Stratospheric Wave–Mean Flow Feedbacks and Sudden Stratospheric Warmings in a Simple Model Forced by Upward Wave Activity Flux

JEREMIAH P. SJÖBERG AND THOMAS BIRNER
Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 28 April 2014, in final form 22 July 2014)

ABSTRACT

A classic result of studying stratospheric wave–mean flow interactions presented by Holton and Mass is that, for constant incoming wave forcing (at a notional tropopause), a vacillating stratospheric response may ensue. Simple models, such as the Holton–Mass model, typically prescribe the incoming wave forcing in terms of geopotential perturbation, which is not a proxy for upward wave activity flux. Here, the authors reformulate the Holton–Mass model such that incoming upward wave activity flux is prescribed. The Holton–Mass model contains a positive wave–mean flow feedback whereby wave forcing decelerates the mean flow, allowing enhanced wave propagation, which then further decelerates the mean flow, etc., until the mean flow no longer supports wave propagation. By specifying incoming wave activity flux, this feedback is constrained to the model interior. Bistability—where the zonal wind may exist at one of two distinct steady states for a given incoming wave forcing—is maintained in this reformulated model. The model is perturbed with transient pulses of upward wave activity flux to produce transitions between the two stable states. A minimum of integrated incoming wave activity flux necessary to force these sudden stratospheric warming–like transitions exists for pulses with time scales on the order of 10 days, arising from a wave time scale internal to the model at which forcing produces the strongest mean-flow response. The authors examine how the tropopause affects the internal feedback for this model setup and find that the tropopause inversion layer may potentially provide an important source of wave activity in the lower stratosphere.

1. Introduction

Since the discovery of sudden stratospheric warmings (SSWs) by Scherhag (1952), the middle-atmosphere community has sought to understand the dynamics of the coupled troposphere–stratosphere system leading to such events. Early work in analyzing these events hinted at a strong linkage between SSWs and tropospheric, large-scale planetary waves (e.g., Teweles and Finger 1958). This idea was supported by simple modeling performed by Matsuno (1971). His two-dimensional, quasigeostrophic, and linearized model resolved zonal wind responses to wave forcing generated by bottom boundary geopotential height perturbations at the tropopause. Through this, Matsuno generated a prototype sudden warming and laid the groundwork for our understanding of these events. Matsuno additionally demonstrated that wave-driven zonal wind reversal leads to the formation of a critical layer, beyond which waves cannot propagate. Building on these ideas, Geisler (1974) showed similar prototype SSWs and the existence of a wave–mean flow positive feedback in a geometrically simplified model.

Additional modeling studies ensued (see Schoeberl 1978), where perhaps the most widely known of these early studies was that of Holton and Mass (1976). Their quasigeostrophic, beta-plane channel model of wave–mean flow interaction (hereafter, the Holton–Mass model) presented a numerically simple yet physically sound means of investigating the generation of SSWs. This one-dimensional model prescribes wave forcing through bottom boundary geopotential perturbations that propagate upward and drive the mean flow. The notable result from Holton and Mass (1976) is that, for a constant bottom boundary geopotential perturbation of sufficient amplitude, a vacillating state of the zonal mean zonal wind is forced in their model.

This type of internal variability has been studied both in the context of the Holton–Mass model (Chao 1985;
Yoden 1987; Christiansen 2000) and in general circulation models (Scott and Haynes 2000; Scott and Polvani 2006). In these studies and others (e.g., Plumb and Semeniuk 2003; Hardiman and Haynes 2008), geopotential perturbations are included as the stratospheric bottom boundary forcing. Forcing these models in such a way is sensible from a numerical standpoint as the model fields are either cast in terms of geopotential or readily incorporate specified geopotential on the boundary.

Yet, geopotential perturbations are not good proxies for wave activity fluxes. Edmon et al. (1980) showed that these fluxes are more appropriately captured by Eliassen–Palm (EP) flux vectors. From the transformed Eulerian mean formulation of Andrews and McIntyre (1976), it follows that the convergence of the EP flux vector is the primary wave forcing on the zonal mean zonal wind. As the waves of interest for forcing of SSWs have tropospheric origins, observational studies on this subject principally analyze either upward EP fluxes or their dominant component, heat fluxes (e.g., Coy et al. 1997; Waugh et al. 1999; Newman et al. 2001; Limpasuvan et al. 2004). Upward EP flux then is the better proxy for studying wave forcing in the Holton–Mass model.

In considering the upward EP flux (or heat flux) of the Holton–Mass model, it is found that, under constant bottom boundary geopotential forcing of amplitude large enough to force an oscillating zonal wind state, the upward EP flux (or heat flux) in the model also oscillates. This is due to upward wave propagation being dependent on the state of the zonal wind (Charney and Drazin 1961; Dickinson 1968). Typical stratospheric zonal wind speeds are too fast at mid- and upper levels to allow deep vertical wave propagation. As vertically propagating waves encounter these fast winds, the waves dissipate and the associated upward wave activity flux convergence acts to decelerate the zonal wind, allowing deeper wave propagation. This represents a positive feedback that results in wave activity flux amplification by the waves themselves (Geisler 1974; Plumb 1981, 2010). If the flux convergence is sufficiently large, the positive feedback terminates and a nonlinear negative feedback is induced instead (Matsumo 1971). Here, a critical line to propagation is formed and the waves are shut off through reflection and absorption (though wave amplitudes may be nonzero above because of wave evanescence). As the zonal winds restore through radiative damping, the waves are again able to propagate and force the zonal wind.

The same wave–mean flow feedbacks as above apply at the bottom boundary. Thus, the bottom boundary upward EP flux in the Holton–Mass model also oscillates for large yet constant bottom boundary geopotential perturbation. In this way, when the model is forced with geopotential perturbations, the incoming wave activity flux is far from specified as may be desired on more physical grounds.

Here we introduce a reformulation of the Holton–Mass model so that the bottom boundary wave forcing becomes a specified upward wave activity flux (EP flux). This study explores how this EP flux model impacts prior results utilizing the Holton–Mass model under both constant and transient geopotential forcing. Primary focus is given to the differences in the wave–mean flow positive feedback, which becomes limited when prescribing the lower boundary upward EP flux. Moreover, because the waves relevant here are of tropospheric origin and the EP flux description takes into account changes in stratification, the EP flux model is well-suited for documenting cross-tropopause wave propagation effects in this simple modeling context.

Section 2 covers stability states and their transitions under time-independent, steady wave forcing. Results from an idealized, time-dependent forcing experiment are shown in section 3. We present how two simple tropopause structures affect the wave–mean flow feedback in section 4. We close with our conclusions in section 5.

2. Steady bottom boundary incoming wave activity flux

a. Model reformulation and parameters

Our incoming wave activity specification is achieved by simply prescribing the (time-dependent) geopotential perturbation at the bottom boundary such that the desired upward EP flux is produced. Upward EP flux in this model takes the form

\[ F'(z) = \frac{\rho^2 Rf_0}{N^2 H} \frac{\sigma}{T}, \]

where \( \rho = e^{-(z-z_0)/H} \) is the density, \( z_0 \) is the bottom boundary height, \( H \) is the scale height, \( R \) is the gas constant, \( f_0 \) is the Coriolis parameter, and \( N^2 \) is the buoyancy frequency squared. The final term is the eddy meridional heat flux; the components of which are eddy meridional wind and eddy temperature. These two fields may be cast as

\[ \nu' = \partial_z \psi' \]

and

\[ T' = \frac{Hf_0}{R} \frac{\partial_z \psi'}, \]

where \( \psi' \) is the perturbation streamfunction. Together with the wave assumption as in Holton and Mass (1976), the heat flux may be written as
The vertical spacing is 500 m. Time stepping is 900 s. Integrations with deviations from these parameters are noted. All other parameters—such as Newtonian damping—are as in Holton and Mass (1976). The following results are found to be insensitive either to raising the model lid or to inclusion of a Rayleigh friction layer near the model lid.

### b. Results

To begin, we show the model response to both steady bottom boundary geopotential perturbation and upward EP flux. Displayed in Fig. 1a is the zonal wind (shaded) and EP flux (contoured) for constant bottom boundary geopotential forcing at a point in time far from initialization. We choose the amplitude (300 m) such that the model zonal wind is within the perpetual oscillating (SSW-like) state.

An initial wave propagates into the model (not shown) and dissipates, driving the winds to decelerate. The wave–mean flow positive feedback is then triggered, allowing greater wave propagation. Wind reversal is subsequently forced, producing a critical layer to propagation. This terminates the positive feedback and induces the negative feedback.\(^1\) The critical layer—above which the upward EP flux amplitudes decay to zero—extends to roughly 18 km, resulting in wave reflection and net downward wave activity flux to the model bottom boundary. As the zonal wind restores, the model background state again becomes susceptible to wave propagation, and the process continues. The classic theory on the forcing of SSWs holds in these results.

There are both external and internal features of the wave–mean flow positive feedback manifested in Fig. 1a. The external feedback process is associated with the large burst of upward EP fluxes at the bottom boundary. One may think of this as a fluxing upward, by a favorable mean state, of wave activity from some source existing below the model bottom boundary. The internal part of this feedback behaves in the same way but relies only on that wave activity that exists or may be generated within the model domain. An example of the latter may be seen in Fig. 1a between days 1310 and 1320, where there is EP flux divergence (wave generation) above 18 km.

The external feedback process here may be thought of as the response of tropopause-level waves to perturbations in the state of the stratosphere by prior waves. While the real-atmosphere manifestation of this feedback process will certainly be limited in available tropospheric

\(^1\) Hereafter, references to wave–mean flow feedback refer only to the positive wave–mean flow feedback that acts to amplify any initial wave perturbation.
wave activity, the source of wave activity below the model domain for the standard Holton–Mass model is potentially limitless. The external feedback process represented in the Holton–Mass model forced with bottom boundary geopotential perturbations is therefore likely unrealistic. Hence, it appears desirable to isolate the internal feedback in models with idealized bottom boundary conditions, such as the Holton–Mass model. As the external feedback mechanism does not operate where the incoming wave activity flux is specified, our EP flux bottom boundary condition does just that. Though we are motivated to recast the Holton–Mass model by the desire to specify upward bottom boundary wave activity flux, isolation of the internal feedback mechanism comes along as a favorable conceptual benefit.

Shown in Fig. 1b is the zonal wind (shaded) and upward EP flux (log contoured) for constant bottom boundary EP flux forcing. The amplitude of this specified upward EP flux (153 mPa) is chosen such that the mean bottom boundary geopotential is 300 m (i.e., the same as in Fig. 1a). Rather than the upward EP flux and zonal wind oscillating in the model interior, both fields are now steady with a maintained layer of easterlies between 24 and 32 km. This should not come as a surprise: steady wave forcing, if based on a conservative wave property (such as the EP flux), should lead to steady zonal wind response. Analysis of the representative index of refraction (see Harnik 2009) shows that the layer between 30 and 35 km represents a critical layer to wave propagation. Waves are no longer able to propagate beyond this critical layer, leaving only the imposed radiative damping and small dissipation from evanescent wave amplitudes to force the zonal wind above.

As discussed above, zonal wind oscillation in the original Holton–Mass model is possible because an upward EP flux oscillation ensues in the lower and midlevels of

Fig. 1. (a),(c) Zonal wind (shaded) and contours of upward EP flux (solid black) for the geopotential model forced with a constant bottom boundary perturbation of 300 and 1700 m, respectively. (b),(d) As in (a) and (c), but the EP flux model forced with constant bottom boundary upward EP flux of amplitude such that the mean bottom boundary geopotential is equal to that from the respective geopotential model integrations. The zonal wind contour spacing is 5 m s$^{-1}$ in all plots with the 0 m s$^{-1}$ contour labeled and dashed. The upward EP flux contours here and throughout this manuscript are on a logarithmic scale (0.1, 0.2, 0.5, 1, 2, 5, 10, etc.). Only positive EP flux contours are shown.
the model. With a prescribed constant bottom boundary upward wave activity flux, this oscillation is necessarily dampened at lower model levels. The lower-level upward wave activity flux is sufficient to overcome the effect of wave reflection that results from the critical layer starting at 30 km, preventing the system from reaching an oscillatory state.

An oscillating state is still possible in the EP flux model, but only for extremely large upward EP flux amplitudes. Shown in Figs. 1c and 1d are the same fields as in Figs. 1a and 1b, but for a constant geopotential perturbation and constant upward EP flux of 1700 m and 285 mPa, respectively. Here, both models oscillate, but about a zonal wind less than 0 m s\(^{-1}\). With the zonal winds through a majority of the model depth consistently easterly, this is far from the observed variability. Yet, Fig. 1d shows that the internal feedback alone is in principle able to drive the type of perpetual oscillations observed in the geopotential model.

Figure 2 shows how the bottom boundary geopotential and next-higher model-level geopotential vary as a function of time in the two constant EP flux model integrations (Figs. 1b,d). For the 153 mPa EP flux integration (gray), the geopotential at these two levels is steady by this point in the integration. This is in line with the steady EP fluxes at all levels between the shown times. For the 285 mPa EP flux integration (black), the geopotential at both levels oscillates with the same periodicity as the upward EP flux. Note that the maximum (minimum) of this oscillation relates to the maximum (minimum) in upward EP flux (see Fig. 1d). Both levels oscillate in near unison, acting to keep the upward EP flux along the bottom boundary constant.

We further explore the effects of specifying the bottom boundary EP flux by analyzing the stable states of this new setup. A particular characteristic of the original Holton–Mass model is that a bistable steady state exists within the zonal wind field with respect to constant bottom boundary geopotential perturbation. As explored in previous work with this model (Chao 1985; Yoden 1987) and a simpler prototype (Ruzmaikin et al. 2003; Birner and Williams 2008), when the steady geopotential perturbation lies between the two thresholds of the bistable regime, the zonal wind stably exists either in a lightly perturbed (near to radiative equilibrium) or in a strongly perturbed (far from radiative equilibrium) state. The bounds of this bistable regime represent pitchfork bifurcation points where one of the stable solutions ceases to exist.

Such a behavior may be found by very slowly ramping up and subsequently ramping down the bifurcation parameter (i.e., bottom boundary geopotential or EP flux) so that the zonal wind remains in a quasi-steady state (Birner and Williams 2008). Once the bifurcation point is passed, a sudden transition between stability states will occur. Shown in Fig. 3a is the resultant zonal wind from one such integration with the EP flux model, where the durations of both the ramp up and the ramp down are 50 000 days. For this experiment, the maximum incoming upward EP flux is 150 mPa. This is below the values of steady upward EP flux that result in oscillating zonal wind.

At the start of the integration, the zonal wind is initialized at the radiative equilibrium state. As the bottom boundary EP flux increases, the zonal wind is increasingly, yet smoothly, pushed away from this radiative state. Near day 23 800, the incoming EP flux becomes sufficiently large so as to force a transition from this lightly perturbed zonal wind state to a different, strongly perturbed state. Shown in Fig. 3c is a zoomed-in view of this transition, which appears to be a prototype SSW: wind reversal starting in the upper stratosphere propagates downward to the middle stratosphere. Beyond this transition, the zonal wind evolution again becomes smooth, implying that the model again resides within a stable state. In this (quasi)-steady state, the zonal wind is near to (or less than) zero in the layer between 25 and 35 km—corresponding to a critical layer for wave propagation—and increases above this layer because of the imposed radiative damping and very weak wave forcing.

\[\text{FIG. 2. The geopotential perturbations at the bottom boundary}\]

[![Figure 2](image-url)](image-url)
As the EP flux is linearly decreased beyond 50,000 days, the critical layer shifts upward along with the height of the minimum zonal wind. Near day 88,200, the zonal wind rapidly transitions from the strongly perturbed to the weakly perturbed state. As shown in Fig. 3d, this transition back to the lightly perturbed steady state is analogous and opposite to the transition in Fig. 3c. Beyond this point, the zonal wind gradually relaxes toward the radiative equilibrium zonal wind as the wave forcing wanes.

That the two transitions occur at different values of bottom boundary EP fluxes shows that the model has a bistable regime. To better visualize this, Fig. 3b displays the zonal wind at 30 km from this integration. Given that the bistable regime exists within the EP flux model, the external wave–mean flow feedback is not a necessary component of this bistability. Remarkably, two stable zonal wind solutions exist for the same amount of incoming wave activity flux.

We test the stability of the system by initializing the zonal wind at points between the two zonal wind states and determining to which state the model equilibrates. Through this, we identify where multiple equilibrium solutions exist in the model and whether these are stable (solid black) or unstable (solid gray). Between incoming upward EP flux ranges of 35–36 and 69–71 mPa, only two equilibrium solutions exist. In the former two-solution regime, we find that the upper branch is a stable solution while the flipped branch is an unstable solution, and vice versa for the latter two-solution regime. Between 36 and 69 mPa, three equilibrium solutions exist: two stable solutions and one unstable solution. In this three-solution regime, model initializations with zonal wind greater than (less than) the unstable solution will equilibrate to the upper (lower) steady-state solution. Given that the parameter space where solutions stabilize to the strongly perturbed state is larger than the...
parameter space where solutions stabilize to the weakly perturbed state, it is apparent that the strongly perturbed state is the more stable solution for this model setup.

With this bistability in mind, we attempt to force a transition between the stable states of the system, similar to Birner and Williams (2008) and Hardiman and Haynes (2008). Here, we integrate both models with equivalent constant bottom boundary wave forcing, and we impose a zonal wind deceleration in the model stratosphere independent of the waves. This deceleration was chosen such that the wave–mean flow feedback would allow wave propagation sufficient to drive a transition between stable states and may be thought of as resulting from unresolved waves—for example, gravity waves—or from prior planetary wave breaking.

Shown in Fig. 4 is the zonal wind evolution following the imposed deceleration and the upward EP flux, from both the EP flux model (Fig. 4a) and the geopotential model (Fig. 4b). Incoming upward EP flux is held constant at 50 mPa, while incoming geopotential perturbation is held constant at 273 m. For reference, the bistable regime in the geopotential model exists between bottom boundary geopotential perturbations of ~92 and ~287 m. Note that the initial zonal winds and upward EP fluxes are identical in both integrations.
Up to day 1000 in both integrations, the models are steady and within the lightly perturbed (near-radiative equilibrium) regime. After the onset of the imposed external deceleration at day 1000, the upward EP flux increases in the lower model levels. This anomalous upward EP flux continues to grow at all levels and force the zonal wind to transition into the strongly perturbed state. Following the transition in both models, the zonal wind stabilizes to the strongly perturbed regime.

Examining the SSW-like transition in the EP flux model shows how the internal feedback causes enhanced wave amplitudes in the middle and upper model levels. Following the imposed deceleration, EP flux convergence decreases on average by ~50% in the lowermost 10 km of the model, allowing greater vertical penetration of wave amplitudes from the bottom boundary. Furthermore, following the imposed deceleration and prior to the zonal wind reversal, the index of refraction is positive up to 27 km—6 km deeper than prior to the imposed deceleration—indicating considerably deeper wave propagation.

In the geopotential model, the index of refraction evolves in a nearly identical manner as in the EP flux model. The main difference between the integrations is that the bottom boundary upward EP flux increases by over an order of magnitude from the initial, constant value. The primary effect of the external feedback is to provide a substantial source of additional wave energy to the model domain. In further experimentation, it was found that this source is strong enough that the imposed deceleration in the geopotential model need only be a factor of one-third as strong as that for the EP flux model.

A question then arises: is the induced, transient upward EP flux along the bottom boundary of the geopotential model sufficient to drive the SSW-like transition without the imposed zonal wind deceleration that originally induced it? By injecting the transient bottom boundary upward EP flux from the geopotential model integration into the EP flux model, we find that this is indeed the case (not shown). Given that such transient wave forcing is imperative to understanding wave–mean flow processes in the real world, we now focus attention on the two feedback processes under time-dependent forcing.

3. Time-dependent bottom boundary incoming wave activity flux

In this section, we examine SSW-like transitions in our versions of the Holton–Mass model using time-dependent forcing, with particular emphasis on the wave–mean flow positive feedback. Following Sjoberg and Birner (2012), we determine such a transition to have occurred when the zonal wind at all model levels above and including 30 km is equal to or less than a threshold wind speed for each respective level. We take this threshold zonal wind profile to be the zonal wind at day 24,000 in Fig. 3c, that is, where the zonal wind is within the strongly perturbed steady state. While 30 km is an arbitrary choice of lower level for our threshold, it is near to the 10-hPa pressure surface. This is a commonly used level for identification of major SSWs (e.g., McInturff 1978; Charlton and Polvani 2007). We found that the qualitative nature of the following results remains intact across a range of lower levels for this definition.

We use sine-squared pulses of bottom boundary upward EP flux or geopotential perturbation that are shown schematically in Fig. 5. For our idealized pulses, the baseline (initial and final) bottom boundary upward EP flux and the period of the pulse are specified, while the maximum amplitude (Max) is stepwise increased until a stability-state transition is forced. With amplitude A and period P known, total incoming EP flux is given by $F_{\text{int}} = \int_0^T A \sin^2(\theta/P) \pi dt = (1/2)AP$.

![Fig. 5. Schematic of the sine-squared EP flux forcing utilized in the time-dependent forcing experiments. The period and baseline EP flux (labeled BL) are specified, while the maximum amplitude (Max) is stepwise increased until a stability-state transition is forced. With amplitude A and period P known, total incoming EP flux is given by $F_{\text{int}} = \int_0^T A \sin^2(\theta/P) \pi dt = (1/2)AP$.](image-url)
As these pulses become more transient (going from right to left in the figure), the amplitudes increase while the total incoming EP flux decreases, both at roughly linear rates proportional to the pulse period. This behavior continues down to ~20 days, at which point the total incoming EP flux minimizes. At still shorter time scales, the amplitudes continue to increase with decreasing period, though at a rate more quickly than for periods greater than 20 days. This manifests as an asymptotic approach in the total incoming EP flux to a value larger than that near 20 days.

If the prescribed radiative equilibrium zonal wind (which is also the initialized zonal wind state) is moved, the pulse period of the total incoming EP flux minimum likewise moves. By adding 5 m s$^{-1}$ to the entire radiative equilibrium zonal wind profile, the minimum moves to ~15 days. By subtracting 5 m s$^{-1}$ from the entire profile, the minimum moves to ~30 days.$^{4}$ No such changes in the location of this minimum are found by varying the time scales of radiative damping. This suggests that the minimum is a manifestation of the wave-mean flow positive feedback and not due to externally prescribed (e.g., radiative) time scales.

To further understand how the wave–mean flow feedback affects the period of minimum total incoming EP flux, we remove the time-dependency of zonal wind terms affecting the waves. That is, for the streamfunction $\Psi$ evolution equation (see Holton and Mass 1976), we set the zonal wind $U$ and the meridional gradient of mean potential vorticity $\partial_y q$ to their initial values at all time steps [the form of $\partial_y q$ is given below in Eq. (8)].

Figure 7a shows the zonal wind anomalies in shading and the upward EP flux in log contouring for an integration of the standard EP flux model. Here the model is forced by a pulse of 10 days with amplitude 1.17 Pa, sufficient to drive an SSW-like transition. Figure 7b shows the same fields, but for the EP flux model with no time-dependent internal feedback of the wind on the waves.

For the experiment shown in Fig. 7a, the upward EP flux strongly dissipates, driving zonal wind reversal. In contrast, the experiment shown in Fig. 7b results in only weak EP flux dissipation of evanescent waves above the reflecting layer located near 21 km. Because the internal feedback is fully active in Fig. 7a, the upward EP flux amplitudes of the initial wave pulse are enhanced (relative to those in Fig. 7b) and the duration over which the initial upward EP flux pulse is greater than zero becomes longer with height, especially up to 30 km. In contrast, the duration over which the initial upward EP flux pulse is greater than zero in Fig. 7b decreases with height.

Despite the zonal wind feedback on the waves being independent of time in Fig. 7b, the 10-day pulse of bottom boundary wave activity flux is organized onto an almost 20-day period within the model domain. To test what happens when the bottom boundary pulse period approaches the period onto which the model state organizes the waves (here, 20 days), we repeat the experiments shown in Figs. 7a and 7b, but across a range of bottom boundary pulse periods. The duration where the initial upward EP flux pulse at 30 km is greater than zero is shown in Fig. 7c as a function of bottom boundary pulse period for the standard EP flux model and for the EP flux model with no time-dependent feedback.

In the fully active feedback case, the duration of the initial wave activity pulse increases as a function of bottom boundary pulse period. In the time-independent feedback case, the duration of the initial wave activity flux pulse increases for bottom boundary pulse periods up to 20 days. A sudden decrease in durations occurs for a 21-day pulse, indicating that the pulse durations aloft

\[ (\partial_t + ikeU) \left\{ -\left(k^2 + l^2\right)\Psi + \frac{f_0}{N^2} \left[ \partial_{zz} \Psi - \frac{1}{4H^2} \Psi \right] \right\} + \frac{\alpha}{N^2} \left[ \partial_{zz} \Psi - \frac{1}{4H^2} \Psi \right] + \partial_y ik \Psi = 0, \]  

\[ (7) \]
are actually shorter than the initial bottom boundary pulse in the time-independent feedback case. This results from internal EP flux reflection (e.g., between days ~520 and 540 in Fig. 7b) suppressing positive upward EP fluxes.

The pulse duration behavior found here—increasing for bottom boundary pulse periods up to 20 days and then suddenly decreasing—further validates that the ~20-day bottom boundary pulse period is preferred, given our particular model initial state. Increasing the radiative equilibrium (initial) zonal wind profile by 5 m s\(^{-1}\) results in this jump instead occurring at a 15-day pulse period, while decreasing the radiative equilibrium zonal wind profile by 5 m s\(^{-1}\) moves this jump to a 30-day pulse period. As we previously found, these are exactly the periods at which the total incoming wave activity flux minimum occurs for the respectively altered radiative equilibrium zonal wind.

The above changes to the radiative equilibrium zonal wind produce concomitant changes to both the height of the model internal reflecting layer and the wave group velocity, where larger wind speeds produce a lower reflecting layer [in line with Charney and Drazin (1961)] and larger group velocity. Thus, with faster (slower) background flow, we expect internal propagation to be faster (slower) and over a shorter (longer) distance within the model domain. Note that the simulated waves here have vertical wavelengths of ~50 km. The aforementioned reflecting level (near 21 km in our control setup, that is, 11 km above the model’s bottom boundary) therefore severely limits the vertical propagation depth of the waves relative to their vertical wavelengths, with the bulk of the vertical domain leading to evanescence. We note again that these results are insensitive to the height of the model lid or to inclusion of upper-level Rayleigh friction—that is, insensitive to wave reflection from the model lid.

The minimum in total incoming wave activity flux seems to occur because we force the model with a pulse of duration at which the wave–mean flow positive feedback acts most strongly. We therefore hypothesize that pulses shorter than ~20 days are not as efficient at forcing the zonal wind and thus require increasingly larger maximum amplitudes of upward EP flux (and larger associated total incoming EP flux). Pulses longer than ~20 days are also not as efficient at forcing the zonal wind but have the added effect of larger integrated radiative damping.

Compounding the above functional behaviors of the necessary amplitudes and total incoming EP flux leads us to conclude, as in Sjoberg and Birner (2012), that there is a preferred time scale on the order of 10 days for wave forcing associated with SSWs. To lend credence to this
suggestion, we analyze December–February (DJF) pulses of upward EP flux in the Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim) dataset. Because we utilize zonal wavenumber 2 for the wave forcing in the model, we constrain the presented results to that wavenumber. Here, we define a pulse as a continuous anomaly from the seasonally evolving mean upward EP flux averaged over 45°–75°N.

There is moderate correlation (~0.67) between tropopause-level pulse maximum amplitudes and durations. When only considering those pulses within ±20 days of any SSW (identified using the technique of Charlton and Polvani 2007), pulse duration and maximum amplitude are highly correlated (~0.90). This relationship suggests that large amplitude pulses are typically only associated with long-duration pulses. We do not, however, find these results to be significantly different at 95% confidence from a time series of EP fluxes generated using an autocorrelation model of order 1, suggesting that the relationship between pulse duration and amplitude is consistent with a red noise spectrum.

Figure 8 shows, in shading and as a function of pressure, the percentage of wavenumber-2 upward EP flux pulses in ERA-Interim that have duration at least as long as that given on the abscissa. Very few pulses in the real atmosphere last longer than 20 days. Only above 50 hPa, where radiative time scales are large (Hitchock et al. 2010), does the 5% contour in the stratosphere exceed the 14-day period mark. For pulses in the lower stratosphere—between 100 and 300 hPa—observing a pulse of longer than 14 days is uncommon. Taken in sum, the reanalysis results show that neither anomalously large amplitude yet short (~1 day) duration pulses nor very long (~100 day) pulses are readily observed. The observational analysis is thus consistent with our simple modeling suggestion of an order 10-day preferential wave forcing time scale associated with SSWs.

The above modeling results exclusively consider the EP flux model; accordingly, only the internal wave–mean flow feedback is active. To gain insight into how the external wave–mean flow positive feedback interacts with the transient wave forcing utilized here, we perform the above sine-squared pulse experiment in both the geopotential and EP flux models with nonzero baseline. For a zero baseline in the geopotential model, the external feedback can only operate during the pulse period (i.e., when the bottom boundary geopotential is greater than zero). However, a nonzero baseline bottom boundary geopotential allows the external feedback to operate even once the pulse has ceased.

We find that the external feedback is strengthened as the baseline is increased. Furthermore, the external feedback allows small-amplitude pulses—which are insufficient to force an SSW-like transition on their own—to condition the model state in such a way that sufficient wave activity is fluxed into the model to complete the transition. As the baseline is increased, the geopotential model is more able to utilize the limitless wave source below the model boundary. Thus, the external feedback more strongly impacts the evolution of the wave activity flux within the geopotential model as the baseline is increased.

Figure 9 shows the bottom boundary wave activity flux amplitude necessary to drive an SSW-like transition as a function of baseline EP flux. Results for pulse periods of 20 days are shown in blue while results for pulse periods of 2 days are shown in red. Symbols along the left ordinate represent zero-baseline results, where the plus sign is for the geopotential model and the asterisk is for the EP flux model.
function of the baseline for both models. Results for only two pulse periods are shown—the 20-day period in blue and the 2-day period in red. For direct comparison, we present results from the geopotential model in equivalent wave activity flux values rather than in geopotential values.

When the baseline is zero (symbols), the geopotential model requires at least 25% higher bottom boundary wave activity flux amplitudes than does the EP flux model. That is, the EP flux model more easily forces SSW-like transitions when the external feedback is severely reduced in the geopotential model. We suggest this is so because even simple forms of prescribed bottom boundary geopotential perturbations in the geopotential model result in complex forms of incoming wave activity flux. The magnitudes of wave activity flux in either model rely on the magnitudes of geopotential, on the phase shifting with height of the geopotential, and on the zonal wind (which affects both the magnitudes and phases of the geopotential). Thus, specified bottom boundary geopotential perturbations poorly specify incoming upward wave activity fluxes.

With nonzero baseline, the geopotential model overcomes the poor specification discussed above by drawing wave activity from below the bottom boundary. Incoming wave activity flux amplitudes necessary for SSW-like transition tend to be lower in the geopotential model than in the EP flux model for nonzero baseline. As the baseline approaches the bifurcation points for each model, the incoming wave activity flux amplitudes necessary to force an SSW-like transition collapse to the baseline value. This is evident for the baseline larger than $-0.070 \text{ Pa}$ in Fig. 9, where the amplitudes of both models at both pulse periods are equal to the baseline value.

In the geopotential model, the initial bottom boundary pulse of wave activity flux conditions the model mean state to subsequently flux sufficient bottom boundary wave activity into the model in order to complete the transition. In this way, the external feedback allows an SSW-like transition to be forced with smaller pulse amplitudes than would be needed without the feedback. In line with how the external feedback works, these reductions in necessary amplitudes are, however, accompanied by increases of at least 25% in the total incoming wave activity flux. For the transient forcing results presented here, the total incoming wave activity flux is always higher in the geopotential model than in the EP flux model.

In both models, the period at which the minimum in total incoming wave activity flux is located becomes shorter as the baseline is increased. The external feedback allows pulses to force transitions with smaller amplitudes, which allows the associated total incoming wave activity flux to be smaller as well. Because this reduction is of largest magnitude at the shortest pulse periods, the minimum in total incoming wave activity flux moves toward shorter periods. This is perhaps not surprising, given the strong dependence of necessary amplitudes on pulse period (see Fig. 6) and that there is a minimum amplitude to drive a transition (see Fig. 3). Amplitudes for very long pulses can only be reduced to a certain degree before reaching the steady-state transition amplitude, while amplitudes for very short pulses may be reduced substantially before reaching this point. Through this, we find that the external feedback acts to shorten the wave forcing time scales as the feedback is strengthened (i.e., as the baseline is increased).

4. Tropopause effects

To study the impact of the tropopause on the forcing of SSW-like events in our versions of the Holton–Mass model, we consider three tropopause setups. The first has no vertically varying stratification (as measured by the buoyancy frequency squared, $N^2$); the entire model depth is held at the stratospheric stratification of $4 \times 10^{-4} \text{s}^{-2}$. We refer to this as the stratospheric setup. Note that this is equivalent to the control setup used up to this point. The second is a hyperbolic tangent structure in height starting at 11 km and increasing over a depth of 4 km from tropospheric to stratospheric stratification. We refer to this stratification profile as the tropopause setup. The third profile includes a tropopause inversion layer (TIL), for which $N^2$ near the top of the tropopause jumps to a much higher stratification and then quickly relaxes to stratospheric values (see Birner 2006). For this profile, referred to as the TIL setup, we utilize the tropopause setup, but add a 1-km, squared sinusoid of maximum amplitude $2 \times 10^{-4} \text{s}^{-2}$ to the profile, centered a half kilometer above the center of the hyperbolic tangent. This relates to a maximum TIL stratification of $6 \times 10^{-4} \text{s}^{-2}$, a magnitude similar to that found in climate model output (Birner et al. 2006).

We first analyze the unforced model states under different magnitudes of tropospheric stratification. We find that as the stratification of the model troposphere is lowered (i.e., the jump in stratification at the tropopause is increased), the relevant index of refraction at the tropopause levels becomes increasingly negative, asymptotically approaching $-\infty$. The tropopause then will reflect increasing amounts of upward wave activity flux as the tropopause stratification jump is increased. In the real atmosphere, this reflection would be at least partially overcome by horizontal wave propagation along the tropopause. Meridional wave propagation is inhibited in the Holton–Mass model, however, leading instead to highly
unphysical results as the tropopause stratification jump is increased. We therefore opt for a tropospheric stratification of $3 \times 10^{-4}$ s$^{-2}$ (greater than the observed $\sim 1 \times 10^{-4}$ s$^{-2}$). No such highly unphysical results come from the jump in stratification for our TIL setup, and thus, we feel its magnitude is appropriate for this type of study. Furthermore, we find the below results hold qualitatively both for different maximum stratification of the TIL setup and for different depths of the TIL.

The above idealized tropopause setups are shown schematically in Fig. 10a. Note that the stratification profiles shown are at the model resolution of 500 m. To determine the steady states for each stratification profile, we perform the same slowly ramping, quasi-steady EP flux experiment as before (see section 2). Figure 10b is similar to Fig. 3b, except that the results of all three of our idealized tropopause setups are shown.

Comparing the tropopause setup (dark gray) and TIL setup (light gray) to the stratosphere setup (black), one observes that including a tropopause of either form shifts both the upper bound and the lower bound of the bistable regime to higher magnitudes of bottom boundary upward EP flux. The higher magnitudes of the upper bounds imply that the tropopause and TIL stratification profiles reduce the stratospheric wave activity flux, requiring that more be injected into the model in order to drive a transition between steady states. Similarly, the higher magnitudes of the lower bounds of the bistable regime for both the tropopause and TIL setups denote that the necessary forcing to keep the model state within the strongly perturbed regime has increased. These results are in line with Chen and Robinson (1992), who showed that the tropopause reduces upward wave propagation into the stratosphere. Furthermore, with both the tropopause and TIL setups, the zonal wind at 30 km in the strongly perturbed stable state is approximately 5 ms$^{-1}$ larger than in the stratosphere case, indicating that the level of strongest wave forcing has shifted upward.

If the tropopause acts to decrease the upward EP flux entering the stratosphere and thus the model requires more incoming EP flux for transition, why is the TIL setup upper bound closer to that of the stratosphere case than is the tropopause case? Its higher maximum stratification would seem a priori to result in a higher upper bound than even the tropopause setup.

By analyzing the vertical structure of the wave activity flux in the presence of our TIL, we find EP flux divergence (wave generation) along its upper bound. This divergence comes about from a localized reversal of the meridional gradient of mean potential vorticity, satisfying the Charney–Stern necessary condition for baroclinic instability (Charney and Stern 1962). In the Holton–Mass model, this term takes the form

$$\frac{\partial y}{\partial z} q = \beta + \frac{f_0^2}{N^2} \epsilon \left( \frac{1}{H} \frac{\partial z U}{\partial z} - \frac{2}{N} \frac{\partial z N}{\partial z} \right) \left( \frac{\partial z U}{\partial z} \right),$$  

(8)

where $f_0$ is the Coriolis parameter, $\beta$ is the meridional derivative of the Coriolis parameter, $l$ is the meridional wavenumber of the model channel, $\epsilon$ is a factor resulting from the Fourier sine series expansion of $\sin^2(ly)$, and $H$ is the scale height. Because of the vertical structure of the TIL stratification, the final term in Eq. (8) is large and negative enough such that $\partial_y q < 0$.

This reversed meridional gradient of mean potential vorticity provides a small but important source of EP flux within the model domain, as it helps abate the loss of wave activity by the tropopause. We demonstrate the effect of this source by making use of the imposed zonal wind deceleration experiment of Fig. 4a, but with the tropopause and TIL setups. Figure 11 shows the zonal
wind and upward EP flux evolution for the tropopause setup (Fig. 11a) and the TIL setup (Fig. 11b). The bottom boundary EP flux magnitude and form of the imposed zonal wind deceleration are identical in the two runs. The amplitude of the imposed deceleration is increased to $39 \text{ m s}^{-1} \text{ day}^{-1}$ to account for the higher threshold needed to overcome for transition. In this experiment, the TIL wave source adds sufficiently to the internal wave–mean flow feedback so that a transition is forced, whereas no such transition occurs in the tropopause setup.

It is interesting to note that such mean potential vorticity gradient reversals may be forced in our TIL setup even when the model parameters do not produce a perpetual gradient reversal. For instance, increasing the radiative equilibrium (initial) zonal wind at all levels increases the positive contribution of the second term in Eq. (8) while keeping the other terms the same, allowing $\partial_z \varphi$ to be greater than zero. By perturbing this model state with a relatively small bottom boundary upward EP flux, we find that a meridional potential vorticity gradient reversal may be forced. The dominant term in such a reversal becomes the zonal wind curvature term of Eq. (8), which is zero here when the model is unperturbed but becomes negative and large as tropopause-level wave activity flux convergence decelerates the wind. In this way, only a reduced meridional potential vorticity gradient magnitude along the top of the TIL is necessary to provide an in situ wave source within the lower stratosphere.

In seeking to verify the applicability of these results outside of the simplified modeling environment, we have searched for such potential vorticity gradient reversals in other sources. ERA-Interim data show no clear signs of
these reversals near the tropopause. Thus, the perpetual reversal we find above is not a robust condition of the real atmosphere. However, vertical smoothing of fields due to data assimilation is known to decrease the maximum stratification of the TIL (Birner 2006) and could thereby affect how these gradient reversals are resolved. Idealized general circulation modeling with sufficiently high resolution to capture the TIL is therefore a next step in determining the potential impact of such reversals.

5. Conclusions

Wave–mean flow interaction theory shows that upward EP flux (sometimes approximated by its heat flux component) represents the appropriate measure to describe the planetary wave forcing of SSW-like events. This theory as well describes a wave–mean flow positive feedback where wave activity flux convergence within the stratosphere decelerates the zonal wind there, allowing more wave activity to be drawn upward. A quasigeostrophic, beta-plane channel model of such wave–mean flow interaction—the Holton–Mass model—contains this feedback mechanism, yet is forced by bottom boundary geopotential perturbations. Here, we recast the Holton–Mass model so that upward EP flux is specified at its bottom boundary.

In the standard Holton–Mass model with bottom boundary geopotential forcing, the positive feedback exists both internally and externally. The external feedback allows the model to draw wave activity upward from a potentially limitless source below the model domain, which makes this feedback unrealistically strong and may create unphysical results. We therefore isolate the internal portion of this positive feedback by specifying the incoming wave activity flux. Furthermore, our two model bottom boundary setups allow us to distinguish characteristics of both the external and internal feedback components.

Through analysis of the stable steady states, we find that the external component of the positive feedback, active in the geopotential model, is the strongest driver of the oscillating zonal wind state associated with the Holton–Mass model. It is through this external feedback that the model state may “turn on” and “turn off” incoming waves. Model states that are conducive to wave propagation allow incoming wave activity fluxes to decelerate the wind, while model states that are not conducive to propagation (i.e., which result in wave reflections) result in acceleration of the wind. In the EP flux model, the effect of wave reflection is dampened at lower levels, which acts in part to prevent zonal wind oscillations except for at unrealistically large amplitudes of incoming wave activity flux. This is not unexpected as steady, conservative wave forcing should result in a steady zonal wind response.

The external feedback is not, however, the cause of the parameter-dependent bistable regime previously identified in the Holton–Mass model. We find that this bistability is a more fundamental result of this simple model: even for a given incoming upward wave activity flux, the zonal wind may exist at one of two substantially different steady states. This may apply hypothetically to the real world, where a particular seasonal progression into winter (i.e., the state of the zonal wind at the beginning of winter) determines how the stratosphere will be affected by the DJF mean upward EP flux. For very weak or very strong mean upward EP flux, the DJF mean zonal wind state may be determined by the waves, but for moderate upward EP flux (within the bistable regime), the zonal wind may determine its own mean state by affecting how and where the waves propagate and subsequently force the zonal wind itself.

We demonstrate this mean flow determination of its own final state by imposing a perturbation on the zonal wind and allowing the wave–mean flow feedback to drive a stable state transition. In doing so, we find that the EP flux model is able to generate an SSW-like transition by reducing internal EP flux convergences and increasing wave propagation, thereby increasing the amplitudes of wave activity flux at middle to upper model levels. The geopotential model instead drives the same transition with an order of magnitude increase in incoming EP flux. Because the geopotential model can tap into a limitless reservoir of wave activity below the model domain, much smaller levels of wind perturbation are necessary to set off this transition.

Through an idealized time-dependent incoming EP flux forcing experiment, we find, in agreement with previous results, that the stratosphere has a “preferred” wave forcing time scale on the order of 10 days. This preferred time scale manifests as a minimum in the total incoming upward EP flux necessary to drive a transition. At periods longer than 20 days, the total incoming EP flux increases linearly with increasing period. At periods shorter than 20 days, the necessary amplitudes are excessively large. These features present considerable requirements on the atmosphere to force SSW-like transitions. Excessively large-amplitude pulses with very short (~1 day) duration and pulses with very long (~100 day) durations are not found in the reanalysis dataset, consistent with these results.

We suggest that the preferred time scale is related to the interplay between internal wave propagation (e.g., group velocity) and the height of the reflecting layer. The positive feedback appears to be strongest when the model is forced with pulses near this preferred time.
scale. Pulses of shorter duration require larger integrated incoming EP flux to drive the same zonal wind variability. Pulses longer than this preferred time scale are subjected to this requirement as well, but also to large integrated radiative damping.

By performing these time-dependent forcing experiments with the geopotential model, we find that the external feedback significantly affects the results of this transient forcing experiment. As the source of available wave energy at the model bottom boundary increases, the external feedback becomes stronger and the bottom boundary EP flux pulse amplitude necessary to drive a transition becomes smaller. The magnitude of this decrease in necessary pulse amplitude is largest at short pulse periods, thereby shifting the minimum in total incoming EP flux to these short pulse periods. The external feedback, in essence, allows the model to draw up the necessary wave activity to force a transition following these short duration pulses (which concurrently result in much higher total incoming EP flux).

Two idealized forms of the tropopause were included in the EP flux model: a simple tropopause and a TIL. We find, as expected, that the tropopause setup and TIL setup both reduce the EP flux entering the model stratosphere, relative to the model integrated with no vertically varying stratification (the stratosphere setup). It follows that forcing steady-state transitions requires larger EP fluxes when a tropopause is present.

However, the TIL setup requires a smaller increase in incoming EP flux because of in situ wave generation caused by a reversal of the meridional gradient of mean potential vorticity. We find that these gradient reversals may exist in the model either from the selected parameters or may be forced by zonal wind decelerations caused by wave activity flux convergences. This wave generation provides an important addition of wave activity to the internal feedback, increasing EP flux amplitudes in the stratosphere relative to amplitudes in the tropopause setup. Even if the external feedback component is weak or removed, in situ wave generation from the TIL will help decelerate the winds aloft, thus strengthening the wave–mean flow internal feedback process. However, it is not clear at this time how applicable this result is to the real world. Further study of this potential impact is required.

In summary, the reformulated model presented here shows that bistability of the zonal wind is a robust, internal feature of this type of simple model; that there exists internal wave time scales upon which zonal wind is most easily forced (in an integrated wave activity flux sense); and that the TIL may be an important source of large-scale wave activity within the stratosphere. The simplified model used here allows us to perform the idealized experiments needed to eke out these features of the system.

Acknowledgments. The ERA-Interim data were obtained from the ECMWF data server. This research was supported by funding from the Climate and Large-Scale Dynamics Program of the U.S. National Science Foundation. This manuscript benefited from helpful discussions with John Albers. We thank Edwin Gerber and two anonymous reviewers for their comments and suggestions.

REFERENCES


