Part III Essay:
The Stratospheric Quasi-biennial Oscillation

May 5, 2021

1 Introduction

The stratospheric quasi-biennial oscillation (QBO) is an almost periodic, approximately axisymmetric variation of the zonal (longitudinal) wind in the equatorial mid-atmosphere. The oscillation is characterised by alternating, downward-propagating easterly and westerly wind regimes and acts as the dominant mode of interannual variability in the stratosphere, the layer of the atmosphere lying above the troposphere at roughly 15 to 50 km (100–1 hPa) in altitude. The wind regimes form in the upper stratosphere and descend at approximately 1 km per month, dissipating in the lower stratosphere an average of 28 months later, giving rise to a quasi-biennial oscillation (*Baldwin et al.*, 2001).

Application of atmospheric dynamical theory to the QBO started after its identification in the early 1960s, initially with the development of a mechanistic theory utilising wave/mean-flow interaction before later unravelling the intricate connection of the QBO with other atmospheric phenomena. This connection is both direct through the equatorial tropopause and indirect through the wider extratropical atmosphere; the fine details remain under consideration today. The difficulties of gathering observational data far above Earth's surface, particularly at sufficient spatial and temporal resolution, place the representation of the QBO and its teleconnections in global climate models under scrutiny today as satellite data improves and increased computational power allows the direct modelling of physical processes of greater complexity.

In this essay, we will discuss the dynamical behaviour of the QBO and its influence in the extratropical atmosphere, incorporating both mathematical and physical viewpoints in pursuit of understanding why such a phenomenon exists, how it fits into the broader picture of atmospheric dynamics, and where the influences may be felt. The remainder of section 1 provides a contextual discussion of the discovery and phenomenology of the QBO. In section 2, the basic dynamics are described with reference to a prototypical numerical model and wave/mean-flow theory. Finally, section 3 elucidates the extratropical effects of the QBO via a mechanistic approach to three illustrative examples of atmospheric variability modulated by the QBO.

Much of the discussion in this essay assumes familiarity with wave/mean-flow interaction, equatorial wave theory, and the (transformed) Eulerian-mean formalism. The pertinent details are made clear where necessary and full details can be found in the references herein.

1.1 Discovery

The earliest indication of persistent stratospheric winds came from the 1883 Krakatau volcanic eruption. The ash cloud circled the globe east to west in around 2 weeks, implying the existence of strong easterly winds which became known as the 'Krakatau easterlies'. In 1908, data from meteorological balloons launched by A. Berson indicated westerly winds at an altitude of 15km, which became known as the 'Berson westerlies'. Further, *Palmer* (1954) found the transition between these winds varied on monthly and yearly timescales, though there was insufficient data to identify any periodicity. A gradual descent of these wind regimes was identified by *Graystone* (1959) using 2 years of observations of wind speed from Christmas Island at 2°N.

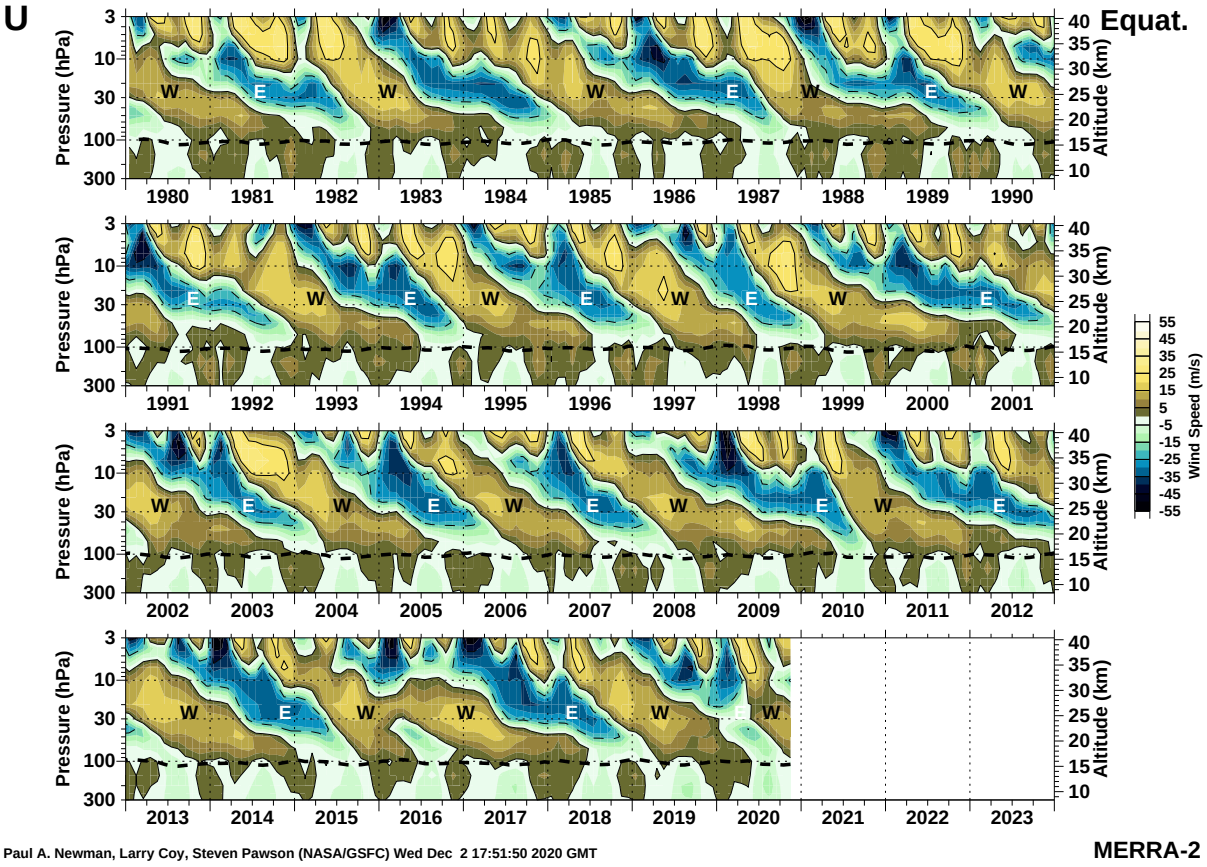


Figure 1: 1980–present MERRA-2 monthly mean zonal wind at the equator. The thick dotted line represents the tropopause. Plot sourced from `https://acd-ext.gsfc.nasa.gov/Data_services/met/qbo/qbo.html`

Discovery of the QBO as it is now understood is accredited to R. J. Reed in the United States and R. A. Ebdon in the United Kingdom. In 1960, using rawinsonde data from Canton Island (2.8°S), Reed discovered alternating bands of easterly and westerly winds, which originate above 30km and descend through the stratosphere at approximately 1 km per month, calculating a period of 26 months (later published as *Reed et al.* (1961)). *Ebdon* (1960) used data also from Canton Island over the period 1954–1959 to identify a period of about 2 years. Later, including data from 1954–1960, a period of 25–27 months was identified by *Ebdon and Veryard* (1961). They concluded wind fluctuations occur simultaneously around the equator, and take 1 year to descend from 10 to 60 hPa. Shortly thereafter, *Veryard and Ebdon* (1961) extended their study and identified similar fluctuations in temperature data.

The term 'quasi-biennial oscillation' was coined by *Angell and Korshover* (1964) following further observations of a quasi-periodic cycle in mid-atmospheric winds with a period close to two years. A mechanistic explanation of the QBO was not readily apparent but understanding grew through the late 1960s and early 1970s with the gradual identification of aspects such as its longitudinal symmetry (*Belmont and Dartt*, 1968) and the development of equatorial wave theory (*Matsuno*, 1966). In particular, ideas of wave/mean-flow interaction, which are now integral to an understanding of atmospheric dynamics, entered the literature in the late 1960s (e.g. *Booker and Bretherton* (1967)) and were soon linked to the QBO.

1.2 Phenomenology

The oscillatory, downward-propagating nature of QBO wind regimes is perhaps best illustrated by a time-height cross-section of zonal wind over the equator; see figure 1. Observational data may be averaged over a monthly period to reduce noise, and similarly may be longitudinally averaged, though this is not necessary owing to the longitudinal symmetry. Therefore data from a single station is sufficient, as in

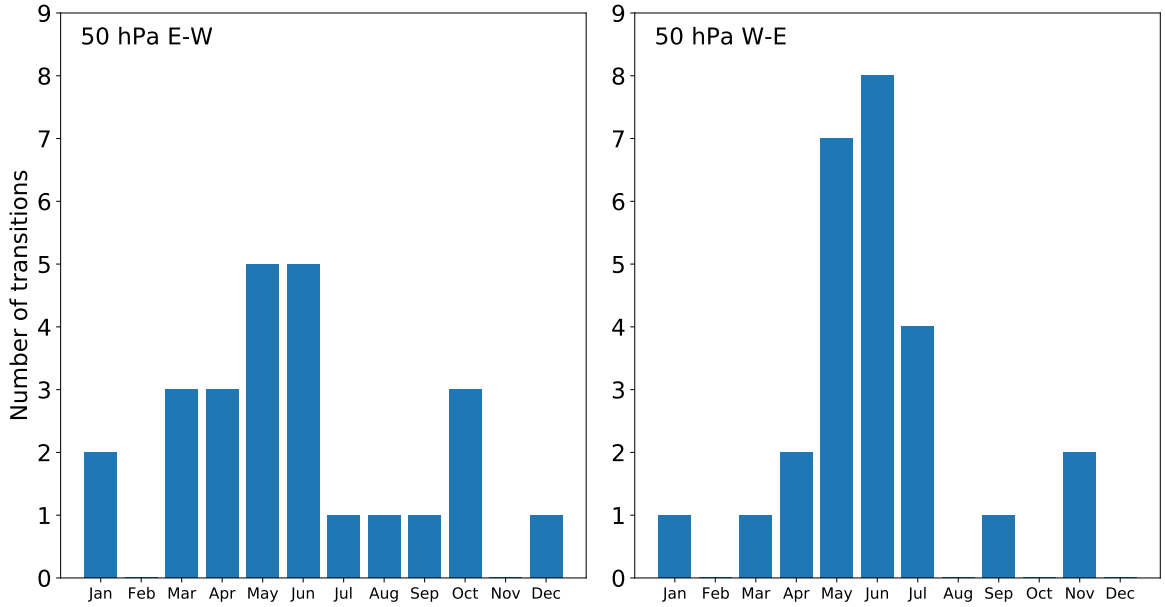


Figure 2: Seasonal distribution of QBO phase transition at 50 hPa. Data from Encyclopedia of Atmospheric Sciences (Second Edition), 2015.

the case of figure 1. A number of features are apparent from a time-height view:

- Alternating regimes of easterly and westerly[1] winds, repeating in intervals between 22–34 months at an average of 28 months from 1953–1995. The wind regimes are particularly well organised between 15–30km (100–10 hPa).
- Westerly shear zones[2] descend more regularly and rapidly than easterly shear zones.
- Easterly wind regimes exhibit greater intensity and longer duration.
- The amplitude of each QBO phase[3] is roughly constant as winds descend from 5 to 50 hPa and decreases rapidly below 50 hPa.
- There is an apparent seasonal preference in the phase reversal; see *Dunkerton* (1990). Using the 50 hPa level as reference, onset occurs mainly during Northern Hemisphere (NH) late spring: see figure 2.
- Above 35 km ($\sim$ 7 hPa), the zonal wind oscillates with an apparently annual period. This is the Semi-Annual Oscillation (SAO) of the upper stratosphere. It is a distinct phenomenon from the QBO, though caused by a similar wave forcing mechanism and related to the onset of QBO wind regimes *Gray* (2010).

Beyond the altitude structure, the QBO displays an approximately Gaussian meridional structure in zonal wind amplitude, with a 12.5° half-width; see *Wallace* (1973) and in particular figure 4 of *Gray* (2010). The precise meridional structure is dependent on the phase of the oscillation; easterly wind regimes are much wider than westerly regimes (*Pahlavan et al.*, 2021a) and exhibit a flatter base whilst westerly regimes exhibit U-shaped bases (*Hamilton*, 1984).

[1] The terms 'easterly' and 'westerly' refer to winds blowing *from* the east and west respectively. In the literature, these terms are used interchangeably with 'westward' (equivalent to easterly) and 'eastward' (equivalent to westerly), particularly with reference to wave speeds. Note also that by convention, westerly winds are positively signed and easterly winds are negatively signed.

[2] The term 'shear zone' refers to the region of zonal wind shear between easterly and westerly regimes. It is most straightforward to refer to the phase of the QBO in terms of an easterly or westerly shear zone since typically two wind regimes are present but only one shear zone of a particular sign. By convention, a westerly shear zone has $\overline{u}_z > 0$.

[3] The 'phase' of the QBO refers to the two distinct arrangements of wind regimes with oppositely signed zonal wind at some reference height (typically 50 hPa, lower stratosphere). A westerly (easterly) phase refers to the arrangement with a westerly (easterly) shear zone, i.e. westerly (easterly) winds in the mid-stratosphere and oppositely signed winds in the lower stratosphere.

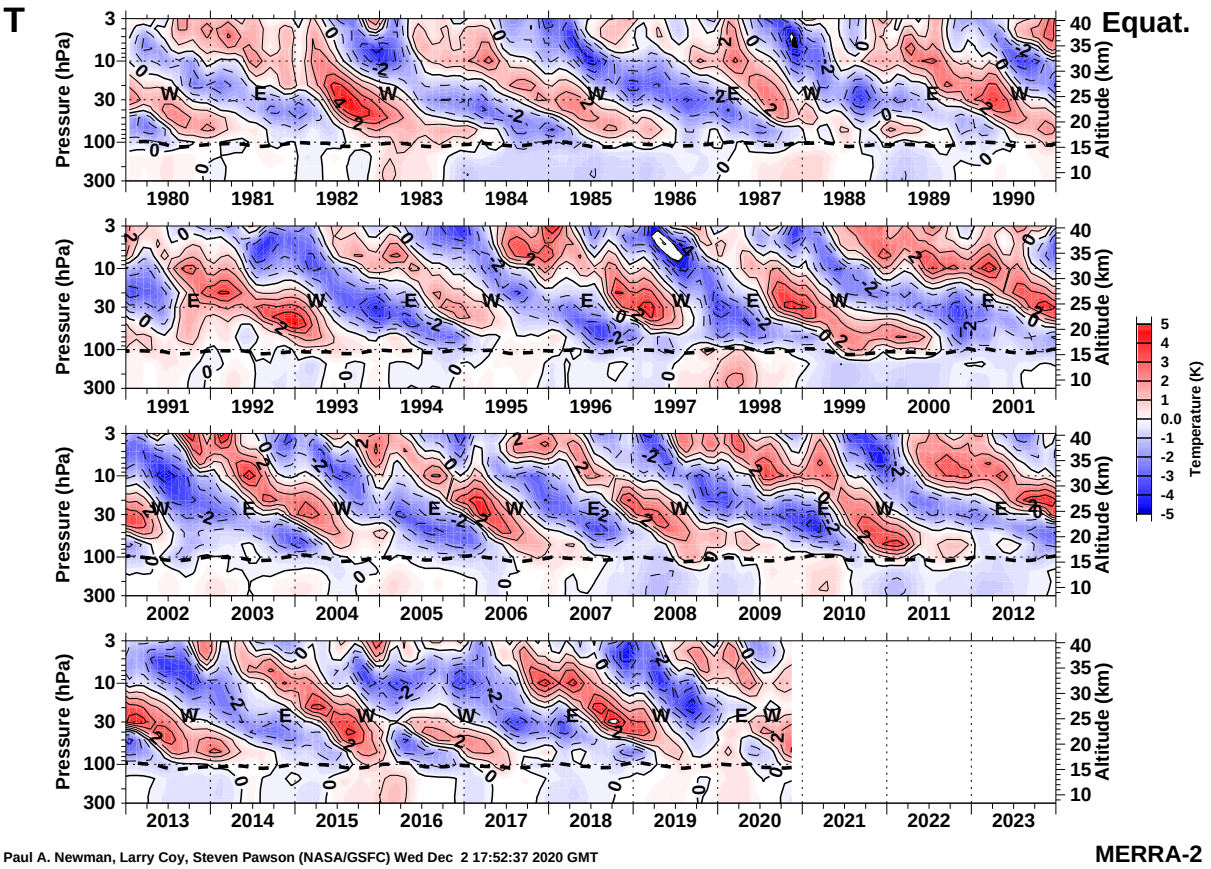


Figure 3: QBO temperature signal from MERRA-2 data. Plot generated by taking the monthly mean of zonal mean temperature, subtracting long-term monthly mean, then performing linear detrending of the residual temperature. Sourced from `https://acd-ext.gsfc.nasa.gov/Data_services/met/qbo/qbo.html`

The QBO has a detectable signal in many dynamical variables which couple with the zonal wind, the most apparent of which is the stratospheric temperature over the equator; see figure 3, later discussed in section 2.6. The amplitude of the temperature oscillation is small, approximately 4K at maximum near 30–50 hPa. Despite the low amplitude, these temperature perturbations penetrate into the tropical tropopause layer and modulate the coverage of high, thin cirrus clouds (*Tseng and Fu*, 2017); negative temperature anomalies favour greater relative humidity and consequently increased cloud cover, and vice versa.

A particularly important QBO signal arises in equatorial ozone concentration as well as other trace gases, which has consequences for their global circulation and hence their contribution to global climate change. This modulation is typically indirect, due to QBO-induced perturbations in temperature and advection (*Pahlavan et al.*, 2021a; *Baldwin et al.*, 2001).

2 Basic dynamics of the QBO

It is now understood that the QBO mechanism relies on interaction between waves and the mean flow, in particular the momentum transport of vertically propagating waves, with the key idea of separating dynamical variables into ‘mean’ and ‘wave’ (or ‘eddy’) components, where the mean is typically longitudinal. Due to some physical process which generates perturbations, energy and momentum can be transferred from the mean flow to the wave component, which may then propagate and redeposit energy and momentum into the mean flow. This provides a mechanism for momentum transport without advection by the mean flow. In this section we discuss the application of this idea to the QBO, first laying out the essential theory then detailing the development of a numerical model which captures the

essence of the mechanism giving rise to the QBO.

2.1 Early insights

Booker and Bretherton (1967) provided the first insight relevant to the QBO with their paper on the absorption of gravity waves at a 'critical level'. The main result is a mechanism by which internal gravity waves can be reabsorbed by the mean flow without the need for turbulence or other dissipative processes. Ignoring viscosity and heat conduction, a sinusoidal wave with horizontal phase velocity c encounters a 'critical level' $z = z_c$ where the horizontal mean flow matches the phase speed; $\overline{u}(z_c) = c$. At this level, the frequency of the wave relative to the surrounding fluid vanishes. Consequently, a wave propagating vertically through a critical level is strongly attenuated and its momentum is deposited in the mean flow.

Mathematically, the critival level for a wave equation of the form $\psi_{zz} + m^2(z)\psi = 0$ arises wherever $m(z)$ is singular. Physically, a wave packet approaching a critical level of its dominant frequency and wavenumber will not reach the critical level in finite time: whilst the intrinsic frequency and total wavenumber remain constant, the wave frequency *relative to the local fluid* vanishes and associated with this, the vertical wavelength decreases. Thus the group velocity becomes closer to horizontal, and all components relative to the local velocity field vanish. In this way, the wave is 'absorbed' by the mean flow.

Utilising the concept of critical levels, *Lindzen and Holton* (1968) suggested the QBO arises through an internal mechanism involving two-way feedback between waves and the mean flow; using a 2D model, they found that the QBO could be driven by a spectrum of vertically propagating waves with eastward and westward phase speeds. In their model, the oscillation is a downward propagating pattern of easterly and westerly winds. A key insight is that the period of oscillation is partly controlled by the wave momentum fluxes, hence a range of periods is possible and the fact that the QBO period is close to a multiple of the annual cycle is a coincidence. However, there is evidence to suggest there is some modulation by the annual cycle due to its influence on the background stratospheric flow (*Gray*, 2010).

Holton and Lindzen (1972) refined the model to demonstrate two-way feedback where the wave forcing is provided by eastward propagating equatorial Kelvin waves and westward propagating equatorial Rossby-gravity waves. The differing properties of these waves gives rise to a zonal asymmetry resembling that of the QBO. Further, *Andrews and McIntyre* (1976) and *Boyd* (1976) showed that the effects of rotation and meridional structure do not change the fundamental mechanism of the oscillation. However, these effects do have consequences for the momentum flux necessary to drive the QBO which is discussed in section 2.4.

2.2 Modelling framework

A natural pathway to understanding the QBO mechanism is a 'toy model' which captures the essence of the wave/mean-flow interaction and the important factors modifying the nature of the zonal wind oscillation. To do so first requires a mathematical framework for modelling the atmosphere.

2.2.1 Density variations & stratification

In the stratosphere, distinct variations in fluid density exist, but the total range of density is small. In this case, the predominant influence on dynamics is of changes in the weight of fluid per unit volume. Consequently the Boussinesq approximation is employed, whereby density changes are acknowledged only as part of the buoyancy force due to differences in fluid weight; to understand the dynamics, the buoyancy force in the vertical direction $\sigma = -g\rho'/\overline{\rho}$ is introduced, where $\rho'(\boldsymbol{x}, t)$ is the wave component of density and $\overline{\rho}(z)$ is the mean density. Formally, this is a decomposition of the density $\rho(\boldsymbol{x}, t) = \overline{\rho}(z) + \rho'(\boldsymbol{x}, t)$ where $\rho'/\overline{\rho} \ll 1$, hence wave/eddy components may also be referred to as 'perturbation' or 'deviation' components. The stratification of the mean flow is described in terms of the Brunt-Väisälä frequency N, usually assumed to be slowly varying in z and defined by

$$N^2(z) = -\frac{g}{\overline{\rho}}\frac{\mathrm{d}\overline{\rho}}{\mathrm{d}z} \tag{2.1}$$

in the case of *Booker and Bretherton* (1967), though definitions vary depending on the specific decomposition of ρ used.

The Boussinesq approximation is self-consistent only if the fluid is regarded as incompressible at leading order. The atmosphere is compressible, but the incompressible approximation is valid when the local vertical scale is small compared to the density scale height, though $\overline{\rho}$ must be replaced by the mean potential density[4] in this case. In the literature, this subtlety is often overlooked to avoid obfuscation and the buoyancy frequency is used with reference to the mean density, even if mean *potential* density is meant.

The importance of vertical variations in mean horizontal velocity $(\overline{u}, \overline{v}, 0)$ compared with ambient stratification is characterised by the Richardson number

$$\mathrm{Ri} = \frac{N^2}{\overline{u}_z^2 + \overline{v}_z^2} \tag{2.2}$$

which may be simplified to exclude $\overline{v}_z^2$ if only zonal velocity is under consideration, as will be the case henceforth. A flow with $\mathrm{Ri} \geq 1/4$ everywhere is stable to small disturbances (*Miles*, 1961; *Howard*, 1961) and Ri must be taken to be arbitrarily large for concepts such as wave packets and group velocity to be consistent.

2.2.2 Eddy effects

The effects of the eddy component of the flow on the mean component are typically associated with a wave activity conservation relation of the form

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \boldsymbol{\mathcal{F}} = \mathcal{D} \tag{2.3}$$

which relates a density of wave/eddy activity $\mathcal{A}$ with the divergence of the eddy flux $\boldsymbol{\mathcal{F}}$ and dissipative effects $\mathcal{D}$. The form of $\mathcal{A}, \mathcal{F}$ and $\mathcal{D}$ depends on the framework under which the relation is derived; a common form (which will be used henceforth) is the Eliassen-Palm (EP) wave activity and flux (*Eliassen and Palm*, 1961) which arises from the quasi-geostrophic framework in which the dynamics are assumed to be in near-geostrophic balance.

Combined with this wave activity relation, the transformed Eulerian mean (TEM) formalism is used as a useful framework for discussing eddy effects. This formalism combines the separate eddy forcing terms for each velocity component into a single forcing term and simplifies the buoyancy equation. Under the TEM formalism, the zonal momentum and buoyancy equations are

$$\frac{\partial \overline{u}}{\partial t} = f_0 \overline{v}^* + \nabla \cdot \boldsymbol{\mathcal{F}} + \overline{F} \tag{2.4}$$

$$\frac{\partial \sigma}{\partial t} = -N^2 \overline{w}^* + \overline{S} \tag{2.5}$$

where $\overline{F}$ and $\overline{S}$ represent frictional and heating terms respectively. It is evident from these equations that the zonally-averaged zonal velocity $\overline{u}$ is forced by the Coriolis force, frictional forces, and eddy forces represented by $\nabla \cdot \boldsymbol{\mathcal{F}}$, so that the divergence of the eddy flux can be considered as a force. In the literature and remainder of this essay, the EP flux divergence is used as a convenient proxy for the momentum deposition & consequent forcing of the mean flow by waves.

A crucial interpretation of the wave activity conservation relation (2.3) is the importance of disssipation. First note that if $\nabla \cdot \boldsymbol{\mathcal{F}} = 0$ then the TEM equations are independent of eddy effects. Now consider a region where dissipation is present: a wave/eddy entering this region has $\mathcal{A}_t > 0$ and hence (assuming dissipation initially small) $\nabla \cdot \boldsymbol{\mathcal{F}} < 0$ according to (2.3). Without dissipation, as the wave leaves then $\nabla \cdot \boldsymbol{\mathcal{F}} > 0$ (with amplitude identical to when arriving) and there is no net eddy forcing, with $\nabla \cdot \boldsymbol{\mathcal{F}} = 0$ in total. With dissipation present, $\mathcal{A}_t = \mathcal{D}$ hence $\nabla \cdot \boldsymbol{\mathcal{F}} = 0$ and in total, $\nabla \cdot \boldsymbol{\mathcal{F}} < 0$ so there is a net forcing and hence acceleration of the mean flow, captured by the divergence of eddy flux $\nabla \cdot \boldsymbol{\mathcal{F}}$ which is non-zero only if dissipation (or wave breaking) is present. Thus waves propagating through a fluid interact with the mean flow when they are dissipated, effectively depositing the momentum they carry and therefore accelerating the mean flow. The magnitude of this forcing is represented by the divergence of the EP flux as above.

[4]The mean potential density is the density that a fluid parcel at height z would have if it were reduced adiabatically to some reference pressure.

2.3 QBO mechanism

Based on the work of Lindzen and Holton in their 1968 and 1972 papers, the underlying dynamics and parameter dependence of the QBO was clarified and well illustrated by *Plumb* (1977) using a numerical 'toy model' which we now elaborate on. Consider 2D motion in the (x, z) plane of a Boussinesq fluid with buoyancy frequency N, viscosity ν and thermal dissipation rate μ.

2.3.1 Governing equations

Owing to (approximate) incompressibility following from the Boussinesq assumption, we may employ a streamfunction ψ such that $(u, w) = (\psi_z, -\psi_x)$. The dynamics are governed by two coupled equations, firstly the vorticity equation

$$\frac{\partial}{\partial t}\nabla^2\psi + \frac{\partial \sigma}{\partial x} - \nu\nabla^4\psi = J(\psi, \nabla^2\psi) \tag{2.6}$$

where $\sigma = -g\rho'/\overline{\rho}$ is the buoyancy with the mean density component $\overline{\rho} = \rho_0$ assumed constant and $\nabla^2 \equiv \partial_x^2 + \partial_z^2$. Here, J is the Jacobian defined by $J(a, b) = a_x b_z - a_z b_x$. We also have the governing equation for buoyancy as

$$\frac{\partial \sigma}{\partial t} - N^2\frac{\partial \psi}{\partial x} + \mu\sigma = J(\psi, \sigma) \tag{2.7}$$

Following the standard procedure for investigating wave/mean-flow interaction dynamics we decompose the streamfunction ψ into mean and perturbation components

$$\psi = \overline{\psi} + \psi' \tag{2.8}$$

where an overbar denotes the zonal mean. We wish to derive an equation for the response of the mean zonal flow $\overline{u}$ to wave forcing. We will employ the following assumptions.

(i) Apply the 'mean field approximation' following *Herring* (1963) whereby the perturbation component

$$J(a', b') - \overline{J(a', b')} \tag{2.9}$$

is neglected. This component represents non-linear wave self-interaction, which is of a frequency twice that of the basic wave so produces a small response in comparison. Equivalently, this is a third-order quantity, whilst we will retain only first order perturbation quantities.

(ii) Neglect viscous dissipation of the wave, on the basis that thermal dissipation is more effective in this problem. Note that viscous dissipation of the mean flow is still present.

(iii) Changes in the mean flow occur on a timescale much larger than the timescale of wave motion. Consequently, we neglect terms involving $\frac{\partial \overline{u}}{\partial t}$.

(iv) The thermal dissipation rate μ is small compared to the doppler shifted frequency, in particular $\mu \ll k(\overline{u} - c)$.

(v) The Richardson number $\mathrm{Ri} \equiv N^2/(\partial\overline{u}/\partial z)^2$ is large, representing the assumption buoyancy dominates the mean flow shear in driving wave motion.

(vi) Ambient stratification N^2 is unaltered by wave activity, which follows from the assumptions (iii), (iv) on Ri and μ.

(vii) As is realistic in the atmosphere, the aspect ratio $D/L \ll 1$ where D and L are characteristic vertical and horizontal lengthscales respectively.

The decomposition $\psi = \overline{\psi} + \psi'$ applied to the vorticity equation, retaining first order disturbance quantities only, yields

$$\frac{\partial}{\partial t}\left(\nabla^2\psi' + \frac{\partial \overline{u}}{\partial z}\right) + \frac{\partial \sigma}{\partial x} - \nu\nabla^4\left(\overline{\psi} + \psi'\right) = \frac{\partial \psi'}{\partial x}\frac{\partial^3\overline{\psi}}{\partial z^3} - \frac{\partial \overline{\psi}}{\partial z}\frac{\partial}{\partial x}\nabla^2\psi' \tag{2.10}$$

By assumption (iii) we neglect $\partial^2\overline{u}/\partial t\partial z$ and by assumption (ii) we neglect $\nu\nabla^4\psi'$ to get

$$\left[\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right]\nabla^2\psi' + \frac{\partial \sigma}{\partial x} = \frac{\partial \psi'}{\partial x}\frac{\partial^2\overline{u}}{\partial z^2} \tag{2.11}$$

The same decomposition applied to the buoyancy equation combined with the mean field approximation gives

$$\frac{\partial \sigma}{\partial t} - N^2 \frac{\partial \psi'}{\partial x} + \overline{u}\frac{\partial \sigma}{\partial x} = -\mu\sigma \tag{2.12}$$

The coupled equations (2.11) and (2.12) admit plane wave solutions of the form

$$\psi' = \Re\left[\hat{\psi}(z)e^{ik(x-ct)}\right], \qquad \sigma = \Re\left[\hat{\sigma}(z)e^{ik(x-ct)}\right] \tag{2.13}$$

Substituting into (2.11) and (2.12) and eliminating $\hat{\sigma}$ yields a governing equation for $\hat{\psi}$

$$\frac{\mathrm{d}^2\hat{\psi}}{\mathrm{d}z^2} + \left[\frac{N^2(1+i\mu/k(\overline{u}-c))}{(\overline{u}-c)^2+\mu^2/k^2} - k^2 - \frac{1}{\overline{u}-c}\frac{\partial^2\overline{u}}{\partial z^2}\right]\hat{\psi} = 0 \tag{2.14}$$

We now simplify the factor in square brackets by applying assumption (v) to neglect $\partial^2\overline{u}/\partial z^2$ (equivalently, take the leading order contribution after expansion in orders of $\mathrm{Ri}^{-1/2}$), assumption (iv) to neglect μ^2/k^2, and assumption (vii) to neglect k^2. This amounts to a WKB approximation which can be applied to

$$\frac{\mathrm{d}^2\hat{\psi}}{\mathrm{d}z^2} + m(z)^2\hat{\psi} = 0 \tag{2.15}$$

where now

$$m(z) \approx \left(\frac{N^2(1+i\mu/k(\overline{u}-c))}{(\overline{u}-c)^2}\right)^{1/2} \approx \frac{N}{\overline{u}-c}\left[1+\frac{i\mu}{2k(\overline{u}-c)}\right] \tag{2.16}$$

The WKB solution representing upward propagating waves is

$$\hat{\psi}(z) = Am(z)^{-1/2}\exp\left(is\int^z m(z')\,\mathrm{d}z'\right) \tag{2.17}$$

where A is an arbitrary constant determined by boundary conditions and $s = \mathrm{sgn}(c_g^z)$ is the sign of the vertical group velocity. The vertical flux of zonal momentum associated with this wave is

$$\overline{F}(z) = \overline{u'w'} = -\frac{k}{2}\Re\left[i\hat{\psi}_z\hat{\psi}^*\right] \tag{2.18}$$

$$= -\frac{ik}{4}\left(\hat{\psi}\frac{\mathrm{d}\hat{\psi}^*}{\mathrm{d}z} - \hat{\psi}^*\frac{\mathrm{d}\hat{\psi}}{\mathrm{d}z}\right) \tag{2.19}$$

$$= \overline{F}(0)\exp\left[i\int_0^z (m-m^*)\,\mathrm{d}z'\right] \tag{2.20}$$

$$= \overline{F}(0)\exp\left[-s\int_0^z \frac{N\mu}{k(\overline{u}-c)^2}\,\mathrm{d}z'\right] \tag{2.21}$$

Note that we have $\mathrm{sgn}(\overline{F}(0)) = \mathrm{sgn}(c)$ for upward propagating waves. The integrand is the wave attenuation rate, equal to the dissipation rate divided by the vertical group velocity. Physically, this is the loss of wave flux intensity per unit distance of wave propagation. As the wave approaches the critical line of the flow where $\overline{u} = c$, the effects discussed in section 2.1 take place and the wave is absorbed by the mean flow, depositing its momentum. The process of wave absorption close to a critical line is captured mathematically here by the fact the integrand (i.e. the attenuation rate) is greatest where $\overline{u}-c$ is small and singular when $\overline{u} = c$.

The forcing provided by the wave momentum flux exerts an acceleration on the mean flow, assuming μ is non-zero (i.e. dissipation is present, see discussion in section 2.2.2). Accounting for viscous dissipation of the mean flow and generalising for multiple waves, we have

$$\frac{\partial \overline{u}}{\partial t} = -\sum_n \frac{\partial \overline{F}_n}{\partial z} + \nu\frac{\partial^2\overline{u}}{\partial z^2} \tag{2.22}$$

where n is used as an index for the individual waves. We see from (2.22) that an eastward wave with phase speed $c > 0$ exerts a positive (eastward) acceleration and vice versa. Note that in representing multiple

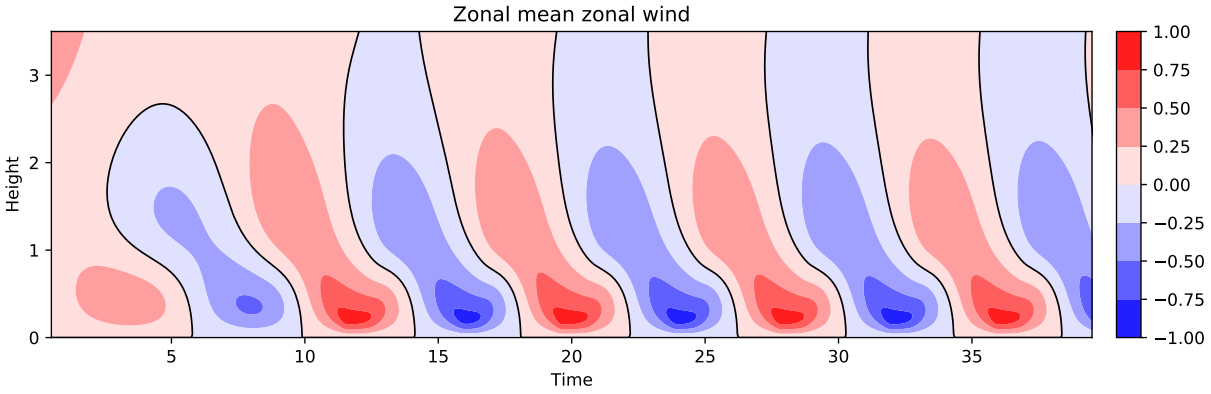


Figure 4: Numerical integration of (2.27)–(2.29) with boundary conditions $\overline{u}(\zeta, 0) = \zeta/10$, $\overline{u}_\tau(4, \tau) = 0$ where the upper boundary is $\zeta = 4$ and 100 spatial grid points are used. Wave forcing is provided by 2 waves with (non-dimensionalised) amplitude 1, wavenumber 1, and phase speeds ± 1. Integrated using 3$^{\text{rd}}$ order Adam-Bashforth method. See *Vallis* (2017) for a similar implementation.

waves by simply adding the momentum flux shear $\partial \overline{F}_n/\partial z$, we are implicitly neglecting interaction between the waves. Physically, this is acceptable since the timescale of such interaction is that of the wave period, assumed to be much smaller than the timescale of the mean flow variation, and consequently the response of the mean flow would be small.

The governing equation may be non-dimensionalised by definining

$$F_n(0) = \frac{\overline{F}_n(0)}{\hat{F}} \tag{2.23}$$

$$\alpha = \frac{N\mu}{\hat{N}\hat{\mu}} \tag{2.24}$$

$$\Lambda = \frac{\nu \hat{N}\hat{\mu}}{\hat{k}\hat{c}\hat{F}} \tag{2.25}$$

where the characteristic scale for each quantity is denoted with a hat. We then use non-dimensional vertical and time coordinates

$$\zeta = \frac{\hat{N}\hat{\mu}}{\hat{k}\hat{c}^2} z, \qquad \tau = \frac{\hat{N}\hat{\mu}\hat{F}}{\hat{k}\hat{c}^3} t \tag{2.26}$$

so that the governing equation is

$$\frac{\partial \overline{u}}{\partial \tau} - \Lambda \frac{\partial^2 \overline{u}}{\partial \zeta^2} = -\sum_n \frac{\partial F_n}{\partial \zeta} \tag{2.27}$$

where

$$F_n(\zeta) = F_n(0) \exp\left[-s \int_0^\zeta g_n(\zeta')\, \mathrm{d}\zeta'\right] \tag{2.28}$$

$$g_n(\zeta) = \frac{\alpha}{k_n(\overline{u} - c_n)^2} \tag{2.29}$$

The non-dimensionalised mean zonal flow governing equation (2.27) may be integrated numerically. Necessarily, the domain includes a vertical boundary which ideally would be at infinity to prevent a reflected wave which invalidates the WKB approach, but the vertical boundary may be chosen suitably large that the WKB approximation still holds. An example of the numerically simulated QBO under the Plumb model with forcing from two waves with opposite phase speeds can be seen in figure 4.

2.3.2 Oscillation mechanism

Numerical integration of the Plumb model (figure 4) clearly demonstrates downward propagating alternate wind regimes qualitatively similar to the QBO. It is therefore natural to consider an analytic

view of the governing equations (2.27)–(2.29) to understand the essential mechanism responsible for this oscillation and hence that of the QBO.

In the case of forcing by a single wave, Plumb found that for $0.005 \leq \Lambda \leq 1$, i.e. all but extremely small dissipation rates, a steady state is eventually reached where viscous dissipation balances wave-driven acceleration. The transient evolution is a jet, forming either eastward or westward according to the phase speed of the wave forcing, which is diffused vertically so that the steady state has a strong shear layer near $\zeta = 0$ and for large z, we have $\overline{u} \to (1+\Lambda)^{-1}$. In this case, the single source of forcing is not opposed by other forcing, only limited by viscous dissipation, hence no oscillation arises.

When a second wave is added, we see from the following linear stability analysis that a steady state is possible but unstable to perturbations[5]. We will then see that this instability gives rise to oscillation. Consider a static basic state $\overline{u} = 0$ with $F_1(0) = -F_2(0) = F$, $k_1 = k_2 = k$, $c_1 = -c_2 = c$ subject to a small perturbation

$$|\overline{u}| \ll 1, \qquad \left| \int \overline{u} \, \mathrm{d}\zeta' \right| \ll 1 \tag{2.30}$$

The wave attenuation rate $g_n(\zeta)$ becomes

$$g_n(\zeta) = \frac{\alpha}{k_n}(c_n - \overline{u})^{-2} \tag{2.31}$$

$$= \frac{\alpha}{k_n c_n^2}\left(1 + \frac{2\overline{u}}{c_n} + \dots\right) \tag{2.32}$$

Consequently the momentum flux under such a perturbation is

$$F_n(\zeta) = F_n(0) \exp\left[-\frac{s\alpha}{k_n c_n^2}\left(\zeta + \frac{2}{c_n}\int_0^\zeta \overline{u} \, \mathrm{d}\zeta' + \dots\right)\right] \tag{2.33}$$

$$\approx F_n(0) \exp\left[-\frac{s\alpha\zeta}{k_n c_n^2}\right]\left(1 - \frac{2\alpha s}{k_n c_n^3}\int_0^\zeta \overline{u} \, \mathrm{d}\zeta' + \dots\right) \tag{2.34}$$

Hence the mean flow perturbation is governed by

$$\frac{\partial \overline{u}}{\partial \tau} - \Lambda \frac{\partial^2 \overline{u}}{\partial \zeta^2} = \frac{4\alpha F}{kc^3} \exp\left[-\frac{\alpha\zeta}{kc^2}\right]\left(\overline{u} - \frac{\alpha}{kc^2}\int_0^\zeta \overline{u} \, \mathrm{d}\zeta'\right) \tag{2.35}$$

We now define the mean stratification

$$\phi = \int_0^\zeta \overline{u} \, \mathrm{d}\zeta' \tag{2.36}$$

so that

$$\frac{\partial \phi}{\partial \tau} = \frac{4\alpha F}{kc^3}\phi \exp\left[-\frac{\alpha\zeta}{kc^2}\right] + \Lambda\left[\frac{\partial^2 \phi}{\partial \zeta^2}(\zeta, \tau) - \frac{\partial^2 \phi}{\partial \zeta^2}(0, \tau)\right] \tag{2.37}$$

The purpose of introducing ϕ is that (2.37) is generally easier to solve analytically than (2.35). However, it is easiest to describe the dynamics in the context of (2.35).

The governing equation (2.35) relates the dynamics of the mean flow $\overline{u}$ at some level ζ_0 to the mean flow profile in $\zeta \leq \zeta_0$. Hence the behaviour at any level is independent of the flow profile above it, and it suffices to discuss the influence of the flow profile below said level. Plumb explicitly demonstrates this numerically by integrating (2.37) in a standard case as well as with $\phi = 0$ in $\zeta > \zeta_0$ (for arbitrary ζ_0). The flow evolution in $\zeta \leq \zeta_0$ is identical in each case.

The evolution of the mean flow may be split into two phases based on the importance of viscosity. In the initial phase, viscosity is unimportant so we first consider the inviscid dynamics with initial conditions where the mean flow profile is a small westerly (positive) perturbation; see figure 5(a). For simplicity, we also set $k = c = 1$, $F = 1/2$ and $\alpha = 1$. The inviscid governing equations are then

$$\frac{\partial \overline{u}}{\partial \tau} = 2e^{-\zeta}\left(\overline{u} - \int_0^\zeta \overline{u} \, \mathrm{d}\zeta'\right) \tag{2.38}$$

$$\frac{\partial \phi}{\partial \tau} = 2\phi e^{-\zeta} \tag{2.39}$$

[5]Note that whilst the model includes only two distinct waves, the discussion references multiple waves with a spectrum of phase speeds, to aid relevance to the QBO.

with general solution

$$\phi(\zeta,\tau) = \phi(\zeta,0)\exp\left[2\tau e^{-\zeta}\right] \tag{2.40}$$

Therefore ϕ is amplified with a growth rate decaying with height. This 'differential amplification' results in the downward propagation of ϕ maxima, which are zeroes of $\overline{u}$. In the context of (2.38), for an initial positive perturbation;

- at small ζ, the 'local attenuation term' $\overline{u}e^{-\zeta}$ dominates and the positive flow is amplified;
- at large ζ, the 'shielding term' $-e^{-\zeta}\int_0^\zeta \overline{u}\,\mathrm{d}\zeta'$ dominates and the acceleration is negative regardless of the sign of $\overline{u}$ at high levels.

Note that the local attenuation term is so named as it represents the amplification of the mean flow resulting from asymmetric attenuation of the wave forcing. The shielding term represents the effect of wave absorption by the mean flow near critical lines, which shields regions above the critical line from the wave forcing. It is interaction between these terms which is crucial to the development of oscillation.

Early in the evolution, the shielding term creates a region of easterly (negative) flow at large ζ, which is amplified and extends in ζ. Below this region, the initial westerly perturbation is amplified by the local attenuation term to form a westerly jet; see figure 5(b). A zero of $\overline{u}$ separates the two opposing wind regimes, at which point only the shielding term contributes, so the zero moves downward and the westerly jet narrows. At upper levels, the easterly jet broadens as the westerly jet inhibits waves with westerly wave speeds better than the easterly jet inhibits waves with easterly wave speeds, owing to the greater amplitude of the westerly jet. This is captured by the negative acceleration of the shielding term. The westerly jet has greater amplitude since the amplification by the local attenuation term, i.e. the local deposition of westerly momentum as waves with westerly phase speeds encounter their critical level, is focused in an increasingly narrow range of ζ.

The change of sign in wave-driven acceleration occurs at level $\zeta = \zeta_0$ defined by

$$\overline{u}(\zeta_0) = \int_0^{\zeta_0} \overline{u}\,\mathrm{d}\zeta' \tag{2.41}$$

Below $\zeta = \zeta_0$, the acceleration $\overline{u}_\tau$ is positive and hence the shielding term increases in magnitude with time, whilst at $\zeta = \zeta_0$, the acceleration vanishes and $\overline{u}_\zeta < 0$ since this level lies between the jets, in a region of easterly shear. To preserve the balance (2.41), ζ_0 descends. Thus we see the downward propagation of the change in sign of wave forcing is a consequence of the flow evolution at lower levels. In the absence of viscous dissipation, the evolution described above continues indefinitely. The westerly jet amplifies and narrows, with westerly wave forcing confined to an increasingly narrow layer, whilst the easterly jet amplifies and broadens; see figure 5(c).

In the case $\Lambda > 0$ with viscous dissipation, the dynamics are modified only where $\Lambda\,|\overline{u}_{\zeta\zeta}|$ is non-negligible; in such regions, the viscous term acts to limit maxima of $\overline{u}$. The evolution up to the establishment of a narrow jet at lower levels is therefore qualitatively the same, only with reduced shear.

As the westerly jet narrows, the $\overline{u}_{\zeta\zeta}$ term eventually becomes locally dominant (compared to wave driving) and viscosity acts to dissipate the jet, whilst the easterly jet continues to broaden and amplify; see figure 5(d). Note that viscosity is unimportant at this stage for the easterly jet as the broadness means $\overline{u}_{\zeta\zeta}$ is small. Once the westerly jet is largely dissipated by viscous diffusion, the acceleration at large ζ becomes positive: in terms of (2.38), the factor $\int_0^\zeta \overline{u}\mathrm{d}\zeta'$ of the shielding term becomes negative in the absence of a westerly (positive) jet at low levels, so the acceleration changes sign to become positive. Similarly, at low levels $\overline{u}$ becomes largely negative so the local attenuation term provides negative acceleration.

The dynamics now proceed in a similar but opposite manner, with a new zero of $\overline{u}$ established at upper levels and consequently an upper westerly jet forms which broadens and amplifies, whilst the (now) lower easterly jet amplifies and narrows. Mathematically, the governing equation (2.35) is antisymmetric with respect to $\overline{u}$ in the equal and opposite wave forcing case, i.e. if $\overline{u}$ is subject to positive acceleration, then $-\overline{u}$ is subject to negative acceleration, so in either case the same amplification of $\overline{u}$ arises. Hence the above explanation of the dynamics applies to each phase in a qualitatively similar but opposite way. Physically, the stronger easterly jet now inhibits waves with easterly phase speeds better than the reduced westerly jet inhibits waves with westerly phase speeds, hence the wave forcing at upper levels drives positive acceleration, with negative acceleration at low levels, in the opposite sense to the previous phase.

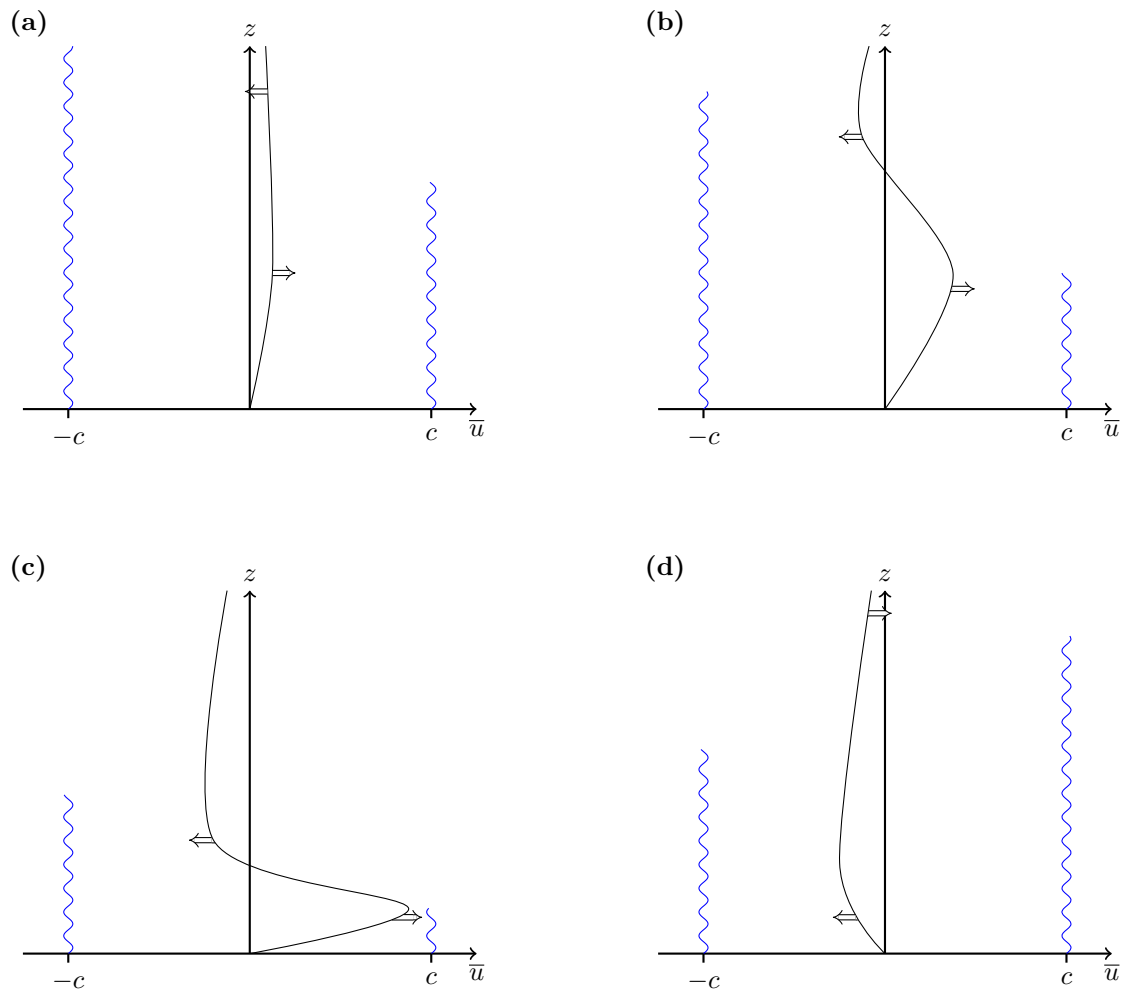


Figure 5: Representation of the mean flow evolution in Plumb's QBO model. One half-cycle is shown. Double arrows indicate wave-driven acceleration, wavy lines broadly indicate the penetration of westward and eastward waves. Adapted from *Plumb* (1984).

The above discussion constitutes one half-cycle of the QBO, arising from the two-way feedback between momentum deposition by waves dissipating close to critical lines and the resulting mean flow acceleration which modifies said critical lines. To summarise:

1. As waves propagate vertically through the stratosphere, they are damped via mechanical dissipation, thermal dissipation, or otherwise. Such dissipation results in momentum deposition and local acceleration of the mean flow in the direction of the dissipated waves' phase speed.
2. Dissipation of waves is amplified close to critical layers, so that the mean flow inhibits the vertical propagation of waves with phase speed c where $|\overline{u} - c| \ll 1$.
3. Some perturbation of the static mean flow basic state experiences differential amplification due to asymmetric dissipation of waves, as above.
4. Oppositely signed jets form, which are amplified by wave forcing. The upper level jet broadens and the lower level jet narrows due to downward propagation of the effective wave forcing resulting from feedback by the mean flow.
5. The narrowing jet is eventually removed by viscous dissipation, changing the sign of the differential amplification of the mean flow and hence setting up the next phase of oscillation.

Plumb found that this instability ceases at $\Lambda = 0.112$, since viscosity becomes strong enough to dissipate perturbations before any interesting dynamics occur. Over the full cycle, the effect of viscosity is largely unimportant (as $|\overline{u}_{\zeta\zeta}|$ is small throughout most of the domain), hence the only relevant timescale is that

used to define τ, namely

$$T = \frac{\hat{k}\hat{c}^3}{\hat{N}\hat{\mu}\hat{F}} \tag{2.42}$$

Thus the period is proportional to the timescale T, i.e. $P = \eta T$, with the constant of proportionality η determined numerically. Plumb found $\eta \approx 14$. Note that the period P is not directly dependent on the wave periods, but *is* dependent on the wave momentum flux amplitudes.

In the full non-linear case, the dynamics are once again qualitatively the same, except $\overline{u}$ is confined to the range $c_2 < \overline{u} < c_1$, so no critical levels which fully dissipate waves develop, but momentum deposition via attenuation is present in the same manner as the linearised case. This is consistent with the result that under the WKB approximation and assuming the mean flow is approximately steady, forcing by internal waves cannot generate a critical level if none exists originally (*Plumb*, 1975). However, the maxima of $\overline{u}$ do approach the wave forcing phase speeds asymptotically, resulting in almost complete dissipation of some waves.

The actual stratospheric QBO demonstrates clear asymmetry in wind amplitude as well as other characteristics, which can be imitated in Plumb's model. In the case of unequal wavevectors or unequal phase speeds, the vertical penetration of each phase of the oscillation is asymmetric. Since $\overline{u}$ is restricted to the range $c_2 < \overline{u} < c_1$, the flow maxima are also asymmetric in the case of unequal phase speeds. Asymmetry also arises when the momentum fluxes are unequal; if $F_1 > 0$ is larger in magnitude than $F_2 < 0$ (where F_1 (F_2) is associated with a westerly (easterly) phase speed), the positive flow regime moves downward faster and with stronger associated shear. Consequently, the westerly (positive) jet is longer lived and as a result of shielding, the higher level flow is largely negative. This is reflective of the real QBO; see figure 1. Note that *Plumb and McEwan* (1978) demonstrated that the mechanism in *Plumb* (1977) may be realised physically in a laboratory experiment.

2.3.3 Model limitations & extensions

The model described thus far assumes a Boussinesq fluid with uniform ambient density. This is not strictly valid in the mid-atmosphere; the decay of mean density with height gives rise to stratification, and has distinct effects on the influence of momentum flux and consequently the vertical structure of the oscillation. Following *Lindzen* (1971), we may adopt a Boussinesq approximation locally while letting the mean density depend on ζ, so that (2.27) becomes

$$\frac{\partial \overline{u}}{\partial \tau} - \Lambda \frac{\partial^2 \overline{u}}{\partial \zeta^2} = -\frac{\overline{\rho}(0)}{\overline{\rho}(\zeta)} \sum_n \frac{\partial F_n}{\partial \zeta} \tag{2.43}$$

with attenuation rates g_n and momentum fluxes F_n unchanged. Further, assuming an isothermal atmosphere, we have $\overline{\rho}(\zeta) = \overline{\rho}(0)e^{-\varepsilon\zeta}$ where $\varepsilon = h/H$ with $h = \hat{N}\hat{\mu}/\hat{k}\hat{c}^2$ the height scale and H a constant. Formally, this is an application of the method of multiple scales. Then (2.27) becomes

$$\frac{\partial \overline{u}}{\partial \tau} - \Lambda \frac{\partial^2 \overline{u}}{\partial \zeta^2} = -e^{\varepsilon\zeta} \sum_n \frac{\partial F_n}{\partial \zeta} \tag{2.44}$$

In the limit $\varepsilon \to 0$ we recover the previous system. In general, the effect of non-zero ε is increased acceleration of the mean flow at high levels (where $\varepsilon\zeta$ is non-negligible) and hence the vertical scale of the oscillation is increased. Also, the magnitude of viscous effects relative to wave forcing is reduced at high levels, resulting in two changes to the dynamics. Firstly, the period of oscillation is physically determined by the time taken for the interior shear layer in the lower jet to become narrow enough for viscous diffusion to become significant. For small viscosity the layer is located at small ζ where the effects of non-zero ε are small. However, as the viscosity Λ increases, the interior shear layer moves to greater ζ where acceleration is more rapid under non-zero ε, so the period decreases. Secondly, the transition to stability occurs at larger Λ, since the transition occurs when diffusion becomes dominant throughout the interior of the flow, and $\varepsilon > 0$ decreases the effective viscosity.

In addition to density variations, it is more realistic to use a variable damping coefficient α defined by

$$\alpha = \begin{cases} 0.55 + 0.56\zeta & \zeta \leq 1.97 \\ 1.65 & \zeta > 1.97 \end{cases} \tag{2.45}$$

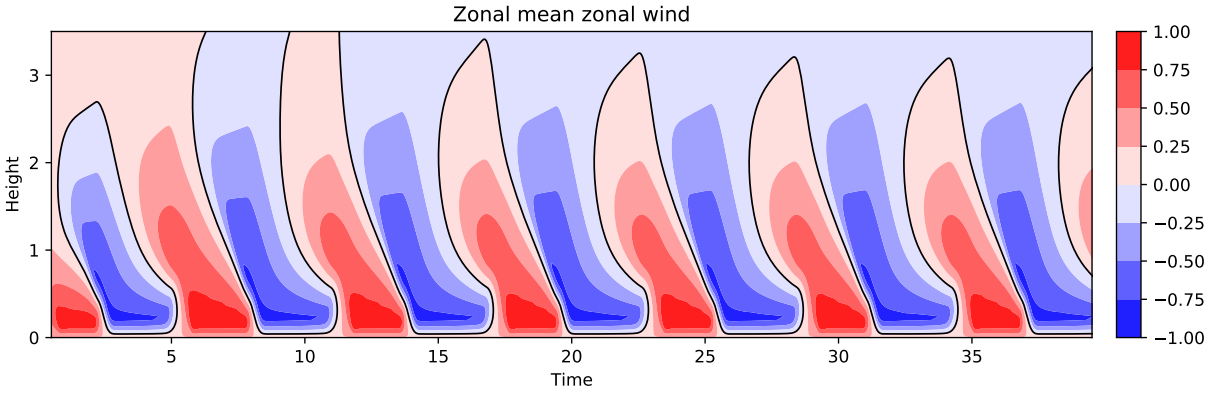


Figure 6: Numerical integration of (2.27), (2.28) and (2.47), (2.48) with variable damping α defined by (2.45) and boundary conditions $\overline{u}(\zeta, 0) = \zeta/10$, $\overline{u}_\tau(4, \tau) = 0$ with the upper boundary at $\zeta = 4$ and 100 spatial grid points. Integrated using 3^{rd} order Adam-Bashforth method.

which is a better representation of the actual damping coefficient in the atmosphere compared with a constant α. The precise form of damping coefficient and its variation with height remains a factor in generating a realistic QBO in modern models (*Dunkerton*, 1997). For this dissipation profile, the level $\zeta = 0$ corresponds to the level $z = 17$km in the real atmosphere: see *Holton and Lindzen* (1972) for further details. In the model of Holton & Lindzen, effects such as the semi-annual oscillation (SAO) are also included via parameterisation. The SAO is dominant in the upper stratosphere, as noted in section 1.2 and may be parameterised by a forcing term

$$G(\zeta, \tau) = \begin{cases} 0 & \zeta \leq 1.66 \\ 4.37(\zeta - 1.66)\sin(9.76\tau) & \zeta > 1.66 \end{cases} \tag{2.46}$$

added to the right-hand side of (2.44). At lower levels, the wave-driven oscillation remains dominant. More complex numerical models of the QBO need not parameterise the SAO as it is driven by a qualitatively similar mechanism to that of the QBO, in particular wave forcing is provided by waves with phase speeds lying outside the range of QBO wind speeds (*Gray*, 2010), hence the SAO may arise spontaneously with suitable representation of these waves.

Perhaps the most crucial alteration necessary to represent a more realistic QBO is the type of waves which force the mean flow. It was first thought that the QBO forcing is provided by eastward propagating Kelvin waves and westward propagating Rossby-gravity waves, in which case the attenuation rates are modified to

$$g_K = \frac{\alpha}{k_K(\overline{u} - c_K)^2} \tag{2.47}$$

$$g_{RG} = \frac{\alpha}{k_{RG}(\overline{u} - c_{RG})^2}\left[\frac{\beta'}{k_{RG}^2(\overline{u} - c_{RG})} - 1\right] \tag{2.48}$$

where $\beta' = \beta/\hat{k}^2\hat{c}$ is the modified beta parameter. The different attentuation rates introduce an asymmetry in the oscillation; the easterly phase of the simulated QBO has greater accelerations and greater vertical penetration. The precise wave forcing spectrum is an active area of research which is discussed in the following subsection.

An example of a simulated QBO with variable damping coefficient and forcing provided by Rossby-gravity and Kelvin waves can be seen in figure 6. Note that density variation and SAO forcing is not included in this simulation. This relatively simple theory of the QBO mechanism accounts for alternating regimes of easterly and westerly winds and the downward propagation of such shear zones. The asymmetric nature of the wind regimes, for example the increased amplitude of easterlies and the faster descent of westerlies, is well represented with the inclusion of complicating factors such as rotational effects (i.e. quasi-geostrophic equatorial wave theory from *Matsuno* (1966)), which alter the attenuation rates of particular waves, as well as asymmetries in phase speeds and wavenumbers. Some aspects of the QBO are not enlightened by this model, for example the irregular descent of easterly wind regimes, seasonal onset preference, and meridional structure.

2.4 Wave forcing spectrum & momentum budget

The theory developed by *Lindzen and Holton* (1968), *Holton and Lindzen* (1972) and *Plumb* (1977) attributes the wave forcing driving the QBO to Kelvin and Rossby-gravity waves. Whilst these waves succesfully generate a QBO in 1D models, *Gray and Pyle* (1999) found that the momentum flux from these waves alone is insufficient to drive a realistic QBO in a more complete 2D radiative-dynamical-photochemical model. Further, *Dunkerton* (1991, 1997) and *McIntyre* (1994) note that the equatorial arm of the Brewer-Dobson circulation (BDC), also referred to as the residual meridional overturning circulation (RMOC), creates upwelling of around 1km per month in the tropics which roughly doubles the momentum flux necessary to drive the QBO as observed.

A wide spectrum of atmospheric waves exist in the tropics (*Matsuno*, 1966) and it is now thought that the wave-driving of the QBO is provided by a spectrum of these equatorial waves. It has also been suggested that planetary-scale Rossby waves propagating equatorward from the winter hemisphere may influence the upper levels of the QBO when the SAO is in its easterly phase (*Ortland*, 1997), whilst the lower region of the QBO around 20–23km is relatively well shielded from these waves (*O'Sullivan*, 1997). Nonetheless, the majority of the momentum flux is contributed by waves which are predominantly generated by deep convection in the tropics and propagate upwards. Those relevant to the QBO (*Dunkerton*, 1997) are vertically propagating waves with either:

- slow vertical group propagation, which are absorbed due to radiative or mechanical damping at a rate such that their momentum is deposited at QBO altitudes, or;
- faster vertical group propagation, which are absorbed at critical levels that lie within the range of QBO zonal wind speeds.

Waves with very short horizontal wavelengths on the order of 10km and less are thought to be vertically trapped in the troposphere close to the altitude at which they are generated so these are often disregarded (*Baldwin et al.*, 2001). It should be noted however that short wavelength gravity waves are known to make significant contributions to the QBO momentum budget and characterising the contribution from smaller wavelengths is difficult due to constraints in the spatial resolution of observational data.

It is difficult to constrain the period of waves which may contribute to the QBO as long-period waves dominate spectra of zonal wind and temperature, but higher frequency waves contribute more to momentum fluxes than may be expected from temperature considerations. The QBO wave-driving is now attributed to three categories of waves (*Gray*, 2010):

1. Equatorially-trapped Kelvin and Rossby-gravity waves with eastward and westward phase speeds respectively; zonal wavenumbers ∼1–4; zonal wavelengths $\geq 10^4$km; periods ≥ 3 days.
2. Inertia-gravity waves which may or may not be equatorially trapped and may propagate eastwards or westwards; zonal wavenumbers ∼4–40; zonal wavelengths $\sim 10^3$–10^4km; periods ∼1–3 days.
3. Gravity waves which may propagate in either direction and are found at all latitudes; zonal wavenumbers > 40; zonal wavelengths $< 10^3$ km; periods ≤ 1 day.

Numerical results from *Dunkerton* (1997) suggest that the flux in all vertically propagating waves (planetary-scale equatorial modes, intermediate-scale inertia-gravity waves and mesoscale gravity waves) is sufficient to drive a realistic QBO with Brewer-Dobson upwelling present, provided the total wave flux is $2-4$ times that obtained from large-scale, long-period Kelvin and Rossby-gravity equatorial modes. It is found that planetary-scale equatorial modes account for only around 40% of the observed Eliassen-Palm flux within 15° degrees of the equator. Small wavenumbers are almost entirely responsible for the flux in fast waves, whilst slower phase speeds (which dominate the contribution to the QBO) account for most of the Eliassen-Palm flux, though most of this flux is attributable to wavenumbers much greater than 20.

Details on the spectra of waves which drive the QBO can also be derived empirically, by various regression analyses (*Yang et al.*, 2011). Dominant propagation parameters such as zonal wavenumber, zonal phase speed, period, vertical wavelength and vertical group velocity can be found. Identifying the spectra of each wave mode is difficult due to effects such as Doppler shift, meaning that separate waves are not distinct in the frequency and zonal wavenumber domain due to the varying windspeeds through the stratosphere.

By far the greatest difficulty with determining the wave forcing spectrum is a lack of spatial and temporal resolution (*Baldwin et al.*, 2001; *Yang et al.*, 2011) though the continuing growth in available satellite observations and more powerful atmospheric models allow better insight today. *Pahlavan et al.* (2021a) present an analysis of the QBO momentum budget (i.e. contributions to the necessary momentum flux) using the ERA5 reanalysis[6]. Compared with older reanalyses, the higher spatial resolution in ERA5 allows a broader spectrum of waves to be resolved, and better representation of wave/mean-flow interaction in the model, which increases the resolved momentum forcing. Improving vertical resolution is known to have a larger effect than improved horizontal resolution (*Holt et al.*, 2016).

The momentum budget is evaluated in terms of three contributions: meridional advection, vertical advection, and wave forcing. The primary contribution of advection terms is to damp the outer edges of QBO wind regimes, which is consistent with energy considerations; meridional motions are known to be thermally indirect, converting kinetic energy into available potential energy, which is subject to radiative damping (*Wallace*, 1973). The contribution from wave forcing is split into waves with wavenumber less than 20 and greater than 20. The total EP flux divergence is consistent with general circulation models (GCMs) where gravity waves are modelled explicitly, indicating the reanalysis succesfully captures the total momentum budget. It is found that planetary scale waves (wavenumber less than 20) and mesoscale gravity waves (wavenumber greater than 20) have comparable contributions. Planetary-scale waves make a stronger contribution to the descent of westerly regimes, with Kelvin waves dominating the momentum flux (*Ern and Preusse*, 2009), whilst mesoscale gravity waves are primarily responsible for periods of rapid descent of easterly regimes. The intermittency of gravity waves is therefore of importance to QBO dynamics; it remains an open question how the intermittency arises and in what sense the QBO itself modulates gravity wave activity (*Ern et al.*, 2014).

The unresolved wave forcing, from waves with wavelengths below the grid scale of the reanalysis data, accounts for $\sim 40\%$ of the total forcing in easterly zones, agreeing qualitatively with previous studies which conclude small scale waves are essential for driving the QBO (*Dunkerton*, 1997; *Ern and Preusse*, 2009). This unresolved forcing may be further split into that captured by gravity wave parameterisation and a re-analysis correction from errors in the numerical representation of the QBO momentum balance. A significant area of future study therefore lies in improving the modelling of gravity waves, either spontaneously generated as in GCMs or parameterised as in ERA5.

2.5 Meridional confinement

The theory presented thus far does not seek to explain the meridional structure of the QBO. The confinement of the QBO to the tropics may be attributed to two reasons: first, that a large source of momentum flux, from equatorial waves, is generated only near the equator; secondly and perhaps most importantly, that the response of the mean flow to wave-forcing is latitude dependent (*Haynes*, 1998). To see this, consider the TEM equations governing the longitudinally symmetric mean flow in their full form, including the effects of rotation, sphericity and compressibility.

$$\frac{\partial \overline{u}}{\partial t} - f\overline{v}^* = \nabla \cdot \boldsymbol{\mathcal{F}} \tag{2.49}$$

$$f\frac{\partial \overline{u}}{\partial z} + \frac{R}{aH}\frac{\partial \overline{T}}{\partial \theta} = 0 \tag{2.50}$$

$$\frac{\partial \overline{T}}{\partial t} + \overline{w}^* \frac{HN^2}{R} = \mu(T_0 - \overline{T}) \tag{2.51}$$

$$\frac{1}{a\cos\theta}\frac{\partial}{\partial \theta}\left[\overline{v}^* \cos\theta\right] + \frac{1}{\rho_0}\frac{\partial}{\partial z}\left[\rho_0 \overline{w}^*\right] = 0 \tag{2.52}$$

These equations express, respectively, the zonal momentum equation, thermal wind balance, the thermodynamic equation and the mass continuity equation. Here, z is in log-pressure coordinates with H

[6] Reanalysis datasets combine observational data with atmospheric models, so care must be taken in drawing conclusions from the results as they are neither fully empirical or fully theoretical, meaning errors may lie in our theoretical understanding or a defficiency (e.g. of resolution) in the data. For example, *Pahlavan et al.* (2021a) assume all unresolved forcing in the reanalysis is related to mesoscale gravity waves, but also note inaccuracies in the parameterisation of gravity waves in the atmospheric model. Further study is therefore essential in order to identify where the misrepresentation lies.

the associated scale height, R is the gas constant, N^2 is the usual buoyancy frequency (a function of z only), and T_0, ρ_0 are reference temperature and density both dependent on z only. We can form an equation relating $\overline{u}$ and $\nabla \cdot \boldsymbol{\mathcal{F}}$; (2.52) and (2.51) eliminate $\overline{w}^*$, (2.50) then eliminates $\overline{T}$, and finally (2.49) eliminates $\overline{v}^*$. It is illustrative to assume harmonic time-dependence in line with the idea of an oscillating zonal wind, hence we assume $\overline{u} = \Re(\hat{u}e^{i\omega t})$ and $\nabla \cdot \boldsymbol{\mathcal{F}} = \Re(\hat{F}e^{i\omega t})$. We have

$$\frac{\partial}{\partial\theta}\left[\frac{1}{\cos\theta}\frac{\partial}{\partial\theta}\left(\hat{u}\frac{\cos\theta}{\sin\theta}\right)\right] + \frac{4\Omega^2 a^2}{\rho_0}\left(1 - \frac{i\mu}{\omega}\right)\sin\theta\frac{\partial}{\partial z}\left[\frac{\rho_0}{N^2}\frac{\partial\hat{u}}{\partial z}\right] = \frac{1}{i\omega}\frac{\partial}{\partial\theta}\left[\frac{1}{\cos\theta}\frac{\partial}{\partial\theta}\left(\hat{F}\frac{\cos\theta}{\sin\theta}\right)\right] \tag{2.53}$$

The response of the mean flow to the wave-forcing is determined by the relative sizes of the left-hand side (LHS) terms. If the response is largely in $\overline{v}^*$ or $\overline{w}^*$, the dominant behaviour is a mean meridional circulation. If the response is largely in $\overline{u}$ then the dominant behaviour is an acceleration of the mean zonal flow. Now note that if the first LHS term dominates the second, then the response is in $\overline{u}$. If the second LHS term dominates the first, then the response is in $\partial\overline{u}/\partial z$ and from (2.50) this induces a response in $\overline{T}$. Then, from (2.51) and (2.52), the response is in $\overline{v}^*, \overline{w}^*$ and a mean meridional circulation is the dominant behaviour.

Haynes (1998) invokes a latitudinal scale L and vertical scale D for the mean flow response and the applied force. Note that for an oscillation with period around 2 years, $\omega \sim 10^{-7}s^{-1}$ and in the lower stratosphere $|\mu| \sim 5 \times 10^{-5}s^{-1}$, so the factor $1 - i\mu/\omega$ may be approximated by $-i\mu/\omega$. Consider low latitudes where $\Omega \ll 1$ so that $\sin\theta \sim \theta \sim L/a$ and $\cos\theta \sim 1$. The ratio of the first to the second term is then

$$\frac{D^2N^2a^2\omega}{4\Omega^2L^4\mu} \tag{2.54}$$

Thus when there is 'deep forcing', i.e. the latitudinal scale is sufficiently large that $L \gg \sqrt{aDN/2\Omega}(\omega/\mu)^{1/4}$ the second term is dominant and the response is a mean meridional circulation. If instead the latitudinal scale is sufficiently small that $L \ll \sqrt{aDN/2\Omega}(\omega/\mu)^{1/4}$, known as 'shallow forcing', the first term is dominant and the response is a mean zonal flow acceleration.

The distinction in responses to an applied force at small and large latitudinal scales arises due to the close link between temperature and mean meridional circulation, expressed by thermal wind balance (2.50) and the thermodynamic relation (2.51). A shallow force generates shear which via thermal wind balance induces a temperature anomaly. Such an anomaly is necessarily associated with a meridional circulation from which the Coriolis force balances the applied force. Further, the temperature anomaly is thermally damped which limits the response amplitude. A deep force produces a much smaller temperature response so the amplitude is not limited and the response is dominated by zonal flow acceleration.

Close to the equator, the Coriolis parameter is small so that essentially all forces are deep. Therefore, in the tropics, wave forcing typically produces a response in the mean zonal flow which lasts much longer, as seen in the QBO, rather than a meridional circulation which cannot be sustained for long periods due to damping.

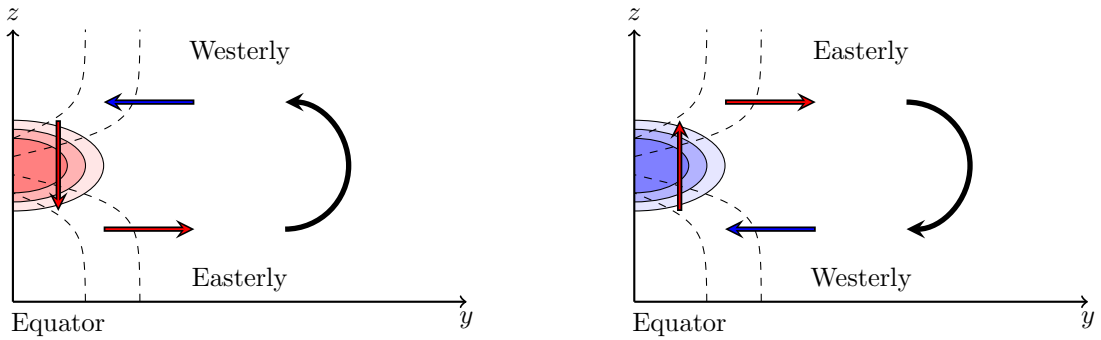
2.6 Temperature anomalies & meridional circulation

The temperature anomaly associated with the QBO exists owing to thermal wind balance. This balance is present since QBO dynamics have a long period and are equatorially symmetric; (2.50) is modified for the equatorial β-plane using the (equatorial) local Cartesian coordinate $y = \theta a$, approximating $\sin\theta \approx \theta$, and finally assuming that $\overline{T}$ is maximised at the equator, hence $\frac{\partial\overline{T}}{\partial\theta} \approx \frac{\partial^2\overline{T}}{\partial\theta^2}\theta$. Then we have

$$\frac{\partial\overline{u}}{\partial z} = -\frac{R}{H\beta}\frac{\partial^2\overline{T}}{\partial y^2} \tag{2.55}$$

where z is in log-pressure co-ordinates and $H \approx 7$ km is the associated scale height.

Temperature anomalies of approximately 4 K have have been observed, maximising near 30–50 hPa (*Randel et al.*, 1999). Assuming a meridional scale L, observations of $\partial\overline{u}/\partial z$ imply $L \sim 1000 - 1200$ km which is around 10° of latitude, broadly consistent with the observed Gaussian structure of the wind ampltiude with a half-width of 12.5°.



(a) Descending westerly phase

(b) Descending easterly phase

Figure 7: Schematic illustration of mean meridional circulation, QBO wind and associated temperature anomaly in latitude-altitude plane. Solid contours indicate surfaces of constant temperature, dashed contours indicate surfaces of constant zonal wind. Arrows indicate motion of mean meridional circulation with red/blue colouring indicating positive/negative zonal wind acceleration. Temperature anomalies coloured blue are cold, and red are warm. Note that the z axis as shown represents the low-to-mid stratosphere.

Due to thermal wind balance between temperature and vertical wind shear, warm anomalies are associated with the descending westerly[7] (DW) phase where the vertical zonal wind shear is positive (figure 7a) and cold anomalies are associated with the descending easterly (DE) phase (figure 7b). Note that radiative damping means temperature anomalies require some form of heating or cooling to be sustained. According to *Plumb and Bell* (1982a,b), adiabatic heating in the DW phase is provided by the sinking motion of an induced (weak) meridional circulation and similarly in the DE phase, cooling is provided by the rising motion of this QBO-induced meridional circulation. In fact, meridional motions may be deduced to exist from the meridional structure of zonal wind alone; according to Hide's theorem, a conserved quantity (such as angular momentum) in steady state cannot reach a maximum in the fluid interior, as diffusion dissipates any maximum and advection cannot replace it. Hence the maximum in zonal wind at the equator must be facilitated by a meridional circulation.

In section 1.2 it was noted that westerly shear zones descend more rapidly and regularly than easterly shear zones. It has been suggested that the sinking motion during the DW phase may explain the greater rate of descent of this phase; the downward advection of momentum associated with the QBO-induced meridional circulation enhances the previously discussed mechanism by which the zonal winds descend. A further effect of the momentum advection by the meridional circulation is a broadening of the latitudinal width of the DE phase and a narrowing of the DW phase (*Gray*, 2010).

The irregular descent of easterly shear zones compared to westerly shear zones may be attributed to the residual meridional circulation (Brewer-Dobson circulation) with annual variability. The apparent stalling in downward propagation (e.g. early 1989, 30–50 hPa in figure 1) may be inhibited by the upwards advection of momentum associated with the residual circulation. The annual variability of the circulation, which results in a minimum in the strength of equatorial upwelling during Northern Hemisphere spring–summer, leads to step-like behaviour in the descent of the shear zone. This effect of annual variability may also contribute to the seasonal onset preference; see figure 2 and *Gray* (2010). Further, mesoscale gravity wave activity is at maximum during periods of rapid descent of easterly shear zones (*Pahlavan et al.*, 2021a) and the intermittency of gravity waves, potentially as a result of modulation by the QBO itself (*Ern et al.*, 2014; *Pahlavan et al.*, 2021a), adds to the irregular descent.

3 Extratropical influence of the QBO

Whilst the QBO is an interesting case study of atmospheric dynamics in and of itself, it is perhaps most important to understand due to its far-reaching influence in the extratropical atmosphere. Owing

[7]Whilst QBO phase is typically in terms of the sign of 50 hPa winds in the lower stratosphere, alternative naming is used here to aid relation with the dynamical motions. The descending westerly (easterly) phase corresponds to westerly (easterly) winds in the mid-stratosphere and easterly (westerly) winds in the lower stratosphere.

to the equatorial confinement of the QBO, pathways of influence on the wider atmosphere are indirect and consequently discussed in terms of the *modulation* of variability otherwise independent of the QBO. Contextualising the role of the QBO among global atmospheric dynamics therefore requires understanding which modes of variability are modulated by the QBO and the mechanisms underlying this modulation. In this section we discuss three examples, two of which relate to the stratospheric polar vortex: the Holton-Tan effect demonstrates influence via modification of wave propagation, and the combined influence of westerly QBO phases and solar maximum conditions demonstrates modulation via the strengthening of an independent mode of variability. Finally, we discuss the modulation of El Niño-related wave propagation via modification of tropospheric winds.

3.1 Stratospheric variability

The extratropical stratosphere exhibits much greater seasonal variability compared to the tropical stratosphere or extratropical troposphere. In winter months, reduced solar heating creates a strong meridional temperature gradient and hence strong shear of the zonal wind to maintain thermal wind balance. The cooling of the high-latitude stratosphere forms a deep, strong westerly vortex known as the polar (stratospheric) vortex. During summer, the return of solar heating reverses the temperature gradient and the zonal wind tends to become easterly.

In addition to seasonal variations in solar heating, the dynamics are modified by the effects of planetary Rossby waves[8] which typically have zero horizontal phase speed (quasi-stationary) and are generated by land-sea contrasts and orography/topography. These waves propagate vertically and meridionally, depending on the background flow which may be thought of as refracting the waves. Their effects are mostly confined to the winter hemisphere as planetary waves can only propagate in a westerly mean zonal wind and become evanescent in the mean easterly winds of summer. Further, the NH has much greater land-sea contrasts and larger amplitude orography hence NH winter exhibits stronger effects of planetary Rossby waves compared to SH winter.

Dynamical variability from planetary waves predominantly arises in two ways: waves propagating up from the troposphere, so that stratospheric variability derives from tropospheric variability where the dynamics are more unstable (e.g. baroclinic instability); or from oscillatory/complex behaviour as a result of wave–mean-flow interaction and also wave self-interaction. These variabilities occur on large scales, as only wavenumbers 1 and 2 tend to reach the stratosphere (*Charney and Drazin*, 1961). The most striking influence of planetary Rossby waves interacting with the mean flow (for which the essential details may be found in *Andrews and McIntyre* (1976)) is sudden stratospheric warming (SSW). In winter, planetary waves of large amplitude dissipate or break in the stratosphere, weakening the westerly flow, allowing more waves to propagate upwards into the stratosphere. Eventually, this positive feedback causes the winds to reverse, forming a critical layer $\overline{u} = 0$ since most planetary waves are quasi-stationary. This layer completely inhibits upward propagation, greatly increasing wave breaking in the layer and rapidly inducing an easterly flow. The reversed flow is associated with a reversed meridional temperature gradient by thermal wind balance, and hence the high-latitude mid-stratosphere rapidly warms (*Dunkerton et al.*, 1981). These events can either displace the polar vortex or split it in two, which has important consequences for tropospheric dynamics which are responsible for the weather we experience day-to-day on smaller scales.

The NH stratosphere is sensitive to the effects of vertically propagating planetary waves; typical amplitudes are just large enough to trigger SSWs in some years but not others, leading to variable polar vortex strength year-to-year. This sensitivity allows the QBO to influence the polar stratosphere by acting as a waveguide for planetary waves propagating upwards and polewards from mid-latitudes. However, the sensitivity means the polar vortex is also influenced by multiple external factors: the 11-year solar cycle, the El Niño Southern Oscillation, sea surface temperature anomalies, and volcanic eruptions all play a role (*Anstey and Shepherd*, 2014) which makes isolation of a QBO modulation signal difficult. It has also been noted by *Camp and Tung* (2007) that once the polar vortex becomes disturbed due to one of these external factors, its sensitivity to other influences is decreased.

[8]Planetary Rossby waves may be referred to simply as planetary waves. Note that 'planetary-scale waves' are not necessarily planetary Rossby waves, they are large wavelength variants of some wave mode which is usually clear from context.

3.2 Holton-Tan relationship

Despite the complex nature of the polar vortex in the Northern Hemisphere, observations clearly demonstrate a relationship between QBO phase and vortex variability. The first evidence was provided by *Holton and Tan* (1980, 1982) using observations of NH 50 hPa geopotential heights[9] during 16 NH winters from 1962–1977. They found a statistically significant decrease during the DE phase of the QBO, which corresponds to a stronger westerly polar vortex, resulting in colder and less disturbed winters. In the DW[10] phase, the vortex is weaker owing to increased disruption from planetary waves, which is associated with warmer and more disturbed winters.

Holton and Tan (1980) proposed a mechanism for this relationship based on modulation of the subtropical zero-wind line $\overline{u} = 0$ which acts as a critical line for the quasi-stationary planetary waves propagating upwards from the troposphere. They suggest that the critical line is shifted polewards during a DW phase, narrowing the mid-latitude waveguide and enhancing reflection/refraction of planetary waves towards the pole. This results in deceleration of the polar vortex. Under a DE phase, the critical line is instead located in the summer hemisphere so that planetary waves can propagate into the tropics more easily and are less confined in the winter hemisphere, leaving the polar vortex less disturbed.

According to *Lu et al.* (2014) and *Garfinkel et al.* (2012) the QBO modulation of the subtropical critical line is only a small part of the mechanism. Instead, they provide observational evidence to support a mechanism involving the QBO-induced meridional circulation: composite differences between QBO phases demonstrate that significant changes of EP flux occur in the middle to upper stratosphere and correspondingly, modification of the refractive index (*Garfinkel et al.*, 2012). The anomalies in the refractive index expand poleward in winter and create a 'barrier' to planetary wave propagation which enhances wave convergence in the polar region and destabilises the polar vortex. To elucidate this mechanism further, it is valuable to consider the planetary wave refractive index in detail and how the mean flow modifies it.

3.2.1 Planetary wave propagation dynamics

Planetary waves arise from conservation of potential vorticity and may be captured mathematically by perturbations of vorticity ζ'. Following *Matsuno* (1970), we consider small amplitude perturbations to a basic state (the mean flow in our case) in geostrophic balance, i.e. using the quasi-geostrophic framework. Assume the motion is adiabatic, frictionless, slowly varying in time, and of planetary scale in the zonal direction. We will use spherical coordinates (z, θ, λ) where θ is latitude, λ is longitude and $z = -H \log p/p_0$ is in log-pressure coordinates. The equation describing planetary waves is derived from the vorticity equation written in the form

$$\left(\frac{\partial}{\partial t} + \overline{\omega}\frac{\partial}{\partial \lambda}\right)\zeta' + \frac{1}{a}\frac{\partial \overline{\zeta_a}}{\partial \theta} - \frac{f}{p}\frac{\partial}{\partial z}\left(pw'\right) = 0 \tag{3.1}$$

and the thermodynamic equation

$$\left(\frac{\partial}{\partial t} + \overline{\omega}\frac{\partial}{\partial \lambda}\right)\frac{\partial \phi'}{\partial z} - af\cos\theta\frac{\partial \overline{\omega}}{\partial z}v' + N^2 w' = 0 \tag{3.2}$$

where ζ' is the vorticity perturbation, $\overline{\zeta_a}$ is the absolute vorticity of the basic state, ϕ' is the geopotential perturbation and $\overline{\omega}$ is the angular speed of the basic zonal flow, defined via $\overline{u} = a\overline{\omega}\cos\theta$. We wish to eliminate the meridional and vertical velocity perturbations (v', w') and vorticity disturbance ζ'. First, w' is eliminated to give the potential vorticity equation

$$\left(\frac{\partial}{\partial t} + \overline{\omega}\frac{\partial}{\partial \lambda}\right)\left[\zeta' + \frac{f}{p}\frac{\partial}{\partial z}\left(\frac{p}{N^2}\frac{\partial \phi'}{\partial z}\right)\right] + \frac{f^2}{N^2}a\cos\theta\frac{\partial \overline{\omega}}{\partial z}\left(\frac{\partial v_g}{\partial z} - \frac{\partial v'}{\partial z}\right) + \frac{v'}{a}\frac{\partial \overline{q}}{\partial \theta} = 0 \tag{3.3}$$

where v_g is the geostrophic meridional velocity and $\overline{q}$ is the quasi-geostrophic potential vorticity of the basic state. To eliminate ζ' and v' we use geostrophic wind balance and the zonal equation of motion,

[9] In vertical pressure co-ordinates, pressure gradients become gradients of geopotential height (GPH), hence GPH anomalies correspond to pressure anomalies.

[10] *Holton and Tan* (1980) refer to this phase as an easterly phase, since they define the phase based on the 50 hPa (lower stratosphere) zonal wind sign.

assuming a stationary state, to find

$$f\zeta' = \frac{1}{a^2}\left[\tan\theta\frac{\partial}{\partial\theta}\left(\cot\theta\frac{\partial\phi'}{\partial\theta}\right) + \frac{1}{\cos^2\theta}\frac{\partial^2\phi'}{\partial\lambda^2}\right] \tag{3.4}$$

$$v' = \frac{1}{f}\left[\frac{1}{a\cos\theta}\frac{\partial\phi'}{\partial\lambda} + \overline{\omega}\frac{\partial u'}{\partial\lambda}\right] \approx \frac{1}{fa}\left[\frac{1}{\cos\theta}\frac{\partial\phi'}{\partial\lambda} - \frac{\overline{\omega}}{fa}\frac{\partial^2\phi'}{\partial\lambda\partial\theta}\right] \tag{3.5}$$

Combining (3.3), (3.4) and (3.5) we have

$$\overline{\omega}\frac{\partial}{\partial\lambda}\left[\frac{\sin^2\theta}{\cos\theta}\frac{\partial}{\partial\theta}\left(\frac{\cos\theta}{\sin^2\theta}\frac{\partial\phi'}{\partial\theta}\right) + \frac{1}{\cos^2\theta}\frac{\partial^2\phi'}{\partial\lambda^2} + \frac{f^2a^2}{p}\frac{\partial}{\partial z}\left(\frac{p}{N^2}\frac{\partial\phi'}{\partial z}\right)\right] + \frac{1}{\cos\theta}\frac{\partial\overline{q}}{\partial\theta}\frac{\partial\phi'}{\partial\lambda} = 0 \tag{3.6}$$

The coefficients of ϕ' in (3.6) are independent of λ, so we can expand in a Fourier series

$$\phi'(\lambda,\theta,z) = \sum_{k=1}^{\infty}\phi_k(\theta,z)e^{ik\lambda} \tag{3.7}$$

Further, assuming an isothermal atmosphere which is a useful approximation in the stratosphere (also used in section 2.3.3), we define

$$\psi_k(\theta,z) = e^{-z/2H}\phi_k(\theta,z) \tag{3.8}$$

Finally we have an equation describing ψ_k, which encapsulates planetary Rossby wave disturbances to the mean flow. The wave equation is

$$\mathcal{L}(\theta)\psi_k + n_k^2\psi_k = 0 \tag{3.9}$$

where the modified meridional Laplacian operator $\mathcal{L}(\theta)$ is given by

$$\mathcal{L}(\theta) = \frac{\sin^2\theta}{\cos\theta}\frac{\partial}{\partial\theta}\left(\frac{\cos\theta}{\sin^2\theta}\frac{\partial}{\partial\theta}\right) + \frac{f^2a^2}{N^2}\frac{\partial^2}{\partial z^2} \tag{3.10}$$

and the refractive index squared n_k^2 is

$$n_k^2 = \frac{a}{\overline{u}}\frac{\partial\overline{q}}{\partial\theta} - \frac{\Omega^2a^2\sin^2\theta}{H^2N^2} - \frac{k^2}{\cos^2\theta} \tag{3.11}$$

The latitudinal derivative of the QGPV can be written for an isothermal atmosphere as

$$\frac{\partial\overline{q}}{\partial\theta} = 2\Omega\cos\theta - \frac{1}{a}\frac{\partial}{\partial\theta}\left[\frac{1}{\cos\theta}\frac{\partial}{\partial\theta}\left(\overline{u}\cos\theta\right)\right] + \left(\frac{f^2}{HN^2} + \frac{f^2}{N^2}\frac{\mathrm{d}N^2}{\mathrm{d}z}\right)\overline{u}_z - \frac{f^2}{N^2}\overline{u}_{zz} \tag{3.12}$$

Planetary waves can propagate where $n_k^2 > 0$ and become evanescent where $n_k^2 < 0$ (*Plumb*, 2010). The dominant contributions to n_k^2 are the mean zonal wind $\overline{u}$, the meridional shear term $-\frac{1}{a^2}(\frac{(\overline{u}\cos\theta)_\theta}{\cos\theta})_\theta$, the vertical shear term $(\frac{f^2}{HN^2} + \frac{f^2}{N^2}\frac{\mathrm{d}N^2}{\mathrm{d}z})\overline{u}_z$, and the vertical curvature term $-\frac{f^2}{N^2}\overline{u}_{zz}$ (*Lu et al.*, 2014). It is anomalies of these terms which give rise to the anomalous refractive index in the QBO-induced meridional circulation mechanism of the Holton-Tan effect of *Lu et al.* (2014) and *Garfinkel et al.* (2012).

3.2.2 Meridional circulation mechanism of polar vortex modulation

The mechanism by which the easterly phase of the QBO disturbs the stratospheric polar vortex can be understood in terms of the induced temperature, zonal wind, circulation anomalies and their effect on refractive index, and consequently planetary wave propagation. Whilst it is clearest to describe the mechanism in 'steps', it should be stressed that these anomalies form almost simultaneously via (for example) thermal wind balance and geostrophic balance rather than in succession. Note that the mechanism is subject to deviation based on stratospheric circulation, which may disrupt and weaken the Holton-Tan effect, as observed during 1978–1997 (*Lu et al.*, 2008, 2014).

During winter, the change in temperature gradient due to solar heating strengthens the QBO-induced meridional circulation discussed in section 2.6. This circulation acts to form a quadrupole[11] of temperature anomalies: see figure 8. In this figure, the QBO-induced circulation discussed previously is shown by solid black arrows.

[11]There are in fact further temperature anomalies in the upper stratosphere, but these are not necessarily relevant to this mechanism.

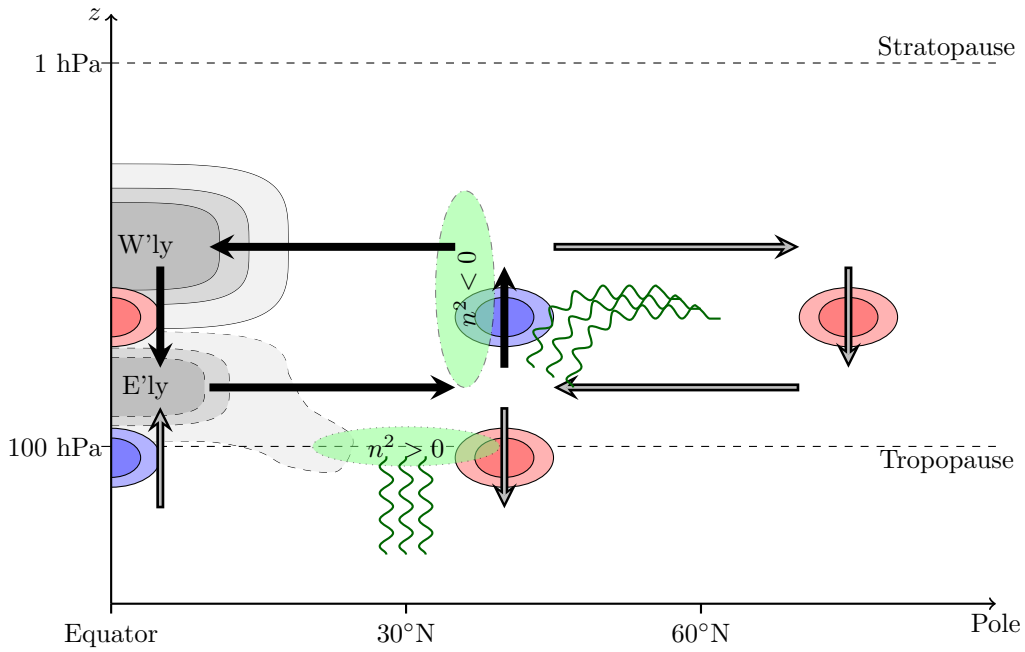


Figure 8: Schematic diagram showing QBO-induced meridional circulation mechanism of the Holton-Tan relationship. QBO zonal winds and associated zonal wind anomalies are shaded: solid (dashed) lines indicate westerly (easterly) winds. Temperature anomalies shown in red (warm) and blue (cold). QBO-induced meridional circulation shown by solid black arrows, anomalous circulations shown in shaded arrows. Planetary wave refractive index anomalies shown in green: dotted (dash-dotted) lines indicate positive (negative) anomaly. Planetary waves indicated by green wavy lines.

Owing to thermal wind balance, the lower stratosphere temperature dipole causes the easterly zone of QBO winds to branch down and polewards (see *Pahlavan et al.* (2021b), section 3). Therefore, the quasi-stationary planetary Rossby wave critical line $\overline{u} = 0$ moves polewards, enhancing wave breaking in the lower subtropical stratosphere between $20° - 30°$N, indicated in the figure by wavy green lines. The resulting easterly acceleration of the mean flow propagates the branching of the easterly QBO wind further polewards. Moreover, the acceleration strengthens the curvature of the flow $\propto \overline{u}_{zz}/\overline{u}$ so that a positive n^2 anomaly forms, indicated by the dotted green region.

The poleward propagation of the easterly branching and the resulting easterly acceleration due to wave breaking induces a meridional circulation with upwelling near the equator and downwelling between $30° - 50°$N. This anomalous circulation is represented by shaded arrows near the tropopause in figure 8 and reinforces the QBO-induced circulation (solid arrows) higher in the stratosphere. The meridional and vertical gradients of the zonal flow generate a negative n^2 anomaly around $35° - 45°$N, 5–50 hPa, indicated by the dash-dotted green region.

Planetary waves are evanescent in regions of negative n^2 and the anomaly tends to refract upward propagating planetary waves poleward, that is, a mid-stratosphere waveguide (MWG) is formed. The decreased mid-latitude stratosphere wave forcing as a result of the MWG increases wave forcing in the polar stratosphere. According to the 'downward control' principle of *Haynes et al.* (1991), this drives an anomalous circulation (shaded arrows at high latitude) which acts to warm the polar stratosphere. The polar vortex is weakened by the easterly acceleration of the increased planetary wave forcing, which causes descent of the critical line and hence a descent of the anomalous wave forcing and circulation, so that the stratospheric warming descends as observed (*Garfinkel et al.*, 2012).

The mechanism can be compared with observations by considering the effect on the Eliassen-Palm (EP)

flux divergence, given in terms of the potential temperature θ (using ϕ to denote latitude) by

$$\nabla \cdot \boldsymbol{\mathcal{F}} = \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left(\mathcal{F}^{(\phi)}\cos\phi\right) + \frac{\partial \mathcal{F}^{(z)}}{\partial z} \tag{3.13}$$

$$\mathcal{F}^{(\phi)} = \rho_0 a\cos\phi\left(\frac{\overline{v'\theta'}}{\theta_z}\overline{u}_z - \overline{v'u'}\right) \tag{3.14}$$

$$\mathcal{F}^{(z)} = \rho_0 a\cos\phi\left[\left(f - \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left(\overline{u}\cos\phi\right)\right)\frac{\overline{v'\theta'}}{\theta_z} - \overline{w'u'}\right] \tag{3.15}$$

Lu et al. (2014) presents such an analysis of QBO phase composites which illustrates the effect of the easterly QBO phase on the EP flux divergence. The foregoing mechanism described suggests increased EP flux convergence should be observed in the polar stratosphere and lower subtropical stratosphere, with increased EP flux divergence in the mid-latitude and subtropical mid-stratosphere due to the refraction of planetary waves poleward. Additionally, the QBO-induced circulation is associated with EP flux divergence regions due to the poleward arm, and a stronger EP flux convergence region due to the equatorward arm. These regions are observed in the QBO phase composites of *Lu et al.* (2014) and also found in numerical simulations using quasi-geostrophic and linear theory (*Garfinkel et al.*, 2012).

Lu et al. (2014) suggest the QBO phase composites support the meridional circulation mechanism over the critical line mechanism of *Holton and Tan* (1980, 1982). Instead of polewards anomalous wave forcing due to a shift in the critical line, observed EP fluxes indicate increased planetary wave propagation toward the subtropics in the upper stratosphere, as a result of positive refractive index anomalies at $30--45°N$ above the MWG. These anomalies arise due to branching of the westerly QBO winds polewards and upwards into the region. However, it could be argued that whilst the evidence suggests the critical line is unimportant in the middle to upper stratosphere, the branching of easterly QBO winds in the subtropical lower stratosphere is the key to the meridional circulation mechanism. This is indeed a modification of the zero-wind critical line, as put forward by Holton & Tan. In this sense, the meridional circulation mechanism is an extension/elaboration of the critical line mechanism rather than a replacement.

An important area of study mentioned but not clarified in *Lu et al.* (2014) is the dependence of QBO composites on the length of temporal average (*Lu et al.* (2014) and previous studies use 3-monthly averages), which is highlighted by *Watson and Gray* (2013) as the extratropical dynamics can influence the QBO-induced meridional circulation at later times which can obfuscate the proposed Holton-Tan mechanism. Further, studies such as *Yamashita et al.* (2011) (and mentioned in *Gray* (2010)) highlight the sensitivity of the Holton-Tan effect to the reference level used to define QBO phases, i.e. QBO anomalies at 10 hPa demonstrate correlation with polar vortex behaviour, but anomalies lower in the stratosphere are less correlated. There is clear disagreement between studies: *Naito and Yoden* (2006); *Hitchman and Huesmann* (2009); *Garfinkel et al.* (2012) find the strongest HT effect when using lower stratospheric QBO wind anomalies, whilst *Gray et al.* (2001); *Gray* (2003) highlight the importance of the upper stratospheric QBO, and the above discussion due to *Lu et al.* (2014) focuses on the entire mid to low stratosphere.

3.3 Influence of westerly QBO and solar-maximum conditions

The Holton-Tan relationship demonstrates that the westerly QBO tends to leave the stratospheric polar vortex less disturbed compared with during an easterly QBO phase. As discussed in section 3.1, many external influences such as solar activity can further modulate the Holton-Tan relationship. During the westerly phase of the QBO under solar maximum conditions, observations suggest that the polar vortex is not only less disturbed but in fact strengthened during December–January, followed by a weaker polar vortex during February and March (*Anstey and Shepherd*, 2014). Note that *Yamashita et al.* (2015) suggest the weakened polar vortex in late winter may not be specific to westerly QBO and solar maximum conditions, but may simply require a strengthened polar vortex in early winter. The dynamical response leading to a strengthened polar vortex is thought to rely on the combined westerly QBO and solar maximum dynamical responses and their effect on planetary wave propagation. The effects of each of these conditions are of the same 'sign', so when combined the response is reinforced.

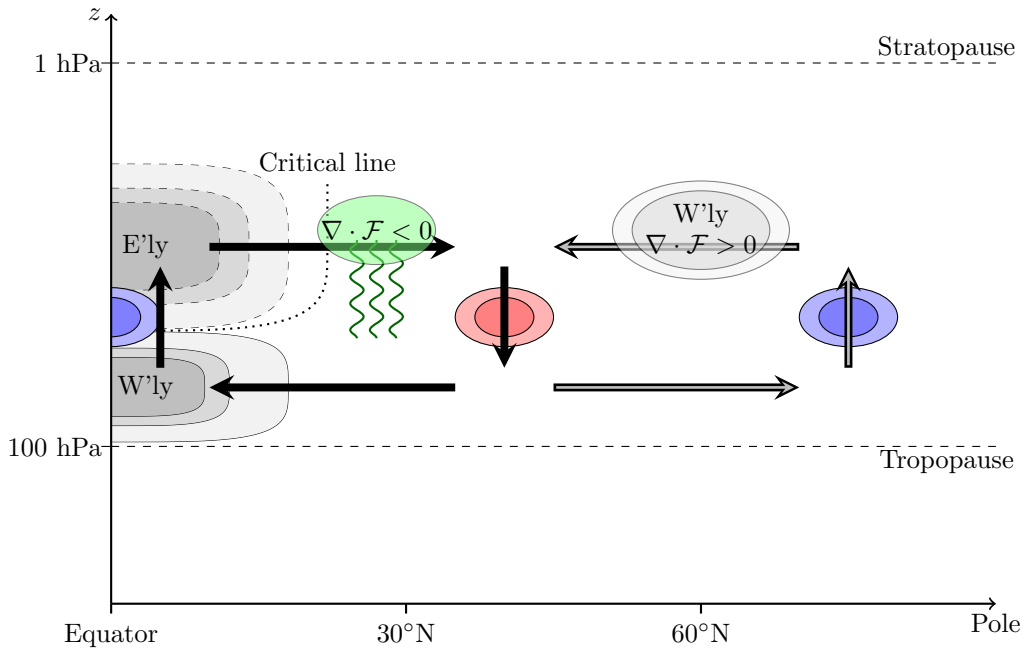


Figure 9: Schematic diagram of the dynamical response to westerly QBO phase. Zonal winds are shaded, solid outlines are westerly and dashed outlines are easterly. QBO-induced meridional circulation shown by solid arrows, additional circulation anomaly shown by shaded arrows. Temperature anomalies indicated by blue (red) regions indicating cold (warm) anomaly. Anomalies of Eliassen-Palm flux divergence shown in green regions. Green wavy lines indicate anomaly of planetary wave propagation.

3.3.1 Dynamical response of the upper extratropical stratosphere

Solar activity varies on an 11-year cycle. The variable heating and resulting temperature gradient in the winter hemisphere is strengthened during solar maximum and reinforced by other temperature anomalies associated with ozone heating (*Gray*, 2010). The mechanism by which westerly QBO–solar maximum (wQBO/Smax) conditions influences the polar vortex relies on the presence of a zonal wind anomaly in the upper stratosphere/lower mesosphere, but the cause of this anomaly is subject to debate. Nonetheless, given the presence of this anomaly, a mechanism is proposed by *Yamashita et al.* (2015) which is based on the subsequent tropospheric response to wQBO/Smax conditions.

Yamashita et al. (2011) note a dynamically consistent relationship between zonal wind, temperature, planetary wave forcing, and circulation anomalies in response to a poleward shift of the critical line $\overline{u} = 0$ in the upper stratosphere during the westerly QBO phase, shown schematically in figure 9. The poleward movement of the critical line enhances planetary wave forcing, propagating the QBO-induced meridional circulation further north, indicated by the green region. The circulation drives a warm anomaly at mid-latitudes, which acts to support a westerly anomaly (shaded region) around 60°N. This anomaly may also be attributed to the dipole of wave forcing with increased EP flux convergence near the critical line and a corresponding EP flux divergence anomaly at 60°N. As a result of the westerly anomaly, upward propagating planetary waves in the high-latitudes are suppressed which reinforces the anomaly further. The westerly anomaly and mid-latitude warm anomaly are associated with an anomalous circulation at mid–high latitudes (shaded arrows) which cools the polar stratosphere. The combination of polar cooling and westerly wind anomaly strengthens the polar vortex. The dynamical effect leading to the strengthened polar vortex are supported by observations of a mid–high latitude upper–mid stratosphere (10–50 hPa) EP flux divergence anomaly.

The dynamical response to solar maximum conditions is described by *Kodera and Kuroda* (2002) and shown schematically in figure 10. The strong latitudinal temperature gradient during solar maximum in the winter hemisphere drives a westerly anomaly in the mid-latitude upper stratosphere. The westerly anomaly suppresses upward propagating planetary waves, driving a circulation in the same manner as under westerly QBO conditions (solid arrows). Observations and models detailed by *Kodera and Kuroda* (2002) show a mid–high latitude downward EP flux and divergence anomaly, corresponding to suppressed

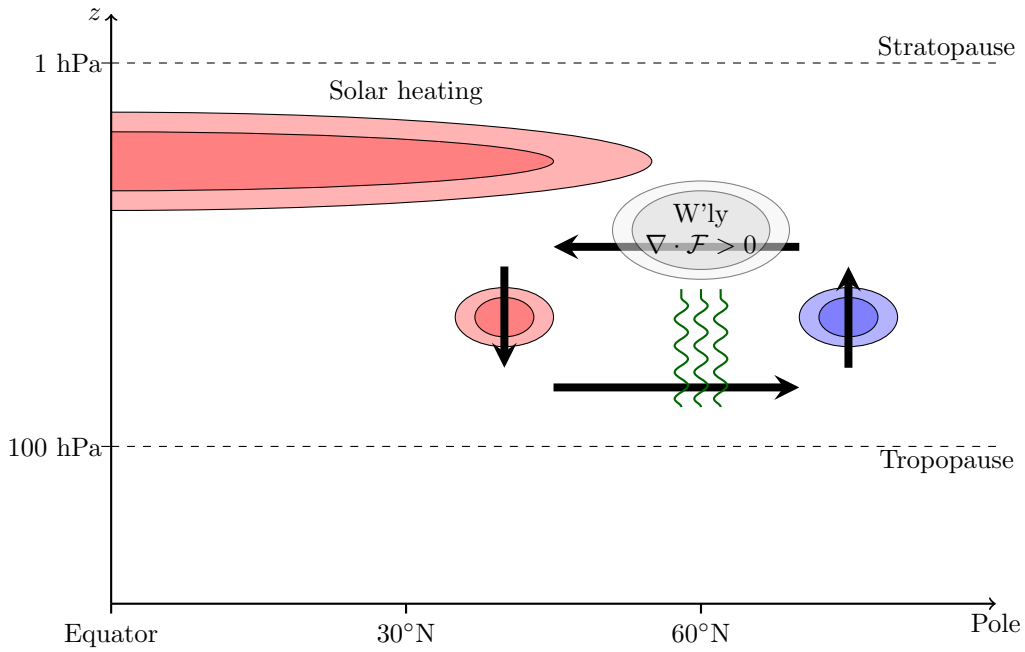


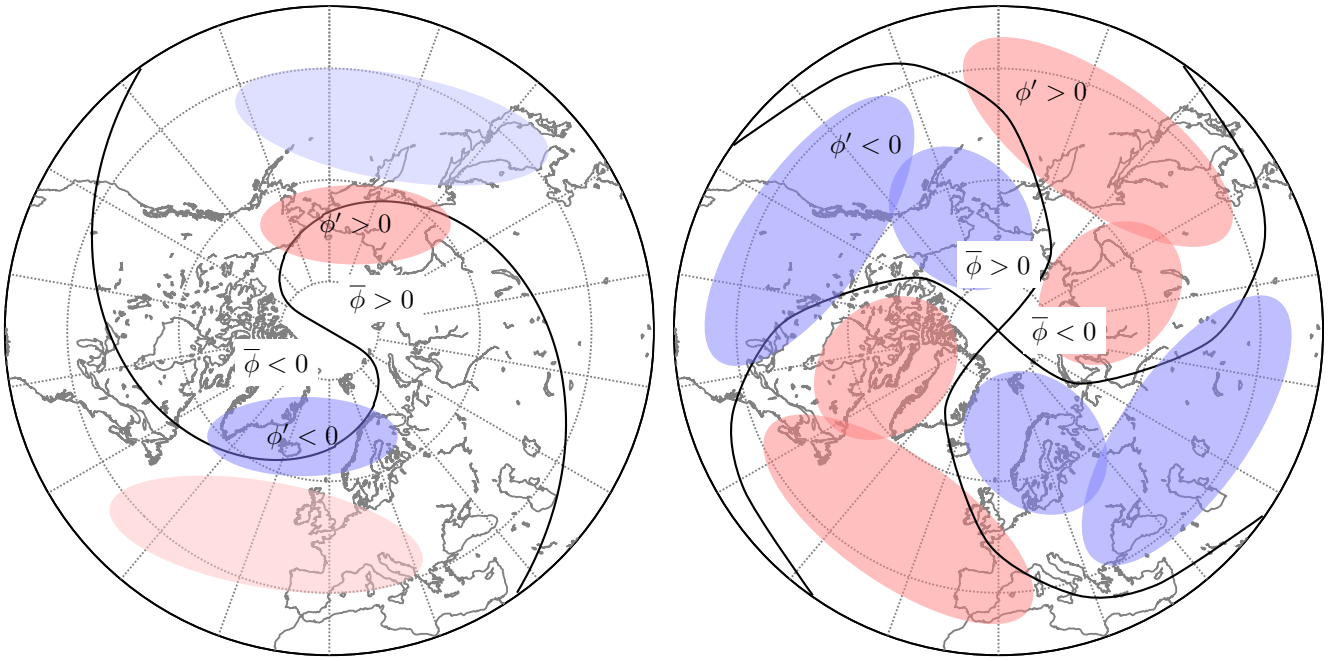
Figure 10: Schematic diagram of the dynamical response to solar maximum in winter hemisphere. Zonal wind anomalies are shaded, solid outlines are westerly and dashed outlines are easterly. Anomalous circulation shown by solid arrows. Temperature anomalies indicated by blue (red) regions indicating cold (warm) anomaly. Green wavy lines indicate anomaly of planetary wave propagation.

planetary wave propagation. The residual circulation driven by the EP flux anomaly is consistent with the downward control principle (*Haynes et al.*, 1991) and a cooling of the polar stratosphere, which by thermal wind balance strengthens the polar vortex zonal winds.

3.3.2 Tropospheric response to a strengthened polar vortex

Following the strengthened polar vortex in December, a negative anomaly of sea level pressure (SLP) at high latitudes is observed with increasing magnitude in December and January during wQBO/Smax conditions. The anomaly is a deviation from the mean state in which a low pressure region persists over Iceland, known as the 'Icelandic low', in the far North Atlantic. The negative anomaly strengthens this low, producing a zonally asymmetric mean state resembling the North Atlantic Oscillation (NAO) which exhibits a similar pressure anomaly. A positive pressure anomaly is also seen in the mid-latitude Atlantic, intensifying the 'Azores high' in December and January. The observed north-south dipole intensifies surface zonal wind and from (3.11) and (3.12), the increased meridional shear generates a positive n^2 anomaly. *Yamashita et al.* (2015) note that the NAO-like dipole response is seen in 500 hPa geopotential height as well as surface geopotential height, indicating the anomalies are present throughout the depth of the troposphere. They note that such a zonally asymmetric response is characteristic of the tropospheric circulation response to an anomalous polar vortex, with a strengthened polar vortex producing a positive NAO-like anomaly.

The zonal asymmetry is important since the $k = 1$ planetary wave (henceforth referred to as wavenumber 1 or WN1) has similar zonal asymmetry; see figure 11a. The constructive interference of the climatological and anomalous geopotential height (from WN1) drives an upward EP flux anomaly in the mid-latitudes. Consequently, the mid–high latitude stratosphere exhibits anomalous upward propagation of planetary waves and hence the EP flux convergence around the polar vortex is strengthened. In the same manner as easterly QBO conditions, the increased wave forcing weakens the polar vortex. The dynamical response to wQBO/Smax demonstrates how planetary waves of different wavenumber can play different roles: the $k = 2$ planetary wave (henceforth WN2) has a geopotential height anomaly out of phase with the climatology, hence the destructive interference drives an EP flux divergence anomaly in the mid-latitude troposphere; see figure 11b. The resulting westerly anomaly supports the zonal wind anomaly driven by the strengthened polar vortex in December and January, maintaining the NAO-like



(a) (b)

Figure 11: Schematic diagram of climatological 500 hPa geopotential height $\overline{\phi}$ (a) WN1 and (b) WN2 components given in *Yamashita et al.* (2015) figure 7. Zero contour of $\overline{\phi}$ shown in solid black lines. Red (blue) regions show positive (negative) anomalies ϕ'.

response into February and March whilst WN1 planetary waves propagate upwards and disturb the polar vortex.

3.4 Modulation of El Niño influence

Whilst the modulation of solar maximum conditions is best observed in the North Atlantic region, modulation of another important recurrent phenomenon, the El Niño Southern Oscillation (ENSO), is best observed in the North Pacific (NP) region. ENSO is an episodic variation in winds and sea surface temperatures (SST) over the tropical eastern Pacific ocean, composed of two phases: warm ENSO or El Niño (EN), and cold ENSO or La Niña (LN).

EN periods are associated with a positive pressure anomaly in the tropical western Pacific. The changes in SST and surface pressure are related to extremes of weather in many extratropical regions; one mechanism by which the North American continent is influenced is via a low pressure anomaly in the North Pacific ocean during EN periods. The influence of North American weather via this anomaly is referred to as a teleconnection and has been shown to be modulated by the QBO phase (defined at 50 hPa); *Garfinkel and Hartmann* (2008), *Barnston et al.* (1991) show in particular that under an easterly QBO phase, the low pressure anomaly associated with EN is weaker than under a westerly QBO phase, hence the NP teleconnection is stronger under westerly QBO.

El Niño teleconnections with the North Pacific are also related to planetary Rossby wave propagation. The increase in SST associated with El Niño periods drives anomalously strong convection in the tropical pacific, acting as a source of vorticity which generates a 'Rossby wave train' propagating northwards to influence NP and Canada (*Hoskins and Karoly*, 1981).

Garfinkel and Hartmann (2010) suggest 3 potential pathways for the QBO to influence EN teleconnections:

- Direct effects on divergent outflow from EN anomalous convection.
- Direct weakening of the North Pacific low pressure anomaly under easterly QBO.

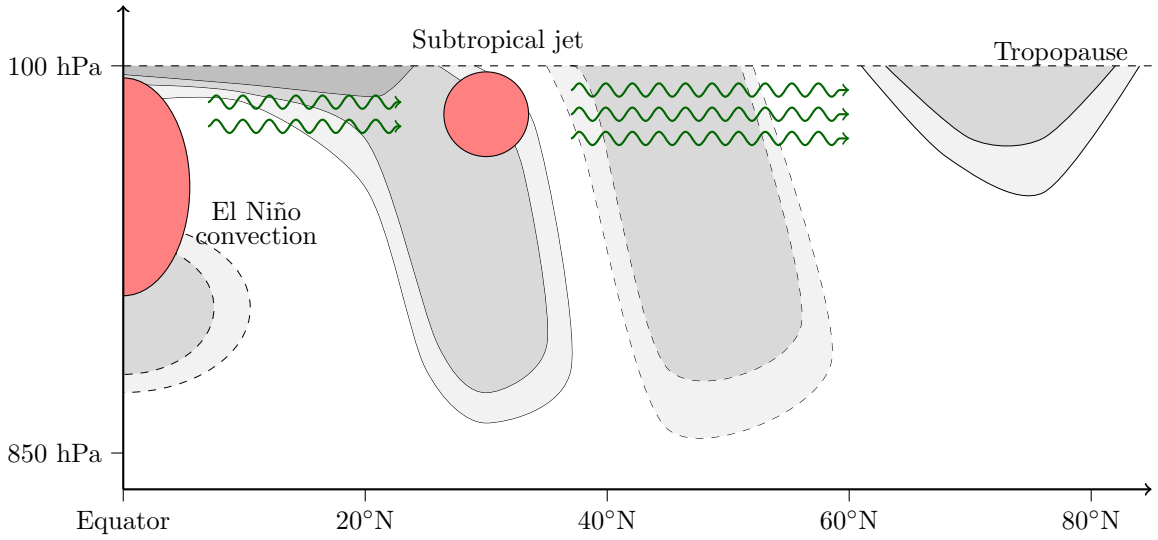


Figure 12: Schematic diagram of zonal wind anomalies during westerly QBO phase. Solid (dashed) shaded regions are positive (negative) anomalies. Anomalous El Niño convection indicated by red region at equator. Subtropical jet indicated by red region at 30°N. Propagation of planetary Rossby waves indicated by green wavy lines.

- Influence on Pacific tropospheric winds, hence modulating planetary wave propagation.

The first two pathways are shown to account for at most half of the geopotential height anomalies in the North Pacific, hence we focus on the mechanism by which the QBO can modify planetary wave propagation in the troposphere. The branching of QBO zonal wind anomalies due to thermal wind balance was discussed with reference to the Holton-Tan relationship in section 3.2.2, and here the extension of wind anomalies into the troposphere is important again. *Crooks and Gray* (2005) identify easterly anomalies in the deep tropics and mid latitudes, with westerly anomalies in the subtropics (near 35°N) and high latitudes during westerly QBO. These wind anomalies are broadly in geostrophic balance with geopotential height anomalies. *Garfinkel and Hartmann* (2010) use a shallow water model to investigate the effect of these wind anomalies; a shallow water model is used as it is simplistic but still resolves Rossby waves. The anomalous convection associated with El Niño is represented by a source of vorticity in the deep tropics, and QBO wind anomalies are parameterised at the approximate locations seen in observational data. Their model replicates the weakened response seen under easterly QBO, but not the exact location of the NP pressure anomaly. The modulation provided by wind anomalies and their effect on NP anomaly growth can be explained via energy considerations of the eddy component of the flow (representing Rossby waves) and also via wave activity density.

3.4.1 Eddy kinetic energy view of EN teleconnection modulation

The eddy kinetic energy equation in local Cartesian coordinates is derived and simplified by *Simmons et al.* (1983) and describes the growth of kinetic energy E associated with deviations from the basic state:

$$\frac{\partial E}{\partial t} = -\overline{(u'^2 - v'^2)\frac{\partial u_0}{\partial x}} - \overline{u'v'\frac{\partial u_0}{\partial y}} \tag{3.16}$$

where (u_0, v_0) is the basic horizontal flow. Dissipation effects are neglected. The second term is of particular interest here; given a zonally-uniform basic state, it represents the exchange of kinetic energy (KE) between mean and eddy form. The eddy energy grows when the meridional transfer of zonal momentum by eddies, represented by $\overline{u'v'}$, is in the direction of a weakening westerly component of the basic flow (*Kuo*, 1951).

In the context of (3.16), the modulation of EN teleconnections by the QBO can be understood as enhancement of the eddy KE under westerly QBO. Figure 12 shows the approximate locations of westerly QBO tropospheric wind anomalies and the location of anomalous convection at the equator, and the subtropical jet at $\sim$ 35°N. The subtropical jet, much like the polar tropospheric jet, plays a role in

propagating weather systems around the planet and its modification and intensity is therefore particularly relevant to understanding EN teleconnections. Initially, eddy kinetic energy is generated by anomalous EN convection at the equator. The planetary waves carrying this eddy KE propagate northwards. The subtropical jet acts as an intense gradient of vorticity of the background flow, consequently $\frac{\partial u_0}{\partial y}$ is large and since the meridional transfer of zonal momentum in this region is from a westerly to an easterly region, the eddy KE increases and the planetary wave anomaly is enhanced. The enhanced anomaly propagates northward into the North Pacific and the low pressure anomaly is strengthened as the planetary waves dissipate and accelerate the zonal flow.

3.4.2 Wave activity view of EN teleconnection modulation

An equivalent description of the EN teleconnection modulation can be given in terms of wave activity density $\mathcal{A}$, proportional to the enstrophy $\overline{q'^2}/2$ which may be interpreted as measuring the kinetic energy of the flow which is as a result of turbulence or eddy components. Strictly speaking, to consider wave activity density we require a zonally symmetric basic state, whilst in the real atmosphere we are concerned only with a specific zonal region. The argument is admissible since it can be shown that the pacific sector variability dominates the total wave activity density and enstrophy (*Garfinkel and Hartmann*, 2010). Wave activity density $\mathcal{A}$ is governed by the enstrophy equation, which is a wave activity conservation relation in the form of (2.3):

$$\frac{\partial \mathcal{A}}{\partial t} + \overline{v'q'} = \frac{\overline{q'\mathcal{D}'}}{\overline{q}_y} \tag{3.17}$$

where $\mathcal{D}$ represents dissipation of wave activity and the wave activity density is given by

$$\mathcal{A} = \frac{1}{2}\frac{\overline{q'^2}}{\overline{q}_y} \tag{3.18}$$

Anomalous EN convection near the equator generates vorticity, which imposes $\overline{q'^2}$, therefore generating wave activity in the region. The local eddy flux of potential vorticity, represented by $\overline{v'q'}$, propagates $\mathcal{A}$ northward. Near the subtropical jet, the gradient of background vorticity $\overline{q}_y$ is large, hence $\overline{q'^2}$ increases to balance it, representing the enhanced planetary wave anomaly discussed above. The modulation by the QBO arises through changes in the background vorticity gradient structure due to zonal wind anomalies. Under easterly QBO, the opposite wind anomalies are seen to westerly QBO, i.e. opposite to those shown in figure 12. The easterly anomaly close to the subtropical jet reduces $\overline{q}_y$, hence the increase in enstrophy $\overline{q'^2}$ caused by the subtropical jet is smaller and a comparatively weaker anomaly propagates into the North Pacific. Therefore the teleconnection is weaker under easterly QBO.

To summarise, the structure of QBO zonal wind anomalies during the westerly phase and the resulting strengthened subtropical jet enhances Rossby wave propagation from anomalous El Niño convection into the North Pacific, by requiring more energy conversion from the background state to the eddy component of the flow via stronger vorticity gradients compared to easterly QBO conditions.

4 Conclusion

The quasi-biennial oscillation is an essential element of atmospheric dynamics as the dominant mode of variability in the equatorial stratosphere, with consequences for numerous other atmospheric phenomena. The oscillation is simple in essence but complex in detail, perhaps best illustrated by the difficulty in identifying the underlying wave/mean-flow mechanism; precise details on the spectrum of wave forcing constituting the QBO momentum budget are still under active research. More than 50 years since its discovery, the QBO remains challenging to study owing to the multiple scales involved and the complex feedback of the oscillation with the wider atmosphere. Representation of the QBO in global climate models is still lacking (*Bushell et al.*, 2020), though a realistic oscillation and sufficiently representative global effects can be modelled via parameterisation.

In this essay, we have discussed the basic dynamics of the QBO, using the model of *Plumb* (1977) to illustrate the essential wave/mean-flow interaction mechanism whereby a spectrum of waves with opposing phase speeds provide vertical transport of momentum to accelerate the mean zonal wind, which

modifies the location of critical layers where the waves dissipate. This two-way feedback gives rise to the quasi-biennial oscillation. From this model, we saw how asymmetries in phase speed, momentum flux and complicating factors such as rotation, external forcing and variable damping contribute to generating the QBO as it is observed. Decomposing the spectrum of waves which contribute to the QBO momentum budget is a complex issue that highlights the limitations in current atmospheric models and the inferences which can be made from re-analysis data. Whilst the QBO is an oscillation of mean zonal wind at its core, a signal is apparent in temperature and other coupled dynamical fields.

Moreover, the physical balances present in the equatorial stratosphere induce meridional motions which couple with the wider stratosphere. These motions, subject to radiative damping, explain the equatorial confinement of the oscillation and through modification of the background fields on which planetary waves propagate, provide pathways for extratropical influence of the QBO. The particular sensitivity of the stratosphere to external factors such as solar heating and planetary wave activity allows the subtle dynamical changes between QBO phases to yield far greater effects, from weakening of the polar vortex to modification of pressure anomalies, and in extreme cases these effects may contribute to far more dramatic events such as sudden stratospheric warmings.

4.1 Incomplete discussions

Naturally, many discussions in this essay are left somewhat incomplete. In particular, the complexities of the QBO meridional structure and interaction with the climatology of the equatorial stratosphere have been omitted as they constitute essays in their own right.

Beyond the Gaussian structure in zonal wind amplitude, the meridional structure of the QBO has not been discussed as it depends on the particular phase of the oscillation as well as the time of year and climatology of the subtropics. Studies such as (*Pahlavan et al.*, 2021a) highlight the influences of separate components of the dynamics, such as maintenance of the poleward flanks of easterly wind regimes and the bottom edges of westerly wind regimes by meridional advection. Further, seasonal asymmetry is readily apparent in reanalysis data sets and can be related to the seasonality of trace species of ozone (*Randel and Wu*, 1996) but the seasonality is not clearly explained in the context of stratospheric climatology.

Another aspect frequently mentioned but not detailed is the apparent seasonal onset preference, since it requires a significant discussion of the interaction between the QBO and semi-annual oscillation. In the upper stratosphere, the QBO and SAO share westerly shear zones (*Smith et al.*, 2017) and once every few annual cycles, SAO wind regimes propagate downwards via the same wave/mean-flow mechanism leading to the downward propagation of QBO wind regimes (*Garcia et al.*, 1997; *Gray*, 2010). However, the interaction is complex to describe as during a QBO phase with weakening westerlies in the lower stratosphere, more eastward momentum reaches the upper stratosphere, increasing forcing of SAO shear zones and leading to deeper penetration into the stratosphere (*Krismer et al.*, 2013). There is thus a further element of feedback between the QBO and SAO present.

4.2 Further study

The QBO remains a very active area of research in the modern literature. In this essay, particular topics subject to further study are the representation of the QBO in global climate models and the parameterisation of subgrid scale gravity waves.

A discussion of the difficulties of spontaneously generating a QBO in climate models can be found in *Baldwin et al.* (2001). Recently, an initiative to diagnose the defficiencies in QBO representation has been set up, with preliminary results supporting the need for better parameterisation of mesoscale gravity waves (*Bushell et al.*, 2020). Part of the complexity in determining to what extent current representations of gravity waves falter is the cross-correlation of reanalysis corrections and gravity wave parameterisation (*Pahlavan et al.*, 2021a). Nonetheless, these subgrid scale waves are known to be large contributors to the QBO momentum budget. Another aspect of study is their intermittency, which is related to the irregular descent of easterly shear zones (*Ern et al.*, 2014).

As the availability and resolution of satellite data continues to improve, future studies may focus on verifying mechanistic explanations of QBO extratropical influence as discussed in section 3. Studies such as *Garfinkel et al.* (2012); *Lu et al.* (2014); *Kodera and Kuroda* (2002) emphasise that their discussion

is *diagnostic*, with the aim of presenting a coherent description of the dynamical response to certain atmospheric conditions, without necessarily providing an underlying causal relationship. The need for further research in this area is highlighted by the contradictory conclusions on the vertical region of the QBO responsible for modulation of the stratospheric polar vortex.

References

D. G. Andrews and M. E. McIntyre. Planetary waves in horizontal and vertical shear: The generalised Eliassen-Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, 33:2031–2048, 1976.

J. K. Angell and J. Korshover. Quasi-biennial variations in temperature, total ozone, and tropopause height. *J. Atmos. Sci.*, 21:479–492, 1964.

J. A. Anstey and T. G. Shepherd. High-latitude influence of the quasi-biennial oscillation. *Q. J. R. Meteorol. Soc.*, 140:1–21, 2014.

M. P. Baldwin, L. J. Gray, T. J. Dunkerton, K. Hamilton, P. H. Haynes, W. J. Randel, J. R. Holton, M. J. Alexander, I. Hirota, T. Horinouchi, D. B. A. Jones, J. S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi. The Quasi-Biennial Oscillation. *Revs. Geophys.*, 39:179–229, 2001.

A. G. Barnston, R. E. Livezey, and M. S. Halpert. Modulation of Southern Oscillation-Northern Hemisphere midwinter climate relationships by the QBO. *J. Clim.*, 4:203–217, 1991.

A. D. Belmont and D. G. Dartt. Variation with longitude of the quasi-biennial oscillation. *Mon. Weather Rev.*, 96:767–777, 1968.

J. R. Booker and F. P. Bretherton. The critical layer for internal gravity waves in a shear flow. *J. Fluid Mech.*, 27:513–539, 1967.

J. P. Boyd. The noninteraction of waves with the zonally-averaged flow on a spherical earth and the interrelationships of eddy fluxes of energy, heat and momentum. *J. Atmos. Sci.*, 33:2285–2291, 1976.

A. C. Bushell, J. A. Anstey, N. Butchart, Y. Kawatani, S. M. Osprey, J. H. Richter, F. Serva, P. Braesicke, C. Cagnazzo, C.-C. Chen, H.-Y. Chun, R. R. Garcia, L. J. Gray, K. Hamilton, T. Kerzenmacher, Y.-H. Kim, F. Lott, C. McLandress, H. Naoe, J. Scinocca, A. K. Smith, T. N. Stockdale, S. Versick, S. Watanabe, K. Yoshida, and S. Yukimoto. Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBO-initiative. *Q. J. R. Meteorol. Soc.*, 2020. doi: https://doi.org/10.1002/qj.3765.

C. D. Camp and K. K. Tung. The influence of the solar cycle and QBO on the late-winter stratospheric polar vortex. *J. Atmos. Sci.*, 64:1267–1283, 2007.

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

S. A. Crooks and L. J. Gray. Characterization of hre 11-year solar signal using a multiple regression analysis of the ERA-40 data set. *J. Clim.*, 18:996–1015, 2005.

T. J. Dunkerton. Annual variation of deseasonalized mean flow acceleration in the equatorial lower stratosphere. *J. Meteorol. Soc. Jpn.*, 68:499–508, 1990.

T. J. Dunkerton. Nonlinear propagation of zonal winds in an atmosphere with Newtonian cooling and equatorial wavedriving. *J. Atmos. Sci.*, 48:236–263, 1991.

T. J. Dunkerton. The role of gravity waves in the quasi-biennial oscillation. *J. Geophys. Res.*, 10:26,053–26,076, 1997.

T. J. Dunkerton, C. P. F. Hsu, and M. E. McIntyre. Distribution of major stratospheric warmings in relation to the quasi-biennial oscillation. *J. Atmos. Sci.*, 38:819–843, 1981.

R. A. Ebdon. Notes on the wind flow at 50 mb in tropical and subtropical regions in January 1957 and in 1958. *Q. J. R. Meteorol. Soc.*, 86:540–542, 1960.

R. A. Ebdon and R. G. Veryard. Fluctuations in equatorial stratospheric winds. *Nature*, 189:791–793, 1961.

A. Eliassen and E. E. Palm. On the transport of energy in stationary mountain waves. *Geophys. Publ.*, 22:1–23, 1961.

M. Ern and P. Preusse. Wave fluxes of equatorial Kelvin waves and QBO zonal wind forcing derived from SABER and ECMWF temperature space-time spectra. *Atmos. Chem. Phys.*, 9:3957–3986, 2009.

M. Ern, F. Ploeger, P. Preusse, J. C. Gille, L. J. Gray, S. Kalisch, M. G. Mlynczak, J. M. Russel III, and M. Riese. Interaction of gravity waves with the QBO: A satellite perspective. *J. Geophys. Res. Atmos.*, 119:2329–2355, 2014.

R. R. Garcia, T. J. Dunkerton, R. S. Lieberman, and R. A. Vincent. Climatology of the semiannual oscillation of the tropical middle atmosphere. *J. Geophys. Res. Atmos.*, 102:26019–26032, 1997.

C. I. Garfinkel and D. L. Hartmann. Different ENSO teleconnections and their effects on the stratospheric polar vortex. *J. Geophys. Res.*, 113, 2008.

C. I. Garfinkel and D. L. Hartmann. The influence of the Quasi-Biennial Oscillation on the North

Pacific and El Nino teleconnections. *J. Geophys. Res. Atmos.*, 115:D20116, 2010.

C. I. Garfinkel, T. A. Shaw, D. L. Hartmann, and D. W. Waugh. Does the Holton-Tan mechanism explain how the quasi-biennial oscillation modulates the arctic polar vortex? *J. Geophys. Res. Atmos.*, 116:1713–1733, 2012.

L. J. Gray. The influence of the equatorial upper stratosphere on stratospheric sudden warmings. *Geophys. Res. Lett.*, 30(4):1166, 2003.

L. J. Gray. Stratospheric equatorial dynamics. *Washington DC American Geophysical Union Geophysical Monograph Series*, 190:93–107, 2010.

L. J. Gray and J. A. Pyle. A two-dimensional model of the quasi-biennial oscillation in ozone. *J. Atmos. Sci.*, 56:977–993, 1999.

L. J. Gray, S. J. Phipps, T. J. Dunkerton, M. P. Baldwin, E. F. Drysdale, and M. R. Allen. A data study of the influence of the equatorial upper stratosphere on northern-hemisphere stratospheric sudden warmings. *Q. J. R. Meteorol. Soc.*, 127(576):1985–2003, 2001.

P. Graystone. Meteorological office discussion – Tropical meteorology. *Meteorol. Mag.*, 88:113–119, 1959.

K. Hamilton. Mean wind evolution through the quasi-biennial cycle in the tropical lower stratosphere. *J. Atmos. Sci.*, 41:2113–2125, 1984.

P. H. Haynes. The latitudinal structure of the quasi-biennial oscillation. *Q. J. R. Meteorol. Soc.*, 124:2645–2670, 1998.

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

J. R. Herring. Investigation of problems in thermal convection. *J. Atmos. Sci.*, 20:325–338, 1963.

M. H. Hitchman and A. S. Huesmann. Seasonal influence of the quasi-biennial oscillation on stratospheric jets and Rossby wave breaking. *J. Atmos. Sci.*, 66:935–946, 2009.

L. A. Holt, M. J. Alexander, L. Coy, A. Molod, W. Putman, and S. Pawson. Tropical waves and the quasi-biennial oscillation in a 7-km global climate simulation. *J. Atmos. Sci.*, 73:3771–3783, 2016.

J. R. Holton and R. S. Lindzen. An updated theory for the quasi-biennial cycle of the tropical stratosphere. *J. Atmos. Sci.*, 29:1076 – 1080, 1972.

J. R. Holton and H. C. Tan. The influence of the equatorial quasi-biennial oscillation on the global circulation at 50mb. *J. Atmos. Sci.*, 37:2200–2207, 1980.

J. R. Holton and H. C. Tan. The quasi-biennial oscillation in the Northern Hemisphere lower stratosphere. *J. Meteorol. Soc. Japan*, 60:140–147, 1982.

B. J. Hoskins and D. Karoly. The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, 38:1179–1196, 1981.

L. N. Howard. Note on a paper of John W. Miles. *J. Fluid Mech.*, 10:509–512, 1961.

K. Kodera and Y. Kuroda. Dynamical response to the solar cycle. *J. Geophys. Res.*, 107(D24): 4749–4761, 2002.

T. R. Krismer, M. A. Giorgetta, and M. Esch. Seasonal aspects of the quasi-biennial oscillation in the Max Planck Institute Earth System Model and ERA-40. *Journal of Advances in Modelling Earth Systems*, 5:406–421, 2013.

H.-L. Kuo. A note on the kinetic energy balance of the zonal wind systems. *Tellus*, 3:205–207, 1951.

R. S. Lindzen. Equatorial planetary waves in shear. Part I. *J. Atmos. Sci.*, 28:609–622, 1971.

R. S. Lindzen and J. R. Holton. A theory of the quasi-biennial oscillation. *Q. J. R. Meteorol. Soc.*, 123:2041–2068, 1968.

H. Lu, M. P. Baldwin, L. J. Gray, and M. J. Jarvis. Decadal-scale changes in the effect of the QBO on the northern stratospheric polar vortex. *J. Geophys. Res. Atmos.*, 113(D10), 2008.

H. Lu, T. J. Bracegirdle, T. Phillips, A. Bushell, and L. J. Gray. Mechanisms for the Holton-Tan relationship and its decadal variation. *J. Geophys. Res. Atmos.*, 119:2811–2830, 2014.

T. Matsuno. Quasi-geostrophic motions in the equatorial area. *J. Meteorol. Soc. Jpn.*, 44:25–43, 1966.

T. Matsuno. Vertical propagation of tationary planetary waves in the winter northern hemisphere. *J. Atmos. Sci.*, 27:871–883, 1970.

M. E. McIntyre. The quasi-biennial oscillation (QBO): Some points about the terrestrial QBO and the possibility of related phenomena in the solar interior. In E. Nesme-Ribes, editor, *The Solar Engine and Its Influence on the Terrestrial Atmosphere and Climate*, pages 293–320, New York, 1994. Springer-Verlag.

J. W. Miles. On the stability of heterogeneous shear flows. *J. Fluid Mech.*, 10:496–508, 1961.

Y. Naito and S. Yoden. Behaviour of planetary waves before and after Stratospheric Sudden Warming events in several phases of the equatorial QBO. *J. Atmos. Sci.*, 63:1637–1649, 2006.

D. A. Ortland. Rossby wave propagation into the tropical stratosphere observed by the High Resolution Doppler Imager. *Geophys. Res. Lett.*, 24: 1,999–2,002, 1997.

D. O'Sullivan. Interaction of extratropical Rossby waves with westerly quasi-biennial oscillation winds. *J. Geophys. Res.*, 102:19,461–19,469, 1997.

H. A. Pahlavan, Q. Fu, J. M. Wallace, and G. N. Kiladis. Revisiting the Quasi-Biennial Oscillation as seen in ERA5. Part I: Description and Momentum Budget. *J. Atmos. Sci.*, 78:673–691, 2021a.

H. A. Pahlavan, Q. Fu, J. M. Wallace, and G. N. Kiladis. Revisiting the Quasi-Biennial Oscillation as seen in ERA5. Part II: Evaluation of Waves and Wave Forcing. *J. Atmos. Sci.*, 78:693–707, 2021b.

C. E. Palmer. The general circulation between 200 mb and 10 mb over the equatorial Pacific. *Weather*, 9:3541–3549, 1954.

R. A. Plumb. Momentum transport by the thermal tide in the stratosphere of Venus. *Q. J. R. Meteorol. Soc.*, 101:763–776, 1975.

R. A. Plumb. The interaction of two internal waves with the mean flow: Implications for the theory of the quasi-biennial oscillation. *J. Atmos. Sci.*, 34:1847–1858, 1977.

R. A. Plumb. The quasi-biennial oscillation. In J. R. Holton and T. Matsuno, editors, *Dynamics of the Middle Atmosphere*, pages 217–251, Tokyo, 1984.

R. A. Plumb. Planetary waves and the extratropical winter stratosphere. *Washington DC American Geophysical Union Geophysical Monograph Series*, 190:23–41, 2010.

R. A. Plumb and R. C. Bell. Equatorial waves in steady zonal shear flow. *Q. J. R. Meteorol. Soc.*, 108:313–334, 1982a.

R. A. Plumb and R. C. Bell. A model of the quasi-biennial oscillation on an equatorial beta-plane. *Q. J. R. Meteorol. Soc.*, 108:335–352, 1982b.

R. A. Plumb and A. D. McEwan. The instability of a forced standing wave in a viscous stratified fluid: A laboratory analogue of the quasi-biennial oscillation. *J. Atmos. Sci.*, 35:1827–1839, 1978.

W. J. Randel and F. Wu. Isolation of the ozone QBO in SAGE II data by singular-value decomposition. *J. Atmos. Sci.*, 53:2546–2559, 1996.

W. J. Randel, F. Wu, R. Swinbank, J. Nash, and A. O'Neill. GLobal QBO circulation derived from UKMO stratospheric analyses. *J. Atmos. Sci.*, 56:457–474, 1999.

R. J. Reed, W. J. Campbell, L. A. Rasmussen, and R. G. Rogers. Evidence of a downward propagating annual wind reversal in the equatorial stratosphere. *J. Geophys. Res.*, 66:813–818, 1961.

A. J. Simmons, J. M. Wallace, and G. W. Branstator. Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. Atmos. Sci.*, 40:1363–1392, 1983.

A. K. Smith, R. R. Garcia, A. C. Moss, and N. J. Mitchell. The semiannual oscillation of the tropical zonal wind in the middle atmosphere derived from satellite geopotential height retrievals. *J. Atmos. Sci.*, 74:2413–2425, 2017.

H.-H. Tseng and Q. Fu. Temperature control of the variability of tropical tropopause layer cirrus clouds. *J. Geophys. Res. Atmos.*, 122:11–62, 2017.

G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation. Second Edition.* Cambridge University Press, 2017.

R. G. Veryard and R. A. Ebdon. Fluctuations in tropical stratospheric winds. *Meteorol. Mag.*, 90: 125–143, 1961.

J. M. Wallace. General circulation of the tropical lower stratosphere. *Rev. Geophys.*, 11:191–222, 1973.

P. A. G. Watson and L. J. Gray. How does the quasi-biennial oscillation affect the stratospheric polar vortex? *J. Atmos. Sci.*, 71:391–409, 2013.

Y. Yamashita, H. Akiyoshi, and M. Takahashi. Dynamical response in the Northern Hemisphere midlatitude and high-latitude winter to the QBO simulated by CCSR/NIES CCM. *J. Geophys. Res.*, 116(D6), 2011.

Y. Yamashita, H. Akiyoshi, T. G. Shepherd, and M. Takahashi. The combined influences of westerly phase of the Quasi-Biennial Oscillation and 11-year solar maximum conditions on the Northern Hemisphere extratropical winter circulation. *J. Meteorol. Soc. Japan*, 93:629–644, 2015.

G. Y. Yang, B. J. Hoskins, and Julia M. Slingo.

Equatorial waves in opposite QBO phases. *J. Atmos. Sci.*, 68:839–862, 2011.