year.

It is not surprising that such an increase in transience is observed immediately prior to a warming event. Sudden warming events are typically preceded by a sudden increase in wave amplitude of the wavenumber matching the warming type (Charlton and Polvani, 2007). Though this result would seem to contradict the hypothesis that wave amplitude transience decreases the occurrences of SSWs, this result may include the exception to the stated hypothesis: the amplitude of a wave can increase to a critical magnitude where the transience of the wave amplitude does not affect the initiation or suppression of a sudden warming.

The statistics discussed here produce convoluted evidence supporting the hypothesis. If one considers sudden warmings events irrespective of type, then the transience leading up to the event is less than for periods where there is no warming. Though upon breaking the statistical sets up into event type, the data do not match this observation. The former result supports the transience hypothesis while the latter result may or may not.

The statistical analysis of the reanalysis datasets is not the main body of this work, however, and is not intended to prove or disprove the existence of the stratospheric sensitivity to transient wave amplitudes. The statistical analysis portion of this work encompassed only the basic statistics included above. Further analysis is required to fully document whether the observations contain proof of our model transient forcing effect on SSWs.

6 SUMMARY

6.1 CONCLUSIONS

Two low-order models of stratospheric wave-mean flow interaction are developed. Following Holton and Mass (1976) and Ruzmaikin (2003), we derive both the Holton-Mass model and a highly truncated version of that model which we have labeled the low-order model. It is shown that in this simple framework, inclusion of multiple wavenumbers is straightforward so long as the assumption of no wave-wave interaction is made. From Charney and Drazin (1961), we know that the two principal wavenumbers in the stratosphere are wavenumbers 1 and 2. Resolving multiple wavenumbers allows us to model the effects on the zonal wind of these two lowest wavenumbers.

Equilibrium states and stability regions are identified in these models. Bifurcations exist within the equilibrium states of the low-order model — excluding wavenumber 1-only forcing — such that for quasi-stationary forcing, we may model a primitive SSW. Linear stability analysis techniques identifies that this bifurcation occurs between two stable states: a radiative solution state and a sudden warming state. It has been shown (Chao 1985, Christiansen 2000, Yoden 1987b) that the Holton-Mass model contains remarkably similar bifurcations and equilibrium states. As well, it has been shown that the existence of two distinct stable states in the stratosphere exist in GCMs (Boville 1986) and in the observations (Monahan et al. 2003).

We choose a simple parameterization of small-scale variability to include in our low-order model where the strength of the stochastic variability is chosen to be small relative to the wave forcing. Doing this allows us to include zonal wind forcing from high wavenumbers and to check for statistical significance. We use Fokker-Planck analyses to study the

behavior of the additive noise with respect to wavenumber, wave forcing amplitude, and noise strength. This noise allows for the perturbation of the zonal wind to the sudden warming solution when the wave forcing amplitude is less than the stationary wave amplitude needed to force the deterministic model to the sudden warming solution. Integrations of the Fokker-Planck equation also demonstrate that by including small-scale variability with wavenumber 1 forcing greatly reduces the amplitude of the forcing needed to force the model into the arbitrary sudden warming solution that was defined for wavenumber 1.

We analyze the effects of transiently forced planetary waves on the models. A highly idealized forcing was included in integrations of the models to identify the necessary amplitude to force the model to the sudden warming solution as a function of the wave amplitude transience. It is shown that for wave periods less than the radiative time scale of the model, large amplitudes of the transient wave forcing are required to force the model to the sudden warming solution. We use an ensemble of wave forcing profiles derived from observations to force the models. We increase the transience of this ensemble while holding the maximum amplitude constant for one experiment, and while holding the integrated wave activity constant for another experiment. In both sets of experiments with this ensemble of increasing wave amplitude transience, the modeled number of final solutions residing in the sudden warming solution decreased with increasing transience. Observational evidence from ECMWF Re-Analysis datasets is shown to contain some degree of evidence in support of these modeled results.

6.2 OUTLOOK

There are a number of experiments that remain to be completed with regards to the low-order model and the Holton-Mass model. We briefly analyzed the system with inclusion of the term \dot{U}_B . This term allows for both the inclusion of bottom boundary noise and the inclusion of a time dependent bottom boundary. If this term is resolved as a stochastic term, then \dot{U}_B represents a state-dependent noise term. Whereas the state-independent

noise we included typically aided perturbing the system to the sudden warming solution, we observed that inclusion of state-dependent noise in the low-order model resulted in increased stability about the radiative solution.

Likewise, we also would like to develop and test a meaningful way of including small-scale variability within the Holton-Mass model. Some experiments were run with a simple additive noise at various levels and depths of the model, though it was found that even for very small noise strength relative to the low-order model noise strength, the additive noise in the Holton-Mass model would strongly perturb the zonal wind towards the sudden warming solution. Inclusion of the small-scale variability may require a more complicated parameterization than the simple additive noise included in the low-order model runs.

For the transient forcing of the models, we would like to further explore the enhancement of sudden warming solutions around five to six replications of the forcing ensemble. Removal of this feature may be as simple as deseasonalizing the forcing profiles so that they do not contain a low-frequency variation in the amplitudes. Another possibility for removal of this is to randomly reshuffle the values of the forcing amplitudes within each forcing profile. This should remove any subseasonal cycle of amplitudes within the profiles.

The observational analysis of wave transience must be furthered. While we demonstrate a proof of concept in the reanalysis datasets, the provided statistics leave much to be desired. There are many ways to go about producing a more demonstrative analysis of the observations and thus choosing the proper method will require careful thought.

A more complex idea for future work revolves around developing a stratosphere-troposphere coupled model to truly include tropospherically forced transient planetary waves. The principle behind this is to resolve the wave-mean flow interactions in the stratosphere with our low-order model, but include a simple troposphere system that models time-varying wind blowing over orography. Doing so would remove the steady bottom boundary condition

in our low-order model as well as provide the ability to include the observed effects of the stratospheric circulation on the troposphere.

References

- Alexander, S. P. and M. G. Shepherd, 2010: Planetary wave activity in the polar lower stratosphere. *Atmos. Chem. Phys.*, **10**, 707–718.
- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.
- Andrews, D. G. and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear: The generalized Eliassen-Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, **33**, 2031–2048.
- Baldwin, M. P. and T. J. Dunkerton, 2001: Stratospheric harbringers of anomalous weather regimes. *Science*, **294**, 581–584.
- Birner, T. and P. D. Williams, 2008: Sudden stratospheric warmings as noise-induced transitions. *J. Atmos. Sci.*, **65**, 3337–3343.
- Boville, B. A., 1986: Wave-mean flow interactions in a general circulation model of the troposphere and stratosphere. *J. Atmos. Sci.*, **43**, 1711–1725.
- Chao, W. C., 1985: Sudden stratospheric warmings as catastrophes. *J. Atmos. Sci.*, **42**, 1631–1646.
- Charlton, A. J. and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Clim.*, **20**, 449–469.
- Charney, J. G. and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205–1216.
- Charney, J. G. and P. G. Drazin, 1961: Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Rev.*, **66**, 83–109.
- Christiansen, B., 2000: Chaos, quasiperiodicity, and interannual variability: Studies of a stratospheric vacillation model. *J. Atmos. Sci.*, **57**, 3161–3173.
- Dickinson, R. E., 1969: Theory of planetary wave-zonal flow interaction. *J. Atmos. Sci.*, **26**, 73–81.
- Durrant, D. R., 1999: *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Springer, 465 pp.
- Egger, J., 1981: Stochastically driven large-scale circulations with multiple equilibria. *J. Atmos. Sci.*, **38**, 2606–2618.

- Fritts, D. C. and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.*, **41**(1), 1003. doi:10.1029/2001RG000106.
- Gardiner, C. W., 1985: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Science*. Springer, 442 pp.
- Geisler, J. E., 1974: A numerical model of the sudden stratospheric warming mechanism. *J. Geophys. Res.*, **79**, 4989–4999.
- Hamilton, K., Ed., 1997: *Gravity Wave Processes: Their Parameterization in Global Climate Models*. Springer, 414 pp.
- Harnik, N., 2009: Observed stratospheric downward reflection and its relation to upward pulses of wave activity. *J. Geophys. Res.*, **114**, D08120. doi:10.1029/2008JD010493.
- Held, I. M., 2005: The gap between simulation and understanding in climate modeling. *Bull. Amer. Meteor. Soc.*, **86**, 1609–1614.
- Holton, J. R., 2004: *An Introduction to Dynamic Meteorology*. Academic Press, 535 pp.
- Holton, J. R. and T. Dunkerton, 1978: On the role of wave transience and dissipation in stratospheric mean flow vacillations. *J. Atmos. Sci.*, **35**, 740–744.
- Holton, J. R. and C. Mass, 1976: Stratospheric vacillation cycles. *J. Atmos. Sci.*, **33**, 2218–2225.
- Horsthemke, W. and R. Lefever, 1983: *Noise-Induced Transitions*. Springer, 318 pp.
- Limpasuvan, V., D. W. J. Thompson, and D. L. Hartmann, 2004: The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Clim.*, **17**, 2584–2596.
- Matsuno, T., 1971: A dynamical model of the sudden stratospheric warming. *J. Atmos. Sci.*, **28**, 1479–1494.
- Monahan, A. H., J. C. Fyfe, and L. Pandolfo, 2003: The vertical structure of winter-time climate regimes of the Northern Hemisphere extratropical atmosphere. *J. Clim.*, **16**, 2005–2021.
- Pulido, M. and J. Thuburn, 2006: Gravity wave drag estimation from global analyses using variational data assimilation principles. Part II: A case study. *Quart. J. Roy. Meteor. Soc.*, **132**, 1527–1543. doi:10.1256/qj.05.43.
- 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.
- Randel, W. J., D. E. Stevens, and J. L. Stanford, 1987: A study of planetary waves in the southern winter troposphere and stratosphere. Part II: Life cycles. *J. Atmos. Sci.*, **44**, 936–949.

- Ruzmaikin, A., J. Lawrence, and C. Cadavid, 2003: A simple model of stratospheric dynamics including solar variability. *J. Clim.*, **16**, 1593–1600.
- Scherhag, R., 1952: Die explosionsartigen stratosphärenerwärmungen des spätwinters 1951-52. *Ber. Dtsch. Wetterdienst (US Zone)*, **6**, 51–63.
- Schoberl, M. R., 1978: Stratospheric warmings: observations and theory. *Rev. Geophys. Space Phys.*, **16**, 521–538.
- SPARC CCMVal (2010), SPARC Report on the Evaluation of Chemistry-Climate Models, V. Eyring, T. G. Shepherd, D. W. Waugh (Eds.), SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, <http://www.atmosph.physics.utoronto.ca/SPARC>.
- Sura, P., 2002: Noise-induced transitions in a barotropic β -plane channel. *J. Atmos. Sci.*, **59**, 97–110.
- Thompson, D. W. J., M. P. Baldwin, and J. M. Wallace, 2002: Stratospheric connection to Northern Hemisphere wintertime weather: Implications for prediction. *J. Clim.*, **15**, 1421–1428.
- Tung, K. K. and A. J. Rosenthal, 1985: Theories of multiple equilibria - a critical reexamination. Part I: Barotropic models. *J. Atmos. Sci.*, **42**, 2804–2819.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 627 pp.
- Yoden, S., 1987a: A new class of stratospheric vacillations in a highly truncated model due to wave interference. *J. Atmos. Sci.*, **44**, 3696–3709.
- Yoden, S., 1987b: Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.*, **44**, 1723–1733.