How Does the Quasi-Biennial Oscillation Affect the Stratospheric Polar Vortex?

PETER A. G. WATSON
Department of Physics, University of Oxford, Oxford, United Kingdom

LESLEY J. GRAY
Department of Physics, and National Centre for Atmospheric Sciences, University of Oxford, Oxford, United Kingdom

(Manuscript received 22 March 2013, in final form 2 September 2013)

ABSTRACT

The stratospheric polar vortex is weaker in the easterly phase of the quasi-biennial oscillation (QBO-E) than in the westerly phase (QBO-W), but the mechanism behind the QBO’s influence is not well understood. The composite difference of the atmospheric state between QBO-E and QBO-W is found to closely resemble the structure of the northern annular mode, the leading empirical orthogonal function of stratospheric variability, including its wave components. Studies of dynamical systems indicate that many different forcings could give rise to this response, and therefore this composite difference does not provide much information about the forcing mechanism. It is argued that the full transient response of a system to an applied forcing is likely to be much more informative about the dynamics of the forcing mechanism, especially the response on time scales shorter than the dynamical time scale, which is about a week for vortex variability. It is shown that the transient response of the vortex to forcing by the QBO in a general circulation model is consistent with the proposed mechanism of Holton and Tan, indicating that this mechanism has a role in the QBO modulation of vortex strength, in contrast to the conclusions of several recent studies. This novel approach of examining the transient response to a forcing on short time scales may be useful in various other outstanding problems.

1. Introduction

The most prominent feature of the wintertime polar stratosphere is the westerly vortex that forms around the pole. The Northern Hemisphere (NH) winter vortex is more variable than its Southern Hemisphere counterpart, with midwinter breakdowns of the vortex, known as major stratospheric sudden warmings (SSWs), happening about six times per decade on average (Charlton and Polvani 2007). As well as being theoretically interesting, understanding how external factors influence, this variability may help improve seasonal forecasts of the NH troposphere as it has become realized that weakenings of the vortex give rise to a more negative tropospheric northern annular mode (NAM) (e.g., Baldwin and Dunkerton 1999, 2001; Jung and Barkmeijer 2006).

Holton and Tan (1980) showed that the vortex is influenced by the quasi-biennial oscillation (QBO) (Baldwin et al. 2001; Gray 2010; Anstey and Shepherd 2013). The QBO is a phenomenon that dominates variability in the equatorial lower stratosphere whereby the zonal-mean zonal wind (ZMZW) direction on a given pressure level alternates between being easterly and westerly, with the easterly and westerly wind regimes descending with time from the middle to the lower stratosphere. The QBO phase can be defined as easterly (QBO-E) or westerly (QBO-W)—for example, according to the sign of the ZMZW in the lower stratosphere (e.g., Holton and Tan 1980). The average period is 28 months. The vortex is weaker on average in the easterly QBO phase than in the westerly phase by over 10 m s⁻¹ (Holton and Tan 1980; Pascoe et al. 2005). This “Holton–Tan (HT) relationship” has also been found in atmospheric models of varying complexity (e.g., O’Sullivan and Young 1992; Hamilton 1998; Gray et al. 2003; Calvo et al. 2007).

Understanding the mechanism behind this relationship is important for having confidence in observations...
of apparent nonlinear interactions with other forcings, such as with the solar cycle (e.g., Labitzke 2005; Camp and Tung 2007) and the El Niño–Southern Oscillation (ENSO) (e.g., Garfinkel and Hartmann 2007; Wei et al. 2007) and in the seasonal timing of the effect, which is still not well reproduced by models (Anstey and Shepherd 2013). It is also important for knowing what models must represent well in order to reproduce the HT relationship and exhibit realistic vortex variability.

The explanation for the HT relationship put forward by Holton and Tan (1980) involved the equatorial winds influencing the waveguide for extratropical planetary waves. Low-wavenumber stationary planetary waves dominate wave forcing of the extratropical NH stratosphere. The surface in the tropics where the ZMZW is zero [also referred to as the “zero wind line” (ZWL)] is a critical surface for these waves. Holton and Tan (1980) referred to the work of Tung (1979), who argued that this surface ought to reflect planetary waves back toward the pole if their amplitudes are small. In QBO-E, the critical surface in the lower stratosphere is positioned in the NH subtropics, so Holton and Tan (1980) suggested that this would concentrate wave activity in the NH polar region, weakening the vortex. This will be referred to as the “Holton–Tan mechanism.” Killworth and McIntyre (1985) showed that the critical surface ought to reflect eddy zonal momentum flux in the time-averaged sense even if wave amplitudes become large. This depends on several assumptions, however, that are not strictly met in the real stratosphere. It is not clear from theory whether a meridional shift of the ZWL would directly affect high latitudes by this mechanism.

In observations the geopotential height (GPH) wavenumber-1 amplitude and upward component of the Eliassen–Palm (EP) flux (Andrews et al. 1987), which is commonly used as an indicator of planetary wave propagation, are greater in November and December in QBO-E composites, but in January and February these values are greater during QBO-W (although the difference in these months is not highly statistically significant) (Holton and Tan 1980; Ruzmaikin et al. 2005). Holton and Tan (1982) and Hu and Tung (2002) considered the January–February data not to be consistent with the HT mechanism.

Modeling studies using GCMs indicate that on average in winter there is a greater upward component of the EP flux into the high-latitude stratosphere from the troposphere and greater EP flux convergence in the stratosphere during QBO-E than during QBO-W (e.g., Hamilton 1998; Calvo et al. 2007), although the locations of these effects differ among models. Holton and Austin (1991) and O’Sullivan and Dunkerton (1994) found that the amplitude of planetary waves peaks faster in QBO-E in perpetual winter runs in primitive equation models. These results have been interpreted as being broadly consistent with the HT mechanism.

Kodera (1991) suggested that the ZMZW anomalies associated with the QBO meridional circulation (Baldwin et al. 2001) may also affect planetary wave propagation, but this has been given less attention until fairly recently. Ruzmaikin et al. (2005) suggested the meridional circulation may directly affect the vortex by advection of potential temperature.

Naoe and Shibata (2010) argued that according to the HT mechanism the midlatitude lower-stratospheric EP flux ought to be more poleward in QBO-E; yet this is not the case. Their analysis of composite differences of EP flux between QBO-E and QBO-W in a chemistry–climate model led them to argue that the QBO meridional circulation has an important role in the HT relationship and that the shift of the critical surface in the lower stratosphere is not important. Garfinkel et al. (2012) reached a similar conclusion by examining the transient response to nudging equatorial winds toward QBO-E in a general circulation model (GCM) without coupled chemistry. Yamashita et al. (2011) also argued that the HT mechanism was not consistent with composite differences of EP flux between QBO-E and QBO-W in their chemistry–climate model, and proposed that the southward critical surface shift in the middle stratosphere is more important than the northward shift in the lower stratosphere.

However, experiments with primitive equation models showed that the vortex is more disturbed when equatorial winds are relaxed toward a constant easterly value at all heights, which would not be expected to produce a strong meridional circulation owing to the lack of vertical wind shear in the tropics (Gray et al. 2003) or toward easterlies in a limited height range without there being tropical westerlies (Naito et al. 2003; Gray et al. 2004). These results suggest that neither the meridional circulation nor the southward/northward shift of the middle-stratospheric critical surface in QBO-E/W are necessary to produce the HT relationship. This raises the questions of whether GCMs behave differently to the primitive equation models or whether the effects of shifts in the critical surface at different heights do not combine linearly.

Here we further analyze the mechanism behind the HT relationship using observational data and a GCM, described in section 2. We demonstrate in section 3 that the composite difference of the atmospheric state between the QBO phases closely resembles the signature of the stratospheric NAM and argue that this is not likely to be helpful for understanding the mechanism behind the HT relationship, in observations or models.
In section 4 we argue that the full transient response of a system to an applied forcing is likely to be much more informative about the dynamics of the forcing mechanism, especially the response shortly after this forcing is applied. We examine the transient response of the vortex to imposing a QBO-E state at the equator in a GCM in section 5 and show that the easterly acceleration of the ZMZW in the tropical lower stratosphere can directly cause increased EP flux convergence and zonal wind deceleration in the high-latitude NH stratosphere, consistent with the HT mechanism. We do not see the responses in the eddy momentum flux predicted by other proposals in which the QBO meridional circulation or the middle-stratospheric critical surface plays an important role. We conclude that these results and the results of Gray et al. (2003, 2004) and Naito et al. (2003) indicate that the HT mechanism contributes to causing the HT relationship (section 6).

2. Data and methods

a. Observational data

We use the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al. 2005) on standard pressure levels from September 1957 to August 2002 to examine the HT relationship in observations. Randel et al. (2004) found that ERA-40 matches stratospheric measurements of the zonal mean circulation derived from radiosonde, rocketsonde, and lidar measurements quite closely and performs quite well compared to other analyses and reanalyses, although errors may be substantial in the upper stratosphere above about 5 hPa. Baldwin and Gray (2005) found the ERA-40 QBO to agree well with independent rocketsonde data. ERA-40 agrees well with other analyses in its representation of SSWs (Charlton and Polvani 2007).

For analysis of our ERA-40 data we define the QBO as being in its easterly (westerly) phase when the 5°S–5°N November–February mean ZMZW is easterly (westerly) at 50 hPa—the pressure at which the correlation between the November–February mean ZMZW averaged over 5°S–5°N and that at 60°N, 10 hPa is greatest at 0.58.

b. Northern annular mode index

We compare the QBO-E minus QBO-W EP flux and GPH differences to the EP flux signature of the stratospheric NAM—the leading empirical orthogonal function (EOF) of the extratropical stratosphere. We index the NAM by the leading principal component of the monthly mean 3D GPH north of 20°N between 1 and 100 hPa, calculated using the method of Baldwin et al. (2009) using the product of pressure and the cosine of the latitude as weighting matrix elements. This is similar to the NAM index of Thompson and Wallace (2000), but restricted to the stratosphere. We also reverse the usual sign convention so that, when the index is positive, the vortex is weaker, to more easily compare the NAM signature with the QBO-E minus QBO-W differences. Restricting the pressure domain in the calculation to 10–100 hPa makes very little difference to the calculated index, so it is not sensitive to the representation of the upper stratosphere. Then at each grid point the EP flux components and its divergence and GPH are linearly regressed against the index, such that the presented signatures correspond to a one standard deviation increase in our NAM index. This is done separately for each calendar month. The correlation between this index and the leading principal components of GPH on individual pressure levels 10 and 50 hPa is 0.97 or greater in each calendar month November–February and the correlation with that at 5 hPa is 0.75 or above, so this index captures variability throughout the NH stratosphere well.

c. GCM simulations

We have performed experiments using the Met Office coupled carbon cycle configuration of the vertically extended Hadley Centre Global Environment Model, version 2 (HadGEM2-CCS) GCM. This coupled ocean–atmosphere model has a well-resolved stratosphere, with 60 atmospheric levels in the vertical up to 84-km altitude (corresponding to a pressure of approximately 0.01 hPa) and atmospheric horizontal resolution 1.25° latitude and 1.875° longitude. The model includes parameterized orographic gravity wave drag up to 40-km height, using the scheme of Webster et al. (2003), and nonorographic gravity wave drag (NOGWD), using the scheme of Warner and McIntyre (1999) as implemented by Scaife et al. (2002). The NOGWD causes the model to exhibit a spontaneous QBO. The radiation scheme is described by Edwards et al. (2004) and the model does not include stratospheric chemistry apart from methane oxidation. For full model details see Martin et al. (2011). Osprey et al. (2013) found that HadGEM2-CCS exhibits a realistic stratospheric climatology and realistic variability.

Results from a 240-yr preindustrial control run—which has a constant CO2 mass mixing ratio of 4.35 × 10⁻⁴, prescribed zonally symmetric monthly-mean climatological ozone that is the 1850–60 mean seasonal cycle of the dataset of Cionni et al. (2011), volcanic aerosol optical depth 0.0097, and solar constant 1365 W m⁻²—were used to confirm that the model reproduces the HT relationship reasonably well (section 5). Historical runs over 1860–2005 show a similar but weaker HT relationship (not shown). The correlation between the
November–February mean ZMZW averaged over 5°S–5°N and that at 60°N, 10 hPa is greatest for equatorial winds at 30 hPa, so the sign of the 5°S–5°N November–February mean ZMZW on this level is used to define the QBO phase for analysis of model data, which seems the fairest way to compare the modeled and observed HT relationships.

Section 5 presents results from experiments designed to examine the transient response of the vortex to nudging toward a QBO-E zonal wind pattern in the tropical stratosphere. We performed a “climatological tropical wind” (ClimEq) control run, which was set up identically to the 240-yr preindustrial control run except that the zonal wind in the tropical stratosphere was nudged toward the Interim ECMWF Re-Analysis (ERA-Interim) monthly mean climatology between January 1979 and December 2010 (with the climatology at each model time step calculated by linear interpolation between the middle of each month). ERA-Interim is likely to have a better representation of the stratosphere above 10 hPa than ERA-40 (Simmons et al. 2007), so it is used in preference to create target equatorial zonal wind profiles. We then performed 120 “QBO-E” branch runs of length one month, taking initial conditions at 1 January and 1 February of 60 different years from the ClimEq run (the first two years of this run were not used to allow the model to adjust to the nudging). In these runs the zonal wind in the tropical stratosphere was nudged toward the ERA-Interim climatology plus a typical QBO-E profile. The QBO-E profile was taken as the mean 3D zonal wind anomaly of the 30 months in ERA-Interim with the most negative anomalies in the 5°S–5°N mean ZMZW at 30 hPa, multiplied by a factor of 3 in order to raise the signal to noise ratio of the vortex response (Fig. 1)—the equatorial ZMZW anomaly does not become larger than that in observed QBO-E phases in the time scale of 8 days considered in section 5, so this just affects the rate at which equatorial winds adopt a QBO-E profile. This method is similar to that used by Garfinkel et al. (2012) but, importantly, we focus on the vortex response at shorter times after nudging toward the QBO-E profile is begun.

Nudging was carried out between 21.25°S and 21.25°N and was implemented by subtracting \(\alpha(\phi)(\mathbf{u} - \mathbf{u}^T)\Delta t\) from the change in the zonal wind calculated at the end of each time step at each gridpoint, where \(\mathbf{u}\) is the zonal wind, \(\mathbf{u}^T\) is the target nudging profile, \(\phi\) is the latitude, and \(\Delta t\) is the model time step. The nudging parameter \(\alpha(\phi)\) is given by

\[
\alpha(\phi) = \frac{1}{20 \text{days}} e^{-2(\phi/16\text{°})^2}
\]

between heights of 17.4 and 39.1 km (corresponding to pressure range 3.3–84.3 hPa). At one model level below and above this range (16.3 km, 103 hPa and 40.9 km, 2.6 hPa) \(\alpha\) was set to one-half its value within the range and is zero at other heights. The nudging code was adapted from that developed as part of the United Kingdom Chemistry and Aerosols project (Telford et al. 2008). In the ClimEq run, the nudging had the effect of eliminating the QBO, but the extratropical ZMZW climatology and variance are similar to that in the 240-yr preindustrial control run (not shown), indicating that the
extratropical circulation is not strongly affected by the nudging, though the amplitude of equatorial waves will have been reduced. The vertical profile of the 5°S–5°N mean ZMZW differences between the QBO-E runs and the ClimEq run is very similar to that in Fig. 1 between about 3 and 100 hPa over the first 8 days, although the meridional width of the nudged QBO-E winds is only two-thirds that of the target profile owing to the weakness of the nudging away from the equator (not shown)—this would most likely cause the influence on the vortex to be weaker than for a perfect imposed QBO-E profile.

d. Diagnostic tools and statistical methods

We present the influence of the QBO on the EP flux $F = (F_y, F_z)$ (Andrews et al. 1987), which is usually taken to show the negative of the zonal-mean zonal momentum flux associated with zonal asymmetries. The variable $F_y$ is analogous to the negative of the eddy zonal momentum flux $-u'\bar{v}$ that was considered by Killworth and McIntyre (1985) in their analysis of the reflectivity of the critical line in their analytical model, where $u'$ and $v'$ are departures from the zonal-mean zonal and meridional components of the wind, respectively, and the overbar indicates zonal averaging. We present the acceleration term $D_F = (\rho_0 a \cos \theta)^{-1} \nabla \cdot F$ to show where the EP flux is convergent or divergent, where $\rho_0(z)$ is a reference density profile and $a$ is Earth’s radius.

Statistical significances of the QBO-E minus QBO-W composite differences in the EP flux components and in the GPH were calculated according to a Monte Carlo (MC) permutation test. Each year was assigned to a surrogate QBO-E or QBO-W group at random, and the composite difference between these random groups was calculated. This was repeated 1000 times to find the probability that the magnitude of the difference would exceed that of the difference in the data at each grid point under the null hypothesis that there is no dependence of these variables on the QBO.

Statistical significances of the regression coefficients of EP flux and GPH onto the NAM index were also calculated according to an MC permutation test. The time series of the NAM index for each calendar month was shuffled and the regression coefficients against this random time series were calculated. This was repeated 1000 times to find the probability that the magnitude of the regression coefficient would exceed that of the coefficient in the data under the null hypothesis that the variables have zero correlation with the NAM index.

To test the statistical significance of the mean differences between the model branch runs nudged to QBO-E and the ClimEq run, an MC bootstrap technique was used (Efron and Tibshirani 1993). At each gridpoint a surrogate data sample was generated according to the null hypothesis that the mean difference is zero but other moments of the true distribution of differences equal those in the data. The mean of the differences for all pairs of branch and control runs was subtracted from the difference for each pair, and the results were resampled with replacement. The probability of the mean of this resampled data being larger than that for the real data was estimated using 1000 data resamplings. All significance tests are two tailed.

3. The observed influence of the QBO on the wave part of the stratospheric circulation

Figure 2 shows the QBO-E minus QBO-W composite difference in the monthly mean NH EP flux and $D_F$ from November to February in ERA-40. In agreement with the findings of previous studies (e.g., Dunkerton and Baldwin 1991; Ruzmaikin et al. 2005), the EP flux is more upward in November and December near 55°N, but this signal is weak in January when the 30°–90°N mean upward EP flux difference at 70 hPa is positive, but not statistically significant, and the EP flux is more downward at high latitudes in February. It has been argued that the late-winter signal is not consistent with the HT mechanism (Holton and Tan 1982). In all months there is a poleward EP flux difference in the tropical lower stratosphere, indicating that the equatorward flux is less in QBO-E. This is restricted to latitudes south of about 25°N with increased equatorward flux to the north, and Naoe and Shibata (2010) and Yamashita et al. (2011) argued that this means reflection of eddy zonal momentum flux from the lower-stratospheric easterlies in QBO-E cannot be directly influencing the vortex.

Figure 3 shows the regression of the EP flux against our NAM index, showing the anomaly associated with a weaker vortex (note our choice of sign of the index as explained in section 2b). There is a good resemblance between the main qualitative features of this and the QBO-E minus QBO-W composite differences in each calendar month in the extratropics. The flux is more upward north of 55°N in November and December near 45° and 80°N in January and more downward near 60°N in January and February (though the signal to noise ratio in January and February is not very high). The NAM regression does not show a poleward EP flux difference in the tropics, indicating that this feature in the differences is not strongly associated with extratropical variability.

There is a similar correspondence between the QBO-E minus QBO-W composite difference and the NAM signature of GPH. Figure 4 shows the QBO-E minus QBO-W composite difference in the monthly mean...
NH 10-hPa GPH from November to February in ERA-40. The climatological eddy component, defined as the climatological GPH with the zonal mean subtracted, is also shown. Figure 5 shows the signature of the NAM in the NH 10-hPa GPH, which bears a very good resemblance to the GPH differences between QBO-E and QBO-W. Anomaly correlations\(^1\) with the NAM signature north of 20°N are indicated below the composite GPH differences in Fig. 4. The correlations are between 0.79 and 0.95 in November–January, with a lower correlation in February when the differences are not highly statistically significant. The correlations are all greater if only anomalies north of 60°N are considered, so this is not simply arising from a direct influence of the QBO on the subtropics.

In November, the top left panel of Fig. 4 shows GPH is greater in QBO-E over the Canadian Arctic and less over northern Europe. This represents positive interference with the wavenumber-1 part of the climatological wave pattern, so the wavenumber-1 amplitude is greater while the wavenumber-2 amplitude is slightly less (Fig. 6). Over the course of winter, however, the pattern shifts so that more positive GPH is found over the Arctic and North Atlantic with lower GPH over the North Pacific in January and February. This gives weak destructive interference with the climatological waves.

\(\text{1 Anomaly correlation between 2D anomaly patterns } x \text{ and } y \text{ is defined here as}
\)

\[
\frac{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_i x_i y_i}{\sqrt{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_i^2 \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} y_i^2}},
\]

where \(n_i\) and \(n_j\) are the number of grid points along the latitude and longitude dimensions and \(w_i\) is the cosine of the latitude.
so wavenumber-1 and -2 amplitudes are both slightly reduced. Figure 6 also shows the change in the wavenumber-1 and -2 amplitudes at 60°N, 10 hPa associated with the QBO influence on the NAM only. To calculate this, the QBO-E minus QBO-W composite difference in the NAM index is multiplied by the NAM GPH signature in Fig. 5. This is then added to the climatological GPH and the change in wave amplitudes computed. This shows that the observed changes in the wave amplitudes in QBO-E versus QBO-W correspond closely to the seasonal evolution of the NAM signature.

Previous work has noted that the QBO-E minus QBO-W composite difference of extratropical ZMZW and zonal-mean GPH is very similar to the NAM signature (Dunkerton and Baldwin 1991; Kodera 1995; Ruzmaikin et al. 2005), but here we show that this is true for the EP flux and GPH wave amplitude differences as well, which is important given that these differences have been used to try to understand the mechanism of the QBO’s influence.

A similarity between the leading EOF of a system and its response to an applied forcing is a commonly observed feature of dynamical systems, if the response is averaged over time scales that are long compared to the dynamical time scales. Palmer and Weisheimer (2011) illustrate that in the simple system of Lorenz (1963) (that which gives rise to the famous Lorenz butterfly attractor), applying a steady force in any direction in the \( x-y \) plane gives rise to a shift in the system’s mean state that is very nearly aligned with the system’s leading EOF, so the spatial pattern of the response closely resembles that of the leading EOF. Ring and Plumb (2008) found in a tropospheric GCM that the steady-state response to various mechanical and thermal forcings applied in the extratropics closely resembles the tropospheric NAM. Branstator and Selten (2009) examine the reasons why the response to greenhouse gas forcing in a tropospheric GCM is NAM-like. They conclude that it largely results from a linear effect whereby anomalies in the NAM tend to persist for a long time, so
the NAM is a prominent pattern in natural variability and in the response to a forcing after time averaging. This behavior is also predicted by the fluctuation–dissipation theorem (Gritsun and Dymnikov 1999), which has been argued to apply approximately to the atmosphere (e.g., Leith 1975; Gritsun and Branstator 2007).

Invoking this behavior, which is present in atmospheric models and also in simpler systems like that of Lorenz (1963), can then explain the seasonal evolution of the pattern of EP flux differences in QBO-E and QBO-W, and supports the suggestion of Dunkerton and Baldwin (1991) that the QBO “excited a fundamental . . . mode of

---

**FIG. 4.** QBO-E minus QBO-W composite differences in ERA-40 of the wintertime GPH at 10 hPa north of 20°N (grayscale). White contours show the climatological zonally asymmetric component of GPH with contour values ±200 and ±600 m, with negative contours dashed. NAM correlation values indicate the anomaly correlation of the composite differences with the NAM signature north of 20°N shown in Fig. 5. Stippling shows where GPH differences are statistically significant above the 95% level.
variability in the extratropical atmosphere.” We do not know of any studies that have examined this phenomenon in the stratosphere. It does, however, seem consistent with previous literature identifying the stratospheric NAM as a prominent pattern that appears not just in response to the QBO but also to other important natural influences on the vortex, including volcanic eruptions (e.g., Kodera 1995; Stenchikov et al. 2006),

**FIG. 5.** Regression of 10-hPa GPH north of 20°N onto our NAM index in ERA-40 (grayscale) showing a close resemblance to the QBO-E minus QBO-W composite differences in Fig. 4. White contours show the climatological zonally asymmetric component of GPH with contour values ±200 and ±600 m and with negative contours dashed. Stippling shows where the regression coefficients are statistically significant above the 95% level.
ENSO (e.g., Sassi et al. 2004), and the solar cycle (e.g., Kodera 1995; Labitzke 2005).

Importantly, this behavior implies that examining the response to a forcing averaged over a long time or using compositing does not in general yield much information about the forcing mechanism. Hence, if an anomaly in response to an unknown forcing resembles the leading EOF, it does not allow the nature of the forcing to be deduced since many different forcings may give rise to this response. Therefore, the HT mechanism is consistent with the vortex response to the QBO as it simply predicts that the vortex will be weaker in QBO-E, as is observed, and the above discussion indicates that this may manifest itself as a modulation of the NAM in composite analysis. However, these observations could also be consistent with other mechanisms that predict a weakening of the vortex during QBO-E. Furthermore, structure in the extratropical QBO-E minus QBO-W composite differences seems to be primarily related to the NAM signature rather than the forcing mechanism, and is not a reliable indicator of the mechanism.

Another way to understand the difficulty in using these composite differences to infer the forcing mechanism behind the HT relationship is that the differences in the wave components of the flow have contributions not only from the QBO-E-W equatorial wind pattern but also from the effect of the weaker vortex, as changes in the zonally symmetric component of the flow will cause changes in the wave components. What is required is a way of computing the changes due to QBO-E-W forcing while the vortex state is close to constant.

4. Motivating examination of the short-term transient response

Here we argue that, while the QBO-E minus QBO-W composite difference cannot be relied on to show the mechanism of the QBO influence, the full time-dependent transient response should be much more useful.

If the forcing mechanism is simple (meaning it does not influence the system in question in a series of intermediate steps), then the response on short time scales following application of the forcing can be expected to show the mechanism clearly. Consider a system described by state vector \( \mathbf{x}(t) \) that evolves according to equations

\[
\dot{\mathbf{x}}(t) = \mathcal{L}[\mathbf{x}(t), t] + \mathbf{f}[\mathbf{x}(t), t],
\]

where the equations may be nonlinear and are explicitly time dependent for generality. The equations governing atmospheric motion may be expressed in this form. Consider also a forced variant of this system described by state vector \( \mathbf{x}'(t) \) that evolves according to similar equations with the addition of a state- and time-dependent forcing term,

\[
\dot{\mathbf{x}}'(t) = \mathcal{L}[\mathbf{x}'(t), t] + \mathbf{f}[\mathbf{x}'(t), t]
\]

with \( \mathbf{x}(0) = \mathbf{x}'(0) = \mathbf{x}_0 \). In the context of the HT relationship, \( \mathbf{x}' \) would represent the vortex state and \( \mathbf{f} \) the influence of the QBO. Then as long as the difference between the state vectors \( \delta \mathbf{x}(t) = \mathbf{x}'(t) - \mathbf{x}(t) \) is analytic, which will be the case if both \( \mathbf{x}(t) \) and \( \mathbf{x}'(t) \) are analytic, \( \delta \mathbf{x}(t) \) can be evaluated for small \( t \) by writing its Taylor series (in index notation using summation convention) to give

\[
\delta x_a(t) = f_a(\mathbf{x}_0, 0) t + \left[ \frac{\partial \mathcal{L}_a}{\partial x_b}(\mathbf{x}_0, 0) \delta x_b(0) + \frac{df_a}{dt}(\mathbf{x}_0, 0) \right] \frac{t^2}{2} + O(t^3). \tag{1}
\]

Thus, for short times, the difference between the systems is nearly proportional to the applied forcing at \( t = 0 \) when this is nonzero. State dependence of the equations of motion (nonzero \( \partial \mathcal{L}_a/\partial x_b \) acts to complicate the relationship between \( \delta \mathbf{x} \) and the forcing as \( t \) increases and
may be expected to become important on a time scale of the order of the system's dynamical time scale. [If \( f(x_0, 0) = 0 \), it can similarly be shown that \( \delta x(t) \) is proportional to \( df/dt|_{(x,t)=(x_0,0)} \) to \( O(\tau^2) \) when this is nonzero, which is relevant for the results of the nudging experiments in section 5.] Examining the system response on short time scales shows that the direct effect of the forcing before feedbacks due to the change in the state of the system become large. Equation (1) can be rewritten

\[
\delta x_a(t) = \int_0^t f_a(x, t') dt' + \frac{\partial E_a}{\partial x_b} \bigg|_{(x,t)=(x_0,0)} \delta x_b(0) + O(\tau^2).
\]

Therefore, for forcing mechanisms that unfold in several steps, the transient response would be expected to clearly show the cumulative forcing up to a time scale on the order of the dynamical time scale, which may still give useful information for testing hypotheses.

No study has argued that any feedback of vortex changes onto the QBO is important for the mechanism behind the HT relationship, but, even if it were, it can be shown this would be an \( O(\tau^2) \) contribution to \( \delta x(t) \), so the short-term response would still give information about the one-way influence of the QBO on the vortex.

For studies of the atmosphere, it is probably necessary to use a numerical model to evaluate the transient response. Using observations would require identifying two near-identical atmospheric states with different values of the forcing in question, which is practically impossible.

5. The transient response of the vortex to QBO-E forcing

a. The Holton–Tan relationship in HadGEM2-CCS

Figure 7 shows the QBO-E minus QBO-W composite difference in the January–February mean ZMZW in ERA-40 and in the 240-yr preindustrial control run of HadGEM2-CCS—the monthly mean difference is not statistically significant earlier in winter in HadGEM2-CCS. The model reproduces the weakening of the vortex seen in observations, with ZMZW differences that are somewhat smaller at high latitudes. The lack of a Holton–Tan relationship in November and December in the model may be due to the equatorial winds simply having too weak an influence on the vortex and it taking time for their impact to accumulate and give rise to an appreciable vortex response, following the suggestion of O'Sullivan and Young (1992). It could also be related to the model exhibiting slightly less total variability than in observations in early winter (Osprey et al. 2013), which is a common problem in stratosphere-resolving GCMs. As noted in section 1, few modeling studies have reported a HT relationship in early winter. Nevertheless, the mechanism by which the equatorial winds influence the vortex in the model is likely to be qualitatively similar to the mechanism in the real atmosphere.

Figure 8 shows the QBO-E minus QBO-W composite difference in the monthly mean NH EP flux in January and February in the model. In these months the model reproduces the pattern of the observed influence of the QBO on the EP flux within sampling error, with greater upward flux in the high-latitude stratosphere in January.
in QBO-E and more poleward and downward flux in February. The EP flux differences are somewhat smaller than in ERA-40, however, consistent with the ZMZW differences being smaller.

b. The modeled short-term transient response

We have examined the transient response of the vortex to nudging equatorial winds toward a QBO-E state on a time scale of about a week, which is comparable to the middle-stratospheric dynamical time scale. Figure 9 shows the ensemble-mean ZMZW difference between the QBO-E branch runs and the ClimEq run (described in section 2c). The QBO-E pattern of winds is visible in the tropics, with midlatitude anomalies that arise owing to the Coriolis force acting on the QBO meridional circulation. The extratropical winds in the lower stratosphere initially strengthen slightly. However, between day 5 and 6 following the start of the nudging, there is a weakening of the upper stratospheric ZMZW and between day 7 and 8 there is a weakening of the zonal-mean zonal wind throughout the stratosphere north of 60$^\circ$N, as expected from the HT relationship, though these ZMZW differences are not highly statistically significant.

Figure 10 shows the mean difference in the residual circulation between the branch runs and the ClimEq run and the difference in the associated acceleration

$$-\mathbf{v}^*[(a \cos \phi)^{-1}(\mathbf{u} \cos \phi)_{\phi} - f] - \mathbf{w}^* \mathbf{n}_z,$$

where $\mathbf{v}^*$ and $\mathbf{w}^*$ are the meridional and vertical components of the residual circulation, respectively; $\mathbf{n}$ is the ZMZW; $f = 2\Omega \sin \phi$ with $\Omega$ the Earth’s angular velocity and $a$ its radius; and subscripts denote partial differentiation with respect to the subscripted variable (Andrews et al. 1987). The cells of the QBO meridional circulation develop over time with a more poleward subtropical residual circulation in the branch runs near 30 hPa and a more equatorward circulation near 7 hPa. On days 1–2 the associated acceleration closely resembles the ZMZW differences in Fig. 9 between $\sim$20$^\circ$ and 70$^\circ$N in the lower stratosphere and between $\sim$20$^\circ$ and 60$^\circ$N in the upper stratosphere, so these differences appear to be caused by the acceleration of the meridional circulation. At later times the acceleration due directly to the QBO meridional circulation is confined equatorward of $\sim$40$^\circ$N.

Figure 11 shows the mean difference in the EP flux and $D_f$ between the branch runs and the ClimEq run. On days 1–2 the EP flux is more equatorward in the lower stratosphere and poleward in the middle and upper stratosphere between $\sim$15$^\circ$ and 30$^\circ$N, and the $D_f$ differences are relatively small. Between day 3 and 8, however, EP flux differences become statistically significant at higher latitudes. The EP flux is more poleward in the lower stratosphere and more equatorward in the upper stratosphere, extending to the polar region by days 5–6. This is associated with increased EP flux convergence in the high-latitude stratosphere between $\sim$2 and 20 hPa; therefore, this contributes to the weakening of the winds in this region over days 5–8 (Fig. 9). The EP flux differences are very similar if only zonal wave-numbers 1–3 are considered (not shown). The residual circulation differences are more upward in the middle stratosphere at this time and partially cancel this contribution and contribute to weakening of the winds at high latitudes above $\sim$2 hPa on days 5–8 and below $\sim$20 hPa on days 7–8 (Fig. 10).

The more poleward EP flux from the tropics to the high latitudes is the signal that Naoe and Shibata (2010)
argued should be present if the HT mechanism is correct. Our results are thus consistent with the HT mechanism and indicate that increased reflection of eddy zonal momentum flux in QBO-E may directly affect the polar stratosphere. This interpretation is also supported by the fact that there is a positive correlation between the lower stratospheric meridional EP flux difference in days 5–8 and the initial equatorward flux in midlatitudes in the branch runs, which is statistically significant above the 99% level according to a MC permutation test. In other words, the EP flux difference is stronger when there is more equatorward flux, as expected if the difference is due to increased reflection of this flux. Thus, the HT mechanism cannot be ruled out of playing a part in the HT relationship as argued in recent studies. Note that this signal is unlike the QBO-E minus QBO-W composite differences in the EP flux in any calendar month (Fig. 2), illustrating the necessity of examining the transient response to understand the influence of the QBO.

c. Role of the tropical middle and upper stratosphere

However, it is not just the tropical lower-stratospheric easterly wind acceleration that matters—the westerly acceleration above 10 hPa appears to cause a more equatorward EP flux that decreases its convergence at high latitudes, reducing the QBO influence there. Figure 12 shows the EP flux differences using the “acceleration scaling” defined by Gray et al. (2003) to indicate the size of the acceleration associated with the flux. By days 7–8 there is more poleward flux in the QBO-E runs at 1 hPa and above that appears associated with the tropical ZMZW at these heights becoming more easterly (Fig. 9). Easterly anomalies at these altitudes are also observed in QBO-E in winter (Pascoe et al. 2005). Although the magnitude of this momentum flux is small, it is associated with EP flux
convergence and deceleration of the ZMZW in the midlatitude upper stratosphere. If the vortex state is sensitive to ZMZW changes in this region, then this provides a way for tropical upper stratospheric wind changes to contribute to the HT relationship, as suggested by Gray et al. (2001a,b).

**d. Discussion**

These results do not indicate that the Eliassen–Palm flux convergence in the high-latitude stratosphere arises owing to alternatives to the HT mechanism discussed in section 1—the midlatitude EP flux differences have opposite sign to those predicted by Naoe and Shibata (2010) and Yamashita et al. (2011) and its convergence in midlatitudes has opposite sign to that predicted by Garfinkel et al. (2012). It is not clear if there are reasons why the signatures of these mechanisms would take more than 8 days to appear, but these results make it seem less likely that any of these mechanisms are dominant. The subtropical ZMZW anomalies associated with the QBO meridional circulation may take longer than 8 days to fully develop and affect the EP flux.

The greater reflection of EP flux in the lower stratosphere in the QBO-E runs occurs despite the zero wind line (ZWL) shifting both poleward and equatorward on different levels below 10 hPa. It may therefore be interesting to examine the contributions to the EP flux differences from waves with nonzero phase velocities, though this is beyond the scope of this paper. As remarked in section 2c, the QBO profile in these simulations is too

---

**FIG. 10.** Ensemble-mean residual circulation differences between branch runs nudged toward the QBO-E equatorial wind profile and the ClimEq run averaged over each 2-day interval up to 8 days following branching (arrows, shown only where either the \( \mathbf{v}^* \) or \( \mathbf{w}^* \) differences are statistically significant above the 95% level) and the associated zonal acceleration differences (contours at 0, \( \pm 0.02 \), and \( \pm 0.1 \) m s\(^{-1}\) day\(^{-1}\) with negative contours dotted and the zero contour thickened). The residual circulation vectors are shown at pressures 0.5, 1, 2, 5, 10, 20, 50, and 100 hPa and every 3.75° in latitude. A reference arrow is shown in the top left of each plot with its \( (\mathbf{v}^*, \mathbf{w}^*) \) values. Shading shows where differences in the acceleration term are statistically significant above the 95% level.
narrow meridionally, and the ZWL shift may have been more pronounced if this were not the case. It is not clear that the ZWL should be considered fundamental. The ensemble-mean results that we present show the average effect of nudging toward a QBO-E state but may hide sensitivity to the initial conditions. The theory of critical layers only indicates that they may be absorbing for finite time periods and the influence of the QBO may be different in such periods. The QBO influence may also be reduced if critical surfaces exist in midlatitudes. Additionally, Gray et al. (2003) suggest that the QBO may have a weaker influence if the tropospheric wave forcing is very strong or very weak.

The signal in Figs. 9–12 loses statistical significance after day 8, which appears to be because the standard deviation of the differences between the runs grows exponentially with time (not shown), as expected from exponential perturbation growth in a chaotic system such as the atmosphere. The ZMZW and EP flux differences between the branch and ClimEq runs grow approximately linearly with time over the first week, so the loss of statistical significance is consistent with this signal continuing to grow but the noise growing more quickly so that the signal to noise ratio decreases with time. The magnitude of the noise must saturate after some time at the level of the climatological variability. The fact that there is an HT relationship in the 240-yr preindustrial control run implies that the signal to noise ratio would be large enough again after several months for the signal to be statistically significant, implying that

![Ensemble-mean EP flux differences between branch runs nudged toward the QBO-E equatorial wind profile and the ClimEq run averaged over each 2-day interval up to 8 days following branching (arrows, shown only where either the $F^f$ or $F^z$ differences are statistically significant above the 95% level) and $D_f$ differences (contours at 0, ±0.02, and ±0.1 m s$^{-1}$ day$^{-1}$ with negative contours dotted and the zero contour thickened). The thick dashed line shows the zero wind line in the days 1–8 mean ZMZW in the ClimEq run. EP flux differences are shown at pressures 0.5, 1, 2, 5, 10, 20, 50, and 100 hPa and every 3.75° in latitude. A reference arrow is shown in the top left of each plot along with its ($F^f$, $F^z$) values. Shading shows where $D_f$ differences are statistically significant above the 95% level.](image-url)
the signal would continue to increase in time. Seeing the full evolution of the vortex response to QBO-E forcing in this model would likely require a prohibitive amount of computing resources, and similar experiments with computationally cheaper models may be helpful to understand the transient response fully. The transient response to nudging toward QBO-E presented by Garfinkel et al. (2012) appears to be more statistically significant, but the model that they used has unrealistically low vortex variability, which may be part of the reason for the difference.

It should also be considered whether the use of nudging in the tropical stratosphere would cause differences between the coupling between the tropics and extratropics in our model runs and in the real stratosphere. Previous studies have used nudging of equatorial winds to examine the HT relationship and have obtained realistic results (e.g., Hamilton 1998; Garfinkel et al. 2012), implying that the nudging does not drastically interfere with the coupling between the tropics and extratropics. The nudging used here would tend to dampen wave activity in the tropical region with time scales longer than a couple of weeks. The most unrealistic effect of the nudging may be to create a QBO profile that is too narrow meridionally, which may weaken the extratropical response. We see no reason why these effects would qualitatively change the interaction between the tropics and extratropics.

The atmospheric response to an applied forcing at short times following application of the forcing is indicative of the whole forcing mechanism only if the steps of the mechanism unfold on a time scale less than about the system’s dynamical time scale, so mechanisms whose early stages are consistent with our results, but also involve subsequent steps, cannot be ruled out. Our results also do not rule out the meridional circulation having a role in modifying the background state through which the eddy zonal momentum propagates. Our experiments...
also show the effect of nudging toward QBO-E from a state having close to climatological equatorial winds—as the winds approach a full-strength QBO-E state, the vortex response may depend nonlinearly on further increases in the strength of the QBO-E profile. This could be explored in a similar way to our method, using a control run nudged toward a QBO-E state with branch runs nudged toward a stronger QBO-E state. In our runs the equatorial ZMZW differences do reach the approximate magnitude of the anomalies in the QBO-E phase in observations in the time scale considered here, however. This allows the possibility that nonlinearity will show an effect, so we do not expect that the effects of nonlinearity with respect to the equatorial winds would greatly change our results. Our results leave open the question of whether the HT mechanism can account quantitatively for the HT relationship.

The key innovative part of this work has been to examine the stratospheric response to QBO forcing on time scales shorter than or on the order of the middle stratosphere’s dynamical time scale of ~1 week, so we see the response to QBO-E forcing alone before the circulation has evolved. If our EP flux differences are averaged over the first 16 days, as done by Garfinkel et al. (2012), then the poleward EP flux difference from the tropics to high latitudes is no longer apparent. [However, the model used by Garfinkel et al. does not show a more poleward extratropical EP flux difference in the first 8 days as found in our experiments (C. Garfinkel 2013, personal communication), so there appears to be model dependence of this result, perhaps related to the models’ different spatial resolutions.]

6. Summary and conclusions

We have discussed investigations into the mechanism by which the quasi-biennial oscillation influences the stratospheric polar vortex. Understanding this mechanism is important for improving seasonal forecasts of the vortex, and therefore also of the troposphere (e.g., Baldwin and Dunkerton 2001), for understanding nonlinear interactions of this relationship with other forcings such as the solar cycle (e.g., Labitzke 2005) and ENSO (e.g., Garfinkel and Hartmann 2007) and understanding the seasonal timing of the effect. Previous observational and modeling studies, when considered together, have not come down firmly in favor of any of the proposed mechanisms.

We have shown that composite differences of the wave components of the stratospheric circulation between QBO-E and QBO-W are very similar to the signature of the northern annular mode—the leading EOF in the stratosphere. This behavior is qualitatively similar to that of other dynamical systems, in that different forcings applied to the same system are found to give a response similar to that system’s leading EOF. We have argued that this implies that QBO-E minus QBO-W composite differences are not likely to be informative about the mechanism behind the HT relationship since many different mechanisms could give rise to the NAM-like vortex response.

We then showed that the full transient response of a system to a given forcing should be much more informative about the forcing mechanism. The system’s transient response on short time scales after application of the forcing will show the steps of the forcing mechanism that unfold up to about one dynamical time scale—about 1 week in the middle stratosphere—and will closely resemble the forcing if the forcing mechanism essentially involves only one step.

Our examination of the first few days of the transient response of the vortex to nudging of equatorial stratospheric winds toward a QBO-E state in the HadGEM2-CCS GCM indicates that the EP flux becomes less equatorward between the tropics and high latitudes in the lower stratosphere, there is greater convergence of the EP flux in the high-latitude stratosphere between about 2 and 20 hPa, and the westerly wind in this region decelerates. This is consistent with the hypothesis that more easterly winds in the tropical lower stratosphere cause greater reflection of EP flux toward the polar stratosphere, directly causing deceleration of the westerly winds, as suggested by Holton and Tan (1980). Our results do not show the signatures of the mechanisms suggested by Naoe and Shibata (2010), Yamashita et al. (2011), or Garfinkel et al. (2012), who proposed that the role of the lower-stratospheric zero wind line is less important than that of the QBO meridional circulation or the middle-stratospheric zero wind line. Combining this with the results of Gray et al. (2003, 2004) and Naito et al. (2003), which indicate that a strong meridional circulation and the shift of the zero wind line in the middle stratosphere (in the opposite sense to that in the lower stratosphere) are not essential to produce the HT relationship, we conclude that the mechanism of Holton and Tan (1980) is likely to contribute to the HT relationship. However, the total anomalous poleward EP flux depends on zonal-mean zonal wind changes throughout the depth of the tropical stratosphere, and it is not clear if the role of the zero wind line is fundamental. We have not ruled out contributions from mechanisms that include multiple steps that unfold on a time scale greater than about a week, which may include those for which the ZMZW anomalies associated with the meridional circulation are important if they take more than 8 days to reach their full strength.
To our knowledge, using the transient response of the atmosphere to a forcing on time scales of a few days to understand the forcing mechanism is novel and may be of use in various other outstanding problems, such as for understanding the downward influence of the stratosphere on the troposphere (e.g., Baldwin and Dunkerton 2001) or the dynamical influence of the solar cycle (e.g., Labitzke 2005).

Acknowledgments. We thank D. Andrews, J. Anstey, S. Osprey, D. Mitchell, C. Garfinkel, and T. Shaw for helpful discussions and the anonymous reviewers for their comments and also P. Telford and M. Dalvi for assistance with implementing nudging in HadGEM2-CCS. P. Watson is supported by a Natural Environment Research Council studentship.

REFERENCES


