A New Three-Dimensional Residual Flow Theory and Its Application to Brewer–Dobson Circulation in the Middle and Upper Stratosphere

KAORU SATO, a TAKENARI KINOSHITA, b YUKI MATSUSHITA, a AND MASASHI KOHMA a

a Department of Earth and Planetary Science, The University of Tokyo, Tokyo, Japan
b Japan Agency for Marine-Earth Science and Technology, Kanagawa, Japan

ABSTRACT: This study formulates three-dimensional (3D) residual flow, treating both stationary and transient waves. The zonal and meridional momentum equations contain four terms: the geostrophic wind tendency, Coriolis force for the residual horizontal flow, product of the geostrophic wind and potential vorticity other than the constant planetary vorticity, and friction. The thermodynamic equation contains three terms: the potential temperature tendency, advection of the basic potential temperature by the residual vertical flow, and diabatic heating. The zonal mean of the 3D residual flow equals the time mean of the residual flow of the transformed Eulerian-mean equations. The new residual flow is the sum of that derived by Plumb for transient waves and the quadratic terms of the time-mean fields, which correspond approximately to the Stokes correction due to stationary waves. The 3D residual flow and momentum equations are symmetric in the zonal and meridional directions, in contrast with those formulated by Kinoshita et al., which treat the time-mean zonal-mean zonal wind as the basic wind. The newly derived formulas are applied to the climatology of the 3D structure of the deep branch of the Brewer–Dobson circulation. In the Northern Hemisphere in December–February, the residual flows are directed inward toward the polar vortex strongly over east Siberia, where the downward flow is maximized, and weakly over the Atlantic; meanwhile, they are directed outward from the vortex over North America and Europe. A longitudinal dependence of the poleward flow is also observed in the Southern Hemisphere in June–August.

KEYWORDS: Lagrangian circulation/transport; Mass fluxes/transport; Stratospheric circulation; Waves, atmospheric; Angular momentum

1. Introduction

Brewer–Dobson circulation (BDC) (e.g., Butchart 2014) is a general circulation in the stratosphere that consists of transport via the Lagrangian-mean flow with long time scales (i.e., weeks and longer) and mixing via turbulence with short time scales (i.e., days to weeks). The Lagrangian-mean flow determines the overall spatial structure of BDC. The zonal mean Lagrangian mean flow in the meridional cross section is approximated by the residual mean flow defined in the transformed Eulerian-mean (TEM) equation system (e.g., Andrews and McIntyre 1976, 1978; Andrews et al. 1987). The residual mean flow is expressed as the sum of the Eulerian-mean flow and the quasi-Stokes correction, hereafter referred to simply as the Stokes correction, and is maintained by wave forcing, which is represented by the Eliassen–Palm (EP) flux divergence, and by diabatic heating. The EP flux divergence is zero for steady and conservative waves, that is, waves that do not accelerate the mean flow (nonacceleration theorem: Eliassen and Palm 1961; Charney and Drazin 1961). In the middle atmosphere, diabatic heating is secondary and the most important process driving BDC is forcing caused by atmospheric waves to the mean flow; such forcing occurs as a result of the unsteadiness of the atmospheric waves or nonconservative processes such as wave breaking and radiative dissipation.

BDC consists of deep and shallow branches. In solstitial seasons, the deep branch has a one-celled structure in the middle and upper stratosphere and is driven primarily by planetary waves in the winter hemisphere (e.g., Plumb 2002). The upwelling of the deep branch penetrates into the summer hemisphere (e.g., Rosenlof 1995). The shallow branch is a two-celled circulation in the lowermost stratosphere from the equatorial region to the northern and southern mid- and high-latitude regions and is driven primarily by synoptic-scale waves (e.g., Plumb 2002). In this way, BDC is primarily driven by forcing by quasi-geostrophic waves, which hereafter are simply referred to as Rossby waves. On the other hand, the importance of gravity wave forcing to BDC has also been indicated by recent diagnostic studies using reanalysis datasets (e.g., Okamoto et al. 2011; Seviour et al. 2012; Sato and Hirano 2019) and models (e.g., Polichtchouk et al. 2018). According to the previous studies, gravity waves contribute to the formation of upwelling at low latitudes, to the extension of the deep branch to high latitudes in the middle and upper stratosphere, and to the formation of the midlatitude shallow branch. Sato and Hirano (2019) indicated that these contributions of gravity waves are not necessarily precisely expressed by their parameterizations. It is also shown that gravity wave forcing is important for the dynamics of springtime polar vortex breakdown in the Southern Hemisphere (e.g., McLandress et al. 2012; Gupta et al. 2021).

Even though the TEM equations are powerful theoretical tools to analyze BDC, their application is limited to the zonal-
mean field in the meridional cross section. Various attempts have been made to examine the three-dimensional (3D) structure of BDC. Lagrangian approach using trajectory calculation has been made to examine 3D transport in the troposphere and stratosphere by various studies (e.g., Manney et al. 1994a,b; Stohl 2001; Seo and Bowman 2002; Piani et al. 2002; Ollers et al. 2002). Manney et al. (1994a) examined the sudden stratospheric warming that occurred in 1979 and found that diabatic vertical motions are evident in the baroclinic zone. Manney et al. (1994b) also showed interesting 3D features such as air from the stratospheric polar vortex moving into low latitudes in the form of elongated tongues. Callaghan and Salby (2002) examined the horizontal structure of the trajectory of air parcels and mass-weighted vertical flow for the winter stratosphere as simulated under realistic boundary conditions by a Hough spectral model formulated from the primitive equations in isentropic coordinates. They showed a large longitudinal dependence in the vertical flow, with a maximum downward flow over Siberia at \textasciitilde 60\degree N. It was suggested that this strong downward flow was caused by strong radiative cooling of air parcels advected into the polar night region by the polar night jet surrounding a significantly deformed polar vortex. Sato et al. (2009) examined the dynamical features of the dissipation process of the Antarctic ozone hole using data from intensive ozonesonde observations and satellite observations. They showed that there is significant longitudinal dependence in the descent rate of the ozone mixing ratio contours corresponding to the downward Lagrangian-mean flow. The longitudinal dependence is primarily explained by changes in the undulated isentropic surfaces with time as a result of the amplification and slight phase movement of a quasi-stationary planetary wave while the zonal-mean isentropic surface descends over seasonal time scales. They also reported the contribution of longitudinally dependent horizontal mixing, even though this contribution is secondary. These previous studies suggest that the BDC has significant 3D structures, and the Lagrangian approach provides interesting insights into this. However, this approach needs to compute many Lagrangian trajectories and has an ambiguity of quantitative descriptions. Therefore, it would be useful to establish a quantitative analysis method for such 3D structures by generalizing the TEM equations to three dimensions, which gives an approximation of the 3D Lagrangian mean flow.

Plumb (1986) developed a theory of 3D residual mean flow for the quasigeostrophic (QG) equations, in which the time-mean field and the deviation from the time mean are treated as the mean field and the disturbance field, respectively. Kinoshita and Sato (2013) derived formulas for the 3D residual mean flows, which are the sum of the Eulerian-mean flow and the Stokes correction to the primitive equations. However, the 3D residual mean flows derived by these two studies do not contain the Stokes correction due to stationary waves, which is significant, particularly in the winter stratosphere. Demirhan Bari et al. (2013) examined the effects of zonal asymmetries in the BDC on stratospheric ozone and middle atmospheric water vapor using the formulas by Kinoshita and Sato (2013). Their results, therefore, only include the Stokes correction by transient waves. In the presence of stationary waves, there is a lead current along the isentropic surface that is undulated by the waves themselves. For the adiabatic heating to be balanced with the residual vertical advection, this lead current should not be included in the 3D residual vertical flow because it does not flow through the isentropic surface (e.g., Kinoshita et al. 2019). Kinoshita et al. (2019) showed the lead current can be eliminated by adding the Stokes correction due to stationary waves. Prior to this study, Sato et al. (2013) proposed to apply an extended Hilbert transform to the formulas of Kinoshita and Sato (2013) to include stationary waves. However, it was carefully shown by Kinoshita et al. (2019) that although this method gives an appropriate wave activity flux of stationary waves, it does not provide an accurate estimate of the Stokes correction due to stationary waves.

The residual vertical flow should be such a flow that advects the potential temperature balancing with the diabatic heating. In addition, inherently, the residual mean flow in the TEM equations does not include the geostrophic flow. This means that the geostrophic flow does not contribute to the net mean mass transport in the meridional cross section. Therefore, it is important to examine the 3D residual flow without the geostrophic flow. Kinoshita et al. (2019) extended the 3D residual flow of Plumb (1986) to include the Stokes correction due to stationary waves. The zonal mean of the 3D residual vertical flow formulated by Kinoshita et al. (2019) accords the time-mean TEM residual mean flow for the OG system. However, their formulas for the 3D residual horizontal flow and the meridional momentum equation are somewhat complicated because the time-mean zonal-mean zonal wind is treated as the basic field.

Kanno and Iwasaki (2018) developed a theory for the 3D mass-weighted isentropic time-mean (T-MIM) meridional circulation. This theory is an extension of the theory by Iwasaki (1989), hereafter referred to as the Z-MIM theory, which describes the mass-weighted isentropic zonal-mean meridional circulation corresponding to the Lagrangian-mean flow in the meridional cross section, similar to the TEM residual mean flow. The Z-MIM flow includes the zonal mean bolus velocities of both stationary and transient waves, where the bolus velocity in the MIM system corresponds to the Stokes correction in the TEM system. The Z-MIM theory properly describes the flow near the ground surface, which is advantageous compared with the TEM theory. Another advantage of the Z-MIM theory, which is a variant of the isentropic coordinate system, is that, unlike the TEM system, it can handle finite amplitude disturbances. In the T-MIM theory, the mass-weighted time mean instead of the mass-weighted zonal mean is used, which allows a 3D description of the meridional circulation. However, there is a similar limitation to this theory as to that of Plumb (1986). The T-MIM theory, like that of Plumb (1986), deals only with transient waves and does not describe the bolus velocity of stationary waves. The mass-weighted isentropic zonal mean of the T-MIM flow is the same as the mass-weighted isentropic time mean of the Z-MIM flow, as the zonal-mean bolus velocity of the stationary waves is included in the mass-weighted isentropic zonal mean for the T-MIM flow [Eq. (13) of Kanno and Iwasaki (2018)]. However, the T-MIM flow itself describes only the 3D circulation, which does not include the bolus velocity due to stationary waves.
The purpose of this study is to formulate a new 3D residual flow that includes the ageostrophic flow and Stokes correction due to both transient and stationary waves from the QG equations in log-pressure (log-\( p \)) coordinates. This is an extension of the theory of Kinoshita et al. (2019); however, no basic zonal wind field is assumed in the present study, which makes the derived formulas and equations symmetric in the zonal and meridional directions. As an application of this theory, the winter 3D structure of the BDC deep branch is examined using a reanalysis dataset. In section 2, we formulate the new 3D residual flow and its governing equations. It is shown that the zonal mean of the 3D residual flow matches the time mean of the TEM residual flow. The meaning of the derived 3D formulas is discussed in comparison with the formulas of Plumb (1986). Differences from the 3D theory of Kinoshita et al. (2019) are also discussed. Section 3 describes the data used and the analysis method. Section 4 shows the result of the analysis regarding the climatological horizontal structure of the BDC deep branch. A summary and future perspectives are given in section 5.

2. Formulation of the 3D residual mean flow

We used the QG equations in log-\( p \) coordinates as the basic equations (e.g., Andrews et al. 1987):

\[
\frac{D}{Dt} u_g - f_0 v_g - \beta y u_g = X,  \tag{1}
\]

\[
\frac{D}{Dt} v_g + f_0 u_g + \beta y v_g = Y,  \tag{2}
\]

\[
\Phi_z = \frac{R}{H} (\theta_0 + \theta) e^{-\kappa z/H},  \tag{3}
\]

\[
\frac{D}{Dt} \theta + w_a \theta_{bc} = Q,  \tag{4}
\]

\[
u_{ax} + v_{ay} + \frac{1}{\rho_0} (\rho_0 w_a)_z = 0,  \tag{5}
\]

where

\[
\frac{D}{Dt} \equiv \dot{\theta} + u_g \frac{\partial}{\partial x} + v_g \frac{\partial}{\partial y}.  \tag{6}
\]

\[
u_g = (u_g, v_g) = \frac{1}{f_0} (-\Phi_y, \Phi_x),  \tag{7}
\]

\[
u_{gz} + v_{gy} = 0,  \tag{8}
\]

and \( u, v, \) and \( w \) are the zonal, meridional, and vertical winds, respectively. The suffixes \( g \) and \( a \) denote the geostrophic and ageostrophic components, and the suffixes \( t, x, y, \) and \( z \) denote partial derivatives with respect to the respective independent variables. The reference potential temperature and the deviation from it are denoted by \( \theta_0(z) \) and \( \theta(x, y, z, t) \), respectively; \( \Phi \) is the geopotential; \( \rho_0(z) \) is the basic density; \( f_0 \) is the Coriolis parameter; \( \beta \) is the beta parameter; \( X \) and \( Y \) are unspecified frictions and other nonconservative mechanical forcings; \( Q \) is the diabatic heating; \( H \) is the scale height (=7 km in the present study as commonly used, corresponding to a constant reference temperature of \( \sim240 \) K); \( R \) is the gas constant; and \( \kappa \) is the ratio of \( R \) to the specific heat at constant pressure.

The QG potential vorticity \( q \) is written as

\[
q \equiv f_0 + \beta y - u_{gy} + v_{gx} + \frac{f_0}{\rho_0} \left( \frac{\theta}{\theta_{bc}} \right)_z.  \tag{9}
\]

Note that the deviation from the reference potential temperature \( \theta \) is denoted by \( \theta_a \) in Andrews et al. (1987).

a. Definition of the 3D residual mean flow and governing equations

The time-mean QG thermodynamic equation (4) is written as follows:

\[
\overline{\theta_t} + 3 \overline{\theta_{bc}} + \overline{v_{g} \theta_{bc}} + \overline{w_a \theta_{bc}} = \overline{Q},  \tag{10}
\]

where the overbar indicates the time mean. Note that the time scales of the time mean are longer than the time scales for transient waves but shorter than the time scales for quasi-stationary waves and seasonal changes. The longer and shorter time scales depend on the problem of concern. Here we consider days to weeks as the shorter time scales and months and longer as the longer time scales, because we are interested in seasonal characteristics of the 3D residual flow describing the BDC in the present study. Note that we do not treat the zonal mean state as the background. For example, \( u_g \) in (10) is unfiltered and contains both zonal mean and deviation from the zonal mean. When Eq. (10) is rewriten using the continuity equation [Eq. (8)],

\[
\overline{\theta_t} + 3 \overline{\theta_{bc}} \left( \overline{w_{ax}} + \left( \frac{u_{ax} \theta}{\theta_{bc}} \right)_x + \left( \frac{v_{ax} \theta}{\theta_{bc}} \right)_y \right) = \overline{Q}.  \tag{11}
\]

Here, we define the part inside of the brace as the new 3D residual mean vertical flow:

\[
\overline{w_{ax}} \equiv \overline{w_{ax}} + \left( \frac{u_{ax} \theta}{\theta_{bc}} \right)_x + \left( \frac{v_{ax} \theta}{\theta_{bc}} \right)_y.  \tag{12}
\]

It may be worth noting that the second term on the right-hand side of (12) contains the advection by the zonal mean zonal flow. Using the residual mean vertical flow, Eq. (12), the thermodynamic equation is written as

\[
\overline{\theta_t} + 3 \overline{\theta_{bc}} \overline{w_{ax}} = \overline{Q}.  \tag{13}
\]

When the geostrophic wind equation [Eq. (7)] and the continuity equation [Eq. (8)] are used, the time-mean QG momentum equations [Eqs. (1) and (2)] can be transformed into the following equations:
\[ \nabla \cdot (f_0 + \beta y) \mathbf{v}_g - f_0 \mathbf{v}_g + \nabla \times \mathbf{n} = -\nabla f \times \mathbf{n} + \nabla \cdot \mathbf{X}, \]  
\( (14) \)

\[ \nabla \cdot (f_0 + \beta y) \mathbf{v}_g + f_0 \mathbf{v}_g + \nabla \times \mathbf{n} = -\nabla f \times \mathbf{n} + \nabla \cdot \mathbf{Y}. \]  
\( (15) \)

Meanwhile, the time-mean QG potential vorticity flux can be rewritten as:

\[ \overline{\nabla \cdot q} = \overline{\nabla \cdot \mathbf{v}_g} = \overline{\mathbf{v}_g} \left( f_0 + \beta y - \mathbf{u}_g + \mathbf{v}_g + \frac{f_0}{\rho_0} \left( \frac{\theta}{\theta_{0z}} \right)_z \right), \]  
\( (16) \)

\[ \overline{\nabla \cdot q} = \overline{\mathbf{v}_g} \left( f_0 + \beta y - \mathbf{u}_g + \mathbf{v}_g + \frac{f_0}{\rho_0} \left( \frac{\theta}{\theta_{0z}} \right)_z \right). \]  
\( (17) \)

Transforming the time-mean QG potential vorticity flux [(16) and (17)] using (3) and (7) yields

\[ \overline{\nabla \cdot q} = \left( f_0 + \beta y \right) \overline{\mathbf{v}_g} + \left( \overline{\mathbf{u}_g \mathbf{v}_g} \right)_x + \left( \overline{\mathbf{v}_g^2} \right)_y - \frac{1}{2} \left( \overline{\mathbf{u}_g^2} + \overline{\mathbf{v}_g^2} - g \frac{\theta'}{\theta_{0z}} \right)_y \]  
\[ + \frac{f_0}{\rho_0} \left( \theta \frac{\overline{\mathbf{v}_g}}{\theta_{0z}} \right)_z, \]  
\( (18) \)

\[ \overline{\nabla \cdot q} = \left( f_0 + \beta y \right) \overline{\mathbf{v}_g} + \left( \overline{\mathbf{u}_g \mathbf{v}_g} \right)_x + \left( \overline{\mathbf{v}_g^2} \right)_y - \frac{1}{2} \left( \overline{\mathbf{u}_g^2} + \overline{\mathbf{v}_g^2} - g \frac{\theta'}{\theta_{0z}} \right)_y \]  
\[ - \left( \overline{\mathbf{u}_g \mathbf{v}_g} \right)_y + \frac{f_0}{\rho_0} \left( \theta \frac{\overline{\mathbf{v}_g}}{\theta_{0z}} \right)_z. \]  
\( (19) \)

By combining Eqs. (14) and (19) and Eqs. (15) and (18), the zonal and meridional momentum equations can be rewritten as

\[ \nabla \cdot f_0 \left( \frac{\mathbf{v}_g}{f_0} \right)_j - \frac{1}{\rho_0} \left( \theta \frac{\overline{\mathbf{v}_g}}{\theta_{0z}} \right)_j + \nabla \times \mathbf{v}_g = \nabla \times \mathbf{X}, \]  
\( (20) \)

\[ \nabla \cdot f_0 \left( \frac{\mathbf{v}_g}{f_0} \right)_j - \frac{1}{\rho_0} \left( \theta \frac{\overline{\mathbf{v}_g}}{\theta_{0z}} \right)_j + \nabla \times \mathbf{v}_g = -\nabla \mathbf{q} + \nabla \times \mathbf{Y}, \]  
\( (21) \)

respectively, where

\[ \mathbf{S} = \frac{1}{2} \left( \overline{\mathbf{u}_g^2} + \overline{\mathbf{v}_g^2} - g \frac{\theta'}{\theta_{0z}} \right)_y \]  
\( \mathbf{S} = \left( \frac{\mathbf{S}}{f_0} \right)_j - \frac{1}{\rho_0} \left( \theta \frac{\overline{\mathbf{v}_g}}{\theta_{0z}} \right)_j, \]  
\( (22) \)

We define the new 3D residual-mean zonal and meridional flows (\( \overline{\mathbf{v}_g^{***}} \), \( \overline{\mathbf{v}_g^{***}} \)) as follows:

\[ \overline{\mathbf{v}_g^{***}} = \overline{\mathbf{v}_g} \left( f_0 + \beta y - \mathbf{u}_g + \mathbf{v}_g + \frac{f_0}{\rho_0} \left( \frac{\theta}{\theta_{0z}} \right)_z \right), \]  
\( (23) \)

Inserting (22) and (23) into the momentum equations gives

\[ \nabla \cdot f_0 \overline{\mathbf{v}_g^{***}} = -\nabla \times \mathbf{X}, \]  
\( (24) \)

\[ \nabla \cdot f_0 \overline{\mathbf{v}_g^{***}} = -\nabla \times \mathbf{Y}, \]  
\( (25) \)

or

\[ \nabla \cdot f_0 \overline{\mathbf{v}_g^{***}} = \mathbf{v}_g \left( q - f_0 \right) + \mathbf{X}, \]  
\( (26) \)

\[ \nabla \cdot f_0 \overline{\mathbf{v}_g^{***}} = -\mathbf{u}_g \left( q - f_0 \right) + \mathbf{Y}. \]  
\( (27) \)

The newly obtained 3D residual flow (\( \overline{\mathbf{v}_g^{***}}, \overline{\mathbf{v}_g^{***}}, \overline{\mathbf{w}_a^{***}} \)) [(22), (23), and (12)] satisfies the continuity equation as follows:

\[ \nabla \cdot \overline{\mathbf{v}_g^{***}} = \overline{\nabla \cdot \mathbf{v}_g^{***}} + \rho_0 \left( \frac{\partial \overline{\mathbf{v}_g^{***}}}{\partial z} \right)_z = 0. \]  
\( (28) \)

Note that the derived formulas of the 3D residual flow (\( \overline{\mathbf{v}_g^{***}}, \overline{\mathbf{v}_g^{***}}, \overline{\mathbf{w}_a^{***}} \)) [(22), (23), and (12)] are written as the sum of the ageostrophic flow (\( \overline{\mathbf{v}}, \overline{\mathbf{v}_g}, \overline{\mathbf{w}_a} \)) and the quadratic terms. The quadratic terms are regarded as the Stokes correction in this theory. The 3D Lagrangian mean flow is approximated by the sum of the geostrophic flow and the 3D residual flow, namely, (\( \overline{\mathbf{v}_g} + \overline{\mathbf{v}_g^{***}}, \overline{\mathbf{v}} + \overline{\mathbf{v}_g^{***}}, \overline{\mathbf{w}_a} + \overline{\mathbf{w}_a^{***}} \)). The meanings of the 3D residual flow and governing equations [(26), (27), and (13)] are discussed in sections 2c and 2d.

b. Relation with the TEM residual mean flow

The zonal means of the 3D residual mean meridional and vertical flows [(23) and (12)] and zonal momentum and thermodynamic equations [(26) and (13)] are identical to the time-mean TEM equations for the QG system, as shown below. The zonal mean of the 3D residual vertical flow (12) is as follows:

\[ [\overline{\mathbf{v}_g^{***}} = \overline{\mathbf{v}_g} + \left[ \overline{\mathbf{v}_g \theta} \right] \frac{\theta}{\theta_{0z}} + \left[ \overline{\mathbf{v}_g \theta} \right] \frac{\theta}{\theta_{0z}} + \left[ \overline{\mathbf{v}_g \theta} \right] \frac{\theta}{\theta_{0z}} = \overline{\mathbf{v}_g}, \]  
\( (29) \)

where \([\cdot]\) indicates the zonal mean and \(\cdot')\) indicates the deviation from the zonal mean. Note that \([\cdot')] = 0 and \([\cdot] = 0\). The right-hand-side term \(\overline{\mathbf{v}_g^{***}}\) of the last equality in (29) is the time-mean TEM residual vertical flow for the QG system. Similarly, the zonal mean of the 3D residual meridional flow (23) is

\[ [\overline{\mathbf{v}_g^{***}} = [\overline{\mathbf{v}_g} - \left[ \mathbf{S} \right] \frac{\theta}{\theta_{0z}} + \left[ \mathbf{S} \right] \frac{\theta}{\theta_{0z}} = [\mathbf{v}_g], \]  
\( (30) \)
where \([v_{5}]\) is the time-mean TEM residual meridional flow. In summary, the zonal means of the 3D residual meridional and vertical flows are identical to the time means of the TEM residual mean flows for the OG system. This fact ensures that the 3D residual flow \((\overline{\omega}^{***}, \overline{u}^{***}, \overline{w}^{***})\) \([22], (23), and (12)\) includes the Stokes correction due to stationary waves. The additional terms also include a term to exclude the Stokes correction due to stationary waves. Inter-
flows. Therefore, it is expected that the additional terms include the Stokes correction for transient waves (Kinoshita and Sato 2013; Sato et al. 2013). The 3D residual vertical flow of the present study \((12)\) is expressed as the sum of Plumb’s residual vertical flow \((33)\) and two terms:

\[
\overline{w}^{***} = \overline{w} + \left( \frac{\overline{u}\theta_y}{\theta_{0z}} \right)_x + \left( \frac{\overline{v}\theta_y}{\theta_{0z}} \right)_y.
\]  

\[(34)\]

An interpretation of the two additional terms, namely the second and third terms on the right-hand side of \((34)\), has been given by Kinoshita et al. (2019). Here, we repeat their explanation. Using the differential law of products, the following transformation is made:

\[
\left( \frac{\overline{u}\theta_y}{\theta_{0z}} \right)_x + \left( \frac{\overline{v}\theta_y}{\theta_{0z}} \right)_y = \frac{\overline{u}\theta_y}{\theta_{0z}} \left( \frac{\overline{\theta}_y}{\theta_{0z}} \right)_x + \frac{\overline{v}\theta_y}{\theta_{0z}} \left( \frac{\overline{\theta}_y}{\theta_{0z}} \right)_y
\]  

\[(35)\]

The first term on the right-hand side of the first equality in \((35)\) is zero because of the continuity equation \([\text{Eq. (8)}]\). When we define the time-mean (for the time scale of interest) field as \(\overline{\theta}(x,y,z) \equiv \theta(y,z) + \theta(y,x,z)\) and assume \(\overline{\theta}(x,y,z) \ll \theta(y,z)\), which is generally valid, the following relation is obtained:

\[
\overline{u}\theta_y \approx \overline{u} \left( \frac{\partial \overline{z}(x,y,\overline{\theta})}{\partial x} \right)_{\overline{\theta} = \text{const}},
\]  

\[(36)\]

where \(\left( \frac{\partial \overline{z}(x,y,\overline{\theta})}{\partial x} \right)_{\overline{\theta} = \text{const}}\) expresses the slope of the isentropic surface in the \(x\) direction at a point \((x, y)\) for \(\overline{\theta}\). Therefore, the first term on the right-hand side of the second equality \((35)\) is interpreted as the subtraction from \(\overline{w}^{***}\) of the vertical motion along the \(\overline{\theta}\) surface tilted by the presence of stationary waves in \(\overline{\theta}\), which causes no net transport across the \(\overline{\theta}\) surface \([\text{Fig. 1a}]\). More importantly, the difference between \(\overline{u}\theta_y / \theta_{0z}\) and \(\overline{u} \left( \frac{\partial \overline{z}(x,y,\overline{\theta})}{\partial x} \right)_{\overline{\theta} = \text{const}}\) corresponds to the Stokes correction across the isentropic surface due to stationary waves. A similar interpretation is possible for the second term on the right-hand side of the second equality in \((35)\).

Next, the 3D residual horizontal flow formulas are compared with those formulated by Plumb (1986), which are given as

\[
\overline{u}^{*} \equiv \overline{u} + \left( \overline{u} \theta_y \right)_{x} - \left( \frac{1}{\rho_0} \frac{\overline{u} \theta_y}{\theta_{0z}} \right)_x,
\]  

\[(37)\]

\[
\overline{v}^{*} \equiv \overline{v} - \left( \overline{v} \theta_y \right)_{x} - \left( \frac{1}{\rho_0} \frac{\overline{v} \theta_y}{\theta_{0z}} \right)_x,
\]  

\[(38)\]

where \(\overline{S} \equiv \frac{1}{2} \left( \overline{\omega}^2 + \overline{\omega}^2 - g \left( \frac{\theta}{\theta_{0z}} \right) \right)\).

The new 3D residual horizontal flows \([22] and (23)\) are given by \((37) and (38)\) and additional terms as follows:
conservation around the axis of the rotation of Earth and that the residual flow of interest is the flow across the $\mathcal{M}$ contour. The difference between $\frac{1}{2f_0} \left[ \frac{\partial}{\partial x} (\pi_2^f) \right]$ and $\pi_2^f \mathcal{M}_2/\mathcal{M}_1$ corresponds to the Stokes correction across the absolute momentum contour due to stationary waves.

A similar discussion is possible for the meridional momentum equation. Here, we consider the absolute momentum $\mathcal{M}_2 (\equiv f_0 x + \pi_2^f (x, y))$ around the axis perpendicular to the rotation axis of Earth. In the presence of stationary waves, $\mathcal{M}_2$ contours meander meridionally along the meridian. The zonal motions along the $\mathcal{M}_2$ contours should be removed from $\pi_2^f$ because such motions cause no net transport across the $\mathcal{M}_2$ surface. Such motions are included in the second part of the second term on the right-hand side of (39) as shown in the following:

$$\frac{1}{2f_0} \left( \frac{\partial \psi}{\partial y} \right)_y = \psi_{y} = \psi \frac{\mathcal{M}_{2y}}{\mathcal{M}_{2x}} = -\psi_x \left( \frac{\partial \psi}{\partial y} \right)_{\mathcal{M}_{2x=const}}.$$  \hfill (42)

Here, $\left( \frac{\partial \psi}{\partial y} \right)_{\mathcal{M}_{2x=const}}$ expresses the slope of the absolute momentum contour $\mathcal{M}_{2x}$ at a point $(y, z)$. The difference between $\left[ \frac{1}{2f_0} \left( \frac{\partial}{\partial x} (\pi_2^f) \right) \right]$ and $\psi_{y} \frac{\mathcal{M}_{2y}}{\mathcal{M}_{2x}}$ in (42) corresponds to the Stokes correction across the absolute momentum contour due to stationary waves. Similarly, it is expected that other terms of the second- and higher-order terms on the right-hand sides of (39) and (40) are Stokes corrections due to stationary waves that excludes conservative flows from $\pi_2^f$ and $\pi_2^r$, namely, the flows on the angular momentum surfaces, while maintaining continuity with the flows on the isentropic surfaces discussed above.

d. Interpretation of the horizontal momentum equations

The terms $\frac{\partial ^2}{\partial y^2}$ and $-\frac{\partial}{\partial x} \frac{\partial}{\partial z}$ in (24) and (25) are regarded as the time mean of the Coriolis force for the geostrophic wind by the potential vorticity $\varphi$ instead of the planetary vorticity $f_0$ (Kinoshita et al. 2019). The terms $\overline{\psi} (q_f - f_0)$ and $-\overline{\psi} (q_f - f_0)$ in (26) and (27) are regarded as the time mean of the horizontal flux of perturbation potential vorticity.

Because our formulation is equally treated in the zonal and meridional directions, only $\overline{\psi} q_f$ is discussed in the following:

$$\overline{\psi} q_f = \overline{\psi} q_f' + \overline{\psi} q_f = \overline{\psi} q_f' + \overline{\psi} (q_f - f_0) + \overline{\psi} f_0.$$  \hfill (43)

The term $\overline{\psi} q_f'$ in the horizontal momentum equations of Plumb (1986) expresses transient wave forcing. In contrast, Eq. (43) indicates that $\overline{\psi} q_f$ in Eq. (24) is the sum of $\overline{\psi} q'_f$ and $\overline{\psi} q_f$. Taking into account that the time mean of the TEM potential vorticity flux $\overline{\psi} q_f'$ equals that of the EP flux divergence $(1/\rho_0) \nabla \cdot \mathbf{F}_{TEM}$, and also equals $[\overline{\psi} (q_f - f_0)]$ [see (31)], it is expected that $\overline{\psi} (q_f - f_0)$ or $\overline{\psi} q_f$ includes the stationary wave forcing.

The balance represented by the zonal momentum equation [Eq. (24)] when the tendency of the mean fields is negligible and $\mathbf{X} = 0$ is schematically shown in Fig. 2. The thick vertical arrows indicate the flow direction of $\overline{\psi}$ and $\overline{\psi} q_f'$ in the horizontal plane. Horizontal arrows represent the respective terms,
with positive (negative) values indicated by rightward (leftward) arrows. Figure 2a shows the balance of all the terms in Eq. (24), that is, the balance among the pressure gradient force \(-\Phi_x\) (a dark green arrow), the Coriolis force for the residual meridional flow \(f_0 \vec{v}_a^{***}\) (a light green arrow), and \(\nabla f q\) (a brown arrow). The quantity \(\nabla f q\) is expressed as the sum of the Coriolis force for the geostrophic flow \(f_0 \vec{v}_g\) (a red arrow in Fig. 2b) and \(\nabla f(q-f_0)\) (an orange arrow in Fig. 2c). The balance in Fig. 2a is divided into the two balances shown in Figs. 2b and 2c. Figure 2b shows the geostrophic balance (i.e., the dark green and red arrows), and Fig. 2c shows the balance of the remaining terms. The balance shown in Fig. 2c exactly describes the balance in (26) when the zonal wind tendency is negligible and \(X = 0\) as we assumed. It is also possible that \(\vec{v}_a^{***}\) has the same sign as \(\vec{v}_g\), although the illustration is omitted.

The term \(\nabla f(q-f_0)\) in the term of \(\nabla f(q-f_0)\) can be interpreted as the transport of perturbation potential vorticity due to stationary waves. Therefore, the form (26) for the momentum equation has clearer meaning than the form of Eq. (24), showing the balance that should be satisfied for the residual flow. Note that Eq. (26) is also more tractable for data analyses and interpretations of results because this equation does not include the two terms of the Coriolis force for the geostrophic wind and the pressure gradient force. These two terms are large in magnitude and totally cancel each other out where the QG approximation holds.

The horizontal distribution of the perturbation potential vorticity transport \(\nabla f(q-f_0)\) due to stationary waves is not uniform, but depends on the phase of the waves, and hence the residual flow \(\vec{v}_a^{***}\) also depends on the wave phase. In the steady state (\(\nabla f q = 0\)) with \(X = 0\) (i.e., the state in Fig. 2c), the zonal momentum equation [Eq. (26)] can be interpreted as the balance at each point between the perturbation potential vorticity transport due to stationary waves and the planetary vorticity transport by the residual flow. To put it another way, there must be a residual flow to maintain the steady state against the steady transport of perturbation potential vorticity due to stationary waves at each location. A similar discussion can be made for the meridional momentum equation [Eq. (27)].

### Comparison with Kinoshita et al. (2019)

Kinoshita et al. (2019) derived the 3D residual flow \(\vec{v}_a^{***}, \vec{v}_g^{***}, \vec{w}_a^{***}\) and the momentum and thermodynamic equations by treating the time-mean zonal-mean field as the basic field:

\[
\vec{u}_a^{***} \equiv \vec{u}_x + \left(\frac{\Sigma X}{f_0}\right)_y - \frac{1}{\rho_0} \left(\rho_0 \frac{\vec{u}_x \vec{q}}{\theta_{\text{LZ}}} \right) z, \tag{44}
\]

\[
\vec{v}_a^{***} \equiv \vec{v}_y - \left(\frac{\Sigma X}{f_0}\right)_x \frac{1}{\rho_0} \left(\rho_0 \frac{\vec{u}_y \vec{q}}{\theta_{\text{LZ}}} \right) z, \tag{45}
\]

\[
\vec{w}_a^{***} \equiv \vec{w}_z + \left(\frac{\vec{u}_z \vec{q}}{\theta_{\text{LZ}}} \right) x + \left(\frac{\vec{u}_z \vec{q}}{\theta_{\text{LZ}}} \right) y, \tag{46}
\]

\[
\vec{u}_x + f_0 \vec{u}_a^{***} + \vec{u}_y = \vec{u}_g^{***} + \vec{X}, \tag{47}
\]

\[
\vec{u}_x + f_0 \vec{u}_a^{***} + (\vec{u}_y - [\vec{U}]_y) + \beta y \vec{U} = -\left(\vec{u}_g^{***} - [\vec{U}]_y \vec{q}\right) + \vec{Y}, \tag{48}
\]

\[
\vec{u}_z + \theta_{\text{LZ}} \vec{w}_a^{***} = \vec{Q}, \tag{49}
\]

where
\[
\overline{X^2}_K = \frac{1}{2} \left\{ \overline{u^2}_g - [\overline{u}_g]_0^2 + \overline{v^2}_g - [\overline{v}_g]_0^2 + g \frac{\partial^2}{\partial \theta \partial \phi} \right\}.
\]

Note that two terms, \(\beta Y[\overline{u}_g]\) and \([\overline{\Psi}]_T\), are missing in the original equation in Kinoshita et al. (2019). This correction has been made to Eq. (48).

The 3D residual meridional and vertical flow formulas [(45) and (46)] according to Kinoshita et al. (2019) are the same as our formulas [(23) and (12)], as is easily seen when considering that \([\overline{u}_g]_0 = 0\) and \([\overline{\Psi}] = 0\). The zonal momentum and thermodynamic equations [Eqs. (47) and (49)] are also the same as our Eqs. (24) and (13).

In contrast, the 3D residual zonal flow formula (44) and the meridional momentum equation [Eq. (48)] of Kinoshita et al. (2019) differ from our (22) and (25). Kinoshita et al. (2019) defined the 3D residual zonal flow such that the right-hand side of the meridional momentum equation is \(\overline{u}_g\overline{\Psi} - [\overline{u}_g][\overline{\Psi}]\) instead of \(\overline{u}_g\overline{\Psi}\). This formulation is suitable for situations in which the circumpolar westerly jet is nearly symmetric around the pole. However, this is not necessary the case, particularly for the winter Northern Hemisphere (NH) in which the climatological center of the polar night jet is significantly shifted from the geographic pole. In addition, the meridional momentum equation [Eq. (48)] by Kinoshita et al. (2019) explicitly includes the term \(\beta Y\), which complicates data analyses.

The zonal and meridional momentum equations and the 3D residual flow formulas as described here are concise and have good symmetry between the two horizontal directions, which means that any background field can be treated. It is more important that the forms of the momentum equations [Eqs. (26) and (27)] do not include large terms such as the Coriolis force for geostrophic winds and the pressure gradient force. Accordingly, these equations allow the 3D residual flow and its balance with the wave forcing to be evaluated with high precision depending on the data.

3. Data description and analysis method

To demonstrate the use of this new 3D residual flow theory, the winter 3D climatological structure of the BDC deep branch was examined. Daily mean data obtained using four-times-daily data from the European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim) (Dee et al. 2011) on a horizontal grid of \(1.5^\circ \times 1.5^\circ\) were analyzed. The purpose of the analysis is to demonstrate the theory. However, note that this is the first study showing the climatology of the 3D structure of the deep branch using reanalysis data. The analyzed time period is 37 years from December 1980 to November 2017. Because a near-steady state can be assumed during solstitial seasons, the climatology was obtained for the deep branch over the 3 months from December to February (DJF) in the Northern Hemisphere and from June to August (JJA) in the Southern Hemisphere (SH). The time mean was taken for the adjacent three months of each year. For DJF, we take December of one year and January and February of the next year for the time mean, but we refer this to DJF of each year in this study.

The 3D residual flow of the present study was formulated based on the QG equations. In the QG equations, the Coriolis parameter \(f\) is constant at a certain latitude. However, BDC is a global scale phenomenon. Therefore, we used \(f\) at each latitude as \(f_0\) to obtain the approximate structure of the 3D residual flow at a global scale. The differential operations were performed in spherical coordinates. The validity of this treatment was confirmed via comparison with the time-mean TEM residual flows for the primitive equations as shown in section 4a.

4. Climatology of the 3D structure of the BDC deep branch

a. Meridional cross sections of the zonal-mean residual flow

Before the details of the 3D residual flow are shown, an overview of the two-dimensional structure of the BDC deep branch is given here in the meridional cross section. Figure 3 shows the climatology of the mass streamfunction estimated from zonal mean residual vertical flow of the primitive TEM equations for DJF and JJA for a log-p height \(z\) of 20–45 km (e.g., Sato and Hirano 2019 for details). As is well known, the winter circulation is deeper and stronger than the summer circulation and extends to the summer hemisphere. It may be worth noting here that the summer shallow branch is deeper in DJF than in JJA. This is likely because the mean eastward wind extends deeper into the lower stratosphere in the Southern Hemisphere in DJF than in the Northern Hemisphere in JJA, resulting in higher penetration of planetary waves. See Sato and Hirano (2019) for details. In sections 4b and 4c, we illustrate the horizontal structure of this intricate part of the circulation using our newly derived formulas of the 3D residual flow.

Next, the zonal-mean meridional \([\overline{u}_g]_T\) and vertical \([\overline{v}_g]_T\) components of the 3D residual flow defined in this study were examined to confirm consistency with the TEM residual mean flow for the DJF and JJA climatology. Our formulation is based on the QG approximation. As shown in section 3b, the zonal-mean 3D residual flow ((\([\overline{u}_g]_T\), [\([\overline{v}_g]_T\]) is mathematically equivalent to the time-mean QG TEM residual flow ((\([\overline{u}_g]\), \([\overline{w}_g]\)). Here we compare ((\([\overline{u}_g]_T\), [\([\overline{v}_g]_T\)]) with the time-mean TEM residual flow ((\([\overline{u}_g]\_T\), \([\overline{w}_g]\_T\)) for the primitive equations in the meridional cross section. This comparison provides a validation of the QG approximation. Note that the 3D residual flow near the equatorial region cannot be estimated because of the QG approximation; accordingly, only results at latitudes higher than 15° are shown here.

The latitude and log-p height sections of the climatology of the primitive TEM residual meridional flow ((\([\overline{u}_g]\_T\)) and the zonal-mean 3D residual meridional flow ((\([\overline{u}_g]_T\))) for DJF and JJA are shown in Fig. 4. In both seasons, the distribution and strength of [\([\overline{v}_g]_T\)] generally agree with those of [\([\overline{v}_g]\_T\), even though there are slight differences. The northward flow (i.e., the flow toward the winter pole) is greatest at \(z = \sim 35\) km (7 hPa) at 60°N and above \(z = \sim 40\) km (3 hPa) at 20°N in DJF. Similarly, two maxima of the southward flow (i.e., flow
toward the winter pole) are also observed in JJA, even though they are weaker than those in DJF. The locations of the maxima in JJA are \( z = \sim 35-37 \) km (5–7 hPa) at 30°S and above \( z = \sim 40 \) km (3 hPa) at 5°N. These latitudes for JJA are lower than those for DJF. These differences between JJA and DJF are consistent with the difference in the magnitude of the planetary flow activity and the strength and latitudinal location of the polar night jet in the stratosphere, which modulates the propagation of planetary waves (e.g., Shiotani and Hirota 1985; Randel 1988).

There are a few notable differences between \([\bar{w}_0^{+\infty}]\) and \([\bar{w}_3^{+\infty}]\). One is that the magnitude of the poleward flow is slightly larger for \([\bar{w}_0^{+\infty}]\) than for \([\bar{w}_3^{+\infty}]\) for both maxima at \( \sim 7 \) and above 3 hPa. Another difference is the existence, only for \([\bar{w}_0^{+\infty}]\), of a weak minimum near 45°N in DJF and near 45°S in JJA for the flow above 3 hPa. These weak minima are due to the limitation of the geostrophic approximation. For example, the polar night jet is so strong that it is not in the state of geostrophic wind balance but rather is in the state of gradient wind balance.

**Figure 5** shows the meridional cross sections of the climatology of the time-mean primitive TEM residual vertical flow \([\bar{w}_3^{+\infty}]\) and the zonal-mean 3D residual vertical flow \([\bar{w}_3^{+\infty}]\). The observed features of \([\bar{w}_3^{+\infty}]\) and \([\bar{w}_3^{+\infty}]\) are generally consistent, even though the magnitude of the downward flow in the winter hemisphere is slightly stronger for \([\bar{w}_3^{+\infty}]\) at latitudes higher than 60°N in DJF and at latitudes higher than 60°S and near 30°S in JJA. In addition, there is a significant difference between \([\bar{w}_3^{+\infty}]\) and \([\bar{w}_3^{+\infty}]\) around 60°S in JJA where a clear vertically elongated minimum is seen in \([\bar{w}_3^{+\infty}]\), but not in \([\bar{w}_3^{+\infty}]\). This is mainly due to the limitation of the QG approximation: the Stokes correction component \( \left[ \frac{\bar{v}_0 \delta}{\theta_0} \right] \) in \([\bar{w}_3^{+\infty}]\) in (29) based on the QG equations is calculated using geostrophic meridional winds \( v_0 \), while that in \([\bar{w}_3^{+\infty}]\) based on the primitive equations is calculated using meridional winds \( u_0 \). Thus, the difference between \([\bar{w}_3^{+\infty}]\) and \([\bar{w}_3^{+\infty}]\) is particularly large around 60°S in JJA where a strong polar night jet is located approximately in the state of the gradient wind balance rather than the geostrophic balance.

In summary, it was confirmed that the zonal mean of the 3D residual flow derived in this study (\([\bar{w}_3^{+\infty}]\), \([\bar{w}_3^{+\infty}]\)) is basically consistent with the time-mean TEM residual mean flow (\([\bar{w}_3^{+\infty}]\), \([\bar{w}_3^{+\infty}]\)) in the primitive equations, even though there are slight differences, particularly in the strength of the circulation owing to the limitation of the QG approximation. Note that these differences are also observed between the QG TEM residual flows and the primitive TEM residual flows (not shown).

**b. Horizontal distribution of the 3D residual flow in the Northern Hemisphere in DJF**

**Figure 6a** shows a polar-stereo map of the climatology of the 3D residual flow in the Northern Hemisphere in DJF at 7 hPa (\( z = \sim 35 \) km), where \([\bar{w}_3^{+\infty}]\) is maximized (Fig. 4). The arrows represent the horizontal components (\([\bar{w}_3^{+\infty}]\), \([\bar{w}_3^{+\infty}]\)), and the colors represent the vertical component \([\bar{w}_3^{+\infty}]\) of the 3D residual flow. The climatology of the geopotential contours is shown by the solid curves. As seen in the geopotential distribution, there is a strong polar vortex surrounded by a polar night jet situated over the Arctic, the center of which is slightly offset toward the European sector indicating the climatological existence of a stationary wave with the zonal wavenumber \( s = 1 \). The residual horizontal flow (\([\bar{w}_3^{+\infty}]\), \([\bar{w}_3^{+\infty}]\)) is strong along the polar night jet. The direction of the residual horizontal flow is opposite to that of the polar night jet for 90°E–0°–90°W. This is because the polar night jet
is well described by the gradient wind balance (not by the geostrophic wind balance), as discussed in detail in section 4c.

It is clear that the BDC residual vertical flow $\bar{w}_a^{***}$ is not zonally uniform. Even in the mid- and high-latitude regions where the time-mean TEM residual vertical flow $\bar{w}_a^P$ is downward, an upward flow region is observed along 50°N from 0° to 90°E over Europe and western Eurasia. The downward flow is maximized in the two regions of 120°–140°E to the north of 60°N over east Siberia and near 50°W, 70°N over Greenland. The former is stronger than the latter. Because the magnitude of $\tilde{\theta}_P/\theta_0$ is quite small compared with that of $\bar{w}_a^{***}$ in Fig. 6a during this season (not shown), the distribution of $\bar{w}_a^{***}$ roughly agrees with that of the diabatic heating rate [see Eq. (13)].

A zonal wavenumber-4 structure is observed at low latitudes. This structure may partly be real because some of these regions experience strong convection. However, this structure may also be artificial, arising from the use of the four-times-daily data for the analysis, as discussed in the appendix, and should be ignored. Thus, a further description of the zonal wavenumber-4 structure is not given.

c. 3D residual horizontal flow orthogonal to the geopotential contours

In section 4b, it was shown that the 3D residual horizontal flows are approximately parallel to the geopotential contours and have opposite directions to the geostrophic winds in some regions. This is explained by the fact that the polar night jet is strong and under the condition of the gradient wind balance

---

Fig. 4. Latitude and log-pressure height sections of the climatology of (a) the time-mean primitive transformed Eulerian-mean (TEM) residual meridional flow ($\bar{v}_a^P$) and (b) the zonal-mean 3D residual meridional flow ($\bar{v}_a^{***}$) in DJF and (c) $\bar{v}_a^P$ and (d) $\bar{v}_a^{***}$ in JJA.
rather than the geostrophic wind balance (e.g., Randel 1987). The gradient wind is weaker than the geostrophic wind around a cyclonic vortex such as the polar vortex (e.g., Holton and Hakim 2013). Thus, the 3D residual flow includes the difference between the geostrophic and gradient winds, which points in the opposite direction of the geostrophic wind surrounding the cyclonic vortex.

Incidentally, the gradient wind is parallel to geopotential contours like the geostrophic wind, as shown below. The gradient wind \( (u_r, v_l) \) in spherical coordinates is written as follows:

\[
\begin{align*}
  u_r &= -\frac{\Phi_\lambda}{a \cos \phi} \left( f + \frac{u_l \tan \phi}{a} \right)^{-1}, \\
  v_l &= \frac{\Phi_\lambda}{a f \cos \phi},
\end{align*}
\]

where \( \lambda \) and \( \phi \) represent longitude and latitude, respectively; \( a \) denotes the radius of Earth; the subscripts \( \lambda \) and \( \phi \) denote the partial derivatives with respect to each; and \( f(\phi) \) denotes the Coriolis parameter at a latitude \( \phi \). The geostrophic winds in spherical coordinates are written as

\[
\begin{align*}
  u_g &= -\frac{\Phi_\lambda}{a f}, \\
  v_g &= \frac{\Phi_\lambda}{a f \cos \phi}.
\end{align*}
\]

Fig. 5. As in Fig. 4, but for (a) the time-mean primitive TEM residual vertical flow \( \langle [w_a]_p \rangle \) and (b) the zonal-mean 3D residual vertical flow \( \langle [w_a^{***}] \rangle \) in DJF and (c) \( [w_a]_p \) and (d) \( [w_a^{***}] \) in JJA.
Calculating the cross product of the gradient winds \((u_g, v_g)\) and the geostrophic winds \((u_r, v_r)\) yields zero. This means that the geostrophic wind and the gradient wind are always parallel; therefore, the gradient wind is parallel to the geopotential contours.

The TEM residual mean flow does not include the balanced flow components. Thus, the 3D structure of the flow other than the balanced flow is of interest. Accordingly, we examined the component of the 3D residual structure that is orthogonal to the geopotential contours. Note that there can be residual flows parallel to the geopotential contours. However, we ignored such parallel residual flows in the present analysis, partly because the parallel residual flows must be negligibly weaker than the balanced flows, and more practically because the computational cost is high for the calculation of gradient winds required to remove the difference between the gradient wind and the geostrophic wind from the residual flow. The residual horizontal flows orthogonal to the geopotential contours are more important because they correspond to the flows going into and out of the polar vortex, whose center is often significantly offset from the North Pole as seen in Fig. 6a. The polar night jet at the edge of the polar vortex behaves as a barrier against the horizontal transport of mass and/or minor constituents (e.g., Nakamura and Ma 1997), and hence analyses of the flow across the polar night jet are crucial to ozone science (e.g., Russell et al. 1993; Fisher et al. 1993; Callaghan and Salby 2002).

The horizontal residual flow \(u_{r}^* \parallel \nabla \Phi \equiv \frac{\nabla \Phi}{|\nabla \Phi|} \cdot \nabla \Phi\) is divided into two components, \(u_{r, \perp}^* \equiv \frac{\nabla \Phi \times \nabla \Phi}{|\nabla \Phi|} \cdot \nabla \Phi\) and \(u_{r, \parallel}^* \equiv \frac{\nabla \Phi \times \nabla \Phi}{|\nabla \Phi|} \cdot \nabla \Phi\), which are orthogonal and parallel to the time-mean geopotential \(\Phi\) contours, respectively:

\[
u_{r, \perp}^* \equiv \frac{\nabla \Phi \times \nabla \Phi}{|\nabla \Phi|} \cdot \nabla \Phi, \tag{54}
\]
\[
u_{r, \parallel}^* \equiv \frac{\nabla \Phi \times \nabla \Phi}{|\nabla \Phi|} \cdot \nabla \Phi, \tag{55}
\]

where \(V_{H}\) denotes the horizontal gradient. As previously mentioned, this study mainly examines \(u_{r, \perp}^* \).

d. DJF climatology in the Northern Hemisphere

Figures 6b and 6c show polar-stereo maps of the DJF climatology of \((u_{r, \perp}^*, u_{r, \parallel}^*, \bar{w}_{a}^*)\) in the Northern Hemisphere at 7 and 3 hPa, respectively, where \(u_{r, \perp}^*\) is maximized in the vertical. Even though the two maxima are separated in the vertical direction, the horizontal distributions of \(u_{r, \perp}^*\) are well matched at these two levels, except for a slight westward shift at 3 hPa. It is clear that \((u_{r, \perp}^*, u_{r, \parallel}^*, \bar{w}_{a}^*)\) is not zonally uniform. At 7 hPa, a strong inward flow into the polar vortex is primarily observed around 120°–170°E over east Siberia, while the flow around 60°–130°W over North America is strongly outward from the polar vortex. Interestingly, this strong outward flow is located on the western side of the polar vortex, which is displaced from the North Pole, and roughly corresponds to the region of a tongue extending into low latitudes drawn from the polar vortex associated with the breaking of the \(s = 1\) stationary wave as shown by previous studies (i.e., Manney et al. 1994b; Polvani and Saravanan 2000). A relatively weak outward flow is also observed between 20°W and 30°E over Europe.
FIG. 7. As in Fig. 6b, but for each component contributing to \((u_{\perp}^{ ***}, v_{\perp}^{ ***}, w_{\perp}^{ ***})\): (a) as in Fig. 6b, i.e., \((u_{\perp}^{ ***}, v_{\perp}^{ ***}, w_{\perp}^{ ***})\), (b) ageostrophic wind component \((u_{\perp}, v_{\perp}, w_{\perp})\), (c) Stokes correction \((u_{S}^{ \perp}, v_{S}^{ \perp}, w_{S}^{ \perp})\), (c) that due to stationary waves \((u_{st}^{ \perp}, v_{st}^{ \perp}, w_{st}^{ \perp})\), (f) that due to transient waves \((u_{tr}^{ \perp}, v_{tr}^{ \perp}, w_{tr}^{ \perp})\), and (d) the sum of (b) and (c).
It is important that the primary downward flow maximum is observed in the inward flow region over east Siberia. An inward flow is also observed slightly to the south of the secondary downward flow maximum near 70°N, 50°W over the North Atlantic. This feature is consistent with the strong radiative cooling that the inward flows must undergo inside the cold polar vortex (Callaghan and Salby 2002). In addition, and more importantly, this stronger downward flow over east Siberia is related to the phase structure of the s = 1 stationary wave, namely, a slight westward tilt with height (see Figs. 6b and 6c, and more clearly shown in Fig. 8 later), indicating the temperature field has also the s = 1 structure. A positive temperature anomaly is located over east Siberia, suggesting more effective radiative cooling there.

It is worth noting here that the $\mathbf{u}_a^{\perp}$ vectors are not necessarily orthogonal to the climatological geopotential ($\Phi$) contours in Figs. 6b and 6c. This is because ($\mathbf{u}_a$, $\mathbf{w}_a$) was calculated for $\Phi$ averaged over 3 months of each year and then ($\mathbf{u}_a^{\perp}$, $\mathbf{w}_a^{\perp}$) and $\Phi$ for each year were averaged over 37 years to obtain their climatology.

The 3D residual mean flow in (22), (23), and (12) is divided into the ageostrophic wind ($\mathbf{u}_a$, $\mathbf{v}_a$, $\mathbf{w}_a$) and the Stokes correction $\mathbf{u}_S$, $\mathbf{v}_S$, $\mathbf{w}_S$. The structure of each term constituting the 3D residual mean flow orthogonal to the time-mean geopotential contour is examined (Fig. 7). Note that maps in

Fig. 8. (a),(b) Latitude–height sections of ($\mathbf{u}_a^{\perp}$, $\mathbf{w}_a^{\perp}$) (vectors) and $\mathbf{u}_a^{\perp}$ (colors) for 90°–140°E and 80°–120°W, respectively. (c) Longitude–height section of ($\mathbf{u}_a^{\perp}$, $\mathbf{w}_a^{\perp}$) (vectors) and $\mathbf{u}_a^{\perp}$ (colors) for 50°–70°N in DJF. Black contours show the potential temperature at an interval of 100 K.
Fig. 7 are arranged for easy comparison and not in the order of description.

First, it is clear that $\overline{\mathbf{w}}_n$ and $\overline{\mathbf{w}}_S$ are largely canceled (Figs. 7b,c). This is consistent with the interpretation of $\overline{\mathbf{w}}_S$ described in section 2c. In contrast, such a structure with large cancellation between $(\overline{\mathbf{w}}_{n,1}, \overline{\mathbf{w}}_{n,2})$ and $(\overline{\mathbf{w}}_{S,1}, \overline{\mathbf{w}}_{S,2})$ is not clearly observed. This means that the component of $(\overline{\mathbf{w}}_{S,1}, \overline{\mathbf{w}}_{S,2})$ parallel to the absolute momentum contours is small compared with that orthogonal to it. Second, $(\overline{\mathbf{w}}_{S,1}^*, \overline{\mathbf{w}}_{n,2}^*)$ is largely contributed to by stationary waves \cite{i.e., $(\overline{\mathbf{w}}_{S,1puted to by stationary waves \cite{i.e., $(\overline{\mathbf{w}}_{S,1}, \overline{\mathbf{w}}_{S,2})$]. Thus, the sum of $(\overline{\mathbf{w}}_{S,1}, \overline{\mathbf{w}}_{n,2})$ and $(\overline{\mathbf{w}}_{S,1}^*, \overline{\mathbf{w}}_{n,2}^*)$ (Fig. 7d) roughly agrees with $(\overline{\mathbf{w}}_{S,1}^*, \overline{\mathbf{w}}_{n,2}^*)$ (Fig. 7a). However, $(\overline{\mathbf{w}}_{S,1}, \overline{\mathbf{w}}_{n,2})$ has also significant contribution: There are strong poleward flows in 20°–60°W and in 30°–110°E around 60°N.

Figures 8a and 8b show $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ in the latitude–height sections averaged over 90°–140°E where $\overline{\mathbf{w}}_{1,1}$ is strongly poleward and over 80°–120°W where $\overline{\mathbf{w}}_{1,2}$ is strongly equatorward, respectively. The downward motion is maximized near 70°N in Fig. 8a and near 60°N in Fig. 8b. These latitudes correspond to the polar vortex edge at each longitude region, which is consistent with the finding by Manney et al. (1994b). Another interesting feature in Fig. 8b is that an equatorward flow at high latitudes and a poleward flow at low latitudes converge, and a weak upward flow is observed around 40°N. Figure 8c shows $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ in the latitude–height section averaged over 50°–70°N roughly corresponding to the polar vortex edge. The potential temperature contour is undulating due to stationary planetary waves. The downward flow is strong in the longitude range where the potential temperature contour is lowered (i.e., in the warm range) at each height in the middle and upper stratosphere, which is consistent with the radiative relaxation. The downward flow maximum is tilted westward with height, corresponding to the phase structure of planetary waves that propagate energy upward.

e. JJA climatology in the Southern Hemisphere

Polar projection maps of the climatology of $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ and each component in the Southern Hemisphere at 7 hPa in JJA are shown in Fig. 9. Both residual horizontal and vertical flows shown in Fig. 9a are generally weak compared with those in the Northern Hemisphere in DJF (Fig. 7a). The longitude regions where the poleward flows are strong depend on the latitude. For a relatively low-latitude region of 20°–45°S, strong poleward flows are observed in the longitudinal sector clockwise from 10°W to 40°E to the southwest of Africa and at ~70°W over the Andean mountain range. For a relatively high-latitude region of 45°–70°S, a strong poleward flow, i.e., an inward flow into the polar vortex, is primarily observed in 80°–160°E to the southwest of Australia. On the other hand, an equatorward flow is evident at 10°–45°W near 50°S over the west Atlantic and at 140°–170°W near 40°S over the central Pacific. The downward flow (negative $\overline{\mathbf{w}}_{n}^*$) is maximized at 90°–135°E in the high-latitude region where the poleward flow is strong to the southwest of Australia. This feature is again consistent with the strong radiative cooling that the air flowing into the cold polar vortex experiences. A strong upward flow is observed at 40°–60°S, 0°–60°W, where the horizontal flow is equatorward over the west Atlantic. A similar horizontal distribution of $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ was observed at the levels above 3 hPa (not shown).

It is also worth noting that a strong poleward flow is observed in a narrow area over the Andean mountain range. This flow is likely caused by the orographic gravity waves radiated from the mountain by the prevailing eastward background winds near the surface in this season. This region is known as a hotspot for gravity waves (e.g., Eckermann and Preusse 1999; Wu et al. 2006; Sato et al. 2012; Hoffmann et al. 2013; Hindley et al. 2020; Gupta et al. 2021). The orographic gravity waves have a vertical flux of westward momentum. Thus, the dissipation of the orographic waves causes westward forcing, which is consistent with the observed residual poleward flow. A strong equatorward flow over the Antarctic Peninsula is also interesting, but this flow is hardly explained by the orographic gravity wave forcing. Secondary gravity wave generation in the stratosphere (e.g., Vadas et al. 2018) may be a candidate to give the eastward forcing. Further investigation is needed in this regard in the future.

The geopotential contours in the Southern Hemisphere in JJA are more axisymmetric around the pole than those in the Northern Hemisphere in DJF. However, the presence of stationary waves can be seen in the map of vertical ageostrophic flow $\overline{\mathbf{w}}_S$ in Fig. 9b, which is almost canceled out by the Stokes correction due to stationary waves $\overline{\mathbf{w}}_{n}^*$ (Fig. 9c). The vertical Stokes correction by transient waves $\overline{\mathbf{w}}_{n}^*$ has an axisymmetric structure around the pole (Fig. 9d), which contributes significantly to the residual downward flow $\overline{\mathbf{w}}_{n}^*$ around 60°–70°S (see Figs. 9a,d).

f. Comparison with the 3D residual horizontal flow distribution estimates by Kinoshita et al. (2019)

Here the 3D residual horizontal flow distribution obtained by the present study is compared with that estimated using the formulas of $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ [Eqs. (23) and (12)] by Kinoshita et al. (2019) that were derived by treating the time-mean zonal-mean field as the basic field (see section 2e). As noted, the formulas for $\overline{\mathbf{w}}_{1,1}^*$ and $\overline{\mathbf{w}}_{n}^*$ are equivalent to ours (\overline{\mathbf{w}}_{1,2}^* and $\overline{\mathbf{w}}_{n}^*$) [Eqs. (23) and (12)]. The only difference is in the residual zonal flow, i.e., $\overline{\mathbf{w}}_{1,2}^*$ and $\overline{\mathbf{w}}_{n}^*$ [Eq. (22)].

The DJF result of $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ for the Northern Hemisphere at 7 hPa is shown in Fig. 10 as an example, which should be compared with $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*, \overline{\mathbf{w}}_{n}^*)$ in Fig. 6b. The overall structure of the horizontal distribution of $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*)$ is similar to that of $(\overline{\mathbf{w}}_{1,1}^*, \overline{\mathbf{w}}_{1,2}^*)$. However, there are large differences particularly in the magnitude of the flow at high latitudes. The difference is attributed to the structure of the geopotential contours, which are far from concentric circles around the North Pole due to the presence of the stationary Rossby waves. It may not make much sense to use the time-mean zonal-mean field as the basic field in such a case.
5. Conclusions

A formulation of the 3D residual flow and its governing equations has been made on the basis of the QG equations with log-$p$ coordinates. The zonal mean of the 3D residual flow is the same as the time mean of the QG TEM residual flow, ensuring that the 3D residual flow derived here includes the quasi-Stokes correction due to both stationary and transient waves. By examining the differences in the 3D residual flow from that derived for transient waves by Plumb (1986).

Fig. 9. As in Fig. 7, but for JJA in the Southern Hemisphere.
an interpretation of the 3D residual flow was given. The quasi-Stokes correction includes the conservative flows of its opposite sign along isentropic and iso-angular momentum surfaces, which are undulated by stationary waves. Thus, the 3D residual flow, which is the sum of the ageostrophic flow and the quasi-Stokes correction, does not include such conservative flows.

The horizontal momentum and thermodynamic equations derived as governing equations have simple forms. In the formulation, no basic flow is assumed. Therefore, the zonal and meridional momentum equations are expressed symmetrically. The zonal and meridional momentum equations have simple forms, as do the TEM residual-mean momentum equations, which consist of the tendency term, the Coriolis force term for the residual flow, the wave-forcing term, and other nonconservative terms such as gravity wave forcing. The wave forcing is expressed in the form of meridional and zonal fluxes of the potential vorticity other than the planetary vorticity in the zonal and meridional momentum equations, respectively. The zonal mean of the wave-forcing term equals the time mean of the meridional flux of the potential vorticity in the TEM system, which coincides with the EP flux divergence, known as Taylor’s relation. It is important that the momentum equation includes nonconservative terms such as gravity wave forcing, which is not directly expressed in the QG system. Thus, the governing equations derived in the present study can be used for momentum budget analyses even in regions such as the mesosphere where gravity wave forcing is crucial. The thermodynamic equation also has a simple form that is composed of the tendency term, the vertical advection of the basic potential temperature by the residual flow, and the diabatic heating rate.

This theory is a generalization of that made by Kinoshita et al. (2019), in which the time-mean zonal-mean zonal wind is treated as the basic state. In contrast, the presented theory does not assume a particular basic state; therefore, it can even treat cases where the polar vortex is shifted from the pole. The zonal and meridional momentum equations formulated by Kinoshita et al. (2019) are not symmetric, and there is the problem that a term proportional to $\beta v$ appears explicitly in the meridional momentum equation. Their meridional momentum equation and residual zonal flow are different from those in the present study, while their residual meridional and vertical flows, zonal momentum equation, and thermodynamic equation are the same.

As an application of the derived formulas, we have shown the climatology of the 3D structure of the BDC deep branch using reanalysis data. In the QG approximation, the geostrophic flow is treated as the dominant balanced flow. Thus, the 3D residual flow in the present study is expressed as the sum of the ageostrophic flow as the unbalanced flow and the Stokes correction. However, strong flows, such as the winter polar night jet, are often not in the state of geostrophic balance but rather are in the state of gradient wind balance. In this case, the unbalanced wind in the 3D residual flow should be the deviation from the gradient wind. To avoid this problem, we examined the residual horizontal flow component orthogonal to the geopotential contours, as well as the residual vertical flows, using the property that both the geostrophic and gradient winds are parallel to the geopotential contours. The poleward flows of the deep branch in the winter hemisphere are maximized at the two levels of $\sim 7$ hPa ($\sim 35$ km) and above 3 hPa (40 km) for both DJF and JJA. The results show that the 3D residual flow has a similar horizontal distribution at both levels in DJF and JJA (the figure is not shown for 3 hPa in JJA). In DJF, at high latitudes, the residual horizontal flow is strongly poleward in east Siberia and strongly equatorward in North America. At midlatitudes, the flow is northward in the North Atlantic and southward in Europe. There is a strong downward flow in east Siberia, which is
likely due to radiative cooling experienced by air entering the cold polar vortex. Upward flow is observed in the region from Europe to western Eurasia. In JJA, the poleward flow is strong at high latitudes southwest of Australia and at midlatitudes southwest of Africa. The downward flow is relatively strong in these regions. An upward flow is observed over the South Atlantic near 50°S, where weak equatorward flow is seen. In addition, there is a strong poleward flow in a narrow longitudinal region near the Andean mountain range, as this region is known to be a hotspot for gravity waves. Therefore, gravity wave forcing may be responsible for this flow.

In this study, the formulas of the 3D residual flow and its governing equations were derived based on the QG equations. Thus, gravity waves are not directly treated in this theory. However, the contribution to the 3D residual flow by the gravity wave forcing can be treated as the terms of $X$ and $Y$ in Eqs. (26) and (27). Therefore, it is possible to study the residual flows in the mesosphere, where gravity waves are more crucial than Rossby waves (e.g., Becker 2012; Smith 2012). In this paper, we have examined the 3D characteristics of the deep branch of the BDC. The shallow branch will also be examined in the near future, and the results will be presented elsewhere. It will be also interesting to analyze in future studies how the 3D distribution of the 3D residual flows corresponds to the sudden stratospheric warming associated with the breaking of planetary waves. In this study, we used four-times-daily data and hence we could not analyze the structure at low latitudes and in the summer hemisphere where the amplitude of the solar tides is relatively large as discussed in appendix. The use of recently available reanalysis data will help to overcome this problem.

Acknowledgments. The authors greatly appreciate three anonymous reviewers for their constructive comments. The present study uses the ERA-Interim data at pressure levels, which are available from the European Centre for Medium-Range Weather Forecasts (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim). The GFD Dennou library was used for drawing the figures. This research was supported by the Japan Science and Technology Agency Core Research for Evolutional Science and Technology (Grant JPMJCR1663).

Data availability statement. Analysis was made using the ECMWF ERA-Interim reanalysis dataset available from https://apps.ecmwf.int/datasets/.

APPENDIX

Spurious Zonal Wavenumber-4 Structure in the Climatology from the Four-Times-Daily Data

The horizontal map of the vertical wind ($w$) climatology from the four-times-daily data exhibits a spurious zonal wavenumber ($s$)-4 structure. Figure A1 shows examples of such maps at 2 and 10 hPa. The $s = 4$ structure is clearly seen in the latitudinal region from 30°N to 60°S at both levels, although it is weak at 10 hPa. This structure does not necessarily show the real existence of an $s = 4$ disturbance but instead is largely attributable to migrating tides sampled at a time interval of 6 h (four times a day), which is the case for the ERA-Interim data. In fact, using the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017), at a time interval of 3 h (eight times a day), an $s = 8$ structure appears when the climatological vertical flow is calculated (not shown). Therefore, we need to be careful when interpreting the horizontal distribution of the vertical flow in the summer hemisphere (i.e., the BDC deep branch) in the middle and upper stratosphere. Similar spurious wavenumber-4 structures may exist in other physical quantities, such as $u$, $v$, and $T$, but are not as critical because disturbances other than migrating tides have large amplitudes for these quantities.

REFERENCES


