What Drives the Brewer–Dobson Circulation?

NAFTALI Y. COHEN,* EDWIN P. GERBER, AND OLIVER BÜHLER

Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York

(Manuscript received 29 January 2014, in final form 12 June 2014)

ABSTRACT

Recent studies have revealed strong interactions between resolved Rossby wave and parameterized gravity wave driving in stratosphere-resolving atmospheric models. Perturbations to the parameterized wave driving are often compensated by opposite changes in the resolved wave driving, leading to ambiguity in the relative roles of these waves in driving the Brewer–Dobson circulation. Building on previous work, this study identifies three mechanisms for these interactions and explores them in an idealized atmospheric model. The three mechanisms are associated with a stability constraint, a potential vorticity mixing constraint, and a nonlocal interaction driven by modifications to the refractive index of planetary wave propagation. While the first mechanism is likely for strong-amplitude and meridionally narrow parameterized torques, the second is most likely for parameterized torques applied inside the winter-hemisphere surf-zone region, a key breaking region for planetary waves. The third mechanism, on the other hand, is most relevant for parameterized torques just outside the surf zone. It is likely for multiple mechanisms to act in concert, and it is largely a matter of the torques' location and the interaction time scale that determines the dominant mechanism.

In light of these interactions, the conventional paradigm for separating the relative roles of Rossby and gravity wave driving by downward control is critiqued. A modified approach is suggested, one that explicitly considers the impact of wave driving on the potential vorticity of the stratosphere. While this approach blurs the roles of Rossby and gravity waves, it provides more intuition into how perturbations to each component impact the circulation as a whole.

1. Introduction

The meridional overturning circulation of the stratosphere transports air from the equator to the pole. It is named the Brewer–Dobson circulation (BDC) in recognition of its discovery by Brewer (1949) and Dobson (1956), although hints of the circulation date back to Dobson et al. (1929). The stratospheric circulation affects tropospheric climate and variability across many time scales, through coupling to the chemistry and transport of ozone (e.g., Thompson et al. 2011) and water vapor (e.g., Solomon et al. 2010) on decadal time scales, to dynamical interactions with the jet stream on intraseasonal time scales (e.g., Baldwin and Dunkerton 2001).

The BDC is a wave-driven circulation. Retrograde wave torques allow flow across lines of constant angular momentum, balancing the prograde Coriolis force associated with poleward flow. As more figuratively described by Holton et al. (1995), the wave forcing generates a ''fluid-dynamical suction pump,'' pulling the air poleward in both hemispheres. The dynamical framework for understanding the BDC uses the transformed Eulerian-mean (TEM) equations (Andrews and McIntyre 1976; Dunkerton 1978). The merit of this transformation is that the residual TEM circulation approximates the Lagrangian-mean circulation for steady disturbances (e.g., Bühler 2014, chapter 11). Moreover, in the quasigeostrophic (QG) limit, these equations simplify to provide a clear causality of the wave–mean flow driving, known as the ''downward control'' principle (Haynes et al. 1991).

Figure 1a illustrates the conventional paradigm for the wave-driven BDC dynamics. The downward-control argument states that, given a wave forcing, it is possible to deduce the mean-flow fields from the zonal-mean TEM equations (top arrow in Fig. 1a). The key is that

* Current affiliation: Department of Geology and Geophysics, Yale University, New Haven, Connecticut.

Corresponding author address: Naftali Y. Cohen, Department of Geology and Geophysics, Yale University, P.O. Box 208109, New Haven, CT 06520-8109.
E-mail: naftalic@gmail.com

DOI: 10.1175/JAS-D-14-0021.1

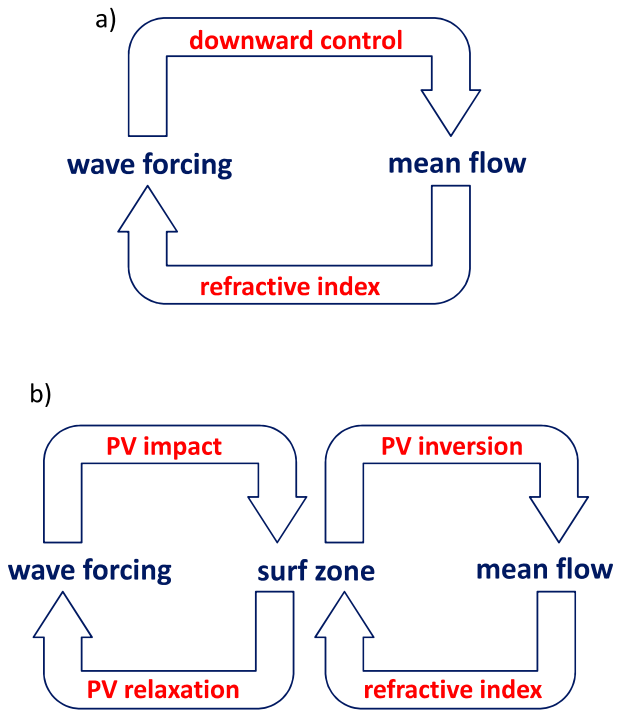

FIG. 1. Two paradigms for interpreting BDC dynamics. (a) The conventional paradigm: the system is coupled by linear-wave and mean-flow dynamics. Given the wave forcing, downward control can be used to infer the mean-flow fields, whereas the index of refraction controls the propagation details of the waves. (b) The modified paradigm: we add an intermediate step, the surf zone, to explicitly consider the impact of the wave forcing and resulting mean flow on the PV of the stratosphere. Given the surf-zone structure, the wave forcing is just that needed to maintain the well-mixed state against radiative restoring forces, and the mean flow can be determined by PV inversion. Conversely, the surf zone is itself a product of the interaction between the mean flow and wave breaking.

the stratosphere is strongly stratified and the sole forcing is the wave activity from below. It is then tempting to consider the wave forcing as independent of the mean flow. However, changes in the mean flow can affect the very existence and propagation of the stratospheric waves, as illustrated first by Charney and Drazin (1961) and later by Matsuno (1970). The index of refraction quantifies whether and how waves can propagate upward and meridionally in the atmosphere. The wave response depends on the mean flow and the source. Thus, any mean-flow change by the wave forcing couples back to change the wave forcing itself (bottom arrow in Fig. 1a). The refractive-index view on the BDC dynamics, however, is equally incomplete in that the flow itself depends intimately on the wave driving.

What waves drive the BDC? Both models and observations consistently show that stratospheric wave driving is dominated by planetary-scale Rossby waves (RW), as summarized, for example, in Butchart et al. (2011). Stationary planetary waves are generated by large-scale orography and land–sea contrasts, but transient planetary-scale and synoptic-scale waves also contribute to the RW driving, particularly in the lower stratosphere (e.g., Shepherd and McLandress 2011, their Fig. 2a). To a lesser extent, small-scale gravity waves (GW), forced by smaller mountains, convection, and frontal instabilities with length scales of roughly 10–1000 km, also contribute to the wave driving. In the upper stratosphere and mesosphere, however, gravity waves begin to play a more dominant role. While planetary waves are sufficiently resolved in state-of-the-art numerical models of the atmosphere, much of the gravity waves spectrum is underresolved and must be parameterized (e.g., Fritts 1984; Fritts and Alexander 2003; Alexander et al. 2010).

Comprehensive chemistry–climate models with well-represented stratospheres largely agree on the total amplitude of stratospheric circulation (Butchart et al. 2010), although in some respect they are tuned to do so. Intermodel comparison shows statistically significant agreement on the annual-mean upward mass flux at 70 hPa (Eyring et al. 2010, their Fig. 4.10a). Somewhat surprisingly, however, this intermodel comparison shows that models do not agree on the contributing wave components—that is, how much is driven by parameterized gravity versus resolved Rossby waves. Some models suggest that GW contribute up to half the wave driving, while others suggest they play a trivial role. The ambiguity is worse for nonorographic GW, where models do not even agree on the sign of their contribution.

In addition, climate models predict, on average, an approximately 2% annual-mean increase per decade at 70 hPa in the BDC in response to anthropogenic forcing in the future (Butchart et al. 2006, 2010; Eyring et al. 2010). In terms of the change, however, the disagreement on the relative contribution of the wave components is more severe. Some models suggest that gravity waves dominated the response, while others rely almost exclusively on Rossby waves (e.g., Garcia and Randel 2008; Li et al. 2008; McLandress and Shepherd 2009; Eyring et al. 2010). Note that observations of stratospheric tracers cannot constrain stratospheric circulation trends, as the sampling uncertainty is too great, and there are concerns about biases and instrument noise from the satellite instruments (Engel et al. 2009; Garcia et al. 2011; Khosrawi et al. 2013). However, Kawatani and Hamilton (2013) suggest that the BDC may be increasing, based on changes in the amplitude of the stratospheric quasi-biennial oscillation.

In a recent paper, Cohen et al. (2013, hereafter CGB13) found that strong compensating interactions

are quite likely between the resolved and parameterized wave driving in the stratosphere. Perturbations to the parameterized wave driving are often canceled by an equal and opposite change in the resolved wave driving (the Eliassen–Palm flux divergence). The phenomenon was first observed by McLandress and McFarlane (1993), and can also be seen in Manzini and McFarlane (1998) and McLandress et al. (2012). Recent work by Sigmond and Shepherd (2014) shows that a comprehensive atmospheric general circulation model exhibits similar phenomenon. In addition, Sigmond and Shepherd (2014) show that compensating interactions are likely for the response to climate change. These compensating interactions may explain why comprehensive models tend to agree more on the total strength of the BDC than on that associated with individual components.

CGB13 suggested, in a proof by contradiction, that compensation between the resolved and parameterized waves is inevitable when the parameterized wave torques, if not compensated, would drive the stratosphere to a physically unrealizable state that is unstable to baroclinic instability. They found that this is likely for strong and/or meridionally narrow torques. Sigmond and Shepherd (2014), however, found no evidence for instability and argued, alternatively, that orographic gravity wave drag (OGWD) tends to weaken the zonal winds in the upper flank of the subtropical jet, thus changing the refractive index in a way that reduces planetary wave vertical propagation. In particular, Sigmond and Shepherd (2014) showed that the compensation by resolved waves was mainly associated with changes to the meridional propagation of the resolved waves (about 70% reduction), while changes in the vertical propagation play a secondary role (about 30% reduction). Both studies, however, examine only the steady, mature state of compensation and do not discuss the development of compensation in time. Here we show that the key to understand the interactions is to investigate their temporal structure and that the most important parameter is the location of the applied torque.

Building on the potential vorticity (PV) analysis by Scott and Liu (2014), we propose an intermediate step to the downward control–refractive index paradigm for the BDC dynamics. The intermediate step emphasizes that Rossby waves respond to the distribution of PV. This perspective is based on the McIntyre and Palmer (1983) conceptual surf-zone model and is used to propose a simple mechanism for the interactions between planetary resolved waves and parameterized gravity waves in the stratosphere. As with downward control or the refractive-index arguments, this lens for viewing the BDC is not complete, but provides new insight into how GW and RW interact with each other. This paradigm is illustrated in Fig. 1b and developed in detail in section 2.

In section 3, these contrasting paradigms are used to identify three possible mechanisms for the interactions, which can be differentiated based on the location and amplitude of the parameterized gravity waves. To validate our hypothesis, we test the conceptual model using an atmospheric general circulation model (AGCM), as described at sections 4 and 5. Last, in section 6, we conclude our study and address the question of what drives the BDC and how the BDC may change in light of these paradigms.

2. A surf-zone paradigm for the BDC

The stratospheric surf zone is, conceptually, a region in the midlatitude stratosphere where PV mixing is substantial (McIntyre and Palmer 1983). This model assumes that there is sufficient planetary wave propagation and breaking to completely mix the PV in the midlatitude stratosphere, sandwiched between two regions of strong PV gradients that inhibit mixing (McIntyre and Palmer 1985). Killworth and McIntyre (1985) showed that for a completely mixed surf zone, any excess of propagating planetary waves will be reflected away, thus the well-mixed region changes from perfect absorber at the initial stages to a perfect reflector. Figures 2a–d shows a latitude–longitude cartoon (for a generic level in the stratosphere) of Rossby wave breaking that results in a PV mixing. Large-scale orography (dashed black) generates the planetary waves that propagate upward and distort horizontally the materially conserved PV isopleths (solid red), thus transporting lower PV values poleward and higher PV values equatorward. The wavy distortion continues, irreversibly, until the wavy pattern breaks, resulting in PV mixing. At this time, the PV gradients necessary for wave propagation are eliminated. Thus, this model suggests that there is sufficient wave breaking to homogenize the surf-zone PV, but no more.

Absent any planetary wave forcing, and so no wave breaking, the mean PV of the flow would return toward an equilibrium or background distribution. Hence, the wave driving can be envisioned as a force that causes the PV to deviate from the background PV, as described most easily with the aid of Fig. 3. In this figure the solid red curve describes the background PV, which increases with latitude. Given the extent of surf-zone width and perfect PV mixing (in blue), the wave driving is essentially constrained by the background PV. There will be exactly enough wave breaking to offset the restoring force, but no more, as long as there is sufficient planetary wave forcing from below. Using this conceptual

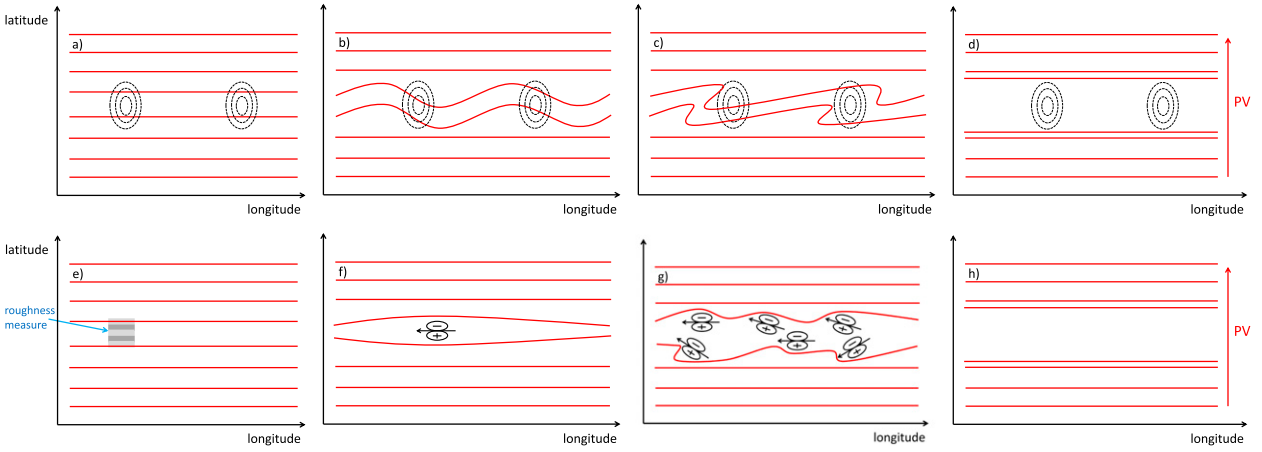

FIG. 2. A latitude–longitude diagram, for a generic level in the midstratosphere, of Rossby wave and gravity wave breaking that result in PV mixing. (a)–(d) Large-scale orography is in dashed black, and the red lines denote PV isopleths increasing poleward. In (a) Rossby waves generated by large-scale orography propagate upward, and in (b) they distort the materially conserved PV, thus transporting lower PV values poleward and higher PV values equatorward. In (c) the wavy distortion continues irreversibly until the wavy pattern breaks, resulting in PV mixing, as shown in (d). (e) Gravity waves generated by a small-scale orography propagate upward, break, and exert a retrograde forcing, (f) thus changing the PV. This PV change acts as a local mixing, effectively separating PV isopleths. (g) We assume many such gravity wave breaking events. (h) Aggregation of them or applying a zonal-mean result in effective PV mixing.

thinking, we provide a quantitative prediction of the wave driving.

The zonal-mean QG PV equation in (x, y, p) Cartesian longitude–latitude–pressure coordinates is [e.g., Andrews et al. 1987, their (3.3.4)]

$$\overline{q}_t = -(\overline{v'q'} + \overline{\mathbf{X}})_y + \overline{S}. \tag{1}$$

Using common notation, an overbar represents a zonal mean, a prime denotes a geostrophic perturbation therefrom, and a subscript denotes a partial derivative. The QG PV is denoted by q, v is the meridional velocity, $\mathbf{X}$ is the unresolved gravity wave driving, and S represents nonconservative terms, mostly due to diabatic heating. Here $\overline{v'q'}$ is the zonal-mean meridional flux of PV and, in the QG limit, is equal to the Rossby wave driving or Eliassen–Palm flux divergence (EPFD). In accordance with QG scaling (Edmon et al. 1980), the PV and its meridional gradient are

$$\overline{q} = f - \overline{u}_y + \left(\frac{f\overline{\theta}}{\overline{\theta}_{0p}}\right)_p \quad \text{and} \quad \overline{q}_y = \beta - \overline{u}_{yy} + \left(\frac{f\overline{\theta}_y}{\overline{\theta}_{0p}}\right)_p, \tag{2}$$

where u is the zonal wind, θ is the perturbed potential temperature about a stratified background state that is independent of latitude, $\theta_0 = \theta_0(p)$, and $f = f_0 + \beta y$ is the Coriolis frequency.

In this framework, we now make a crude approximation, supposing that the nonconservative term can be approximated by a linear relaxation toward a background PV profile. That is,

$$\overline{S} \approx -\frac{\overline{q} - q_b}{\tau}, \tag{3}$$

where q_b is a background PV and τ is a relaxation time scale. We visualize this simple thought experiment using the arrows in Fig. 3. Denoting the total wave forcing with $\overline{\mathbf{G}} = \overline{v'q'} + \overline{\mathbf{X}}$, (1) becomes

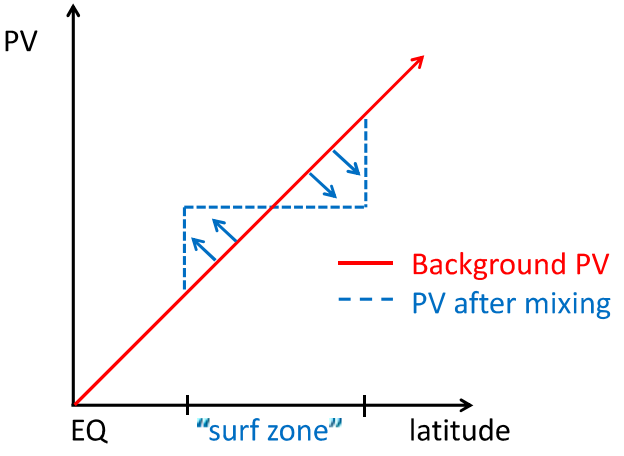

FIG. 3. The conceptual model for the stratospheric surf zone. Planetary wave breaking mixes the PV (dashed blue) and drags the flow away from the background PV (solid red). Given the width of the mixing region, the total wave driving is constrained to be equal to the absolute value of the area between the two curves.

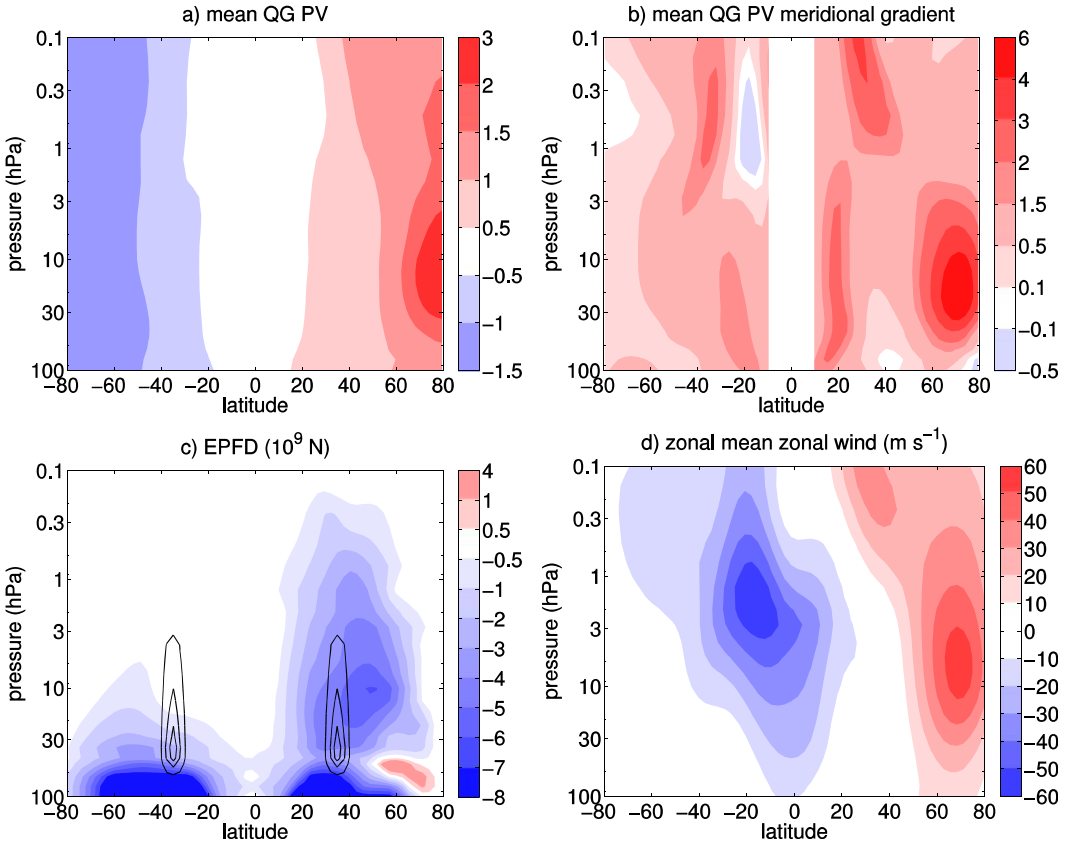

FIG. 4. Time- and zonal-mean climatology of the control integration. (a) The PV is scaled relative to the value of the planetary PV associated with f_0 at 45°N, and (b) its meridional gradient is scaled relative to the planetary value $a\beta_0$ at 45°N. (c) EPFD. In black contours (lines of −1, −5 and −10 are shown) are idealized torques that applied in the SH and NH stratosphere; see section 4 for details. (d) Zonal-mean zonal wind.

$$\overline{q}_t = -\overline{\mathbf{G}}_y - \frac{\overline{q} - q_b}{\tau}. \tag{4}$$

Thus, the time evolution of PV is governed by competition between the wave driving, which tends to homogenize the PV within the surf zone, and nonconservative forces (diabatic forcing, etc.), which continually adjust the PV toward the background state.

Consider a steady state and let us concentrate on the surf-zone region. It follows that $\overline{q}_t = 0$ (steady state) and the balance is

$$\mathbf{G}_y = -\frac{\overline{q} - q_b}{\tau}. \tag{5}$$

Both the mixing $\mathbf{G}_y$ and restoration of PV occur on a finite time scale, and we would expect a balance between these terms that results in nonzero PV gradients—that is, an imperfect surf zone. Indeed, as we observe in our atmospheric model (Fig. 4) and seen in observations, PV gradients are never fully homogenized in the surf zone. If the time scale of mixing is fast relative to the restoration, however, PV gradients will be weak. To make analytic headway, we will assume that the mixing is sufficiently fast such that $\overline{q}_y$ is approximately zero. This means that applying a y derivative on (5) eliminates the $\overline{q}$ term on the right-hand side.[1] That is,

$$\mathbf{G}_{yy} = q_{by}/\tau, \tag{6}$$

where the curvature of the wave driving is proportional to the background PV gradient. To make a back-of-the-envelope estimate, suppose that the background PV is simply the planetary PV. That is, $q_b = f_0 + \beta y$ and $q_{by} = \beta$. Further assume that the wave driving vanishes outside of a surf zone with a fixed meridional extent. It then follows

[1] Even if we relax the QG assumption, the same result applies. For example, retaining the PV advection term modifies (4) to be $\overline{q}_t + \overline{q}_y\overline{v}^* = -\overline{\mathbf{G}}_y - (\overline{q} - q_b)/\tau$, where again the PV homogenization assumption drops the PV advection term.

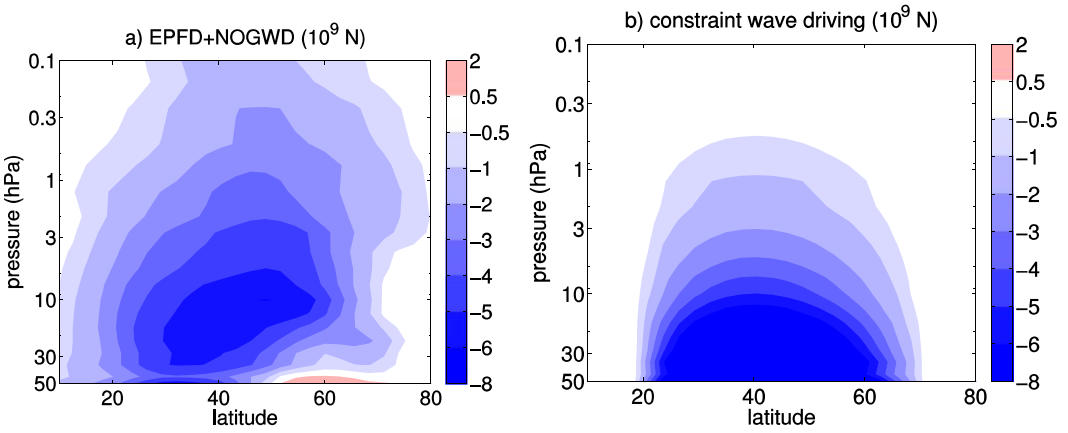

FIG. 5. The actual and estimated wave driving of the NH stratosphere. (a) The sum of the EPFD and the NOGWD from the control integration of the AGCM is shown. (b) The estimated wave driving from (7), obtained using the conceptual model where the surf zone is centered at 45°N, half-width of 26°, $\tau = 40$ days, and $q_{by} = \beta$. The vertical structure of the wave driving follows from the density structure of the atmosphere.

that the wave driving has the following simple parabolic shape

$$\mathbf{G} = -\frac{\beta h^2}{2\tau} + \frac{\beta(y - y_0)^2}{2\tau}, \tag{7}$$

where y_0 is the center of the surf zone and h is the half-width. Its minimum value is $-\beta h^2/2\tau$, the negative value consistent with our expectation of a retrograde forcing. We note that the residual vertical velocity, which is proportional to the meridional gradient of this parabolic wave drag, would be infinite right at the boundaries. In a more realistic context, diffusion and advection of the relative vorticity would smooth this out. We stress that the wave driving in (7) only meant to provide a rough, zeroth-order estimate.

This model for the wave drag in (7) requires some knowledge of the surf-zone location and extent. We use an idealized AGCM to make an estimate. Briefly, the AGCM used in this study is a dynamical core developed by the Geophysical Fluid Dynamics Laboratory. We are using exactly the same configuration as the "default" configuration in CGB13 ("control run"). The model integrates the dry hydrostatic primitive equations with pseudospectral numerics and relaxes the temperature equation with a 40-day time scale toward a simplified perpetual-January radiative-equilibrium temperature field (Polvani and Kushner 2002). This yields realistic tropospheric and stratospheric conditions without the need for convective or radiative schemes (Held and Suarez 1994; Polvani and Kushner 2002). We use the Gerber and Polvani (2009) configuration with most realistic troposphere–stratosphere coupling, including zonal wavenumber-2 topography of 3-km amplitude, centered between 25° and 65°N. We have also replaced the crude GW parameterization, a Rayleigh friction above 0.5 hPa that acts on the uppermost layers of the model in the Polvani and Kushner (2002) configuration, with a momentum-conserving parameterization scheme for nonorographic gravity wave (NOGW) (Alexander and Dunkerton 1999); see CGB13 for details. Note that no orographic gravity wave parameterization was used in the AGCM for this study.

Figures 4a and 4b show the detailed structure of the time-mean $\overline{q}$ and $\overline{q}_y$ (computed in spherical coordinates) from a 10 000-day control integration of the AGCM. For the purposes of computing the QG PV gradient of the model, $\overline{\theta}_0$ in (2) was defined as the latitudinal average from 60° to 30°S and from 30° to 60°N, and $\overline{\theta}$ was computed as the deviation from this background value. Figures 4c and 4d show the integrated time- and zonal-mean EPFD and zonal wind, respectively.

Figure 5 compares an estimate of the wave driving in (7) with that in the AGCM integrations. Figure 5a shows the time- and zonal-mean model-integrated EPFD and nonorographic gravity wave drag (NOGWD), while Fig. 5b shows the wave driving estimate of (7) with the parameters $q_{by} = \beta$, $h = 26°$, $y_0 = 45$°N, and $\tau = 40$ days, which is the thermal relaxation time scale in the AGCM (Polvani and Kushner 2002). The surf-zone width and y_0 parameters were chosen according to Figs. 4b and 4c. Note the reasonably good agreement between the model integrations and the analytic wave driving in both amplitude and shape.

The success of the simple analytic framework suggests that Rossby waves are quite efficient in opposing the restoring force on the PV, as described in (5). Relative to the actual wave driving in the AGCM, however, our

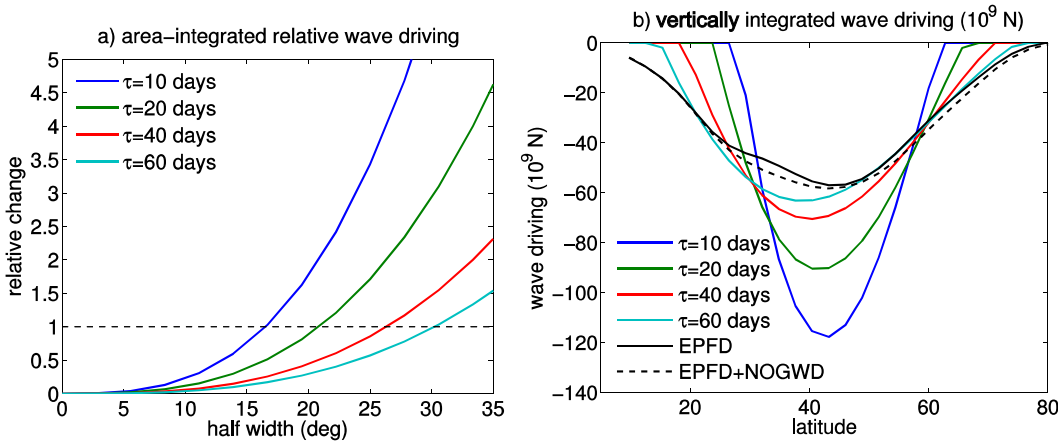

FIG. 6. (a) The area-integrated wave driving estimated by the surf-zone model, as a function of the half-width and time scales. The wave driving is computed as in (7), integrated between 1.25 and 35 hPa and between 10° and 80°N, and scaled by the climatological wave driving (EPFD+NOGWD). Hence, a value of 1 implies an ideal match. The chosen parameters for (7) are $y_0 = 45°$N and $q_{by} = \beta$. (b) The vertically integrated constraint wave forcing as a function of the time scale, with the "best fit" half-width according to (a).

simple framework underestimates the wave driving between 0.1 and 3 hPa and overestimate it between 30 and 50 hPa. This is due, in part, to the fact that (7) is based on a 1D surf-zone conceptual model. The stratosphere in the AGCM is further from its thermal equilibrium at higher elevations, leading to a higher effective-damping time scale. In principle, it would be possible to tune the vertical structure of the constraint wave forcing by taking into account changes in the stratification, but the model is only intend to provide an order of magnitude estimate.

This result should be compared with a recent study by Scott and Liu (2014). Using the TEM shallow-water system, they diagnosed the EPFD needed to establish a fixed stratospheric surf-zone region against the restoring effect of radiative relaxation. In agreement with the constraint estimated provided here, they found, using an iterative scheme, that a parabolic shaped EPFD is needed to maintain a surf zone with homogenized PV. Their model also allows for wave breaking outside the surf zone, providing a more realistic circulation in surrounding regions, but emphasizes the same balance within the surf zone.

The surf zone constrains the total wave forcing only to the point that the parameters can be determined a priori. Inspection of (7) suggests that the dominant parameter is the half-width, as it regulates, to a large degree, the amplitude of the parabolic constraint wave driving. In Fig. 6 we show the total hemispheric wave driving estimated by (7), scaled by the model's climatology, as a function of the half-width, where different colors correspond to different damping time scales. A value of 1 corresponds to the true value of integrated wave driving. Each curve in the area integration presented in Fig. 6a corresponds, with the same color, to a curve in the vertical integration presented in Fig. 6b. It is clear that the conceptual model is highly sensitive to the surf-zone width and the time-scale parameter. Note that the red line uses the natural time scale in the model. In addition, notice that the surf-zone width can be estimated using Figs. 4b and 4c. For example, defining the half-width to be from the polar night jet maximum to the location of minimum meridional PV gradient, we estimated the half-width to be around 20°–30° latitude. It follows, according to Figs. 6a and 6b, that the conceptual model plausibly estimates the total wave driving.

The ability to get a plausible estimate for the total wave driving, based on just the gross properties of the surf zone, shows the advantage in PV thinking on the BDC. This leads us to suggest a slight modification to the standard paradigm for the BDC dynamics that explicitly considers the impact of wave driving on the potential vorticity. The key for the modified paradigm, as illustrated in Fig. 1b, is to consider the fact the Rossby waves mix the PV as an intermediate step to the conventional paradigm. This mixing constraint connects the total wave forcing to the surf-zone extent and background PV (bottom arrow on the left-hand side of Fig. 1b). For example, given the surf-zone width, the total wave forcing is just that needed to maintain a homogenized PV against restoring forces, as crudely estimated in (7). On the other hand, given the surf-zone PV structure, one can invert the PV and compute the zonal-mean-flow fields (top arrow on the right-hand side of Fig. 1b). This ensures that the zonal-wind field satisfies a positive refractive index for planetary wave propagation (Charney and Drazin 1961; Matsuno 1970), as the surf-zone model assumes available

RW for PV mixing at all times (bottom arrow on the right of Fig. 1b).

In the next section we use this model to suggest a new mechanism for stratospheric interactions between parameterized gravity waves and planetary Rossby waves, a mechanism that complements the time-mean stability constraint suggested by CGB13. The key is that (i) the modified view forces one to think about the PV impact sooner and (ii) does not differentiate which wave is mixing the PV.

3. Three mechanisms for interactions between resolved and unresolved waves

Gravity waves that originate from small-scale mountains, convection, and frontal instabilities cannot be captured in most AGCMs, at least at resolutions that permit long-term climate integrations, and need to be parameterized. CGB13 found that the torques produced by GW parameterizations are often sufficiently strong and/or narrow to drive the flow toward an unstable state if they are not compensated by the resolved flow. Thus, barring some other interactions, baroclinic instability can generate resolved waves to compensate the parameterized torque. We associate this potential mechanism with a stability constraint on the flow. But what if the GW break in a region with significant RW breaking? Here, we use the conceptual model of the surf zone to motivate a second interaction mechanism that can lead to compensation by the resolved waves.

Breaking gravity waves change the PV structure on stratification surfaces in a manner that is linked to the pseudomomentum content of the waves. This can be incorporated into our surf-zone paradigm. The PV impact of GW breaking was pursued in Bühler and McIntyre (2005), who argued that a conservation law holds for the sum of the waves' pseudomomentum and the vortex impulse based on the PV distribution. The upshot was a generalization of the usual dissipative pseudomomentum rule familiar from zonal-mean theory: breaking gravity waves result in a dipolar PV change on stratification surfaces such that the horizontal impulse of the new PV distribution is equal to (minus) the dissipated amount of horizontal pseudomomentum of the waves. The sum of horizontal pseudomomentum and of horizontal PV impulse is constant under wave breaking. This process is illustrated in Figs. 2e–h: the PV impact exerted by a retrograde force generated by gravity wave breaking in the surf zone is a local reduction of the PV gradient.

OGWs and planetary waves are generated approximately at the same location (large-scale orography is associated with smaller-scale orography), and both tend to break when their zero phase speeds match the zero background zonal flow. Thus, one can expect much OGW breaking in the same region where the planetary waves break: the surf zone. But, as suggested in the surf-zone conceptual model, the mean PV is being homogenized there. It follows that any PV change induced by the GW will result in less resolved wave breaking (or more, as the case may be) in order to maintain the well-mixed surf zone. Thus, there will be less planetary wave breaking if the GW tend to reduce the PV gradient or more if the GW tend to increase the PV gradient. More generally, Rossby waves will keep mixing the PV until there is no gradient in the surf zone. Therefore, for any perturbation to the PV gradient affected by gravity waves, the Rossby waves alter their behavior to keep the gradient flat.

This mechanism for compensation is illustrated in Figs. 7a and 7c. Figure 7a shows the background PV and the surf zone. A retrograde gravity wave forcing exerts a change on the PV structure that results in a negative perturbation to the north and positive perturbation to the south (dashed green). Thus, as shown in Fig. 7c, the Rossby waves (dashed blue) alter their behavior in order to keep the PV meridional gradient flat.

A third mechanism may occur when the GW breaking occurs outside, along the border of the surf zone. At the edges of the surf zone strong gradients in the PV inhibit mixing, as expressed in the "PV staircase model" (e.g., Dritschel and McIntyre 2008). Any perturbation to the wave driving here can modify the mixing properties of the Rossby wave breaking, potentially extending (and thereby amplifying the BDC) or contracting the wave breaking. Figure 7d illustrates a case where the resolved waves extend the surf zone, in response to GW perturbation on its flank (Fig. 7b). This is a highly nonlinear mechanism, with the potential for nonlocal effects, as discussed in further detail in section 5.

Compared to OGWs, NOGWs can more readily propagate outside of the surf zone, as they may have nonzero phase speed and are not restricted to the refractive index for planetary waves.[2] In addition, the broad phase-speed spectrum of NOGWs naturally provides much wider torques than the orographic ones, leading to vertical spread in their breaking levels, as opposed to the tendency for localized critical level ($\overline{u} = 0$) breaking with stationary OGWs. Thus, when the NOGWs break near the surf zone, they modify the surf-zone width in addition to the refractive index for planetary wave propagation, resulting in nonlinear interaction.

[2] The refractive index that governs GW propagation is described by the Taylor–Goldstein equation (e.g., Haynes 2003).

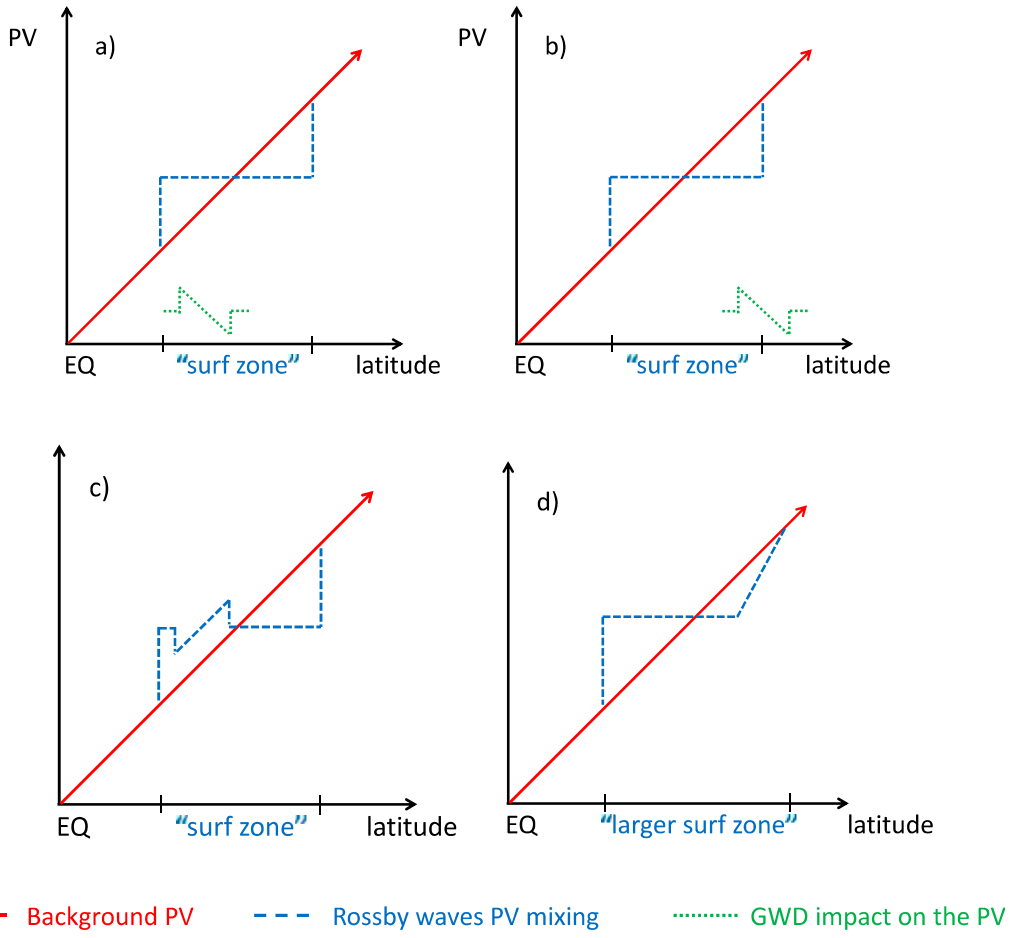

FIG. 7. Two mechanisms for interaction between GWD torques and the existing planetary waves in the flow. (a) The gravity wave breaking occurs inside the surf zone and results in a PV change (dotted green) that will flatten out on a fast time scale by rearrangement of the existing planetary waves, which result in (c) less PV mixing. Here, the resolved waves compensate the GWD perturbation. (b) When the breaking is outside/on the borders of the surf zone, an expansion of the surf-zone width may occur, (d) as the local mixing by the GW weakens the PV barrier, resulting in more PV mixing. Here, the GW perturbation may be amplified by the resolved waves.

Two key points differentiate these mechanisms. First, the "refractive-index interaction" provides a pathway for nonlocal changes in the resolved waves that can amplify the impact of the parameterized wave driving. In contrast, the "stability constraint" and "mixing constraint" are mechanisms for compensation, leading the resolved waves to counteract the parameterized wave driving. Second, the stability constraint differs in that it is largely independent of the existing wave forcing, while the other two mechanisms depend critically on the response of the resolved wave activity in the stratosphere. In particular, the relative importance of stability versus mixing depends on the time scale of the interaction: can the resolved waves effectively mix away the GW perturbation before the flow goes unstable?

In the next sections we use the idealized AGCM to illustrate the three mechanisms for compensation: instability, PV mixing, and refractive-index modification. The key difference between them is the torque's meridional location and time scale of the interaction. In regions where the PV mixing is weak, the flow has time to go unstable and the dominant constraint is stability. In the surf zone, we expect interaction through PV mixing to be fast, and the second mechanism will be most important. For broad torques on the boundaries of the surf zone, we expect the potential for nonlinear interaction.

4. PV mixing versus instability

We first establish two experiments to differentiate the instability and mixing mechanisms, creating two cases designed to favor one or the other. In one case we take an idealized time- and zonal-mean stratospheric torque that mimics the OGWD and place it in the summer stratosphere,

where there is no large-scale topography in the model and easterly winds limit wave propagation: hence there is little mixing and no surf zone. We compare this case with another in which the same torque is placed in the winter stratosphere, inside the surf zone. The torque $\overline{\mathbf{X}}$, as in CGB13, satisfies the following equation:

$$\overline{\mathbf{X}}(y,p) = \begin{cases} -\dfrac{A}{2}\{1+\cos[\pi(y-y_0)/L]\}, & \text{if} \quad |y-y_0| \le L, \quad p_1 \le p \le p_2, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $A = 2\times10^{-5}\,\mathrm{m\,s^{-2}}$, the total meridional width is $h = 2L = 10°$, centered around $y_0 = 35°\mathrm{N}$ with $p_1 = 3\,\mathrm{hPa}$ and $p_2 = 50\,\mathrm{hPa}$. This compactly supported wave forcing is shown by the black contours in Fig. 4c. The amplitude of the stratospheric torque resembles the amplitude of the OGWD that we get from the time- and zonal-mean OGW parameterization (see CGB13, their section 4, for more details). We chose a relatively narrow torque, so that we could place it completely within the surf zone.

CGB13 suggested that the torque's amplitude, meridional and vertical scales, and the mean PV state are sufficient to predict compensation. The parameters that we use for our torque are sufficient to drive instability in the flow (cf. CGB13, their Fig. 9), so we expect compensation in both the SH and NH cases. The key difference between the experiments will lie in whether mixing can compensate the torque before the flow realizes this instability.

a. Time-mean view

We consider two 20 000-day integrations with the torques, compared against a control integration of the same length. We start by exploring the time-mean compensation for both cases, defined in CGB13 as the scaled covariance between the perturbation and the response,

$$C = -\frac{2\sum_i [P(x_i)R(x_i)]}{\sum_i P^2(x_i) + \sum_i R^2(x_i)}, \tag{9}$$

where C is the degree of compensation, P denotes the perturbation [the torque applied as in (8)], R denotes the response of the resolved Rossby waves (the change in the EPFD), and x_i is a generic spatial coordinate. With this definition, if the response is equal and opposite to the perturbation (i.e., $R = -P$), we have perfect compensation and $C = 1$. If $R = 0$ or more generally is uncorrelated with the perturbation, then $C = 0$ (there is no compensation) and if $R = P$, then $C = -1$: the system amplifies the perturbation.

Computation of the compensation for the integration with the SH torque averaged over 20 000 days yields $C = 0.74$ with negligible standard deviation (less than 1% change in C). The standard deviation was computed by bootstrapping (reshuffling with repetition) the data at random, repeatedly, 200 times. The NH torque is applied inside the surf zone, where there is weaker background PV gradient. As expected, in the region of substantial PV mixing the compensation is higher, $C = 0.88$, with negligible standard deviation. In both cases the torque is well compensated by the resolved waves, albeit more strongly in the NH.

While compensation through instability would involve the local generation of Rossby waves, compensation through PV mixing would be associated with a change in the planetary wave propagation and breaking and, therefore, a rearrangement of the existing resolved wave fluxes. This is explored in Fig. 8, which shows the Eliassen–Palm flux budgets for the steady integration with the SH and NH torques. Similar to Edmon et al. (1980) and Kushner and Polvani (2004), we consider the rate of change in the total angular momentum within an annular ring for different sectors of the stratosphere in latitude and pressure, using the natural units of torque (N m). By Green's theorem, the net Eliassen–Palm flux divergence in the region must be balanced by the fluxes across the lateral, top, and bottom boundaries; that is,

$$\begin{aligned} \frac{2\pi a^2}{g}\int_{p_1}^{p_2}\int_{\phi_1}^{\phi_2}\cos\phi\nabla\cdot\mathbf{F}\,d\phi\,dp = {} & \frac{2\pi a\cos\phi}{g}\int_{p_1}^{p_2}F^{(\phi)}\,dp|_{\phi_2} - \frac{2\pi a\cos\phi}{g}\int_{p_1}^{p_2}F^{(\phi)}\,dp|_{\phi_1} \\ & + \frac{2\pi a^2}{g}\int_{\phi_1}^{\phi_2}\cos\phi F^{(p)}\,d\phi|_{p_2} - \frac{2\pi a^2}{g}\int_{\phi_1}^{\phi_2}\cos\phi F^{(p)}\,d\phi|_{p_1}, \end{aligned} \tag{10}$$

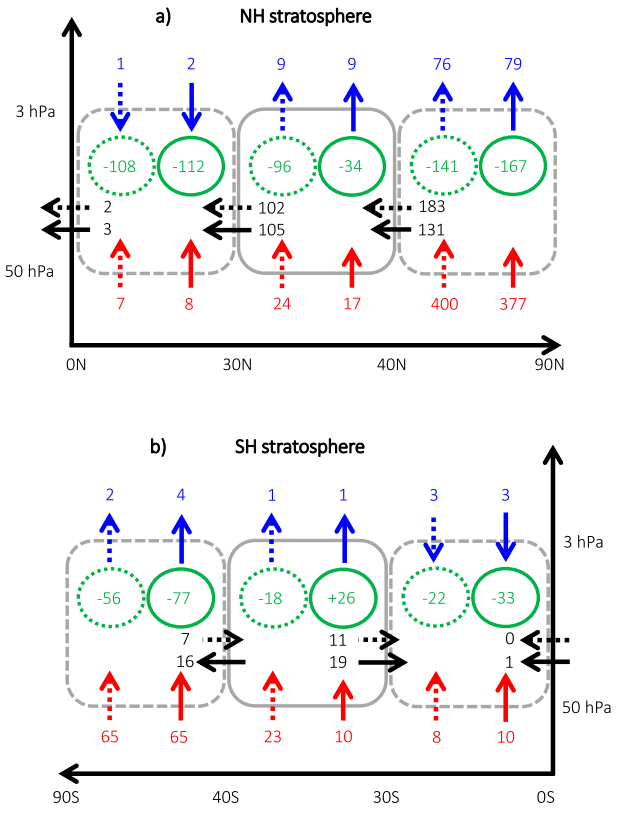

FIG. 8. Eliassen–Palm flux budgets ($\times 10^{16}$ N m) for the (a) NH and (b) SH torque experiments. Dashed arrows characterize the fluxes in the climatological (unperturbed) integrations and solid arrows characterize the integrations with the applied torques. An upward or equatorward arrow indicates a negative flux of momentum, so that the arrows illustrate the direction of wave propagation. The solid gray boxes mark the region surrounding the torque, and dashed gray boxes mark adjacent regions. Green numbers at the center of each box denote the net EPFD within the region, and a dashed (solid) circle denotes the unperturbed (applied torque) integrations. The net fluxes across each boundary are evaluated using the right-hand side of (10) on the circumference of each box, and the total integrated EPFD is evaluated as the sum of the fluxes across all sides.

where the flux is the vector $\mathbf{F} = (F^{(\phi)}, F^{(p)})$ comprising $F^{(\phi)} = a\cos\phi(-\overline{u'v'} + \overline{u}_p\overline{v'\theta'}/\overline{\theta}_p)$ and $F^{(p)} = a\cos\phi[(f + \overline{\xi})\overline{v'\theta'}/\overline{\theta}_p - \overline{u'\omega'}]$. Besides terms that were defined in section 2, ω is the vertical velocity, $\xi = -(u\cos\phi)_\phi/(a\cos\phi)$ is the relative vorticity, a is Earth's radius, and $\nabla \cdot \mathbf{F} = (F^{(\phi)}\cos\phi)_\phi/(a\cos\phi) + F_p^{(p)}$ is the divergence of the Eliassen–Palm flux (e.g., Andrews et al. 1987, chapter 3). In addition, p_1, p_2, ϕ_1, and ϕ_2 are the top, bottom, left, and right boundaries of the region, respectively. The integrals are evaluated over the region surrounding the applied torque (the solid gray box in Fig. 8) and adjacent regions, poleward and equatorward of the torque (dashed gray boxes).

Inspection of the NH Eliassen–Palm budgets in Fig. 8a shows that resolved waves respond to the torque by reducing the local EPFD. Compensation is not perfect; the applied torque of -76×10^{16} N m is associated with a 62×10^{16} N m reduction in the retrograde forcing of the planetary waves—that is, −96 before the torque but only −34 afterward (here and following, we omit the units, always 10^{16} N m, for the sake of brevity). This is achieved primarily by a reduction of the momentum flux from the poleward side (183 before, 131 afterward) and, to a lesser extent, less heat flux from below (24 before the torque, 17 after), while the equator and top sides exhibit only minor changes. Thus the reduction in the local PV mixing is associated with a large change of the flux from the pole side, the chief source of wave activity in the climatological integration. This poleward region exhibits an increase in wave driving as the resolved waves redistribute the torque (the net EPFD here increases from −141 to −167), and, most importantly, less net flux into the stratosphere from below (from 400 to 377): less wave activity enters the high-latitude stratosphere, consistent with the reflection hypothesis of a saturated surf zone. The budget analysis thus supports our hypothesis of surf-zone interaction through PV mixing.

We also note the similarity in the response of our idealized model with the middle- and high-latitude budget analyses presented at Sigmond and Scinocca (2010, their Fig. 6) and Sigmond and Shepherd (2014, their Fig. 3). These studies examined the response of resolved Rossby waves to perturbations of the OGW driving that fell largely in the middle of the surf zone. In addition to a compensating decrease in the resolved wave breaking in the surf zone, they observed an increase in the resolved wave breaking poleward of the OGW perturbation, associated with a decrease in the zonal winds. Comparison with Fig. 8 shows that this effect is quite similar to the response of our idealized model to a torque in the surf zone. Once PV gradients are flattened in the surf zone, excess Rossby wave activity will be reflected poleward (reducing the net meridional propagation). This leads to increased breaking on the poleward flank of the surf zone and, ultimately, less wave activity allowed into the stratosphere (reducing the net upward wave propagation).

The SH Eliassen–Palm budgets in Fig. 8b, on the other hand, tell a different story. Changes are limited to the torque region and its boundaries, with negligible changes to the fluxes entering or leaving the adjacent regions. The local EPFD changes sign in the torque region, from −18 to +26, indicating that Rossby waves are being locally generated to compensate for the applied torque. This leads to less heat flux from below, more momentum flux that leaves on the equatorward side, and, most importantly, momentum flux directed poleward, opposite to the climatological

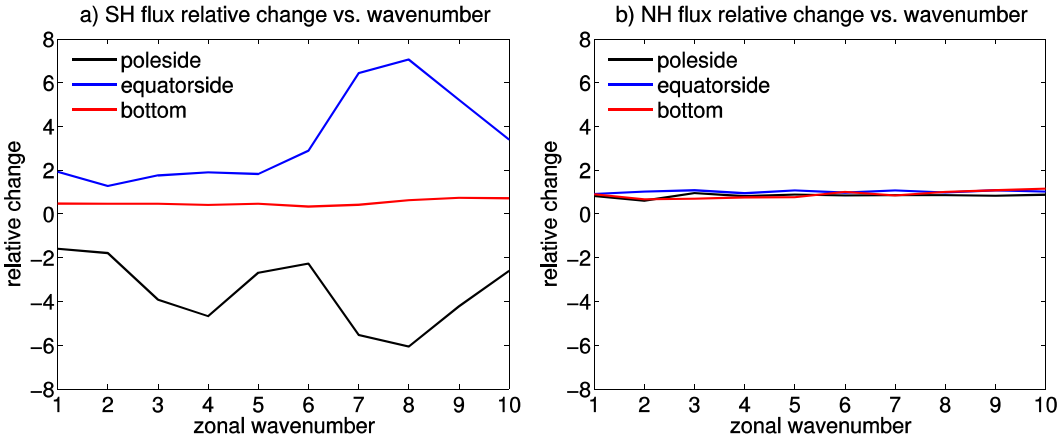

FIG. 9. The change in the mean fluxes encircling the (solid gray) boxes in Fig. 8, relative to their climatological values, as a function of the zonal wavenumber for (a) the SH and (b) the NH. The curves correspond to the different sides of the region, as noted by the labels. When the value is 1 (or 100%), it means that there is no change in the wavenumber contribution to the net flux. Huge relative change in the fluxes are observed in the SH case, whereas in the NH there is a uniform reduction of about 10%.

direction of wave propagation. The observation that the momentum flux changes direction from its climatological state is a clear sign for instability being involved in the compensation process.

Figure 9 complements this view with spectral analysis of the fluxes leaving/entering the torque region. For simplicity, we focus on the sides with the largest net-flux change and consider the first 10 wavenumbers; higher wavenumbers have negligible contribution to the flux. Figure 9 shows the relative change (compared to climatological values) in the wavenumber contribution to the fluxes encircling the torque region (solid gray boxes). It is apparent from Fig. 9a, which shows the analysis of the SH box, that except for the bottom side, there is an amplification of the wavenumber contribution, across all wavenumbers, with a change in the sign of the pole-side flux. The most significant amplification is observed with wavenumbers 7–9. The abundance of small wavenumbers is consistent with the instability hypothesis, as one would expect the zonal scale of the unstable waves to be related to the meridional scale of the jet perturbation. Thus, generation of new waves is playing a role in the compensation process. Figure 9b, which shows the analysis of the NH box, is markedly different. The wavenumber contribution of the surf-zone fluxes does not change relative to their climatological values. This is consistent with the hypothesis that less mixing is needed by the available planetary waves to compensate for the applied torque. Close examination shows a net reduction of about 10% in all wavenumbers at the pole and bottom sides of box.

As will be shown below, consistent with the instability mechanism, a region of negative PV gradient is observed in the SH case (Fig. 10c, blue contours). However, the bottom region of the NH case also exhibits a weak reversal of PV gradient (Fig. 10d, blue contours), a necessary but not sufficient condition for instability. Note, however, that the PV reversal in the NH is about 10 times weaker than in the SH. To ensure that PV mixing is sufficient to explain the compensation, we further explore the transient response to the applied torques for evidence of instability (or lack thereof). In addition, real GW events are transient in nature, and considering only the steady response might obscure the physical nature of the interaction. Thus, in the following we study the transient response to the applied torque, using an ensemble of switch-on torque experiments.

b. *Temporal structure*

We run an ensemble of 1000 integrations, each for 50 days, initialized from a long control run. The initial condition for each ensemble member is taken directly from the control run, but each experiment is separated by 50 days to ensure statistical independence. At day 0 of each experiment, the applied torque is switched on and remains steady. We start by examining the SH region. Figure 11a shows that the ensemble-mean zonal wind is accelerating, so that, after about 8–9 days, the ensemble-mean PV gradient changes sign, thus satisfying the necessary condition for instability. At the same instant, the ensemble-mean EPFD changes sign. Figure 11c shows that the ensemble-mean compensation (solid blue) increases exponentially, mirroring the perturbation to the winds. After 40–50 days the compensation saturates to its steady value (dashed black).

We compare these results with the integration in which the torque is applied in the NH stratospheric surf

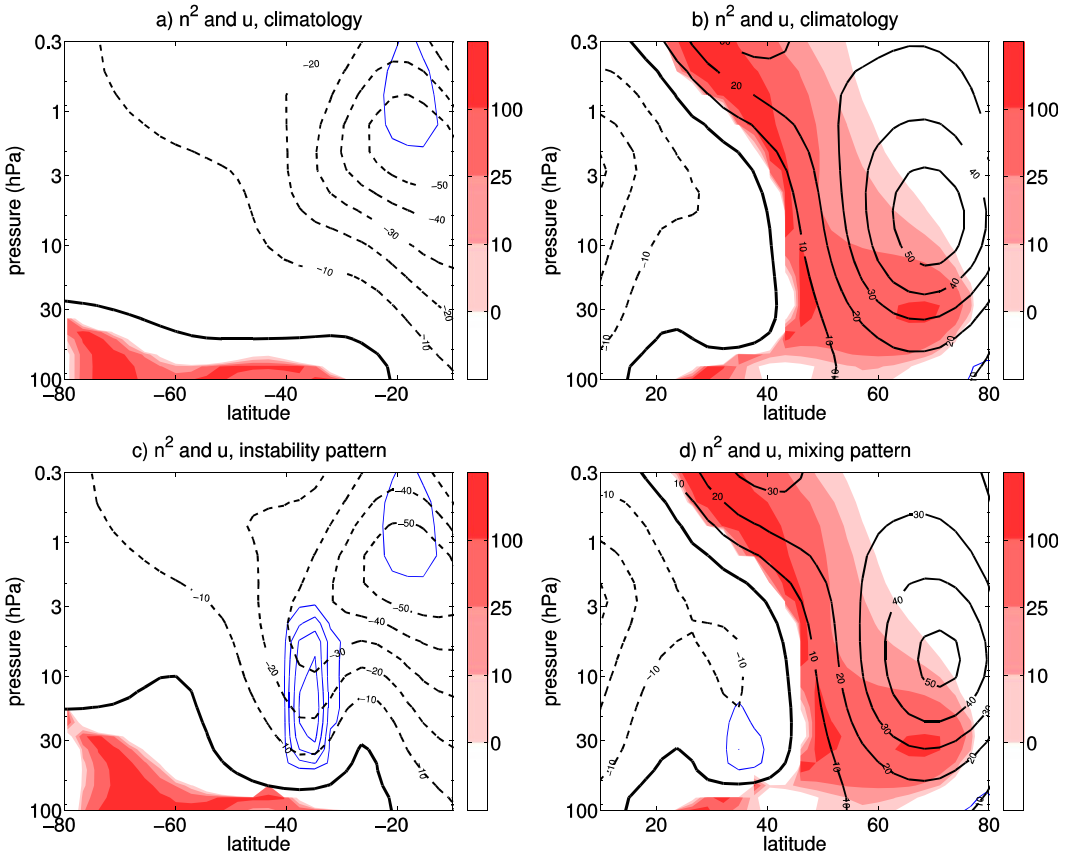

FIG. 10. The dimensionless refractive index for wavenumber 2 (red shading) and the zonal-mean zonal wind (black contours; solid lines for positive values and dashed lines for negative values). The refractive index is computed as in Simpson et al. [2009, their (4)]. (a),(b) Climatological values and (c),(d) ensemble-mean values, averaged over days 40–50, with (a),(c) SH and (b),(d) NH torques. The blue contours show the negative PV gradient (−2.5, −2, −1.5, −1, −0.5, −0.1, scaled by the value of $a\beta_0$ at 45°N).

zone. Figure 11b shows that the ensemble-mean PV gradient does not change sign during the first 50 days: the necessary condition for instability is not satisfied. Throughout the 50-day period the zonal winds slowly respond to the torque (orange line in Fig. 11b), but the ensemble-mean compensation (blue line in Fig. 11d) increases rapidly over the first 5 days and then saturates to its steady value (dashed black). This 5-day time scale is consistent with the life cycle for planetary wave breaking illustrated in Ueyama et al. (2013). The key difference between the SH and NH cases is that the time scale for the interaction is much faster for the torque in the surf zone and much slower for the SH torque, which is consistent with faster times for PV mixing and slower times for instability.

A complementary view on the different mechanisms can be obtained by analyzing the changes in the refractive index in both scenarios (Matsuno 1970). The refractive index indicates the affinity for planetary wave propagation, such that waves can propagate when the refractive index is positive. In addition, the planetary wave amplitude gets larger for higher refractive-index values; thus, planetary wave rays have a tendency to propagate up the refractive-index gradient (Karoly and Hoskins 1982). Figures 10a and 10b show the climatological refractive index (shaded red) and the time- and zonal-mean zonal wind (black contours) for the Southern and Northern Hemispheres. It is clear that planetary waves are allowed to propagate upward along the weakly positive zonal wind in the winter hemisphere but are prohibited from doing so in the summer hemisphere. Figures 10c and 10d show the ensemble-mean states for both the Southern and Northern Hemispheres, averaged over the days in which compensation has saturated—thus, over days 40–50 (Figs. 11c,d). Whereas for the ensemble mean with the SH torque, the zero-wind line was shifted upward: there is no change in the ensemble-mean zero-wind line with the surf-zone torque. In both cases, however, compensation has reached saturation by this time.

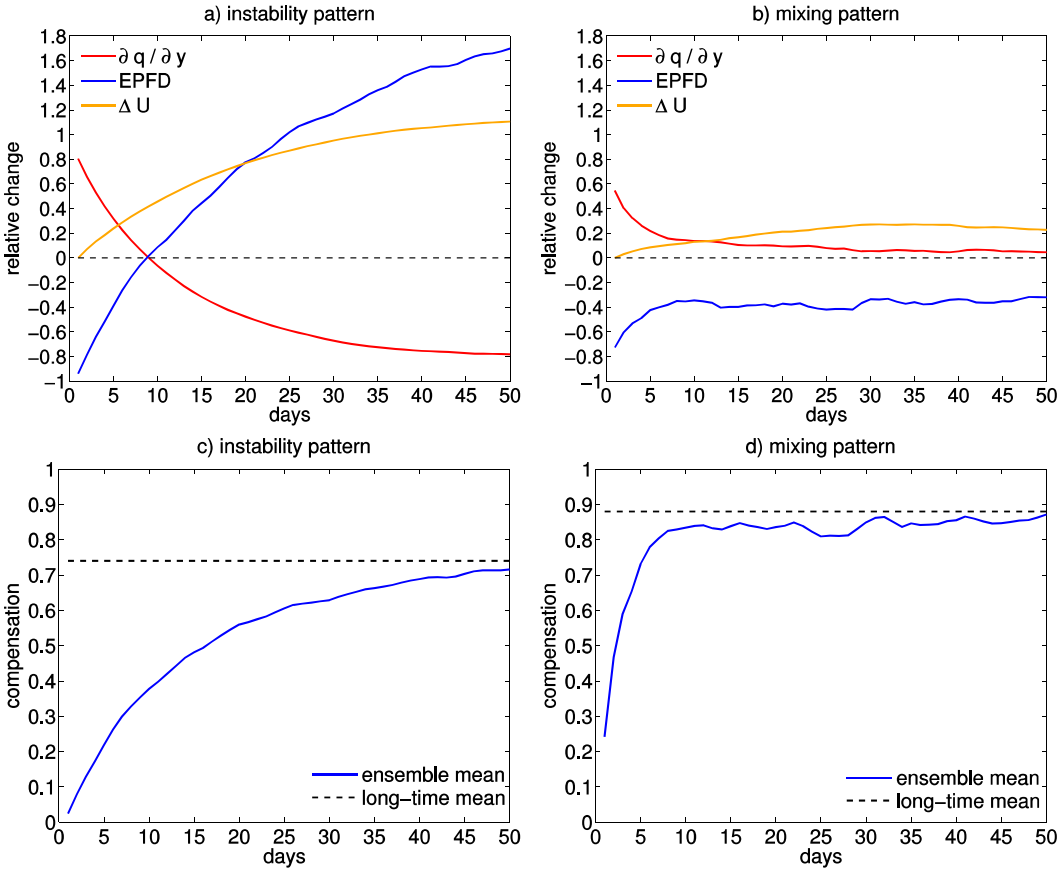

FIG. 11. The ensemble-mean transient response of the flow to the torque placed in (a),(c) the SH stratosphere and (b),(d) the NH stratospheric surf zone, evaluated at the region surrounding the torque. (a),(b) The ensemble-mean meridional derivative of the PV (red), the ensemble-mean EPFD (blue), and the ensemble-mean change in the maximum zonal wind (green), all scaled by their climatological values, as a function of time (the zonal wind change is additionally scaled by 10 for plotting purposes). (c),(d) The ensemble-mean (blue) and long-time-mean (dashed black) compensation as a function of time. The standard deviation was verified to be negligible using the bootstrap method.

5. Amplification of the BDC by GW driving on the edges of the surf zone

To study the third mechanism for interaction through refractive-index modification, we need to generate an idealized torque that is weak and broad enough not to satisfy the condition for instability. In addition, the torque must be applied outside the surf zone, to avoid fast interaction through PV mixing, but near enough to be able to interact with the planetary waves. We found, in practice, that it is very hard to create such a torque. As shown in CGB13, the impact of a torque on the PV is very sensitive to the second and fourth meridional derivative of the implied torque. Thus, any localized torque is likely to generate PV reversals, at least at its edges.

NOGW parameterization is designed to account for gravity waves that originate from a variety of sources. These include moist convection, spontaneous emission from fronts and jets, and instabilities. The precise nature of the wave emission from these sources is largely uncertain. Nevertheless, all these sources have a broad spectrum; they emit waves with a wide variety of wavelengths and frequencies. Thus, these spectral waves have multiple critical layers and are not restricted to break at the surf zone. In addition, the NOGW parameterization is interactive, and so even for large-amplitude breaking, rather than drive potentially unstable flows, the scheme adjusts the breaking level as the zonal wind begins to change. This makes the NOGWD intrinsically broad and weak, less likely to generate instability, and therefore ideal for the purpose of studying this mechanism.

For this experiment, we compare two configurations of our model with different NOGW parameterizations. An integration with the Alexander and Dunkerton (1999) NOGW parameterization (the control integration used in this study) is compared against an integration in which NOGWs are parameterized by a Rayleigh drag, as in the Polvani and Kushner (2002) model. (For reference, these

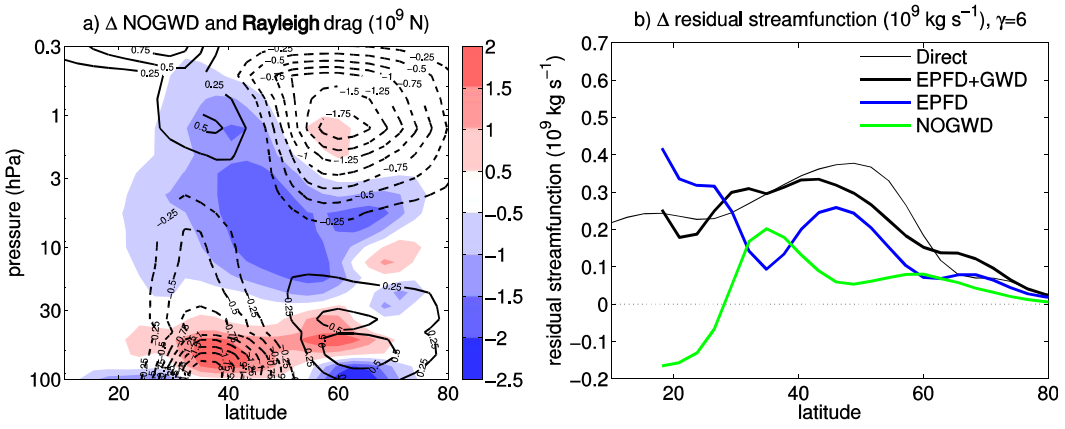

FIG. 12. Analysis of a noncompensating integration, where the perturbation is defined to be the difference between the NOGWD and the Rayleigh drag. (a) The perturbation (contoured) and the EPFD response (shaded) and (b) the change in the streamfunction at 70 hPa, broken down into components using downward control. There is no OGW in these integrations.

two integrations were also compared in CGB13, their sections 2 and 5.) Figure 12a shows the perturbation [the difference between the Alexander and Dunkerton (1999) NOGW torque and the Rayleigh friction torque] with the black contours and the response (the change in resolved wave driving) with shading. The change in the resolved waves is almost orthogonal in space to the perturbation in the parameterized wave driving and therefore is not effective in compensating for the perturbation (the compensation measure is low; $C = 0.21 \pm 0.04$). The orthogonal changes in the resolved waves, however, amplify the perturbation's impact on the residual circulation, as can be seen in the changes to the 70-hPa residual streamfunction in Fig. 12b. The change in circulation associated with the resolved waves (blue) has the same sign as the change induced by the parameterized wave driving (green). Thus, the response of the flow is to amplify the perturbation, as seen in the total change of the residual streamfunction (black). The impact of the increased wave driving by the NOGWs was more than doubled by the associated increase in resolved wave driving.

We now focus on the upper-stratospheric perturbation in Fig. 12a. Note that the meridional extent of the negative perturbation is on the order of 35° and it is about 10^{-6} m s^{-2} in amplitude. Thus the negative perturbation gets compensated only weakly, as it does not satisfy the condition for compensation through instability and is acting outside of the surf zone. Its direct effect on the mean flow is to weaken the polar vortex. This zonal wind weakening allows more upward planetary wave propagation according to Charney–Drazin/Matsuno condition. At the same time, a positive NOGW perturbation centered around 35°N reverse easterlies that exist in the Polvani and Kushner (2002) configuration to westerlies, allowing the planetary waves to propagate farther equatorward. Both of these changes extend the range where planetary waves can propagate, thus enlarging the surf zone.

This process can be seen in the red-shaded refractive-index field in Figs. 13a and 13b. Examination of the change in the PV gradient in Fig. 13c indeed shows a change that supports the changes in the refractive index in Figs. 13a and 13b. In that sense the NOGW positive and negative perturbations act in a joint manner to direct the waveguide of the planetary wave equatorward, where now there is more place for PV mixing. This extension of the surf zone is reflected in the broad increase in the planetary wave driving shown in Fig. 12a, which amplifies the total BDC.

6. Summary and discussion

The BDC is a wave-driven circulation dominated by Rossby waves, and, to a lesser extent, by gravity waves. However, it is nontrivial to quantitatively relate the BDC to its wave contributors because of strong interactions between the wave contributors, as discussed by CGB13 and Sigmond and Shepherd (2014). We have identified three mechanisms for the interaction between parameterized gravity waves and resolved Rossby waves. The first, a stability constraint, applies when the GWD drives the stratospheric flow to go unstable. This mechanism is likely for large-amplitude and meridionally narrow torques and is expected to dominate outside of the surf zone or high in the stratosphere or mesosphere. Observed OGWs are intermittent in nature tending to lead to short, extremely strong bursts of wave forcing (i.e., Hertzog et al. 2012), which may make the potential for instability more likely.

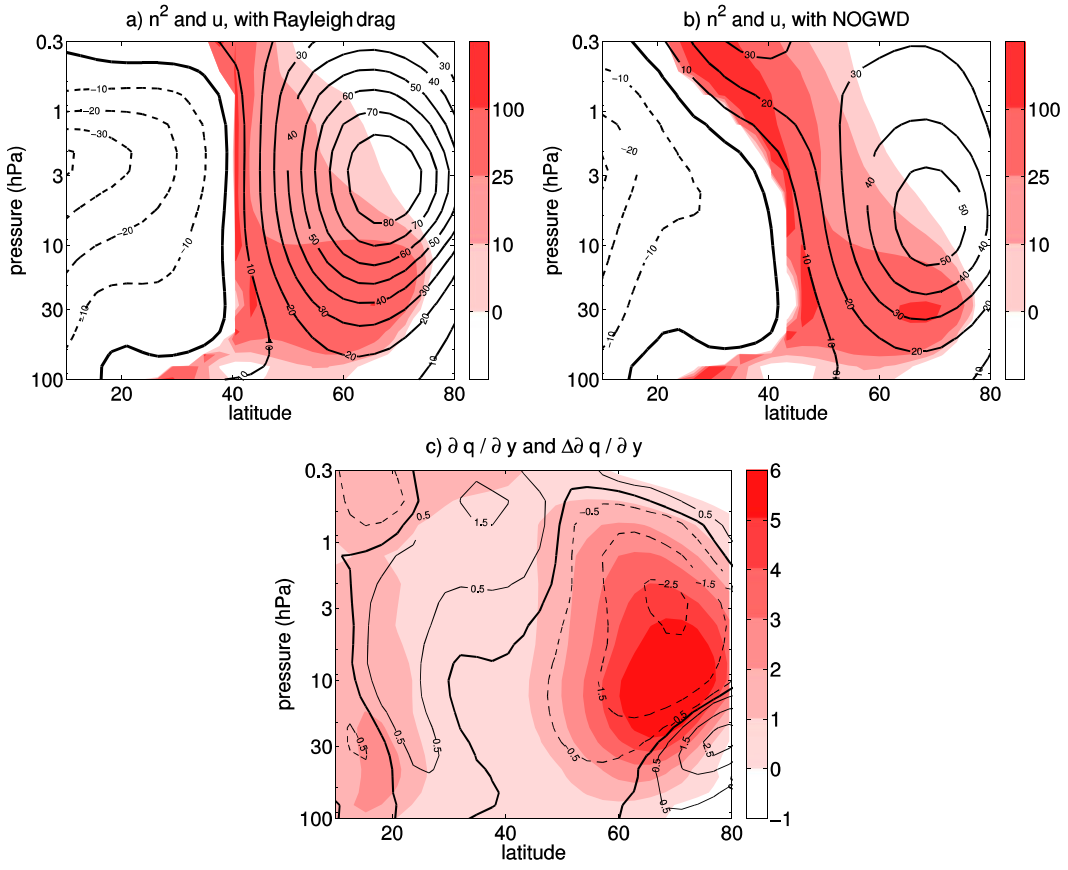

FIG. 13. The dimensionless refractive index for wavenumber 2 (red shading) and the zonal-mean zonal wind (contoured) for the integration with (a) the Rayleigh drag and (b) the NOGW. (c) The PV meridional gradient in the integration with the Rayleigh drag (red shading) and the change therefrom in the NOGW integration (contoured).

The second mechanism, a mixing constraint, was drawn from the conceptual surf-zone model and based on the fact that large-scale Rossby waves mix PV. It is likely to apply for GWD in the winter midlatitude lower stratosphere, where any PV change induced by the GWD is mixed away by the existing planetary wave fluxes, which flatten the PV surfaces in the surf zone. OGWs are stationary and so tend to break at zero-wind critical layers where the planetary waves tend to break. Hence, compensation is likely for these waves. The key difference between this and the instability mechanism is the torques' meridional location and time scale for interaction. In regions where the PV mixing is weak, the flow has time to go unstable and the dominant constraint is stability. In regions where the PV mixing is strong, we generally expect interaction through PV mixing to be fast and so the flow never becomes unstable.

Sigmond and Scinocca (2010) and Sigmond and Shepherd (2014) show results from a comprehensive atmospheric model that are consistent with this efficient PV mixing mechanism. In particular, Sigmond and Shepherd (2014, their Figs. 4b,d) show that changes in the strength of the stratospheric OGWs have almost no effect on the total circulation between the turnaround latitude and 52°N (the region at the heart of the surf zone), both for the current conditions and for a $2 \times CO_2$ atmosphere.

The third mechanism, a refractive-index interaction, is likely when the perturbation is broad and weak and therefore unlikely to drive instability. In this case, the torque can directly change the mean flow and therefore alter the refractive properties of planetary waves. Changes in the refractive index can guide the planetary waves away from the perturbation, leading to nonlocal interactions. This mechanism is more likely for GW perturbations near, but outside, the region of planetary wave breaking. It must be outside the surf zone to avoid fast interaction through PV mixing but close enough to be able to interact with the planetary waves. An example of this nonlinear interaction mechanism from a comprehensive AGCM can be seen in Manzini and McFarlane (1998, their Figs. 13 and 16). They found that perturbations to the

parameterized GW driving near the top of the surf zone, in the vicinity of the stratopause, led to amplification of the resolved wave driving in the mesosphere.

These mechanisms are not mutually exclusive, and it is likely for multiple mechanisms to act in concert. It is largely a matter of the torque's location and the interaction time scale that determines the dominant mechanism. While the first two mechanisms involve strong and local compensating interactions, the third is not necessarily compensating, and nonlocal changes in the resolved waves can, in the end, amplify the net effect of the initial perturbation in the parameterized wave driving. It is important to note, however, that the instability and mixing mechanisms also involve nonlocal effects, as highlighted in the discussion of the idealized torque experiments in section 5.

The potential for interactions between planetary and gravity waves (the latter of which are generally parameterized) suggests a nuanced answer to the following question: what drives the BDC? The conventional paradigm, summarized in Fig. 1a, suggests that, given the wave forcing, it is possible to deduce the zonal-mean-flow fields using mean-flow dynamics through downward control (Haynes et al. 1991). At the same time, changes in the mean flow couple through linear wave dynamics to the wave forcing itself, as expressed by the refractive index (Charney and Drazin 1961; Matsuno 1970). Nevertheless, it is common practice to neglect the coupling in the system by quantifying the wave-driven BDC only through the downward-control branch (top arrow in Fig. 1a). For example, the linear-decomposition view considers the total wave driving, or alternatively the total residual streamfunction, as the sum of its contributions from the resolved Rossby wave driving and the unresolved parameterized orographic and nonorographic gravity wave driving (e.g., Eyring et al. 2010; Butchart et al. 2010, 2011). This view, however, does not consider the system as coupled and could leave a misleading impression on the relative importance of the different wave types.

We argue that the question of what drives the BDC cannot be answered using downward control alone. This is because strong interactions between the wave components are possible for the same total wave forcing. One could evaluate the importance of different wave types by considering the impact of their absence; for instance, how the does the circulation change when the GW parameterizations are turned off? This may not be possible in practice, however, as GWD plays a critical role in stabilizing the polar vortex in the upper stratosphere. In this respect, we believe that the linear-decomposition approach may tend to overestimate the importance of OGWs in the BDC and underestimate that of NOGWs. This is because OGWs tend to break at the surf zone, thus their effect is likely to get canceled by the resolved waves, while NOGWs exhibit nonlocal interactions with the resolved waves, effects that are essential but hard to appreciate in a naive, linear interpretation of downward control.

In light of these interactions, we suggest a modified approach, which explicitly considers the impact of wave driving on the potential vorticity of the stratosphere. As summarized in Fig. 1b, it is helpful to consider the coupled system as one, using a surf-zone PV perspective. In the midlatitude stratosphere where planetary wave breaking is strong, perturbations to the PV gradient affected by breaking gravity waves will tend to be flattened by the planetary waves and, therefore they do not really affect the total wave driving. The total wave forcing is set largely by the surf-zone width. In terms of evaluating the role of GWD, this suggests that their importance is primarily in how they help shape the width and depth of the surf zone. While this approach blurs the roles of Rossby and gravity waves, it provides more intuition into how perturbations to each component impact the circulation as a whole.

Shepherd and McLandress (2011) have shown that the response of the BDC to global warming can be understood in terms of the change in the critical layer for wave breaking associated with the upward expansion of the subtropical jets. This suggests an upward expansion of the stratospheric surf zone, affecting the parameterized OGWs and resolved planetary waves alike. Hence a robust increase in the BDC is expected, independent of the role of OGW and resolved planetary waves in the current climatology. This has been verified by Sigmond and Shepherd (2014), who find that the relative role of the RWs and GWs in driving the response of the BDC to $2 \times CO_2$ varies depending on the basic state.

On the practical side, we urge modeling groups to report the latitude–pressure structure of the parameterized and resolved wave forcing throughout the whole stratosphere. We appreciate the burden of maintaining increasingly complex Earth system models, but this 2D structure of the wave driving is essential for untangling the roles of gravity and resolved waves in driving the BDC.

Acknowledgments. We thank M. Joan Alexander, Olivier M. Pauluis, Alan Plumb, Tiffany A. Shaw, Michael Sigmond, and an anonymous reviewer for inspiring discussions and critical comments on earlier versions of this manuscript. This research was supported in part by Grants AGS-0938325 and AGS-1264195 from the National Science Foundation to New York University.

REFERENCES

Alexander, M. J., and T. J. Dunkerton, 1999: A spectral parameterization of mean-flow forcing due to breaking gravity waves. *J. Atmos. Sci.,* **56,** 4167–4182, doi:10.1175/1520-0469(1999)056<4167:ASPOMF>2.0.CO;2.

——, and Coauthors, 2010: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Quart. J. Roy. Meteor. Soc.,* **136,** 1103–1124, doi:10.1002/qj.637.

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, doi:10.1175/1520-0469(1976)033<2031:PWIHAV>2.0.CO;2.

——, J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics.* International Geophysics Series, Vol. 40, Academic Press, 489 pp.

Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science,* **294,** 581–584, doi:10.1126/science.1063315.

Brewer, A. W., 1949: Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteor. Soc.,* **75,** 351–363, doi:10.1002/qj.49707532603.

Bühler, O., 2014: *Waves and Mean Flows.* 2nd ed. Cambridge University Press, 360 pp.

——, and E. M. McIntyre, 2005: Wave capture and wave-vortex duality. *J. Fluid Mech.,* **534,** 67–95, doi:10.1017/S0022112005004374.

Butchart, N., and Coauthors, 2006: Simulations of anthropogenic change in the strength of the Brewer–Dobson circulation. *Climate Dyn.,* **27,** 727–741, doi:10.1007/s00382-006-0162-4.

——, and Coauthors, 2010: Chemistry–climate model simulations of twenty-first century stratospheric climate and circulation changes. *J. Climate,* **23,** 5349–5374, doi:10.1175/2010JCLI3404.1.

——, and Coauthors, 2011: Multimodel climate and variability of the stratosphere. *J. Geophys. Res.,* **116,** D05102, doi:10.1029/2010JD014995.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.,* **66,** 83–109, doi:10.1029/JZ066i001p00083.

Cohen, N. Y., E. P. Gerber, and O. Bühler, 2013: Compensation between resolved and unresolved wave driving in the stratosphere: Implications for downward control. *J. Atmos. Sci.,* **70,** 3780–3798, doi:10.1175/JAS-D-12-0346.1.

Dobson, G. M. B., 1956: Origin and distribution of the polyatomic molecules in the atmosphere. *Proc. Roy. Soc. London,* **236A,** 187–193, doi:10.1098/rspa.1956.0127.

——, D. N. Harrison, and J. Lawrence, 1929: Measurements of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions. Part III. *Proc. Roy. Soc. London,* **122A,** 456–486, doi:10.1098/rspa.1929.0034.

Dritschel, D. G., and M. E. McIntyre, 2008: Multiple jets as PV staircases: The Phillips effect and the resilience of eddy-transport barriers. *J. Atmos. Sci.,* **65,** 855–874, doi:10.1175/2007JAS2227.1.

Dunkerton, T., 1978: On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci.,* **35,** 2325–2333, doi:10.1175/1520-0469(1978)035<2325:OTMMMM>2.0.CO;2.

Edmon, H. J., Jr., B. J. Hoskins, and M. E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. *J. Atmos. Sci.,* **37,** 2600–2616, doi:10.1175/1520-0469(1980)037<2600:EPCSFT>2.0.CO;2.

Engel, A., and Coauthors, 2009: Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nat. Geosci.,* **2,** 28–31, doi:10.1038/ngeo388.

Eyring, V., T. G. Shepherd, and D. W. Waugh, Eds., 2010: Chemistry-climate model validation. SPARC Rep. 5, WCRP-30, WMO/TD-40, 426 pp. [Available online at http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/.]

Fritts, D. C., 1984: Gravity wave saturation in the middle atmosphere: A review of theory and observations. *Rev. Geophys.,* **22,** 275–308, doi:10.1029/RG022i003p00275.

——, and M. J. Alexander, 2003: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.,* **41,** 1003, doi:10.1029/2001RG000106.

Garcia, R. R., and W. J. Randel, 2008: Acceleration of the Brewer–Dobson circulation due to increases in greenhouse gases. *J. Atmos. Sci.,* **65,** 2731–2739, doi:10.1175/2008JAS2712.1.

——, ——, and D. E. Kinnison, 2011: On the determination of age of air trends from atmospheric trace species. *J. Atmos. Sci.,* **68,** 139–154, doi:10.1175/2010JAS3527.1.

Gerber, E. P., and L. M. Polvani, 2009: Stratosphere–troposphere coupling in a relatively simple AGCM: The importance of stratospheric variability. *J. Climate,* **22,** 1920–1933, doi:10.1175/2008JCLI2548.1.

Haynes, P. H., 2003: Critical layers. *Encyclopedia of Atmospheric Sciences,* J. R. Holton, J. A. Pyle, and J. A. Curry, Eds., Academic Press, 582–589.

——, M. E. McIntyre, T. G. Shepherd, C. J. Marks, and K. P. Shine, 1991: On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.,* **48,** 651–678, doi:10.1175/1520-0469(1991)048<0651:OTCOED>2.0.CO;2.

Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.,* **75,** 1825–1830, doi:10.1175/1520-0477(1994)075<1825:APFTIO>2.0.CO;2.

Hertzog, A., M. J. Alexander, and R. Plougonven, 2012: On the intermittency of gravity wave momentum flux in the stratosphere. *J. Atmos. Sci.,* **69,** 3433–3448, doi:10.1175/JAS-D-12-09.1.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, and L. Pfister, 1995: Stratosphere-troposphere exchange. *Rev. Geophys.,* **33,** 403–439, doi:10.1029/95RG02097.

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

Kawatani, Y., and K. Hamilton, 2013: Weakened stratospheric quasibiennial oscillation driven by increased tropical mean upwelling. *Nature,* **497,** 478–481, doi:10.1038/nature12140; Corrigendum, **501,** 578.

Khosrawi, F., and Coauthors, 2013: Assessment of the interannual variability and influence of the QBO and upwelling on tracer–tracer distributions of N_2O and O_3 in the tropical lower stratosphere. *Atmos. Chem. Phys.,* **13,** 3619–3641, doi:10.5194/acp-13-3619-2013.

Killworth, P. D., and M. E. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect, or over-reflect? *J. Fluid Mech.,* **161,** 449–492, doi:10.1017/S0022112085003019.

Kushner, P. J., and L. M. Polvani, 2004: Stratosphere–troposphere coupling in a relatively simple AGCM: The role of eddies. *J. Climate,* **17,** 629–639, doi:10.1175/1520-0442(2004)017<0629:SCIARS>2.0.CO;2.

Li, F., J. Austin, and J. Wilson, 2008: The strength of the Brewer–Dobson circulation in a changing climate: Coupled chemistry–climate model simulations. *J. Climate,* **21,** 40–57, doi:10.1175/2007JCLI1663.1.

Manzini, E., and N. A. McFarlane, 1998: The effect of varying the source spectrum of a gravity wave parameterization in a middle atmosphere general circulation model. *J. Geophys. Res.,* **103,** 31 523–31 539, doi:10.1029/98JD02274.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.,* **27,** 871–883, doi:10.1175/1520-0469(1970)027<0871:VPOSPW>2.0.CO;2.

McIntyre, M. E., and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature,* **305,** 593–600, doi:10.1038/305593a0.

——, and ——, 1985: A note on the general concept of wave breaking for Rossby and gravity waves. *Pure Appl. Geophys.,* **123,** 964–975, doi:10.1007/BF00876984.

McLandress, C., and N. A. McFarlane, 1993: Interactions between orographic gravity wave drag and forced stationary planetary waves in the winter Northern Hemisphere middle atmosphere. *J. Atmos. Sci.,* **50,** 1966–1990, doi:10.1175/1520-0469(1993)050<1966:IBOGWD>2.0.CO;2.

——, and T. G. Shepherd, 2009: Simulated anthropogenic changes in the Brewer–Dobson circulation, including its extension to high latitudes. *J. Climate,* **22,** 1516–1540, doi:10.1175/2008JCLI2679.1.

——, ——, S. Polavarapu, and S. R. Beagley, 2012: Is missing orographic gravity wave drag near 60°S the cause of the stratospheric zonal wind biases in chemistry–climate models? *J. Atmos. Sci.,* **69,** 802–818, doi:10.1175/JAS-D-11-0159.1.

Polvani, L. M., and P. J. Kushner, 2002: Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophys. Res. Lett.,* **29,** 1114, doi:10.1029/2001GL014284.

Scott, R., and Y. Liu, 2014: On the formation and maintenance of the stratospheric surf zone as inferred from the zonally averaged potential vorticity distribution. *Quart. J. Roy. Meteor. Soc.,* doi:10.1002/qj.2377, in press.

Shepherd, T. G., and C. McLandress, 2011: A robust mechanism for strengthening of the Brewer–Dobson circulation in response to climate change: Critical-layer control of subtropical wave breaking. *J. Atmos. Sci.,* **68,** 784–797, doi:10.1175/2010JAS3608.1.

Sigmond, M., and J. F. Scinocca, 2010: The influence of the basic state on the Northern Hemisphere circulation response to climate change. *J. Climate,* **23,** 1434–1446, doi:10.1175/2009JCLI3167.1.

——, and T. G. Shepherd, 2014: Compensation between resolved wave driving and parameterized orographic gravity-wave driving of the Brewer–Dobson circulation and its response to climate change. *J. Climate,* **27,** 5601–5610, doi:http://dx.doi.org/10.1175/JCLI-D-13-00644.1.

Simpson, I. R., M. Blackburn, and J. D. Haigh, 2009: The role of eddies in driving the tropospheric response to stratospheric heating perturbations. *J. Atmos. Sci.,* **66,** 1347–1365, doi:10.1175/2008JAS2758.1.

Solomon, S., K. H. Rosenlof, R. W. Portmann, J. S. Daniel, S. M. Davis, T. J. Sanford, and G.-K. Plattner, 2010: Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science,* **327,** 1219–1223, doi:10.1126/science.1182488.

Thompson, D. W. J., S. Solomon, P. J. Kushner, M. H. England, K. M. Grise, and D. J. Karoly, 2011: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.,* **4,** 741–749, doi:10.1038/ngeo1296.

Ueyama, R., E. P. Gerber, J. M. Wallace, and D. M. W. Frierson, 2013: The role of high-latitude waves in the intraseasonal to seasonal variability of tropical upwelling in the Brewer–Dobson circulation. *J. Atmos. Sci.,* **70,** 1631–1648, doi:10.1175/JAS-D-12-0174.1.