Eastward-Propagating Planetary Waves in the Middle Atmosphere During Major Sudden Stratospheric Warming Events

Christian Todd Rhodes
Coastal Carolina University

Follow this and additional works at: https://digitalcommons.coastal.edu/etd

Part of the Atmospheric Sciences Commons, and the Climate Commons

Recommended Citation
Rhodes, Christian Todd, "Eastward-Propagating Planetary Waves in the Middle Atmosphere During Major Sudden Stratospheric Warming Events" (2021). Electronic Theses and Dissertations. 133.
https://digitalcommons.coastal.edu/etd/133

This Dissertation is brought to you for free and open access by the College of Graduate and Continuing Studies at CCU Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of CCU Digital Commons. For more information, please contact commons@coastal.edu.
Eastward-Propagating Planetary Waves in the Middle Atmosphere During Major Sudden Stratospheric Warming Events

Christian Todd Rhodes

A Dissertation in Fulfillment of the Requirements for a Doctor of Philosophy Degree in Marine Science: Coastal and Marine Systems Science

Department of Coastal and Marine Systems Science
School of the Coastal Environment
College of Science
Coastal Carolina University

May 2021

Advisor
Varavut Limpasuvan, Ph.D.
Coastal Carolina University, Conway, USA

Committee Members
Shaowu Bao, Ph.D.
Coastal Carolina University, Conway, USA

Yvan Orsolini, Ph.D.
Norwegian Institute of Air Research, Kjeller, Norway

Michael Murphy, Ph.D.
Coastal Carolina University, Conway, USA

Jadwiga Richter, Ph.D.
National Center for Atmospheric Research, Boulder, CO
Abstract

The wintertime polar vortex is characterized by a strong, circumpolar wind identifiable mainly in the stratosphere. The wind structure associated with the polar vortex is shaped by the dissipation of gravity waves (GWs) and planetary waves (PWs). While GW drag maintains a region of weak winds near the stratopause, PW forcing perturbs the polar vortex. Unusually strong PW forcing can result in a vortex breakdown, leading to a phenomenon known as a sudden stratospheric warming (SSW), during which the polar region rapidly warms and the prevailing stratospheric eastward wind reverses direction. SSW-induced perturbations can subsequently alter surface weather patterns, leading to anomalous cold weather outbreaks.

To help improve our climate predictive skills and investigate the dynamical coupling between the stratosphere and mesosphere, planetary-scale features surrounding SSW evolution were investigated in this dissertation using the National Center for Atmospheric Research (NCAR) Whole Atmosphere Community Climate Model. The model was constrained with NASA’s Modern-Era Research Reanalysis to mimic reality below the altitude of 50 km. The simulations above 60 km were verified by satellite observations from NASA’s Microwave Limb Sounder.

In particular, a slow eastward-propagating planetary wave (EPW) identified in the mesosphere prior to the 2009 SSW event motivated a detailed study on the mechanisms responsible for the wave’s presence. The EPW could either propagate into the mesosphere directly from the troposphere or be generated in situ by asymmetric GW drag or shear instability. Prior to the occurrence of this event, a double-maxima configuration in the zonal-mean zonal wind developed and resulted in the northward shift of the polar night jet. The northward movement exacerbated the mesospheric wind shear which generated PWs locally from instability. The background wave geometry favored the growth of PWs with eastward phase speeds. Reasons for this unique characteristic were
understood through the over-reflection perspective which provides a framework to interpret how PWs should interact with atmospheric boundaries like the stratopause.

Further analysis showed that these EPWs were a common feature prior to SSW events and their growth imposes a significant eastward acceleration on the background winds. However, robust results were difficult to ascertain due to the low number of recorded SSWs.

To increase the sample size of SSWs, an ensemble numerical experiment was conducted. The experiment results indicate that EPW growth was common prior to SSWs but could also occur during winters without SSWs. Therefore, while EPWs would be expected prior to SSWs, they would not solely be good predictors of SSW occurrence. The ensemble study was expanded to investigate the developing wave geometries throughout SSW evolution. Just before SSW onset, an enhanced period of over-reflection resulted in the exposure of the mesospheric zero-wind line to quasi-stationary PW forcing. The resulting wave drag caused the zero-wind line to rapidly descend into the stratosphere. During SSW when the stratospheric wind reversed direction over a deep layer, the EPW flux divergence occurred at the upper and lower boundaries of this reversed wind layer. Such features appear consistent with a mechanism in which GW drag and over-reflection operate in tandem to generate and trap EPWs within the reversed wind layer. To our knowledge, this dissertation is the first study to address this feature. Additionally, the abnormal presence of westward-propagating PWs in the mesosphere during SSW was discussed from the over-reflection perspective.

Overall, this dissertation investigated the source mechanisms and characteristics of mesospheric EPWs prior to SSWs through the over-reflection perspective. The over-reflection perspective provided a simplified framework to understand the behavior of PWs within various wave geometries. The success of applying the over-reflection perspective to EPWs prompted further research in the ensemble study to understand the evolution of wave geometries surrounding SSW events.
Acknowledgements

Thank you to my Thesis advisor and mentor, Dr. Varavut Limpasuvan, who cleared the path for me to pursue this research. His dedication and support have greatly improved the quality of this research and advanced my skills as a scholar. Thank you to Dr. Yvan Orsolini who offered guidance to the research presented in Chapters 8 and 9. Thank you to Dr. Jadwiga Richter who significantly advanced my skillset at the beginning stages of my research. Thank you to Dr. Shaowu Bao who improved my understanding of atmospheric dynamics, and Dr. Michael Murphy who advised me on the usage of a high performance cluster. This research would not have been possible without the computing and technical support from the NCAR Computation Information Systems Laboratory as well as the Coastal Carolina University cyberinfrastructure (CCU CI) project, funded in part by NSF award MRI 1624068. This research was also funded by NSF through RUI-AGS 1642232 and REU-AGS 1560210. Finally, I would like to thank the support of Coastal Carolina University and the Department of Coastal and Marine Systems Science.
# Table of Contents

List of Figures................................................................................................................. viii
List of Tables.................................................................................................................... xviii

1. Introduction.................................................................................................................... 1

2. Basic Wind Structure and the Polar Vortex................................................................. 4

3. Planetary Waves........................................................................................................... 10
   3.1 Basic Prognostics for Rossby Waves................................................................. 12
   3.2 Propagation of Rossby Wave in a Simple Flow.................................................. 14
   3.3 Rossby Waves in a Realistic Flow........................................................................ 17
   3.4 Wave-Mean Flow Prognostics.............................................................................. 19
      3.4.1 Refractive Index......................................................................................... 19
      3.4.2 Eliassen-Palm Flux.................................................................................... 21
   3.5 Generation of PWs Through External Forcings............................................... 25
   3.6 Generation of PWs Through Instability............................................................. 27
      3.6.1 Diagnosis for Shear Instability................................................................. 27
      3.6.2 Over-reflection Perspective and Instability.............................................. 29

4. Gravity Waves............................................................................................................. 33
   4.1 Basic Equations Related to Gravity Waves....................................................... 35
   4.2 Pure Internal Gravity Wave Characteristics...................................................... 37
   4.3 Inertio-Gravity Wave Characteristics Incorporating a Constant
      Background Wind....................................................................................................... 39
   4.4 Basic Prognostics of Gravity Waves.................................................................. 40
   4.5 Gravity Wave Parameterization.......................................................................... 43

5. Major Sudden Stratospheric Warming........................................................................ 47
   5.1 Background Winds Associated with Split and Displaced SSWs......................... 47
   5.2 Planetary and Gravity Wave Forcings During ES-SSWs................................. 53

6. Motivation and Research Objectives.......................................................................... 58

7. Data and Methods....................................................................................................... 70
   7.1 Satellite Data....................................................................................................... 70
   7.2 The Model........................................................................................................... 72
   7.3 SSW Identification............................................................................................... 75
   7.4 Ensemble Setup.................................................................................................... 77
List of Figures

Figure 2.1. Polar plots at 10 hPa (left) and 1 hPa (right) on January 1, 1992. Line contours show geopotential height in increments of 0.15 km. Arrows represent the direction of geostrophic winds along the geopotential isolines. Colored-filled contours show temperature (in Kelvin). Data was retrieved from the Whole Atmosphere Community Climate model run in the specified dynamics configuration (WACCM-SD) described in Section 7.2.

Figure 2.2. Five-day running average of $\bar{u}$ (m s$^{-1}$) and zonal-mean temperature (K) on January 1, 1992. Thin solid contours indicate eastward and dashed contours indicate westward $\bar{u}$. The zero-wind line is indicated by a thick black solid contour. Color contours indicate the temperature. The 10 hPa and 1 hPa levels in Figure 2.1 are indicated by the grey horizontal lines. Thick black dashed contour suggests the rim of the polar vortex. Data was retrieved from the WACCM-SD model run described in Section 7.2.

Figure 3.1. Perturbed vorticity field and the induced velocity field (dashed arrows) for a meridionally displaced chain of fluid parcels. The heavy wavy line shows original perturbation position; the light line shows westward displacement of the pattern due to advection by the induced velocity. [From Holton & Hakim, 2013]

Figure 3.2. Schematic of the streamlines (solid) and isotherms (dashed) associated with a large-scale atmospheric disturbance in the mid-latitudes of the Northern Hemisphere. Arrows along the streamline contour indicate the direction of wind velocity. Their streamlines correspond approximately to lines of constant pressure, since the winds are nearly geostrophic. The signs of deviations from zonal-mean are shown at the bottom to illustrate the NE-SW tilt of the streamlines; this tilt indicates a northward zonal momentum transport, and the westward phase shift of the temperature wave relative to the pressure wave gives a northward heat transport. [adapted from Hartmann, 1994]

Figure 3.3. Rays of two stationary planetary wave components introduced into the lower stratosphere in zonal-mean winds (m s$^{-1}$) representative of northern winter on a mid-latitude beta plane. Arrowheads mark uniform increments of time moving along a ray at the group velocity $\vec{c}_g$. The ray for component 1, which initially propagates upward and poleward, encounters a turning line in strong westerlies of the polar night jet, where wave activity is refracted equatorward. As the ray approaches easterlies, it next encounters a critical line ($\bar{u} = 0$), where $\vec{c}_g$ vanishes. Propagation then stalls, leaving wave activity to be absorbed, so the ray terminates. The ray for component 2 is initially directed upward and equatorward, so wave activity encounters the critical line and is absorbed even sooner. Thus, wave activity introduced between tropical easterlies and strong polar westerlies can propagate vertically only a limited distance before being absorbed. [Salby, 1996]
Figure 3.4. Height-latitude plot of EP flux and monthly-averaged winds for January 1983. PW group velocities are indicated by EP flux vectors. EP flux convergence (divergence) are shown in blue (red) contours and increment by 5 m s\(^{-1}\) day\(^{-1}\). Meridional EP flux is scaled by \(5 \times (100\pi R_e \rho)^{-1} \cos \phi\) and vertical EP flux by \(5 \times (R_e \rho)^{-1} \cos \phi\). EP flux convergence/divergence is scaled by \(\rho R_e \cos \phi\). Eastward (westward) winds are shown by solid (dashed) black contours incremented by 20 m s\(^{-1}\) with the zero-wind line in bold. Model data is from WACCM-SD nudged with MLS described in Section 7.2.

Figure 3.5. Anomalous planetary wave field for northern winters during El Niño, when anomalous convection (stippled) is positioned in the tropical central Pacific. The anomalous ridge over western Canada and trough over the eastern United States characterize the so-called Pacific North America pattern that upsets the normal track of the jet stream (wavy trajectory) and cyclone activity during El Niño winters. [Adapted from Salby, 1996 and Horel & Wallace, 1981]

Figure 3.6. A schematic of planetary wave (PW) over-reflection with zonal phase speeds (\(c_x\)), zonal-mean zonal wind (\(\bar{u}\)), meridional potential vorticity gradient (\(\bar{q}_\phi\)), and squared refractive index (\(n^2\)) fields with arbitrary values.

Figure 4.1. Typical mid-latitude zonal winds during (a) winter and (b) summer. The black curve shows observed winds and the grey curve shows model “radiative” winds that result without GWD. Sources of gravity waves with various phase speeds, \(c\), are also depicted, with source and wave breaking symbols. On these plots, waves ascend vertically upwards since \(c\) remains constant, until they reach their critical level \(c = \bar{u}(z_C)\). [Kim et al., 2003; based on a presentation first used by Lindzen, 1981]

Figure 4.2. Zonal-mean orographic GW momentum flux (thin solid line), non-orographic GW momentum flux (thin dashed line), and the combination of both non-orographic and orographic momentum fluxes (bold solid line). Observational data was obtained from long-duration balloons in the stratosphere. [Hertzog et al., 2008]

Figure 4.3. Altered images of an experiment performed by the Geophysical Fluid Dynamics by the Graduate School of Human and Environmental Studies, Kyoto University. A mountain in a background wind is simulated by dragging a roller (effectively a “mountain”) through a tank of stratified fluid. For visualization purposes, the image is rotated such that the roller is now on the bottom. The roller is moving to the right such that a) and b) show sequential moments in time.

Figure 4.4. An illustration of GW propagation as it approaches the critical level (\(z_C\)). As a gravity wave approaches its critical level, (a) vertical wavenumber approaches zero, the \(w'\) approaches zero, \(u'\) approaches infinity, and (b) the vertical wavelength of the wave approaches zero. [Adapted from Nappo, 2002]
Figure 4.5. Illustration of GW propagation constrained by zonal wind profiles of the upper troposphere and lower stratosphere. The phase speed, c, is ground-relative. [Adapted from Plumb, 1977 by Laing & Evans, 2011]

Figure 4.6. (a) Orographic, (b) convective, (c) frontal, and (d) GWD averaged over DJF from 1985 to 2005 in WACCM3.5. Contour intervals are ±0.5, ±1, ±2, ±4, ±10, ±20, ±40, ±60, ±80, and ±100 m s$^{-1}$ day$^{-1}$. [Richter et al. 2010]

Figure 5.1. Polar plots of stratospheric temperature (filled contours) and geopotential height (thick black contours) at 31 hPa (~25 km). Geopotential height is incremented by 0.25 km up to 23.5 km. Three plots illustrate state of the polar vortex before, during, and after onset for (a-c) the January 1987 displaced SSW and (d-f) the January 2009 split SSW. Thin contours outline the continents. Data produced from the WACCM-SD model run described in Section 7.2.

Figure 5.2. $\bar{u}$ anomaly from normal climatology (a & c) 10-20 days and (b & d) 5-10 days prior to SSW onset. Analysis was performed on Japanese Meteorological Agency and Central Research Institute of Electrical Power Industry 25-year Reanalysis data from 1980-2011. Normal climatology is created from a composite of 14 winters without SSWs. Split and displaced SSWs are identified and classified by studies mentioned within Albers & Birner (2014). [Albers & Birner, 2014]

Figure 5.3. Height-time plot of $\bar{u}$ averaged from 70-90°N surrounding the 22 January 2009 split SSW. Grey-shaded regions indicate westward flow outlined by the zero-wind line (bold black line). Westward (blue) and eastward (red) zonal-mean flow are contoured every 10 m s$^{-1}$. Major and minor tick marks along the x-axis indicate 10-day and 1-day intervals, respectively.

Figure 5.4. Altitude-time evolution of composite ES-SSW for (a) PW forcing, and (b) GW forcing, averaged between 40°N and 80°N. The zero-wind (green) line shows the zonal-mean wind reversal from eastward to westward in the stratosphere. Red (blue) contours indicated eastward (westward) acceleration in m s$^{-1}$ day$^{-1}$. Day 0 represents SSW onset. [from Limpasuvan et al., 2016]

Figure 5.5. Vertical and horizontal components of anomalous EP flux integrated poleward of 50° N for all and specified wavenumbers during the composite life cycle of SSWs. Contribution from combined wavenumbers greater than or equal to 4 is denoted by “wave 4+.” The sum of “wave 1,” “wave 2,” “wave 3,” and “wave 4+” produces the result for “all.” Contours of the vertical component and horizontal component are given every $10 \times 10^4$ kg s$^{-2}$ and $20 \times 10^6$ kg s$^{-2}$, respectively. Negative contours are given as dashes. Zero contours are given as a bold solid line. Dark gray shading indicates areas with a 95% confidence level (based on t statistics). Composite calculated with daily-averaged data between 1958 and 2001 from the National Centers for Environmental Prediction –
National Center for Atmospheric Research reanalysis project (Kalnay et al., 1996). [Limpasuvan et al., 2004]

**Figure 5.6.** Percentage of orographic GWD contribution to the total wave forcing (EP Flux forcing + GWD) 5-30 days before the central warming date composited over major SSWs identified between 1980 and 2010. [Albers & Birner, 2014]

**Figure 6.1.** Hovmöller diagram of geopotential height anomalies for zonal wavenumbers 1-6 along 62.5°N at (top) 0.01 hPa and (bottom) 1 hPa. The contour interval is 200 m. The thick broken lines in the top panel denote an eastward traveling feature in the mesosphere before SSW onset. [Adapted from lida et al., 2014]

**Figure 6.2.** Time series from 0-100 km at 60°N of $F_z \exp(z/2H)$ where $F_z$ is the vertical component of the EP Flux. The blue lines indicate the zero zonal-mean zonal wind. Horizontal dashed line at 0.002 hPa marks top data ingestion level. Red arrows depict bursts of PW group velocity propagation. [Coy et al., 2011]

**Figure 6.3.** Geopotential height perturbations at 60°N contoured every 0.16 km during (a) 12, (b) 16, (c) 18, and (d) 20 January 2009. Black horizontal lines denote altitude level of the zero zonal-mean zonal wind. Corresponding vertical profiles of zonal-mean zonal wind (m s$^{-1}$) and stability (10$^{-4}$ s$^{-2}$) are shown to the right of each panel. [Coy et al., 2011]

**Figure 6.4.** Meridional cross sections zonal-mean zonal winds (line contours), Eliassen-Palm (EP) flux (vectors), and EP flux divergence (color-filled contours) of zonal wavenumbers 1-3 in the Northern Hemisphere. The arrow scale denotes $2 \times 10^6$ and $2 \times 10^8$ kg s$^{-2}$ for vertical and horizontal components of the EP flux, respectively, in the region of 100-10hPa. Vectors are magnified by a factor of 5n for the 1$^{st}$, 2$^{nd}$, ..., n$^{th}$ layers between $10^{2-n}$ and $10^{1-n}$ hPa. The shading denotes zonal wind deceleration; see the tone bare where the units are m s$^{-1}$ day$^{-1}$. The contour interval of zonal-mean zonal winds is 10 ms$^{-1}$. [Adapted from lida et al., 2014]

**Figure 6.5.** QGPV meridional gradient prior to the January 2009 split SSW. The cyan areas (with dashed contours) highlight the negative values regions, fulfilling the necessary conditions for shear instability. [from lida et. al, 2014]

**Figure 6.6.** (top) Ageostrophic wind in the latitude-longitude plane induced by breaking of mesospheric GWs whose vertical transmission has been filtered by the stratospheric circulation (bottom). Stratospheric wavenumber-1 PW geopotential height and gradient winds on latitude-longitude plane. [from Lieberman et al., 2013]

**Figure 6.7.** Eastward zonal wavenumber-2 GWD and EP flux on 17 January 2009. EP flux vectors indicate relative magnitude and direction of PW group velocities. The thick magenta and blue contours indicate the zero isopleth of refractive index and meridional PV gradient, respectively. [Adapted from Song et al., 2020]
Figure 7.1. Polar stereographic plots at 31 hPa of geopotential height in km (bold black contours) and temperature in Kelvin (color-filled contours) from the WACCM-SD model (top) and MLS (bottom) data. Geopotential height contours is incremented every 0.2 km. A 9-day running average was performed on the MLS data. Underlying landmasses are outlined by thin black lines. Dates for each plot are before (left) and after (right) SSW onset on 22 January 2009.

Figure 7.2. $T$ in K) averaged from 60-90°N during the winter of 2012-2013 observed in MLS (top) and simulated by WACCM-SD (bottom). Line contours are incremented every 5 K.

Figure 7.3. An example of SSW identification markers overlaid on a zonal-mean zonal wind plot averaged from 70° N to 90° N. Bold black line indicates the location of the stratopause. Thin black contour indicates the zero-wind line.

Figure 7.4. Probability of a normal member vs. ensemble initialization time in days prior to SSW onset. 20 ensemble members were run for each initialization date.

Figure 7.5. Schematic of normalized Gaussian distributions of diagnostic values during winters with SSWS (left curve) and normal winters (right curve). Values are collected with respect to SSW onset date at a specific height, longitude, and latitude. The green shaded region shows the probability of a diagnostic value being abnormal given that the winter contains an SSW or P(Ab|SSW).

Figure 8.1. (a) Altitude-time evolution of $\bar{u}$ at 60°N during the 2009 split SSW event with an onset date of 22 January as indicated by a vertical dashed line. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s$^{-1}$, with the zero-wind line thickened. (b-d) Altitude-latitude sections of $\bar{u}$ for (b) 25 December 2008, (c) 31 December 2008, and (d) the December-February climatology for 1980-2013 (excluding years with SSWs listed in the Chapter 8 introduction).

Figure 8.2. Altitude-latitude sections of (a, b) PW EP flux and its divergence and (c, d) total GWD, both in m s$^{-1}$ day$^{-1}$. The eastward $\bar{u}$ (m·s$^{-1}$) values averaged from 1.0 to 0.1 hPa indicates the polar jet (dotted contours). The eastward $\bar{u}$ values averaged from 0.1 to 0.01 hPa indicates the subtropical jet (solid contours). Vertical gold
and brown lines mark 25 and 31 December 2008, respectively. The latter date corresponds to the formation of a double-maxima wind structure. The dashed vertical line indicates SSW onset on 22 January 2009.

**Figure 8.4.** Altitude-latitude sections of (a, b) $\bar{w}^*$ and (c, d) $N^2_B$ as shading. (a, b) The residual circulation vectors ($\bar{u}^*, \bar{w}^*$), multiplied by (1, 1000), are overlaid. Only eastward $\bar{u}$ is shown and contoured every 10 m·s$^{-1}$ as thin black contours. The zero-wind line is thickened.

**Figure 8.5.** Altitude-latitude sections of (a-c) non-dimensionalized $\bar{q}_\phi$ and (d-f) the dominant terms contributing to $\bar{q}_\phi$. Dates are shown at the upper right of each row. (d-f) In reference to the right-hand side of Equation 1, blue-, grey-, and red-shaded regions indicate the dominance of the first, second, and third term, respectively. Dark (light) colors represent a negative (positive) contribution to $\bar{q}_\phi$. Eastward $\bar{u}$ is contoured every 10 m·s$^{-1}$ as thin black contours, with the zero-wind line thickened.

**Figure 8.6.** Zonal phase speed vs. time plot of wavenumber-2 geopotential height amplitude or $\bar{Z}_g$ (m) at 33, 51 and 67 km averaged between 45°N and 55°N. White contours increment by 5 m. Positive phase speed indicates eastward movement. The vertical brown line corresponds with the formation of a double-maxima wind structure on 31 December 2008. The vertical dashed line indicates SSW onset on 22 January 2009.

**Figure 8.7.** Stereographic polar projections extending from the pole to 10°N of carbon monoxide (CO) at 0.1 hPa. Geopotential height (black contours) is incremented by 0.5 km. The thick black vertical line depicts the Prime Meridian.

**Figure 8.8.** Altitude-latitude sections of PW2 EP flux and EP flux divergence for (a) all phase speeds and (b) eastward phase speeds. Negative (blue) and positive (red) EP Flux divergence is contoured every (a) 5 m·s$^{-1}$·day$^{-1}$ and (b) 2 m·s$^{-1}$·day$^{-1}$. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s$^{-1}$, with the zero-wind line thickened. The negative $\bar{q}_\phi$ region is grey-shaded. The presence of a critical layer inside the negative $\bar{q}_\phi$ regions is shaded in green. The EP Flux vector components are scaled as done in Figure 8.2, but the EPW2 vector magnitudes are multiplied by 10.0.

**Figure 8.9.** Non-dimensionalized squared refractive index, $n^2$ (color-filled contours) for (a-c) slow eastward and (d-f) stationary PW2s with specified frequencies labeled at the left of each row. Columns are organized by reference date. Eastward $\bar{u}$ is contoured every 10 m·s$^{-1}$ as thin black contours, with the zero-wind line thickened. Features A, B, and C depict notable changes to the $n^2$ field and are discussed in Sections 5.2.4 and 5.2.5.

**Figure 8.10.** Altitude-time section of PW2 vertical EP flux (filled-contours) for waves of (a) eastward phase speeds greater than 5 m·s$^{-1}$ and (b) all phase speeds. Regular-sized (thin) black contours of (+-) 0.2 × $2^m$ kg·s$^{-2}$, $m \in [1,2,3,4,5,6,7,8]$ outline upward (downward)
fluxes. Red contours show EP flux divergence and are incremented by (a) 2 m·s⁻¹·day⁻¹ and (b) 5 m·s⁻¹·day⁻¹. Thick black contour depicts the zero-wind line. Locations with negative $\bar{q}_\phi$ are marked by black stippled. Locations where a critical layer exists inside a region of negative $\bar{q}_\phi$ are marked by green stippled. Stippled regions shown in (a) also apply to (b). Latitudinal averaging is between 45°N and 55°N.

**Figure 8.11.** Altitude-latitude sections of $N_B^2$ as shading. Only eastward $\bar{u}$ is shown and contoured every 10 m·s⁻¹ as thin black contours. The zero-wind line is thickened.

**Figure 8.12.** Zonal phase speed vs. time plot of wavenumber-2 GHP amplitude (m) at 33, 51 and 67 km averaged between 45°N and 55°N. White contours increment by 5 m. Positive phase speed indicates eastward movement. The vertical brown line corresponds with the formation of a double-maxima wind structure on 31 December 2008. The vertical dashed line indicates SSW onset on 22 January 2009. Observations were retrieved from MLS.

**Figure 8.13.** Altitude-latitude sections of PW2 EP flux and EP flux divergence for (a) all phase speeds and (b) eastward phase speeds. Negative (blue) and positive (red) EP Flux divergence is contoured every (a) 5 m·s⁻¹·day⁻¹ and (b) 2 m·s⁻¹·day⁻¹. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s⁻¹, with the zero-wind line thickened. The negative $\bar{q}_\phi$ region is grey-shaded. The presence of a critical layer inside the negative $\bar{q}_\phi$ regions is shaded in green. The EP Flux vector components are scaled as done in Figure 7, but the EPW2 vector magnitudes are multiplied by 10.0. Data retrieved from MLS.

**Figure 8.14** Latitude-time sections averaged between 0.1 hPa and 0.01 hPa of ((a-d)a) PW forcing from WACCM-SD, ((a-d)b) GWD from WACCM-SD, ((a-d)c) upwelling or positive $\bar{w}^\ast$ from WACCM-SD, ((a-d)d) upwelling from MLS, ((a-d)e) $N_B^2$ from WACCM-SD, and ((a-d)f) $N_B^2$ from MLS. GWD is calculated by totaling parameterized the GWD and the EP flux of wavenumbers greater than 6. Black contours indicate eastward winds incremented by 10 m·s⁻¹. The zero-wind line is indicated by a bold black contour. Solid vertical line marks SSW onset. Dashed vertical line marks the formation of a double-maxima configuration. Dashed line is missing for the 2005-2006 winter because the double-maxima wind structure forms prior to December.

**Figure 8.15.** GW forcing (m·s⁻¹·day⁻¹) and wavenumber 1-6 PW forcing (m·s⁻¹·day⁻¹). Eastward (red contours) and westward (blue contours) accelerations are incremented by 20 m·s⁻¹·day⁻¹. Thin black contours indicate eastward zonal-mean zonal winds. EP flux vectors (m²·s⁻²) illustrate the direction of the waves’ group velocity. The meridional EP Flux vector component was scaled by $(100\pi R_e \rho)^{-1} \cos \phi$ and the vertical EP Flux vector component was scaled by $(R_e \rho)^{-1} \cos \phi$. Grey regions indicate negative $\bar{q}_\phi$.

**Figure 8.16.** Height vs. latitude sections from negative (blue contours) and positive (red contours) EP Flux divergence (m·s⁻¹·day⁻¹) for EPW2s moving faster than 5 m·s⁻¹; EP Flux
divergence/convergence contours increment by 2 m·s⁻¹·day⁻¹. EP flux vectors (m²·s⁻²) illustrate the direction of the waves’ group velocity. The meridional EP Flux vector component was scaled by $10 \times (100\pi R_e \rho)^{-1} \cos \phi$ and the vertical EP Flux vector component was scaled by $10 \times (R_e \rho)^{-1} \cos \phi$. The EP Flux divergence was scaled by $(\rho R_e \cos \phi)^{-1}$. Eastward zonal-mean zonal winds are incremented by 10 m·s⁻¹ (thin black contours) with the zero-wind line in bold. A region of negative $\bar{q}_\phi$ (grey-shaded region) fulfills the necessary condition for instability. The presence of a critical layer inside a region of negative $\bar{q}_\phi$ is indicated by green shading. Data was retrieved from (a,b) WACCM-SD and (c) MLS.

**Figure 8.17.** Height vs. latitude sections from (a, c) WACCM-SD and (b, d) MLS of negative (blue contours) and positive (red contours) EP Flux divergence (m·s⁻¹·day⁻¹) for EPW1s (top row) and EPW2s (bottom row) moving faster than 5 m·s⁻¹; EP Flux divergence/convergence contours increment by 2 m·s⁻¹·day⁻¹. EP flux vectors (m²·s⁻²) illustrate the direction of the waves’ group velocity. The meridional EP Flux vector component was scaled by $10 \times (100\pi R_e \rho)^{-1} \cos \phi$ and the vertical EP Flux vector component was scaled by $10 \times (R_e \rho)^{-1} \cos \phi$. The EP Flux divergence was scaled by $(\rho R_e \cos \phi)^{-1}$. Eastward zonal-mean zonal winds are incremented by 10 m·s⁻¹ (thin black contours). The zero-wind line is represented by a thick black contour. A region of negative $\bar{q}_\phi$ (grey-shaded region) fulfills a necessary condition for instability. The presence of a critical layer inside a region of negative $\bar{q}_\phi$ is indicated in green shading.

**Figure 8.18.** (a) Altitude-time of $\bar{u}$ at 60°N composited for all SSWs in Table 1. SSW onset at day 0 is indicated by a vertical dashed line. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s⁻¹, with the zero-wind line thickened. (b, c) Altitude-latitude composites of $\bar{u}$ (b) 35 days and (c) 10 days before SSW onset.

**Figure 8.19.** Latitude-time of (a) PW forcing, (b) total GWD, and (c) $N_2^2$ composited of all SSWs in Table 1. Vertical averaging are indicated to the left of each plot. Zonal-mean zonal wind averaged from 1.0 to 0.1 hPa indicates the position and strength of the polar jet (dotted contours) and the zonal-mean zonal wind averaged from 0.1 to 0.01 hPa indicates the position and strength of the subtropical jet (solid contours).

**Figure 8.20.** Altitude-latitude sections of EPW EP flux and its divergence. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s⁻¹, with the zero-wind line thickened. Incremented by 0.5 m·s⁻¹·day⁻¹, PW EP flux divergence are contoured in blue (red) for westward (eastward) forcing. The meridional EP flux vector component was scaled by $20 \times (100\pi R_e \rho_0)^{-1} \cos \phi$ and the vertical component by $20 \times (R_e \rho_0)^{-1} \cos \phi$.

**Figure 9.1.** Height-latitude composites of $\bar{u}$ averaged from 60-70°N during (a) normal and (b) SSW winters. $\bar{u}$ is shown by thin black contours and is incremented by 10 m·s⁻¹. Blue
and green lines emphasize the location of the -5 and 5 m s⁻¹ isotachs, respectively. Grey shading shows regions of negative \( \bar{q}_\phi \).

**Figure 9.2.** Wave geometries of vertical PW propagation between 20 km and 90 km. In our composite, these scenarios roughly occur (a) before day 0, (b, c) days 0 - 10, (d) and days 10 - 20.

**Figure 9.3.** Height-time (a,b) composites and (b) anomalies of \( \bar{q}_\phi \) during SSW winters. (a) Positive (negative) and negative \( \bar{q}_\phi \) values are shown by red (grey and purple) shaded contours. (b) Dominance of the beta, barotropic, and baroclinic terms are indicated by blue, grey and red contours, respectively. Lighter/darker shading indicates a positive/negative contribution of the dominant term. (c) Positive (negative) anomalies of \( \bar{q}_\phi \) are indicated by medium (thin) black contours. The probability that the anomaly is abnormal is given by orange-shaded contours. The zero-wind line is indicated by a thick bold contour.

**Figure 9.4.** Height-time plot averaged between 60°N and 70°N. (a-c) Upward (downward) vertical EP fluxes are outlined by black regular-sized (thin) contours incremented by +(-) 0.2 × 2^m kg s⁻², \( m \in [1,2,3,4,5,6,7,8] \) and filled with tan (blue) shading. EP flux convergence (blue contours) and divergence (red contours) are incremented by 2 m s⁻¹ day⁻¹. (d-f) Positive (negative) anomalies of EP flux divergence are indicated by solid (dashed) contours and the probability that the anomaly is abnormal is given by orange-shaded contours. The zero-wind line is indicated by a thick bold contour.

**Figure 9.5.** Height-time plot averaged between 60°N and 70°N. (a-c) Geopotential height perturbation amplitudes, \( Z_g \), are incremented by 100 m (d-f) Positive (negative) \( Z_g \) anomalies are indicated by solid (dashed) black contours incremented by 100 m s⁻¹. The probability that the anomaly is abnormal is given by orange-shaded contours. In all sub-plots, the zero-wind line is indicated by a thick bold contour.

**Figure 9.6.** Height-latitude plots for (a) day -5 and (b) day 0 of EP flux divergence (convergence) shown in red (blue) contours and incremented by 5 m s⁻¹ day⁻¹. The meridional EP Flux vector component was scaled by \((100\pi\alpha\rho)^{-1}\cos\phi\) and the vertical component by \((\alpha\rho)^{-1}\cos\phi\). Regions of \( \bar{q}_\phi < 0 \) are shaded grey. The zero-wind line is indicated by a thick bold contour.

**Figure 9.7.** Altitude vs. relative phase shift of zonal GWD averaged between 60°N and 70°N. Before compositing, the average phase between 10 and 5 hPa of the wavenumber-1 geopotential higher perturbation are aligned such that the PW trough is centered at 180° longitude. (a,b) GWD is in filled contours and overlaid by the westward (eastward) wind velocity in thin (bold) contours. (b,c) Orange-shaded contours show the probability of an anomaly being significant (i.e., the value found at a coordinate and time with respect to SSW onset is unique to normal or SSW winters) and overlaid by positive (negative) anomalies in solid (dashed) black contours.
Figure 9.8. Altitude vs. relative phase shift of zonal GWD averaged from 60-70°N and 0.1-0.01 hPa. Before compositing, the average phase between 10 and 5 hPa of the wavenumber-1 geopotential higher perturbation are aligned such that the PW trough is centered at 180° longitude. (a) GWD is in filled contours and overlaid by geopotential height in black contours. (b) Orange-shaded contours show the probability of an anomaly being significant (i.e., the value found at a coordinate and time with respect to SSW onset is unique to normal or SSW winters) and overlaid by positive (negative) anomalies in solid (dashed) black contours.
List of Tables

Table 7.1. Number of normal and SSW ensemble members generated with respect to reference SSW onset dates.

Table D1. Identification and classification of SSWs from 1979 to 2013. SSW onset dates were defined by the point of wind reversal at 1 hPa (Limpasuvan et al., 2016). Each event was assessed to be a WMO-SSW and ES-SSW using criteria set by the World Meteorological Organization (WMO) and Limpasuvan et al. (Limpasuvan et al., 2016) respectively. The SSW type was classified by the shape of the vortex at onset and compared to the type of SSWs collected from multiple studies by Albers & Birner (2014).
The Arctic stratospheric polar vortex is represented by the strong high-latitude circumpolar flow that develops during the wintertime as the Arctic region becomes cold and dark. Averaged zonally (i.e., across all longitudes), this circumpolar flow can be characterized by a net eastward wind with a maximum speed at the vortex’s edge. During certain winters, the polar vortex can become greatly distorted when perturbed by planetary-scale atmospheric waves, which originate primarily in the troposphere. The perturbed polar vortex can become greatly displaced off the pole or split into two vortices, leading to a rapid polar warming on the order of tens of degrees over a few days. When accompanied by the reversal of the longitudinally averaged flow to a westward direction, the resultant intense warming episode is referred to as a major sudden stratospheric warming (SSW) event.

Impacts from SSW events extend well beyond the stratosphere. The aforementioned wind reversal can eventually descend toward the upper troposphere and project onto a dominant mode of climate variability called the Arctic Oscillation (e.g., Baldwin & Dunkerton, 2001). Consequently, after a 10-30-day period following the vortex breakdown, the near-surface weather patterns tend to shift equatorward along with anomalously cold conditions over Europe and Northeast America. Furthermore, the changing stratospheric conditions during SSW can affect the state of the mesosphere and lower thermosphere (e.g., Limpasuvan et al., 2016). The climatological polar descent of air into the mesosphere and stratosphere also becomes enhanced, transporting chemical species that can catalytically destroy ozone (Kvissel et al., 2012; Randall et al., 2009). Planetary-scale Rossby waves, or planetary waves (PWs), associated with SSWs can also affect satellite communication through tidal amplification (e.g., Goncharenko et al., 2012; Liu et al., 2010; Oberheide et al., 2020). Ultimately, current forecasts are unable to reliably predict SSW events until ~2 weeks prior (Domeisen et al., 2020), revealing a critical need to better understand PW interactions in the middle atmosphere.
Despite numerous studies of SSW impact on near-surface climate, chemistry, aeronomy, and space physics, the precursory atmospheric conditions and detailed interactions between the atmospheric waves and the vortex are still under debate. The first order vortex breakdown is associated with nearly stationary PWs of zonal wavenumber 1 or 2 (PW1 or PW2 respectively). Displaced SSWs are accompanied by PW1 and split SSWs by PW2. Therefore, the sources, sinks, and propagation paths of these PWs are vital to understanding SSW events. Despite 3 recent events in 2018, 2019, and 2021, only 5 split SSWs have been observed from 1980-2011 (Albers & Birner, 2014). While the 2009 split SSW exhibited two vortices that remained relatively symmetric about the pole during onset, the other split SSW events (e.g., in December 1987) involved a polar vortex that became highly displaced prior to splitting. This ambiguity in what qualifies as a split and displaced SSW motivates a subsection of the present study that uses a new classification system to reveal how the polar vortex evolves in height and time (Appendix D). Through this classification system, the SSW types in our case studies were verified.

Recently, Iida et al. (2014) observed slow eastward-propagating planetary waves (EPWs) prior to the January 2009 SSW. Dissipation of these precursory EPWs may induce an independent mesospheric circulation prior to SSW onset that warms the lower mesosphere and cools the upper mesosphere (Iida et al., 2014). These EPWs may be sourced from instability due to strong mesospheric wind shear (Iida et al., 2014) or they may be linked to wavenumber-2 EPWs originating in the troposphere (Coy et al., 2011).

During the winter, gravity wave (GW) breaking in the mesosphere drives a strong polar downwelling (and associated adiabatic warming) and helps maintain the climatological stratopause (Hitchman et al., 1989). The filtering of upward-propagating gravity waves (GWs) by planetary-scale oscillations in the wintertime stratospheric winds could likewise imprint PW patterns in the mesosphere where these GWs break (Smith, 1997; 2003; Lieberman, 2013). Song et al. (2020) shows that PW growth prior to the January 2009 SSW corresponds to regions of large GW-induced vorticity.
Furthermore, the mechanisms that explain the presence of EPWs prior to SSWs, may also explain the presence of other PWs after SSW onset. Mesospheric westward-propagating PWs (WPWs) have been suggested to result from instability and aid in the recovery of the stratopause (Limpasuvan et al., 2016). These unstable modes have been found to have a significant impact on the background winds and meridional circulation in the mesosphere and lower thermosphere or MLT (e.g., Rhodes et al. 2021, Sassi et al. 2016). Sassi et al. (2016) also identifies WPWs in the MLT and shows that they can significantly impact the meridional circulation, enhancing upwelling in the tropics and downwelling at the pole. This downwelling can result in a descent of nitric oxides produced from energetic particle precipitation over the pole, as shown in a January 2013 case study by Orsolini et al. (2017).

The goal of this study is to understand the wintertime mesospheric evolution of SSWs with respect to the roles of PWs and GWs during SSW. Since SSW can potentially impact surface meteorological conditions, this understanding may help improve our climate predictive skills and provide new insights on the dynamical coupling between the stratosphere and mesosphere. In supporting this goal, our research objectives are to: (1) identify the source of EPWs before SSW onset, (2) assess the impact and uniqueness of mesospheric EPWs leading up to SSW onset, and (3) explore the behaviors of EPWs along with other PWs surrounding SSW and determine their possible source mechanisms.
Chapter 2

Basic Wind Structure and the Polar Vortex

This chapter provides the theoretical background on the stratospheric polar vortex (hereafter, “the polar vortex”). This background is grounded on a set of conservative laws called the primitive equations. In spherical horizontal coordinates and a geometric vertical coordinate representing the radial dimension, the primitive equations can be written as (e.g., Salby, 1996):

\[
\frac{du}{dt} - f v + \frac{1}{\rho R_e \cos \theta} \frac{\partial p}{\partial \phi} = D_\phi
\]  
(2.1a)

\[
\frac{dv}{dt} + f u + \frac{1}{\rho R_e} \frac{\partial p}{\partial \theta} = D_\theta
\]  
(2.1b)

\[
\frac{dw}{dt} + \frac{1}{\rho} \frac{\partial p}{\partial z} + g = 0
\]  
(2.1c)

\[
\frac{dp}{dt} + \rho \nabla \cdot v = 0
\]  
(2.1d)

\[
c_p \frac{dT}{dt} - \frac{1}{\rho} \frac{dp}{dt} = Q
\]  
(2.1e)

Definitions of the variables in the equations above appear in Appendix A. The material (or total) derivatives of the wind vectors represent the material accelerations on a sphere in terms of the local Cartesian accelerations (with subscript \( C \)) and variations in local Cartesian unit vectors.

\[
\frac{du}{dt} = \left( \frac{du}{dt} \right)_C - \frac{uv \tan \theta}{R_e} + \frac{uw}{R_e}
\]  
(2.2a)

\[
\frac{dv}{dt} = \left( \frac{dv}{dt} \right)_C + \frac{u^2 \tan \theta}{R_e} + \frac{vw}{R_e}
\]  
(2.2b)

\[
\frac{dw}{dt} = \left( \frac{dw}{dt} \right)_C - \frac{u^2 + v^2}{R_e}
\]  
(2.2c)

The material acceleration terms describe how the flow evolves. Generally, the horizontal momentum equations can be combined through cross-differentiation to obtain an equation that
accounts for the prognostic budget of absolute vorticity (that is, the sum of local vertical component of vorticity, $\zeta$, and the local vertical component of the Earth’s vorticity, $f$). With the inclusion of continuity (Equation 2.1d) and thermodynamics (Equation 2.1e), a budget for the potential vorticity (PV) can be obtained. PV is a measure of the ratio between the absolute vorticity to the vertical depth of the fluid column. Under adiabatic conditions, PV is conserved.

For typical large-scale motion, scale analysis of the horizontal momentum equations from 2.1 reveals the dominance of a near-steady state balance between the Coriolis effect and pressure gradient force. In the vertical, the dominant balance is nearly hydrostatic. The corresponding equations are:

$$
\nu_g = \frac{1}{f \rho R_e \cos \theta} \frac{\partial p}{\partial \phi} \quad (2.3a)
$$

$$
u_g = -\frac{1}{f \rho R_e} \frac{\partial p}{\partial \theta} \quad (2.3b)
$$

$$
\frac{\partial p}{\partial z} = -\rho g
$$

where the subscript $g$ denotes the geostrophic horizontal wind component. We can relate the horizontal geostrophic winds to geopotential as:

$$
\nu_g = \frac{1}{f R_e \cos \theta} \frac{\partial \Phi}{\partial \phi} \quad (2.4a)
$$

$$
u_g = -\frac{1}{f R_e} \frac{\partial \Phi}{\partial \theta} \quad (2.4b)
$$

Therefore, the geostrophic wind is determined by the horizontal gradient on a constant geopotential surface. Under geostrophic balance, the Coriolis and pressure gradient forces are in balance and perpendicular to the parcel’s velocity.

The density (of dry air) is related to the temperature through the ideal gas law, $\rho = \frac{p}{RT}$. This approximation can be applied to Equations 2.3 to obtain:
\[ v_g = \frac{RT}{fR_e \cos \theta} \frac{\partial \ln(p)}{\partial \phi} \]  
\[ (2.5a) \]
\[ u_g = -\frac{RT}{fR_e} \frac{\partial \ln(p)}{\partial \theta} \]  
\[ (2.5b) \]
\[ \frac{\partial \ln(p)}{\partial z} = -\frac{g}{RT} \]  
\[ (2.5c) \]

For simplicity, it is convenient to use a log-pressure as the vertical coordinate in the primitive equation. This coordinate assumes that pressure exponentially decays away from the ground and is weighted by the constant scale height \( H \equiv \frac{R \langle T \rangle}{k} \). The scale height, \( H \), is the altitude distance for which the atmospheric pressure falls by a factor of \( e^{-1} \) from a reference level. In the middle atmosphere, the reference level is the ground and \( H \) is \( \sim 7 \) km. The reference pressure at the ground is \( p_0 \) (see Andrews et al., 1987).

\[ z \equiv -H \ln \left( \frac{p}{p_0} \right), \quad \text{or} \quad p \equiv p_0 e^{-z/H} \]  
\[ (2.6) \]

Using this approximation, the derivative of \( p \) with respect to log-pressure height is convenient. Taking the derivative with respect to log-pressure height in Equations 2.5a and 2.5b allows the substitution of Equation 2.5c:

\[ \frac{\partial v_g}{\partial z} = \frac{R}{HfR_e \cos \theta} \frac{\partial T}{\partial \phi} \]  
\[ (2.7a) \]
\[ \frac{\partial u_g}{\partial z} = -\frac{R}{HfR_e} \frac{\partial T}{\partial \theta} \]  
\[ (2.7b) \]

These are the thermal wind equations. Equation 2.7a shows that the local northward meridional wind will change with altitude with the eastward change in temperature. For example, as the eastward air of the tropospheric jet flows from a cooler to warmer region, the wind will begin to turn counterclockwise with height, also known as wind backing. Similarly, as the wind circles the globe and encounters the warmer region again, the opposite effect, wind veering, will occur.
Therefore, geostrophic wind is continuously nudged to flow along isotherms and often results in global-scale waves or Rossby waves.

During the wintertime northern hemisphere, the distribution of solar radiation and gas constituents would produce a net heat rate on a steady atmosphere with a negative poleward gradient in temperature, $\frac{\partial T}{\partial \theta} < 0$. By thermal wind, this temperature structure would be manifested as a continuously increasing eastward wind with height. This radiatively determined wind state is however markedly different from the observed wind structure (discussed with Figure 4.1). The more complex observed wind structure is due in large part to the influence of PWs and GWs.

Starting in the fall, as the Earth’s axis tilts away from the Sun, the diminished solar insolation over the polar region enhances the temperature gradient between middle and high latitudes. This temperature gradient manifests a pressure gradient force that is predominantly balanced by the Coriolis force on a large scale, resulting in a nearly geostrophic circumpolar flow with eastward wind speed often exceeding 50 m s$^{-1}$. Figure 2.1 illustrates the circumpolar flow associated with the polar vortex at two pressure surfaces. Faster circumpolar flow, indicated by the tightening of the geopotential height contours, reflects a sharper temperature gradient and a strong tendency for the cold polar air mass to be isolated from the relatively warmer mid-latitude air. In the upper stratosphere and lower mesosphere where the eastward wind peaks, the flow is referred to as the polar night jet and appears on the periphery of the polar vortex. The weakening or strengthening of the polar vortex is synonymous with the weakening or strengthening of the polar night jet.
Figure 2.1. Polar plots at 10 hPa (left) and 1 hPa (right) on January 1, 1992. Line contours show geopotential height in increments of 0.15 km. Arrows represent the direction of geostrophic winds along the geopotential isolines. Colored-filled contours show temperature (in Kelvin). Data was retrieved from the Whole Atmosphere Community Climate model run in the specified dynamics configuration (WACCM-SD) described in Section 7.2.

Another representation of the polar vortex and the polar night jet can be obtained by longitudinally averaging the eastward zonal wind component ($\bar{u}$) and temperature ($T$) at every latitude and altitude. The longitudinal average (or zonal-mean) operation is denoted by an overbar, e.g. $\bar{u}$. Figure 2.2 illustrates an example of a typical $\bar{u}$ and $T$ structure. As marked by the thick dashed line, the polar vortex encompasses a cold region poleward of the eastward wind maxima. Consider the thermal wind equations (Equations 2.7). If we follow the suggested edge of the polar vortex (bold dashed line in Figure 2.2), we find that the meridional temperature gradient, $T_\phi$, remains negative corresponding to an increasing zonal wind with height. The zonal wind increases with height along the polar vortex edge until it reaches a maximum, the polar night jet, in the upper stratosphere around 50 km. Above this point, $T_\phi$ becomes positive and $\bar{u}$ decreases, eventually flowing in the westward direction. While waves can change the wind and thermal structure by net transporting momentum and heat to different levels of the atmosphere, the background structure will dictate where and how the waves propagate.
Figure 2.2. Five-day running average of $u$ (m s$^{-1}$) and zonal-mean temperature (K) on January 1, 1992. Thin solid contours indicate eastward and dashed contours indicate westward $u$. The zero-wind line is indicated by a thick black solid contour. Color contours indicate the temperature. The 10 hPa and 1 hPa levels in Figure 2.1 are indicated by the grey horizontal lines. Thick black dashed contour suggests the rim of the polar vortex. Data was retrieved from the WACCM-SD model run described in Section 7.2.
Chapter 3

Planetary Waves

Planetary waves (PWs) are Rossby waves with global-scale eddies that transport energy. Ranging in zonal wavenumbers roughly 1-6, PWs can impart zonal asymmetry onto the polar vortex. As most PWs are of tropospheric origin, they play a central role in the coupling between the troposphere and stratosphere. In the troposphere, synoptic-scale Rossby waves are more evident and their presence can lead to fluctuations in the jet stream which organize the weather.

In the meridional direction, Coriolis force is the restoring mechanism for fluid parcel oscillation associated with PWs. Considering the atmosphere in bulk as a closed chain of fluid along a latitudinal circle, the horizontal phase propagation of Rossby waves can be qualitatively understood by PV conservation (Holton & Hakim, 2013). For a fluid chain of constant depth, PV conservation reduces to absolute vorticity conservation, expressed as:

\[ \zeta + f = \text{constant} \] (3.1)

Here, \( f \equiv 2\Omega \sin\theta \) and is the Coriolis parameter, where \( \Omega \) is the angular frequency of Earth’s rotation in radians per second and \( \theta \) is latitude. Assuming an initial or reference state (say, state 0) wrapped around a latitude circle with no local vorticity at latitude \( \theta_0 \), this conservation allows the subsequent absolute vorticity to be written as:

\[ \zeta + f = f_0 \] (3.1)

or:

\[ \zeta = f_0 - f \] (3.1)

By definition, \( f_0 = 2\Omega \sin\theta_0 \). Since the Coriolis parameter only varies in the meridional direction, we can use a Taylor expansion to approximate its variation about \( \theta_0 \).

\[ f(\theta) \approx f(\theta_0) + \Delta\theta \frac{df}{d\theta}(\theta_0) = 2\Omega \sin(\theta_0) + \Delta\theta 2\Omega \cos(\theta_0) \] (3.2)
Under β-plane approximation, the Taylor expansion uses Cartesian coordinates on a plane tangent to the reference location \((\theta_0, \phi_0)\).

\[
f(\Delta y) \approx f_0 + \beta \Delta y \quad \text{s.t.} \quad \beta \equiv \frac{df}{dy_T} \bigg|_{y_T=0} = \frac{df}{d\theta} \frac{d\theta}{dy_T} \bigg|_{y_T=0}
\]

(3.3)

Under this approximation, \(\Delta \theta_0 \approx \Delta y/R_e\) such that \(\Delta y\) is the northward distance away from the reference point. A change in local meridional direction (\(\Delta y\)) can be related to latitudinal change (\(\Delta \theta\)): \(dy_T \approx R_e d\theta \rightarrow \frac{d\theta}{dy_T} \approx \frac{1}{R_e}\). Additionally, \(\frac{df}{d\theta} = 2\Omega \cos(\theta_0)\). Considering these relationships, we can write:

\[
\beta \equiv 2\Omega R_e^{-1} \cos \theta_0
\]

(3.4)

and \(\beta\) is known as the planetary vorticity gradient. Thus, Equation 3.1 becomes:

\[
\zeta = -\beta dy_T
\]

(3.5)

Constrained by absolute vorticity conservation, the relative vorticity must compensate for the deviation of planetary vorticity from its basic state. This effect is illustrated in Figure 3.1. If the fluid chain is initially disturbed such that its local displacement is northward, the fluid must subsequently have an anticyclonic vorticity to conserve PV (i.e., \(\zeta < 0\)). Similarly, fluid displaced southward will have a cyclonic tendency with \(\zeta > 0\). Discussed further in the next section, Figure 3.1 shows that the initial perturbation will migrate westward in time as a Rossby wave with \(\beta\) as the restoring effect.
Figure 3.1. Perturbed vorticity field and the induced velocity field (dashed arrows) for a meridionally displaced chain of fluid parcels. The heavy wavy line shows original perturbation position; the light line shows westward displacement of the pattern due to advection by the induced velocity. [From Holton & Hakim, 2013]

3.1 Basic Prognostics for Rossby Waves

The three-dimensional Rossby wave phase propagation in a uniform background $\bar{u}$ and an isothermal atmosphere can be obtained by linearizing the quasi-geostrophic (QG) version of PV conservation under adiabatic condition (see Chapter 3 of Andrews et al., 1987). The Boussinesq approximation further excludes density variation, except when coupled with gravity, to determine the buoyant force. These assumptions lead to QGPV conservation in terms of the streamfunction perturbation, $\psi$, (Houghton, 1977):

$$\left(\frac{\partial}{\partial t} + \bar{u} \frac{\partial}{\partial x_T}\right)\left(\nabla^2 \psi + \frac{f_0^2}{N_B^2} \frac{\partial^2 \psi}{\partial z_T^2}\right) + \beta \frac{\partial \psi}{\partial x_T} = 0$$

(3.6)

Geostrophic horizontal winds are used to define the streamfunction. Winds that depart from geostrophy contribute to horizontal divergence and, therefore, vertical winds as dictated by mass conservation. Although still non-linear, the above equation approximately describes the evolution of a PW (and Rossby wave, in general) given suitable initial and boundary conditions. For simplicity, the rest of Chapters 3 and 4 will refer to the approximated wave solution on a $\beta$-plane such that $[x, y, z] = [x_T, y_T, z_T]$.

A monochromatic wave solution with a complex amplitude $\Psi$ is assumed for the streamfunction perturbation:
\[ \psi = Re\{\Psi e^{i(kx+ly+mx-\omega t)}\} \]  \hspace{1cm} (3.7)

Substituting this into Equation 3.6 yields a dispersion relationship that relates the wave’s angular frequency to other spectral characteristics:

\[ \omega = \bar{u}k - \frac{\beta k}{k^2 + l^2 + \frac{m^2f_0^2}{N_B^2}} \]  \hspace{1cm} (3.8)

The wave phase/trace speed, \( c \), can be defined as:

\[ c \equiv \frac{\omega}{k^2 + l^2 + m^2} \]  \hspace{1cm} (3.9)

The wave phase propagation along the \( x, y, z \) directions can be defined, respectively, as the phase speed by \( c_x = \omega k^{-1}, c_y = \omega l^{-1}, \) and \( c_z = \omega m^{-1} \). Thus, the dispersion relationship can be casted as:

\[ c_x = \bar{u} - \frac{\beta}{k^2 + l^2 + \frac{m^2f_0^2}{N_B^2}} \]  \hspace{1cm} (3.10)

Relative to a given background wind (\( \bar{u} \)), we can define an intrinsic phase speed in the zonal direction as:

\[ c_x - \bar{u} = -\frac{\beta}{k^2 + l^2 + \frac{m^2f_0^2}{N_B^2}} \]  \hspace{1cm} (3.11)

Since the right-hand side of Equation 3.11 is always negative (\( \beta \) is positive definite), \( c_x - \bar{u} \) must therefore be negative. We can then conclude that a Rossby wave’s zonal phase speed (i.e., its phase speed when traced in the \( x \)-direction) is always westward relative to the background wind. For example, if \( \bar{u} \) is eastward and 50 m s\(^{-1}\), then the Rossby wave must have an eastward phase speed less than 50 m s\(^{-1}\) or a westward phase speed. This intrinsic westward propagation is consistent with the conceptual PV model shown in Fig. 2.3. Furthermore, Rossby waves are
dispersive. Perturbations with longer wavelengths (or smaller wavenumbers) tend to have larger zonal phase speed.

3.2 Propagation of Rossby Wave in a Simple Flow

The propagation of wave energy is dictated by the wave’s group velocity. The vertical group velocity can be derived from the dispersion relationship (Equation 3.8) and simplified with Equation 3.11 such that the group velocity is expressed in terms of the horizontal phase speed:

\[ c_{gz} = \frac{\partial \omega}{\partial m} = \frac{2mk}{\beta} \left( \frac{f_0^2}{N_B^2} \right) (\bar{u} - c_x)^2 \]  \hspace{1cm} (3.12)

Without loss of generality, we can make \( k \) positive definite. Hence, for upward energy propagation (\( c_{gz} > 0 \)), the vertical wavenumber \( m \) must be positive. In arranging Equation 3.11, we can solve for \( m \):

\[ m^2 = \frac{N_B^2}{f_0^2} \left[ \frac{\beta}{\bar{u} - c_x} - (k^2 + l^2) \right] \]  \hspace{1cm} (3.13)

With \( m \) and \( k \) both positive for vertical wave propagation, the Rossby wave phase orientation in the \( x-z \) plane dictates that the wave structure tilts westward with altitude. For vertical propagation of phase, \( m \) must be also be real. This means that bracketed terms inside Equation 3.13 must be positive which implies:

\[ c_x < \bar{u} < c_x + \frac{\beta}{k^2 + l^2} \]  \hspace{1cm} (3.14)

Therefore, a (ground-relative) stationary Rossby wave with \( c_x = 0 \) can only propagate vertically if the background wind is:

\[ 0 < \bar{u} < \frac{\beta}{k^2 + l^2} \]  \hspace{1cm} (3.15)

Consequently, the background flow supporting the vertical propagation of a Rossby wave must be eastward relative to the wave phase speed but not too strong.
Generally, upward propagating PWs (with small $k$ and $l$) can exist over a wider range of wind speed (Houghton, 1977). Eastward-propagating planetary waves (EPWs) can tolerate faster eastward background winds than stationary waves but are unable to propagate in eastward flow slower than their phase speed. Westward-propagating planetary waves (WPWs) are possible in eastward and westward flows, but are not able to tolerate as fast of eastward wind speeds like EPWs and stationary waves.

From the dispersion relationship (Equation 3.8), the meridional energy propagation can similarly be derived:

$$c_{gy} = \frac{\partial \omega}{\partial l} = \frac{2kl}{\beta} (\bar{u} - c_x)^2$$  \hspace{1cm} (3.16)

For Rossby wave energy to move equatorward ($c_{gy} < 0$), we must have $l < 0$ resulting in an eastward tilt toward the North Pole from a top-down perspective (see Equation 3.16). Figure 3.2 illustrates the horizontal wave structure for this scenario. The eastward and meridional wind components in the wave presence tend to be spatially correlated. Thus, qualitatively, as Rossby wave energy propagates equatorward, the wave will transport eastward momentum toward the Pole.
Figure 3.2. Schematic of the streamlines (solid) and isotherms (dashed) associated with a large-scale atmospheric disturbance in the mid-latitudes of the Northern Hemisphere. Arrows along the streamline contour indicate the direction of wind velocity. Their streamlines correspond approximately to lines of constant pressure, since the winds are nearly geostrophic. The signs of deviations from zonal-mean are shown at the bottom to illustrate the NE-SW tilt of the streamlines; this tilt indicates a northward zonal momentum transport, and the westward phase shift of the temperature wave relative to the pressure wave gives a northward heat transport. [Adapted from Hartmann, 1994]

Since streamfunction perturbations are related to the horizontal velocity perturbations, the spatial covariance of their product (i.e., the zonal-mean flux of zonal wind in the meridional direction or \(u'v'\)) can be written as (e.g., D. A. Randall, 2015)

\[
\overline{u'v'} = -\frac{A_{\Psi}^2 k l}{2}
\]  

(3.17)

where \(A_{\Psi}\) is the real amplitude of the streamfunction perturbation. Using the meridional group velocity defined in Equation 3.16, we see that:

\[
\overline{u'v'/c_{gy}} = -\frac{A_{\Psi}^2 \beta}{4(\bar{u} - c_x)^2}
\]

(3.18)

Since \(\beta\) is positive, Equation 3.18 shows that the meridional eddy momentum flux has an opposite sign as the meridional group velocity, consistent with the qualitative picture in Figure
3.2. In other words, an equatorward Rossby wave tends to deposit eastward momentum flux in the latitude belt from where it originated. As Rossby waves often originate in regions of eastward wind, equatorward propagating waves tend to accelerate the mean flow.

For completeness, we can derive the zonal group energy using Equation 3.8 and substituting with for the horizontal phase speed given by Equation 3.10:

\[ c_{gx} - \bar{u} = \frac{\partial \omega}{\partial k} = (c_x - \bar{u}) \left[ 1 + \frac{2k^2}{\beta} (c_x - \bar{u}) \right] \] (3.19)

From Equation 3.11, the intrinsic trace wave phase speed is always westward relative to the flow, i.e., \( c_x - \bar{u} < 0 \). For the intrinsic zonal group velocity to be westward \( (c_{gx} - \bar{u} < 0) \), the sum in the bracket of Equation 3.19 must be positive or equivalently:

\[ k^2 < \frac{\beta}{2|c_x - \bar{u}|} \] (3.20)

Hence, only PWs (i.e., small wavenumber) tend to have westward intrinsic zonal group velocity.

In considering Equations 3.12, 3.16, and 3.19, we note that the group velocity vanishes when the background zonal wind speed matches the wave zonal phase speed. The region where this occurs is called the critical layer where the Rossby wave activity is “frozen” in the medium and becomes absorbed when dissipative effects (like radiative damping) are considered.

3.3 Rossby Waves in a Realistic Flow

For a non-isothermal atmosphere, the background zonal-mean zonal wind can vary in the meridional and vertical directions, i.e. \( \bar{u}(y, z) \). The application of the curl operator to the momentum equations can still lead to QGPV conservation given by Equation 3.6 if \( \beta \) is replaced by an effective beta \( \beta_e \) defined as:

\[ \beta_e = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} \left( \frac{f^2}{N^2} \frac{\partial \bar{u}}{\partial z} \right) \] (3.21)
Here, \( \bar{\rho} \), the zonal-mean density taken to be the background density, and \( N_B^2 \) also vary in the meridional and vertical directions. In addition to the Earth’s rotation, the restoring effect now includes the meridional and vertical curvature of background wind and, as shown in Salby (1996), is the same as the meridional gradient of the zonal-mean QGPV, \( \bar{q}_y \) (see Appendix B for more information):

\[
\beta_e \equiv \frac{\partial \bar{q}}{\partial y} \equiv \bar{q}_y
\]  
\[
\therefore \quad \left( \frac{\partial}{\partial t} + \bar{u} \frac{\partial}{\partial x} \right) \left( \nabla^2 \psi + \frac{f_0^2}{N_B^2} \frac{\partial^2 \psi}{\partial z^2} \right) + \bar{q}_y \frac{\partial \psi}{\partial x} = 0
\]

Additionally, the aforementioned Rossby wave dispersion relation of and propagation properties are similar provided we adjust for the restoring effect as above. As such, we have:

\[
c_x - \bar{u} = - \frac{\bar{q}_y}{k^2 + l^2 + \frac{m^2 f_0^2}{N_B^2}}
\]  
\[
c_{gx} - \bar{u} = \frac{\partial \omega}{\partial k} - \bar{u} = (c_x - \bar{u}) \left[ 1 + \frac{2k^2}{\bar{q}_y} (c_x - \bar{u}) \right]
\]  
\[
c_{gy} = \frac{\partial \omega}{\partial l} = \frac{2kl}{\bar{q}_y} (\bar{u} - c_x)^2
\]  
\[
c_{gz} = \frac{\partial \omega}{\partial m} = \frac{2mk}{\bar{q}_y} \left( \frac{f_0^2}{N_B^2} \right) (\bar{u} - c_x)^2
\]

To this end, discussions presented in the previous sub-section about Rossby wave structure and related direction of energy propagation with respect to the background flow still apply.
3.4 Wave-Mean Flow Prognostics

3.4.1 Refractive Index

A more generalized wave solution of the streamfunction perturbation can be assumed in solving the QGPV equation. In doing so, the wave structure and propagation with respect to the spatial variation of the background wind can be mathematically characterized. First, assume:

\[ \psi = \text{Re} \left\{ \Psi(y, z) e^{i(kx - \omega t)} e^{\frac{z}{H}} \right\} \]  

(3.26)

The factor exponential term as a function of altitude accounts for the exponential decay of the atmospheric density in the vertical. Therefore, to conserve wave energy, the wave amplitude must increase exponentially with height. The substitution of this solution into the QGPV equation results in a Helmholtz equation:

\[ \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial}{\partial z} \left( f_0^2 \frac{\partial \Psi}{N_B^2 \partial z} \right) + n^2 \Psi = 0 \]  

(3.27a)

s.t. 

\[ n^2(y, z) = \frac{\bar{q}_y}{\bar{u} - c_x} - k^2 - \left( \frac{f}{2N_B H} \right)^2 \]  

(3.27b)

The two-dimensional index of refraction squared, \( n^2 \), provides a useful diagnostic for Rossby wave propagation. Equation 3.27a describes the streamfunction amplitude as a free, undamped oscillation in which the product term involving \( n^2 \) (i.e., the third term on the left-hand side) represents the restorative forcing on the streamfunction. When the coefficient \( n^2 \) is positive, the streamfunction amplitude oscillates and the amplitude of the streamfunction itself has a wave-like characteristic (in \( y \) and \( z \) directions). Similarly, if \( n^2 \) is zero, then Equation 3.27a becomes a Laplace-like equation in which an already present streamfunction can be maintained, but there can be not amplitude growth or dampening. If \( n^2 \) is negative, then the refractive index is imaginary and propagation is not possible. Therefore the distortion will become evanescent. In other words, linear wave propagation is inhibited in regions of imaginary refractive index and permitted in regions of real refractive index (Charney & Drazin, 1961). Note that the last two
terms in Equation 3.27b have negative contributions to $\eta^2$. Therefore, in order for wave propagation to occur, it is a necessary condition for the meridional gradient of QGPV, $\bar{q}_y$, and the relative background wind, $\bar{u} - c_x$, to have the same sign. Overall, the propagation of Rossby waves tends toward areas with a large $\bar{q}_y$ and low relative background winds (e.g., McDonald et al., 2011; Smith, 1983).

**Figure 3.3** illustrates the energy propagation of stationary PWs ($c_x = 0$) for a typical wintertime zonal-mean zonal wind distribution. Given a generally positive $\bar{q}_y$, a westward wind is expected to have an imaginary refractive index for stationary PWs, inhibiting propagation. However, a decrease in $\bar{u} - c_x$ increases $\eta^2$. Therefore, stationary PWs would veer toward the zero-wind line near the Equator. There, the waves approach their critical line where $\bar{u} - c_x = 0$. Their group velocity approaches zero and their wave activity is absorbed (see Equations 3.25). However, areas of large meridional gradient of QGPV (leading to large refractive index) could allow the PWs to propagate closer towards the pole. For example, enhanced wind shear from the polar night jet can increase the QGPV gradient and lead to more vertically oriented PWs (Albers & Birner, 2014).
Figure 3.3. Rays of two stationary planetary wave components introduced into the lower stratosphere in zonal-mean winds (m s$^{-1}$) representative of northern winter on a mid-latitude beta plane. Arrowheads mark uniform increments of time moving along a ray at the group velocity $\mathbf{c}_g$. The ray for component 1, which initially propagates upward and poleward, encounters a turning line in strong westerlies of the polar night jet, where wave activity is refracted equatorward. As the ray approaches easterlies, it next encounters a critical line ($\bar{u} = 0$), where $\mathbf{c}_g$ vanishes. Propagation then stalls, leaving wave activity to be absorbed, so the ray terminates. The ray for component 2 is initially directed upward and equatorward, so wave activity encounters the critical line and is absorbed even sooner. Thus, wave activity introduced between tropical easterlies and strong polar westerlies can propagate vertically only a limited distance before being absorbed. [Salby, 1996]

3.4.2 Eliassen-Palm Flux

As Rossby waves propagate, the covariance of the associated eddy perturbations can result in meridional or vertical eddy flux of momentum and heat. To calculate the associated fluxes, the zonal perturbations of relevant variables are averaged to generate the transformed Eulerian-mean set of the primitive equations. The resultant average acceleration on the background wind is proportional to the divergence of the flux scaled by the density and the moment arm,
\( \rho_0 R_e \cos \theta \). In other words, a net divergence of flux in a region due to a wave indicates an overall positive acceleration by the wave perturbation. By manipulating the primitive equations in this way, the Eliassen-Palm (EP) Flux vector can be obtained (c.f. Andrews et al., 1987):

\[
\vec{F} \equiv \left[ 0, \rho R_e \cos \phi \left( \frac{\nu \theta'}{\theta_z} - \nu' u' \right), \rho R_e \cos \phi \left( f - (R_e \cos \phi)^{-1} \left( \frac{\nu \theta'}{\theta_z} - \nu' u' \right) \right) \right]
\] (3.28)

Here, \( \theta' \), \( u' \), \( v' \), and \( w' \) represent departures from the zonal mean in potential temperature, zonal wind velocity, meridional wind velocity, and vertical wind velocity respectively. The vertical gradients of the zonal-mean zonal wind and potential temperature are \( \tilde{u}_z \) and \( \tilde{\theta}_z \). In general, energy transport and the propagation path of waves can be both be diagnosed through EP Flux.

For QG approximation suitable for Rossby waves, the meridional component of EP flux is approximately proportional to the meridional eddy momentum flux, \( F_y \propto -u'v' \), and the vertical component of EP flux is approximately proportional to the meridional heat flux, \( F_z \propto \nu' \theta'/\tilde{\theta}_z \).

Shown in Andrews et al. (1987), the linearized version of the QGPV equation is defined as:

\[
q' t + (\bar{u} - c_x) q'_x + v' \bar{q}_y = X' + O(\alpha^2)
\] (3.29a)

where

\[
q' = v'_x - u'_y + \rho_0^{-1} \left( \bar{f} \frac{\theta'}{\theta_{0z}} \right)_z
\] (3.29b)

Note that subscripts of directional unit vectors imply a partial derivative. The variable \( \bar{q}_y \) (or \( \beta_e \)) is previously defined by Equation 3.21. \( X' \) represents vorticity induced by external forcings, including GWD. \( O(\alpha^2) \) is the error from linearization and may become large in regions of PW breaking. \( \theta_{0z} \) is the derivative of the basic state of potential temperature with respect to height and \( \rho_0 \) is the basic state density.

\[
\left( \frac{1}{2} \frac{\rho_0 q'^2}{\bar{q}_y} \right) + \rho_0 v' q' = \frac{\rho_0 X' q'}{\bar{q}_y} + O(\alpha^3)
\] (3.30)

Overall, Equation 3.30 reflects the zonal covariance of terms in Equation 3.29a with \( q' \). The first term in Equation 3.30 is the change in wave activity density over time or \( \partial A/\partial t \), the second term is the poleward flux of QGPV, the third term is the frictional and diabatic effects of the wind or
\( D \), and \( O(\alpha^3) \) encompasses higher order or non-linear effects. Furthermore, multiplying Equation 3.29b and taking the zonal average reveals that the meridional eddy flux of QGPV is equivalent to the EP flux divergence scaled by the inverse of the density. That is,
\[
\overline{v'q'} = -\overline{(v'u')} + \rho_0^{-1} \overline{\rho_0f_0 \frac{v'\theta'}{\theta_0z}}_z = \rho_0^{-1} \overline{\nabla \cdot \mathbf{F}}  
\tag{3.31}
\]
Therefore, EP flux divergence implies the poleward flux of QGPV.

Finally, the generalized EP theorem is obtained by substituting Equation 3.31 into Equation 3.30 and incorporating the new definitions for terms in Equation 3.30.
\[
\frac{\partial A}{\partial t} + \nabla \cdot \mathbf{F} = D + O(\alpha^3)  
\tag{3.32}
\]
Characterizing wave forcing, the flux divergence occurs when Rossby waves break on the horizontal plane as their amplitudes become large or as they approach their critical surface. The effect of the wave forcing tends to accelerate the zonal-mean zonal wind and drive an overturning circulation (Holton & Hakim, 2013). As stated by Equations 3.25, the wave group velocity is inversely proportional to the meridional gradient of QGPV. The PV flux poleward will increase the meridional gradient of PV and thus lessen the group velocity of subsequent wave while inducing an acceleration on the background wind.

For a linear consideration and a conservative flow, the terms \( D \) and \( O(\alpha^3) \) are zero. If the wave is steady, it would not exert a net influence on the background flow since: \( \nabla \cdot \mathbf{F} \equiv 0 \). Hence, for a linear, conservative flow, the unsteady behavior of waves (through dissipation or some form of damping) will lead to a non-zero EP flux divergence and an acceleration of the mean flow via poleward flux of the QGPV.

Figure 3.4 illustrates EP flux (vectors) and its divergence/convergence (red/blue contours) for the monthly-averaged background \( \overline{u} \) of January 1983. The vectors are computed using Equation 3.28. As suggested by Equation 3.14 and illustrated in the Southern Hemisphere of Figure 3.4, Rossby wave propagation across the tropopause is prevented half of the year by westward stratospheric
winds (Charney & Drazin, 1961). However, the wintertime stratospheric wind configuration readily allows PWs to impact the stratosphere. Like in Figure 3.3, PW group velocities in the wintertime hemisphere tend toward the subtropical zero-wind line. Above the stratosphere, only low wavenumber Rossby waves (i.e., PWs) exist since critical background wind speeds for higher wavenumber Rossby waves are too small to permit vertical transmission unless the mean zonal velocities at high levels are small and positive. Therefore, the EP Flux in Figure 3.4 should be dominated by PWs.

![Monthly-Averaged EP Flux for January 1983](image)

**Figure 3.4.** Height-latitude plot of EP flux and monthly-averaged winds for January 1983. PW group velocities are indicated by EP flux vectors. EP flux convergence (divergence) are shown in blue (red) contours and increment by 5 m s\(^{-1}\) day\(^{-1}\). Meridional EP flux is scaled by \(5 \times (100\pi Re \rho)^{-1} \cos \phi\) and vertical EP flux by \(5 \times (Re \rho)^{-1} \cos \phi\). EP flux convergence/divergence is scaled by \(\rho Re \cos \phi\). Eastward (westward) winds are shown by solid (dashed) black contours incremented by 20 m s\(^{-1}\) with the zero-wind line in bold. Model data is from WACCM-SD nudged with MLS described in Section 7.2.

Given the monthly time average, the illustrated PW activity should be nearly stationary. The wave propagation is similar to the stationary PW rays shown in Figure 3.3. Near the zero-wind line, the refractive index is large and EP flux demonstrates PWs being refracted toward the subtropics where they can be absorbed by the critical surface. Around the equatorward side of the polar night jet core, the enhanced QGPV meridional gradient can lead to an increased refractive index.
Subsequently, PWs can also be refracted toward the edge of the polar vortex and break, resulting in the large EP flux convergence as seen in Figure 3.4 around 50°N and 60 km.

The dissipation of upward propagating PWs, i.e. EP Flux convergence, will generally result in a net poleward heat flux (Randall, 2015) as reflected by Equation 3.31; the vertical flux of zonal momentum ($\overline{w'u'}$) shown in Equation 3.28 tends to be small for PWs. Additionally, net eastward momentum will be moved poleward. Regardless, the overall effect is the equatorward movement of PV. If the PWs become more vertical, the dissipation of the PWs will cause more poleward heat flux, weakening the polar vortex.

### 3.5 Generation of PWs Through External Forcings

PWs can be categorized as forced or free, travelling or stationary, and transient or steady. Forced PWs must be continually maintained by a perturbing mechanism whereas free PWs are not. Travelling PWs move relative to the ground while stationary waves maintain a constant phase about a geographical point. Finally, transient waves vary in amplitude over time, often growing or dampening, while steady waves maintain a constant amplitude over time (Andrews et al., 1987). Some of these properties were discussed earlier.

Most forcing mechanisms for PWs reside in the troposphere. Aside from the superposition of travelling PWs, stationary PWs are generated as the background winds flow over large-scale topography (like the Rockies or the Himalayas) or by persistent longitudinally dependent diabatic heating due to the difference in the land-ocean heat capacity or by anomalous sea surface temperature (e.g., El Niño). Large-scale continental elevation can cause the fluid column to stretch or compress and, thereby, perturbing PV. Poleward motion can initiate PWs due to the conservation of PV (Charney & Eliassen, 1949), as illustrated Figure 3.1. Thermal anomalies can likewise induce a PW response by producing anomalous vertical motion, causing the fluid column to stretch (Smagorinsky, 1953). Over a long timescale, diabatic heating tends to be balanced by horizontal temperature advection and adiabatic cooling due to vertical motion. Typically, thermal anomalies near the tropics (related to convection) are more effective at generating PWs (Holton...
& Hakim, 2013). Overall, these mechanical and thermal forcings tend to produce PW perturbations of zonal wavenumbers 1-6 (Charney & Drazin, 1961). Figure 3.5 shows a global stationary PW response to SST warming in eastern tropical Pacific.

![Figure 3.5](image)

**Figure 3.5.** Anomalous planetary wave field for northern winters during El Niño, when anomalous convection (stippled) is positioned in the tropical central Pacific. The anomalous ridge over western Canada and trough over the eastern United States characterize the so-called Pacific North America pattern that upsets the normal track of the jet stream (wavy trajectory) and cyclone activity during El Niño winters. [Adapted from Salby, 1996 and Horel & Wallace, 1981]

Rossby wave phase speed depends on the scale of the forcing mechanism and not on the amount of forcing. In Figure 3.5, PW has a long wavelength and propagates westward relative to the wind over time. Therefore, if the forcing is stationary in space but changes in time, like an SST anomaly, waves can also be forced and travelling. Additionally, if the forcing remains constant but moves in space, like diurnal fluctuations in solar heating, a forced travelling wave is produced. When the forcing is continuous and constant, like a mountain chain, a stationary wave is maintained as the crest/trough of the wave is fixed relative to the forcing. Depending on the boundary and background wind conditions, the disturbance of a continuously produced stationary PW could have a constructive interference with itself, increasing or diminishing the wave amplitude.
In contrast, free PWs can exist in the atmosphere through short-lived or random forcing. By linearizing the primitive equations for a resting atmosphere and separating height dependency from the horizontal and temporal dependency, free atmospheric oscillations can exist on a spherical coordinate without QG scaling arguments (Laplace, 1799). Free oscillations are devoid of gravitational and thermal forcings. The assumed separable solutions lead to two eigenvalue problems described by second-order ordinary differential equations: one related to the meridional structure of the wave known as Laplace’s Tidal Equation (LTE) and the other related to the vertical wave structure. By carefully considering boundary conditions at the poles for LTE, solutions to this eigenvalue problem called Hough functions give the possible horizontal structures (i.e., normal modes) for waves on a sphere (D. A. Randall, 2015). Coupled with the possible vertical structure solution, a family of waves (including travelling and stationary PWs) of various frequencies and wavenumbers can be supported for various flow characteristics and boundary conditions. While they exist mathematically for a resting atmosphere, we would expect these waves to be weak and short-lived. These waves are generally referred to as global normal modes (Andrews et al., 1987).

For a non-resting atmosphere, analytic solutions as derived by Laplace is not possible. To consider solutions in more complex and realistic environments, global normal modes are found by disturbing the atmosphere in different realistic configurations and observing which waves of particular spectral characteristics exhibit a resonant-like response (e.g., Salby, 1981). In this way, the characteristic modes of a complex atmosphere can be reverse-engineered to determine the wave characteristics that create resonance or wave growth.

### 3.6 Generation of PWs Through Instability

#### 3.6.1 Diagnosis for Shear Instability

PWs can grow spontaneously by drawing on the energy of the background flow through instability. Instability wave growth can be mathematically realized through the conservation
equation for the perturbation streamfunction with a realistic background wind \( u(y, z) \) as given by:

\[
\frac{\partial^2 \Psi}{\partial y^2} + \frac{1}{\rho_0} \frac{\partial}{\partial z} \left( \frac{f_0^2}{N_B^2} \rho_0 \frac{\partial \Psi}{\partial z} \right) + n_e^2 \Psi = 0 \tag{3.33a}
\]

s.t.

\[
n_e^2 = \frac{\bar{q}_y}{\bar{u} - (c_R + i c_I)} - k^2 - \left( \frac{f}{2 N_B H} \right)^2 \tag{3.33b}
\]

Here, the assumed wave solution has the form:

\[
\psi = \Psi(y, z)e^{i(kx - (\omega R + \omega_I)t)} e^{\frac{z}{2H}} = [\Psi(y, z)e^{\omega t} e^{i(kx - \omega_R t)} e^{\frac{z}{2H}} \tag{3.34}
\]

where the wave angular frequency is allowed to be complex, \( \omega = \omega_R + i \omega_I \). As shown in Equation 3.34, the complex part of the frequency, if positive, can exponentially amplify the wave amplitude. The effective index of refraction squared, \( n_e^2 \), is essentially the refractive index squared with the incorporation of the complex phase speed, \( c = c_R + i c_I = \frac{\omega_R}{k} + \frac{i \omega_I}{k} \). If the phase speed only consists of a real component, a singularity exists in the refractive index along the critical line where the stable wave energy is absorbed as the PW group velocity tends to zero (Equations 3.25). The addition of the imaginary phase speed (and imaginary frequency) removes this singularity and allows an exponential wave growth in time for an unstable flow.

Following Salby (1996), the imaginary and real components of the frequency can be decoupled by taking the difference between Equation 3.33a multiplied by \( \Psi \) and Equation 3.33a multiplied by \( \Psi^* \), the complex conjugate of \( \Psi \). Finally, this expression can be integrated across the \( y-z \) plane such that the south most and north most point on the meridional plane is at \( y = -L \) and \( y = L \), respectively.

\[
\frac{\omega_I}{k} \left[ \int_0^\infty \int_{-L}^L \bar{\Psi}_y \left( \frac{\rho_0 |\Psi|^2}{|\bar{u} - c|^2} \right) dy dz - \int_{-L}^L \left( \frac{f_0^2}{N_B^2} \frac{\rho_0 |\Psi|^2}{|\bar{u} - c|^2} \frac{\partial \bar{u}}{\partial z} \right) \bigg|_{z=0} dy \right] = 0 \tag{3.35}
\]
By definition in Equation 3.34, the wave must have a non-zero imaginary frequency for instability to occur. For a positive $\omega_I$ leading to instability, then the imaginary zonal phase speed must also be positive $c_{tx} = \frac{\omega_I}{k} > 0$. First, if we assume that the vertical wind shear vanishes at the lower boundary, $z = 0$, Equation 3.35 reduces to:

$$\int_0^\infty \int_{-L}^L \bar{q}_y \frac{\rho_0 |\Psi|^2}{|\bar{u} - c|} dydz = 0$$

(3.36)

Since only $\bar{q}_y$ is able to change sign, it must reverse somewhere in the interior. The middle atmosphere normally has a positive $\bar{q}_y$. Therefore, a region of negative $\bar{q}_y$ values would generally result from a sign change around its perimeter. The presence of negative region of QGPV meridional gradient with the flow (away from the boundaries) constitutes the necessary condition for instability. From Equations 3.21 and 3.22, $\bar{q}_y$ can be defined as:

$$\bar{q}_y \equiv \beta_c = \beta - \frac{\partial^2 \bar{u}}{\partial y_T^2} - \frac{1}{\bar{\rho}} \frac{\partial}{\partial z_T} \left( \frac{f_0^2}{N_B^2} \bar{\rho} \frac{\partial \bar{u}}{\partial z_T} \right)$$

(3.37)

If this necessary condition is met such that horizontal wind curvature in Equation 3.37 largely contribute to $\bar{q}_y$ being negative, then the growing wave disturbances are associated more with \textit{barotropic instability}. If the condition arises through vertical wind curvature (proportional to the horizontal temperature gradient), then the unstable disturbances are associated more with \textit{baroclinic instability} (Salby, 1996).

**3.6.2 Over-reflection Perspective and Instability**

A turning level occurs when $\bar{q}_y$ changes sign while a critical layer occurs when $\bar{u} - c_x$ changes sign. Whenever $\bar{q}_y$ and $\bar{u} - c_x$ are the same sign, wave propagation is possible. Whenever they are opposite signs, $n^2$ is negative and the wave becomes evanescent. In this evanescent region, the perturbation is no longer a wave and its amplitude exponentially decays with height due to shear effect. The ability for the perturbations of a wave to exist outside of regions of $n^2 > 0$ is called tunneling.
If a wave propagates toward a critical layer, its amplitude and group velocity go to zero. In this scenario, the critical layer acts as a singularity such that waves can never reach this layer, and thus can’t exist beyond it). However, if a critical region is embedded in a region of negative $\bar{q}_y$, then a wave can tunnel past the turning level and the effects of the wave can reach the critical layer. In this scenario, the critical layer can act as a wave source in which the perturbation seeds the critical layer and wave growth occurs (Dickinson, 1973).

Over-reflection is illustrated in Figure 3.6. Upward-propagating PWs (thin solid arrow) can over-reflect in an evanescent region (dark gray shading) if a perturbation is able to tunnel from the turning level to the critical layer. This over-reflection produces a reflected wave (thick solid arrow) below the turning level and a transmitted wave (thin dashed arrow) above the critical layer. The reflected wave carries more energy than the incident wave; this wave growth is indicated by EP flux divergence (red contours). The wave geometry that a PW experiences changes depending on the zonal phase speed of the wave. Red, yellow, and blue colors represent eastward, stationary, and westward zonal phase speeds, respectively. For the background wave geometry shown in Figure 3.6, eastward waves would experience a thinner evanescent region, allowing them to tunnel more easily to the critical layer and stimulate PW growth at the critical level. The modulation of phase speed by the background geometry of the instability layer will alter the unstable PW’s future propagation paths and its eventual dissipation.
Figure 3.6. A schematic of planetary wave (PW) over-reflection with zonal phase speeds ($c_x$), zonal-mean zonal wind ($\bar{u}$), meridional potential vorticity gradient ($\bar{q}_y$), and squared refractive index ($n^2$) fields with arbitrary values.

Ultimately, the resulting unstable PW will tend to diminish negative $\bar{q}_y$ by reducing the pre-existing wind shear/curvature through momentum and heat transport. While additional non-conservative considerations (such as asymmetric GW drag or GWD) are necessary, the over-reflection perspective is powerful as it connects incident, transmitted, and over-reflected waves to one another. The incident and over-reflected waves offer a medium in which the stratosphere and MLT can dynamically influence each other. Additionally, the associated wave transmission aloft offers a medium in which disturbances in the stratosphere may indirectly result in disturbances in the ionosphere.

While mathematics support over-reflection through the conservation of momentum and energy, a mechanistic explanation for over-reflection is still obscure. In order to better understand this, we can look at these conditions from a counter-propagating Rossby Wave (CRW) perspective.
(further described in Appendix C). While this perspective is not used in our current study, it may play an important role in future research when considering flows with evolving wave geometries.
Chapter 4

Gravity Waves

With spatial scales on the order of 10-1000 km, gravity waves (GWs) are atmospheric perturbations with buoyancy as a restoring force that opposes vertical displacements (Holton & Hakim, 2013; Nappo, 2002). Their intrinsic frequencies are bounded between the Coriolis parameter ($f$) and the Brunt-Väisälä frequency ($N$). GWs can be generally classified as either orographic or non-orographic.

Orographic GWs are generated by background flow over a stationary disturbance and have near-zero phase speeds. Sources of orographic GWs include mountains, cities, and ‘heat islands’ (e.g., Dörnbrack et al., 1999; Lilly & Kennedy, 1973). With relatively high frequencies, orographic GWs are considered pure internal gravity waves (discussed further below).

Non-orographic GWs are generated predominantly by convection, wind shear, and geostrophic adjustment (Fritts & Alexander, 2003). Non-orographic GWs range in phase speed depending on the source. Multiple sources can overlap making them difficult to decouple. Low-frequency non-orographic GWs can be categorized as inertia gravity waves (IGWs) if their frequency is close to the Coriolis parameter. The appearance of IGWs has been found in the vicinity of surface and mid-tropospheric fronts (Zhang, 2004; Plougonven & Zhang, 2014) and in the jet-exit regions linked to large-scale geostrophic adjustment processes (e.g., Guest et al., 2000; Uccellini & Koch, 1987).

Regardless of sources, the generated GWs can propagate very far away from their source until they reach a critical layer where they are absorbed or break when their amplitudes become large. At those regions, the resulting GW dissipation would lead to momentum and heat deposition that alters the background flow. As in the case for Rossby waves, the critical layer occurs when GWs encounter a background wind that matches its phase speed. Figure 4.1 illustrates how GWs of various origins deposit momentum at different atmospheric levels, driving the winds away from their radiative equilibrium devoid of dynamics. Orographic GWs (or mountain waves), for
example, induce a large drag on the upper-stratospheric and mesospheric winds. However, during the summer, the orographic GWs break in the tropopause as the general wind direction changes from eastward in the troposphere to westward in the stratosphere. By imposing drag on the upper-level winds, GWs cap the stratospheric/mesospheric jet (Holton, 1982) and drive the pole-to-pole mean meridional circulation with upwelling over the summer pole and downwelling over the winter pole (McLandress et al., 2013).

![Figure 4.1](image1.png)

**Figure 4.1.** Typical mid-latitude zonal winds during (a) winter and (b) summer. The black curve shows observed winds and the grey curve shows model “radiative” winds that result without GWD. Sources of gravity waves with various phase speeds, \( c \), are also depicted, with source and wave breaking symbols. On these plots, waves ascend vertically upwards since \( c \) remains constant, until they reach their critical level \( c = \bar{u}(z_C) \). [Kim et al., 2003; based on a presentation first used by Lindzen, 1981]

In the stratosphere, the zonal-mean vertical flux of eastward momentum of non-orographic GWs tend to be less than that of orographic GWs, as shown in **Figure 4.2** for the Southern Hemisphere. The vertical flux of zonal momentum indicates the vertical energy propagation of GWs. The increase in orographic contributions poleward of 70°S occurs because of forcing by the Antarctic continent, particularly the Antarctic peninsula. By comparison, non-orographic GW momentum
flux peaks near the mid-latitudes overlying the jet-front system and, as noted by Hertzog et al. (2008), the Southern Hemisphere storm tracks (Trenberth, 1991).

Figure 4.2. Zonal-mean orographic GW momentum flux (thin solid line), non-orographic GW momentum flux (thin dashed line), and the combination of both non-orographic and orographic momentum fluxes (bold solid line). Observational data was obtained from long-duration balloons in the stratosphere. [Hertzog et al., 2008]

4.1 Basic Equations Related to Gravity Waves

The primitive equations (e.g., Equations 2.1) can also be linearized and simplified for GWs with no explicit external forcing. The resulting equations are:

\[ \frac{du'}{dt} - f v' + \frac{1}{R_e \cos \theta} \frac{\partial \Phi'}{\partial \phi} = 0 \]  

\[ \frac{dv'}{dt} + f u' + \frac{1}{R_e} \frac{\partial \Phi'}{\partial \theta} = 0 \]  

\[ c_p \frac{dT}{dt} - \frac{1}{\rho} \frac{dp}{dt} = Q \]  

We can assume hydrostatic balance such that the vertical acceleration can be eliminated from the vertical momentum equation.

\[ \frac{1}{\rho_0} \frac{\partial p}{\partial z} + g = 0 \]  

(4.2)
Additionally, if we assume an incompressible atmosphere, the mass continuity equation becomes separable such that:

\[
\frac{1}{R_e \cos \theta} \left( \frac{\partial u}{\partial \phi} \frac{\partial (v \cos \theta)}{\partial \theta} \right) + \frac{1}{\rho_0} \frac{\partial (\rho_0 w)}{\partial z} = 0 \quad (4.3a)
\]

&

\[
\frac{D\rho}{Dt} = 0 \quad (4.3b)
\]

Equation 4.3a can be easily linearized since the divergence of the mean winds is zero, leaving the divergence of the wind perturbations to also be zero.

\[
\frac{1}{R_e \cos \theta} \left( \frac{\partial u'}{\partial \phi} \frac{\partial (v' \cos \theta)}{\partial \theta} \right) + \frac{1}{\rho_0} \frac{\partial (\rho_0 w')}{\partial z} = 0 \quad (4.4a)
\]

Equation 4.3b can also be linearized, yielding:

\[
g \frac{D\rho'}{Dt} + \rho_0 N_B^2 w' = 0 \quad (4.4b)
\]

Notice that the density is assumed to be dependent on height only. Additionally, the mean vertical velocity can be approximated as zero. This assumption may also be made for the density perturbation.

\[
\frac{\partial}{\partial t} \left( -\frac{g \rho'}{\rho_0} \right) + N_B^2 w' = 0 \quad (4.5)
\]

or

\[
\frac{\partial}{\partial t} \left( \frac{\partial \Phi'}{\partial z} \right) + N_B^2 w' = 0 \quad (4.6)
\]

Therefore, the linearized form of the primitive equations for GWs are as follows:

\[
\frac{du'}{dt} - f v' + \frac{1}{R_e \cos \theta} \frac{\partial \Phi'}{\partial \phi} = 0 \quad (4.7a)
\]

\[
\frac{dv'}{dt} + f u' + \frac{1}{R_e} \frac{\partial \Phi'}{\partial \theta} = 0 \quad (4.7b)
\]
\[
\frac{1}{R_e \cos \theta} \left( \frac{\partial u'}{\partial \phi} + \frac{\partial (v' \cos \theta)}{\partial \theta} \right) + \frac{1}{\rho_0} \frac{\partial (\rho_0 w')}{\partial z} = 0 \tag{4.7c}
\]

\[
\frac{\partial}{\partial t} \left( \frac{\partial \Phi'}{\partial z} \right) + N_B^2 w' = 0 \tag{4.7d}
\]

### 4.2 Pure Internal Gravity Wave Characteristics

Similar to PWs, we assume wave-like perturbations in wind and geopotential for Equations 4.7.

\[
[u', v', w', \Phi'] = \exp \left( \frac{z}{2H} \right) \Re \left( [\tilde{u}, \tilde{v}, \tilde{w}, \tilde{\Phi}] \exp(i(kx + ly + mz - \omega t)) \right) \tag{4.8}
\]

This assumption allows perturbations to grow exponentially in amplitude due to the decrease in density with height. GWs can be small such that the Earth’s boundaries or rotation does not affect their characteristics. For pure internal GWs, the Coriolis force can be ignored as they, by definition, have frequencies much larger than \( f \). The wave equation for linear GWs can be derived from the assumed wave solution above (Nappo, 2002):

\[
\frac{d^2 \tilde{w}}{dz^2} - \frac{1}{H} \frac{d \tilde{w}}{dz} + \left( \frac{k_{GW}^2 N_B^2}{\omega^2} + \frac{k_{GW} d^2 u_0}{\omega H} - \frac{k_{GW} 1 d u_0}{\omega} - k_{GW}^2 \right) \tilde{w} = 0 \tag{4.9}
\]

For simplicity, the coordinates are rotated with respect to the wave such that \( \tilde{k}_{GW} = \tilde{k} + \tilde{l} \). The second term on the left-hand side accounts for the non-hydrostatic effects. The various terms in the parentheses are associated with buoyancy, curvature, shear, and the horizontal wavenumber, respectively.

The dispersion relationship is determined by solving for the vertical wavenumber in the general solution to Equation 4.9, namely, \( \tilde{w} = A e^{imz} + B e^{-imz} \) where \( A \) and \( B \) are unknown amplitudes. For simplicity, no background wind is assumed such that the ground-relative frequency is equivalent to the intrinsic frequency, \( \omega = \tilde{\omega} \). Additionally, hydrostatic balance can be assumed for vertical wavenumbers that are much larger than \( H \approx 7 \) km such that \( m^2 \gg 1/(4H^2) \). In other words, the small vertical wavelengths of the GWs (~15km or less) allow for a Boussinesq approximation. These assumptions reduce Equation 4.9 to:
\[
\frac{d^2 \tilde{w}}{dz^2} + \frac{k_{GW}^2 N_B^2}{\omega^2} \tilde{w} = 0 \tag{4.10}
\]

\[
\omega^2 = \frac{N_B^2 (k^2 + l^2)}{m^2} \tag{4.11}
\]

Substituting \( k_{GW} \) we obtain:

\[
\omega = \pm N_B k_{GW} / m \tag{4.12}
\]

The sign of the frequency is chosen to be negative such that \( k \) points in the positive direction and the vertical group velocity travels upwards (shown in Equation 4.13d). The phase and group velocity of GW can be calculated from Equation 4.12:

\[
c_x = \frac{\omega}{k_{GW}} = - \frac{N_B}{m} \tag{4.13a}
\]

\[
c_z = \frac{\omega}{m} = - \frac{N_B k_{GW}}{m^2} \tag{4.13b}
\]

\[
c_{gx} = \frac{\partial \omega}{\partial k_{GW}} = - \frac{N_B}{m} \tag{4.13c}
\]

\[
c_{gz} = \frac{\partial \omega}{\partial m} = \frac{N_B k_{GW}}{m^2} \tag{4.13d}
\]

The opposite signs between the vertical phase and group velocity indicate that the energy propagates in the opposite vertical direction to the phase of the wave. Consequently, the group velocity, \( c_{gz} = \frac{\partial \omega}{\partial m} \), will be directed upward if the wave phase travels downward. Since GW sources are located below the mesosphere, mainly in the troposphere, the present study will focus exclusively on GWs with upward group velocities.

Figure 4.3 shows experimentally the relationship between the phase velocity and group velocity. The experiment effectively shows a zonal flow impinging upon a “mountain” (indicated by the black near-semicircle at the bottom of each figure) from the right. Although background wind is not zero, the structures are still consistent with discussion above. The resulting GWs appear
behind the mountain and exhibit phase lines stationary relative to the ground. Relative to the wind, the phase of the GWs propagates to the right and into the background wind. Perpendicular to the phase direction, the group velocity can be seen by the growth of the disturbance vertically. As the wave disperses, the phase line stretches vertically indicating the vertical movement of energy.

![Figure 4.3](image)

**Figure 4.3.** Altered images of an experiment performed by the Geophysical Fluid Dynamics by the Graduate School of Human and Environmental Studies, Kyoto University. A mountain in a background wind is simulated by dragging a roller (effectively a “mountain”) through a tank of stratified fluid. For visualization purposes, the image is rotated such that the roller is now on the bottom. The roller is moving to the right such that a) and b) show sequential moments in time.

### 4.3 Inertio-Gravity Wave Characteristics Incorporating a Constant Background Wind

With frequencies comparable to the Coriolis parameter, IGWs have a more inclusive form of a dispersion relation which is given by:

\[
\tilde{\omega}^2 = \frac{N_B^2(k^2 + l^2) + f^2(m^2 + 1/4H^2)}{k^2 + l^2 + m^2 + 1/4H^2} \quad (4.14)
\]

Since the frequency shifts due to background wind have not yet been considered, the frequency described in the dispersion relationship above is the intrinsic frequency, \( \tilde{\omega} \). Due to the square, it is possible to have a positive or negative frequency which can be chosen based on convention.
Similar to the convention selected in Section 2.3.2, the sign for the intrinsic frequency is chosen to be negative such that the vertical group velocity is positive for upward propagation.

\[
\tilde{\omega} = -\frac{\sqrt{N_B^2(k^2 + l^2) + f^2(m^2 + 1/4H^2)}}{\sqrt{(k^2 + l^2 + m^2 + 1/4H^2)}} \tag{4.14}
\]

However, this is just the intrinsic frequency; the ground-relative or apparent frequency of the wave must incorporate the Doppler shift of the background wind such that \( \omega = \tilde{\omega} + \tilde{u}k + \tilde{v}l + \tilde{w}m \). Incorporating the Doppler shifted frequencies, the apparent group velocity would be:

\[
c_{gx} = \frac{\partial \omega}{\partial k} = \tilde{u} + \frac{k(N_B^2 - \tilde{\omega}^2)}{\tilde{\omega}(k^2 + l^2 + m^2 + 1/4H^2)} \tag{4.15a}
\]

\[
c_{gy} = \frac{\partial \omega}{\partial l} = \tilde{v} + \frac{l(N_B^2 - \tilde{\omega}^2)}{\tilde{\omega}(k^2 + l^2 + m^2 + 1/4H^2)} \tag{4.15b}
\]

\[
c_{gz} = \frac{\partial \omega}{\partial m} = \frac{-m(\tilde{\omega} - f^2)}{\tilde{\omega}(k^2 + l^2 + m^2 + 1/4H^2)} \tag{4.15c}
\]

### 4.4 Basic Prognostics of Gravity Waves

The propagation of GWs is heavily constrained by the background wind. In the previous subsections, the background wind was assumed to be constant and therefore was able to simply be incorporated into the phase speed and frequency equations. In a changing background wind, we can assume a GW has a vertical group velocity such that the changes in horizontal background wind can be ignored. Resultantly, the phase velocity of the GW will propagate in the zonal direction only such that the intrinsic phase speed is \( \hat{c}_h = \frac{\tilde{\omega}}{k} \).

\[
\frac{d^2 \tilde{w}}{dz^2} + \left( \frac{N_B^2}{\hat{c}_h^2} - k^2 \right) \tilde{w} = 0 \tag{4.16}
\]

When positive, the parenthetical term yields an oscillatory solution to this second order differential equation. The physical analogue to this equation can be a mass-spring oscillator with
the collective terms in parenthesis representing the spring’s stiffness. As the stiffness increases, the vertical wavelength would decrease. The limit of the “stiffness” as the background zonal wind speed approaches the phase speed of the wave is infinity. The layer where the phase speed matches the background wind speed is called the critical layer.

As a GW approaches its critical layer, the “stiffness of the oscillator” increases, causing the wavelength of the vertical velocity oscillations to decrease. As the vertical wavelength decreases, the vertical wind amplitude decreases while the horizontal wind amplitude increases until the vertical wind amplitude becomes essentially zero and the group velocity of the wave is imparted as a drag on the background wind. This behavior is illustrated schematically in Figure 4.4. Diminishing $\tilde{w}$ with altitude as the GW approaches its critical level agrees with Equation 4.16.

![Figure 4.4](image)

Figure 4.4. An illustration of GW propagation as it approaches the critical level ($z_c$). As a gravity wave approaches its critical level, (a) vertical wavenumber approaches zero, the $w'$ approaches zero, $u'$ approaches infinity, and (b) the vertical wavelength of the wave approaches zero. [Adapted from Nappo, 2002]

GWs of different horizontal phase speeds will have different critical layers. Therefore, GWs of certain phase speeds may be able to propagate higher due to the background wind configuration. Figure 4.5 provides examples of the tropospheric to lower stratospheric wind profiles for
scenarios when the tropospheric and stratospheric wind flows are in the opposite (Figure 4.5a) and the same (Figure 4.5b) direction. The blue and red wiggly lines represent respectively GWs of eastward (positive) and westward (negative) phase velocity ($\pm c$). Figure 4.5a illustrates GWs propagating upward and dissipating whenever the background wind profile matches the phase speed of the GW. In Figure 4.5a, both GWs dissipate but at different height levels since they have different critical surfaces. In Figure 4.5b, the GW with a westward phase velocity is filtered out while the GW with eastward phase velocity never reaches a critical surface. In this way, GWs can be filtered by phase speed given the background wind configuration.

Figure 4.5. Illustration of GW propagation constrained by zonal wind profiles of the upper troposphere and lower stratosphere. The phase speed, $c$, is ground-relative. [Adapted from Plumb, 1977 by Laing & Evans, 2011]

Several implications of GW filtering by the background wind have been studied including the forcing of the quasi-biennial oscillation (Holton & Lindzen, 1972). During the summertime, orographic GWs are generally expected to be contained in the tropopause as the stratospheric winds flow opposite to the tropospheric winds. During the wintertime, orographic GWs may propagate through stratosphere, eventually depositing in the upper mesosphere, as there is a persistent eastward flow with height. In the upper mesosphere, they can modulate the mesospheric circulation and temperature (McLandress et al., 2013). Additionally, GW filtering can modulate planetary waves by damping them (Miyahara et al., 1986), or exciting them.
Through asymmetric GW forcing (Holton, 1984). Asymmetric GW forcing is further discussed in Chapter 6.

While the full representation of GWD on the background wind can be derived through the Eliassen- Palm theorem (Eliassen & Palm, 1961), a simplified method for analyzing momentum deposition of a pure internal gravity wave is to consider the density-weighted eddy flux of the zonal wind in the vertical direction.

$$F(z) = -\rho_0 \bar{u}'w'$$  \hspace{1cm} (4.17)

This simplification assumes that the GW energy propagation is mainly directed in the vertical, so only the vertical component of the EP is considered. For a conservative, vertically-propagating GWs, the change in vertical eddy flux with height indicates the amount of momentum deposited into the background zonal wind. This flux divergence is expressed as

$$\rho_0^{-1} \nabla \cdot \bar{F} = -\rho_0^{-1} \frac{\partial}{\partial z} (\rho_0 \bar{u}'w')$$  \hspace{1cm} (4.18)

The scaling by the density reflects how a given amount of momentum will decelerate the winds more in a less dense layer of atmosphere. The flux divergence by GW would then act to accelerate the zonal background wind.

4.5 Gravity Wave Parameterization

GWs of small horizontal wavelengths, some as small as 10 km (Plougonven & Zhang, 2014), cannot be resolved explicitly in global climate models (GCMs) whose horizontal grid-point resolution is often too coarse. The impact of these GWs on the background wind must be parameterized. The parameterization of GWs assumes a conservative vertical propagation of GWs with a dampening effect on the background wind as GWs approach their critical levels. GWs of larger horizontal wavelengths, like IGWs, may not be adequately parameterized with these
assumptions. However, given their wide scale range, larger IGW may be resolved in higher resolution GCMs.

Since the position of the critical layer depends on GW phase speed, typical parameterization schemes control the GW breaking region by specifying the amount of GW flux emitted at a source as a function of phase speed. The pseudomomentum flux, $\rho u'w'$ can be expressed as a forcing in terms of the phase speed, $f(c)$, such that the total flux by a spectrum of waves emitted from a source is (Lindzen & Holton, 1968):

$$\rho u'w' = \int_{-\infty}^{\infty} f(c) dc$$ (4.19)

Assuming GW to be conservative such that the momentum flux is only deposited at the critical level, a simple GW parameterization can be created following Lindzen & Holton (1968). First, the phase speeds can be mapped onto their overlying zonal wind speed ($u$) such that $f(c) \rightarrow f(u(z_c))$. For an incremental change in $u$, the momentum flux applied would be $f(u) du$

$$\rho u'w' = \int_{-\infty}^{\infty} f(u) du$$ (4.20)

For a zonal wind that varies negligibly in time, it can be approximated to be only a function of height at a given location. As such,

$$-\frac{1}{\rho} \frac{\partial}{\partial z} (\rho u'w') = -\frac{1}{\rho} f(u) \frac{\partial u}{\partial z}$$ (4.21)

Applying Equation 4.18, the eastward acceleration, $X = \rho_0^{-1} \nabla \cdot \vec{F}$, then becomes:

$$X = -\rho_0^{-1} f(u) \frac{\partial u}{\partial z}$$ (4.22)

There are several caveats to this scheme. In nature, winds do change in time (as implied by the resultant acceleration). GW dissipation does not occur exclusively at the critical layer. While the wave cannot propagate past the critical layer, the point at which the wave begins to break or
dissipate could begin well before the critical layer is reached. Lindzen (1981) theorized that the wave breaking initiates whenever the GW-induced temperature perturbation becomes thermodynamically unstable. Based on this theory, Lindzen (1981) and Holton (1982) describe a method of parameterization in which a GW induces a drag on the background wind starting at the breaking level and continuing until the wave is completely dissipated at the critical layer.

Figure 4.6 shows GWD for various GW sources parametrized in a GCM called the Whole Atmosphere Community Climate Model (WACCM), the same GCM that will be used in this study. For WACCM’s parameterization orographic GWs, a momentum flux phase speed spectrum is not needed since the apparent phase speed zero. However, for convectively generated gravity waves, the spectrum can be asymmetric depending on the tropospheric winds and the strength of the spectrum varies primarily with the convective heating rate (Beres et al., 2004). For GWs generated by weather fronts, the source is specified whenever the frontogenesis function (e.g., Charron & Manzini, 2002) at 600 mb exceeds the threshold of $0.045 \text{ K}^2(100 \text{ km})^{-1} \text{ h}^{-1}$ (Richter et al., 2010). When the frontogenesis threshold is identified in the WACCM model, a Gaussian momentum flux phase speed spectrum centered around the wind at the source level is launched such that Gaussian curve has a half-width of $30 \text{ m s}^{-1}$. As illustrated above, while GWD due to convection is mostly contained in the low latitudes (see Figure 4.6b), frontal and orographic GWD are dominant the mid- to high latitude upper stratosphere and lower mesosphere (Figures 4.6a,c).
In WACCM, much of the parameterized GWD occurs in the middle and upper atmosphere. Therefore, the interaction of GWs with the background wind will be significant in the present study, particularly since the induced acceleration will increase with the exponential decrease of air density with height.
Chapter 5

Major Sudden Stratospheric Warming

Wintertime stratospheric conditions can strongly influence the physical and chemical processes of the middle atmosphere. The cold and isolated air mass inside the polar vortex promotes the necessary conditions for chemical interactions that can catalytically deplete stratospheric ozone; an extreme consequence of which is the major ozone hole observed over the South Pole (Farman et al., 1985). Dynamically, the prevailing eastward stratospheric flow exposes the stratosphere to upward PW disturbances that would otherwise be confined in the troposphere (Charney & Drazin, 1961) while allowing westward propagating GWs to reach the mesosphere. These disturbances can perturb and alter the shape of the polar vortex.

In the extreme case, strong wave perturbations can lead to a polar vortex breakdown event with anomalous warming in the stratosphere before late spring. This event is known as a major sudden stratospheric warming (SSW). A major SSW onset results in (1) the rapid warming of the polar night region on the order of tens of degrees and (2) the reversal of the typical eastward zonal-mean polar winds to the westward direction. The polar warming is accompanied by enhanced downwelling in the polar stratosphere (McLandress et al., 2013) along with the descent of the stratopause (Limpasuvan et al., 2012). During a major SSW, the polar vortex can split or become highly displaced. When the vortex is weakly perturbed, a minor SSW can occur resulting in polar warming and only weakened zonal eastward wind. As the present study does not focus on minor SSWs, major SSWs will hereafter be simply referred to as SSWs.

5.1 Background Winds Associated with Split and Displaced SSWs

The displaced SSW coincides with the dominant presence of a zonal wavenumber-1 PW (or PW1) and a split SSW coincides with a zonal wavenumber-2 PW (or PW2). Figure 5.1 illustrates the evolution of the polar vortex through a displaced SSW in January 1987 (Figures 5.1a-c) and a split SSW in January 2009 (Figures 5.1d-f). Regions of low and high geopotential height regions indicate the position of low- and high-pressure systems, respectively. During the January 1987
SSW, a PW1 feature appeared as a high-pressure system migrated poleward and displaced the low-pressure center typically associated with the polar vortex. During the January 2009 SSW, a PW2 feature occurred as the peripheral high-pressure systems pinched toward the pole, dividing the low-pressure system onto either side. High-pressure systems have anti-cyclonic circulations with associated warmer air near the center. Their appearance in the polar region can increase polar temperature drastically in a matter of a few days. This is exemplified during the January 1987 and 2009 SSWs by a dramatic increase in temperature of more than 40 K (c.f., Figures 5.1b,c and Figures 5.1e,f).

Figure 5.1. Polar plots of stratospheric temperature (filled contours) and geopotential height (thick black contours) at 31 hPa (~25 km). Geopotential height is incremented by 0.25 km up to 23.5 km. Three plots illustrate state of the polar vortex before, during, and after onset for (a-c) the January 1987 displaced SSW and (d-f) the January 2009 split SSW. Thin contours outline the continents. Data produced from the WACCM-SD model run described in Section 7.2.

Apart from the wavenumber-1 and wavenumber-2 eddy features distinguishing split and displaced SSWs, the zonal-mean zonal wind exhibits different strengths prior to vortex
breakdown (see Figure 5.2). During a displaced SSW, the composite jet structure is relatively weaker in the stratosphere, with a distinct core in the midlatitude mesosphere. A weak stratospheric vortex is more susceptible to tropospheric forcings and can be easily displaced off the pole. This suggests that displaced SSWs simply undergo a gradual weakening of zonal wind until reversal (e.g., Albers & Birner, 2014). On the other hand, the composite wind structure leading up to the split SSW onset exhibits a strengthened stratospheric vortex. This stark contrast between split and displacement SSWs suggest that there are key differences in the jet structure and dynamics leading up to split and displaced SSWs.
Figure 5.2. $\bar{u}$ anomaly from normal climatology (a & c) 10-20 days and (b & d) 5-10 days prior to SSW onset. Analysis was performed on Japanese Meteorological Agency and Central Research Institute of Electrical Power Industry 25-year Reanalysis data from 1980-2011. Normal climatology is created from a composite of 14 winters without SSWs. Split and displaced SSWs are identified and classified by studies mentioned within Albers & Birner (2014). [Albers & Birner, 2014]
Stratospheric conditions associated with SSWs can extend to other parts of the atmosphere. The anomalously weak zonal winds and warming can descend toward the tropopause and project onto the dominant mode of climate variability called the Northern Hemisphere Annular Mode or NAM (Baldwin & Dunkerton, 2001). Defined by the leading Empirical Orthogonal Function (EOF) mode of sea level pressure poleward of $10^\circ N$, NAM explains over 20 percent of the wintertime sea-level pressure variance and describes the coupling between the tropospheric jet stream (and related surface weather) with the polar vortex through wave activities (Limpasuvan & Hartmann, 1999; Thompson & Wallace, 1999). Changes induced by SSWs bias the NAM to its negative phase during which with the tropospheric jet and the storm shift equatorward (Limpasuvan & Hartmann, 2000). The negative NAM phase results in cold low-pressure systems over Russia and North America and warm high-pressure systems over the North Atlantic and North Pacific (Thompson & Wallace, 2001) and correlates to cold weather outbreaks over Asia and Europe (Kolstad et al., 2010).

The evolution of the zonal-mean zonal wind illustrates the downward SSW influence (Figure 5.3). Marked by the zero-wind line, the wind reversal appeared around 50 km on 22 January 2019 then descended downward. Upon reaching the surface about a week thereafter, the wind reversal projected onto the negative NAM phase with an equatorward shift of the tropospheric jet. SSWs were specifically implicated in exacerbating the 1962-1963 and 2009-2010 surface wintertime conditions (Fereday et al., 2012; Greatbatch et al., 2015).
**Figure 5.3.** Height-time plot of $\bar{u}$ averaged from 70-90°N surrounding the 22 January 2009 split SSW. Grey-shaded regions indicate westward flow outlined by the zero-wind line (bold black line). Westward (blue) and eastward (red) zonal-mean flow are contoured every 10 m s$^{-1}$. Major and minor tick marks along the x-axis indicate 10-day and 1-day intervals, respectively.

Particular tropospheric configurations may create ideal conditions for SSW occurrence. Quiroz (1986) suggested that the tropospheric blocking phenomenon could precede SSW. Appearing in the middle to upper troposphere, blocking describes the formation of anticyclonic high-pressure systems that persist for days or weeks; its presence locally reverses the zonal wind from eastward to westward. Tropospheric blocking is not a necessary condition for SSW occurrence. However, SSWs coupled with blocking are preceded by larger poleward heat fluxes at 200 hPa (Colucci & Kelleher, 2015), indicative of increased upward PW activity that may provide the initial forcing of the polar vortex breakdown (Polvani & Waugh, 2004). Woollings et al. (2010) noted that displaced SSWs are associated with the increased frequency of tropospheric blocking events over Europe while blocking over the North Atlantic and North Pacific are associated with split SSWs. The development of these tropospheric blocking patterns could be due to changes in the sea surface temperature through the El Niño Southern Oscillation (ENSO), the Madden Julian Oscillation, or sea ice fluctuation (e.g., Barriopedro & Calvo, 2014; Gollan & Greatbatch, 2017; Peings, 2019).
The impact of SSWs also extends above the stratosphere. While the zonal-mean zonal wind remains anomalously westward, the polar mesosphere experiences an abnormal cooling on the order of tens of degrees. This cooling coincides with an unusual polar upwelling or weakening of the climatological downwelling associated with the mean meridional circulation (Limpasuvan et al., 2016). As the polar circulation recovers from SSW, an enhanced polar downwelling initiates above 80 km and transports chemical species abundant in the mesosphere/lower thermosphere (MLT) region down into the polar stratosphere (e.g., Lee et al., 2009; Kvissel et al., 2012). Energetic particle precipitation related naturally to solar activity commonly produces a family of NO$_x$ (consisting of N, NO, and NO$_2$) in the MLT region. Known to be long-lived catalysts for ozone destruction, NO$_x$ concentration increases in the stratosphere by the strong SSW-induced downwelling (Randall et al., 2009). At certain SSW events, the enhanced downwelling can give rise to the reformation of a stratopause above its climatological position. Those SSW events accompanied by an elevated stratopause are referred to as ES-SSWs (Orsolini et al., 2010). The seasonality of SSWs is also of importance; occurrences of SSWs early in the winter season tend to result in a larger injection of NO$_x$ into the stratosphere (Holt et al., 2013). Additionally, the distribution of polar ozone in the mesosphere and MLT is affected by the anomalous vertical motion during ES-SSWs (Tweedy et al., 2013).

### 5.2 Planetary and Gravity Wave Forcings during ES-SSWs

The vortex breakdown during SSW is associated with wave impact on the vortex. Matsuno (1971) proposed that transient behavior through the convergence of quasi-stationary PW activity in mid- to upper-stratosphere can strongly decelerate the background polar wind and induce a mean meridional circulation with enhanced polar downwelling that adiabatically warms the polar region. To date, this model appears to hold true as PW forcing tends to maximize near SSW onset, as illustrated in **Figure 5.4a** for the ES-SSW composite. Confluence of the EP flux vectors leads to an EP flux convergence with decelerative effects and a meridional flux of potential vorticity.
Prior to onset, PW1 and PW2 activity became more vertically oriented with the poleward shift of the EP Flux convergence region, coinciding with the increased net poleward heat flux at 50°N, as suggested in Figure 5.5 (albeit for a specified latitude and a limited altitude range). By analyzing the index of refraction for PWs, Albers & Birner (2014) showed that a waveguide/“leaky” cavity forms in the background conditions prior to SSW that could confine PW2s to ~10 degrees poleward of 50°N. During the SSW, the continual PW dissipation near the zero-wind line led to wind reversal lower in the stratosphere (illustrated by the green line in Figure 5.4).

**Figure 5.4.** Altitude-time evolution of composite ES-SSW for (a) PW forcing, and (b) GW forcing, averaged between 40°N and 80°N. The zero-wind (green) line shows the zonal-mean wind reversal from eastward to westward in the stratosphere. Red (blue) contours indicated eastward (westward) acceleration in m s⁻¹ day⁻¹. Day 0 represents SSW onset. [from Limpasuvan et al., 2016]
Figure 5.5. Vertical and horizontal components of anomalous EP flux integrated poleward of 50° N for all and specified wavenumbers during the composite life cycle of SSWs. Contribution from combined wavenumbers greater than or equal to 4 is denoted by “wave 4+.” The sum of “wave 1,” “wave 2,” “wave 3,” and “wave 4+” produces the result for “all.” Contours of the vertical component and horizontal component are given every $10 \times 10^4 \text{ kg s}^2$ and $20 \times 10^6 \text{ kg s}^2$, respectively. Negative contours are given as dashes. Zero contours are given as a bold solid line. Dark gray shading indicates areas with a 95% confidence level (based on $t$ statistics). Composite calculated with daily-averaged data between 1958 and 2001 from the National Centers for Environmental Prediction – National Center for Atmospheric Research reanalysis project (Kalnay et al., 1996). [Limpasuvan et al., 2004]

Recent studies identified the presence of strong PW forcing in the mesosphere and above after SSW onset (e.g., Limpasuvan et al., 2012; Orsolini et al., 2018). The wind reversal in the polar stratosphere has been suggested to foster the in-situ generation of PWs by instability (Limpasuvan et al., 2012; Tomikawa et al., 2012). Propagating upward from the stratosphere into the MLT, these PWs appear to be westward propagating (Chandran et al., 2013; Limpasuvan et
al., 2016). Above 80 km, the EP flux convergence of these PWs after SSW onset may help promote the polar vortex recovery by initially driving polar downwelling and reforming the stratopause at an elevated altitude.

Given the importance of GWs in the general circulation, the roles of GWs during ES-SSWs are also evident. As shown in Figure 5.6, zonal momentum drag by orographic GWs was significant near the polar night jet, poleward of 45°N and above 30 km before SSW onset (Albers & Birner, 2014). On the other hand, non-orographic GWs could likewise contribute to the wave forcing in this region (McLandress & Scinocca, 2005; Orr et al., 2010). Figure 5.4b shows the zonally averaged GWD in the polar mesosphere during ES-SSW. Based on their model simulations, Limpasuvan et al. (2016) suggests that these GWs were generated by weather fronts. Typically, when the stratospheric and tropospheric winds are eastward, westward GW momentum is deposited around 60-80 km (as seen well before and after SSW onset). In addition to closing off the polar night jet, this westward GWD induces polar downwelling, and the associated adiabatic warming (e.g., Garcia & Boville, 1994) and helps maintain a warm stratopause (Hitchman et al., 1989). With SSW onset, PW forcing becomes dominant and causes the stratopause to descend. The stratospheric wind reversal filters out westward GWs while allowing eastward GWs to reach the mesosphere. Consequently, the GWD in Figure 5.4b exhibits a transition to eastward forcing; the polar downwelling was reduced, and possibly reversed, resulting in mesospheric cooling (Zülicke & Becker, 2013). As the vortex recovers, westward GW forcing returned. Overall, GWs have a dampening effect on the upper atmospheric dynamics, making their interplay with PWs important to SSW evolution.
While models parameterize unresolved GWs as strictly propagating vertically, GWs can deflect toward jet stream cores (Senf & Achatz, 2011). Planetary-scale perturbations contributing to the jet stream meandering could slightly alter GW propagation. However, the idealized study by Senf & Achatz (2011) shows that, while GW refraction can influence the strength of the background wind, it does not play a dominant role in configuring the wind structure. Given the scale disparity between PWs and GWs, wind oscillations associated with PWs are more likely to provide a filtering effect on GWs, controlling the altitudes and longitudes of GWs momentum deposition (e.g., Smith, 2003). On the other hand, GWs may affect the characteristics of PWs. Asymmetric GWD in the mesosphere could generate or amplify PWs, exacerbating the warming event (e.g., Lieberman et al., 2013). The interplay of GWs and PWs in the mesosphere prior to SSWs will be further discussed in Chapter 8.
Chapter 6

Motivation and Research Objectives

Despite over 60 years of literature on SSWs (Butler et al., 2015), our understanding of the dynamical activities surrounding major SSWs is still lacking. Based on the idea initially put forth by Matsuno (1971), we are aware of the concomitant forcing of quasi-stationary PWs and stratospheric wind reversal associated with the sudden polar warming. However, reasons why PW activities become focused in the polar stratosphere and the roles other dynamical features (like traveling PWs, GWs, and the background flow) are unclear. One impediment to our understanding is the range of temporal and spatial scales that must be considered in addressing events surrounding SSWs. Background flow aside, GWs and PWs may interact in the middle atmosphere while, separately, they can strongly impact the circulation over a deep layer when dampened. With advancements of model simulations and observations, new light is being shed on these dynamic features that may advance our understanding of SSWs.

Relatively recent studies identified a precursory mesospheric signature prior to the onset of the 2009 split ES-SSW (Coy et al., 2011; Iida et al., 2014). As shown previously in Figure 5.3, the zonal-mean zonal wind in the polar middle mesosphere switched direction from eastward to westward about 5 days before the stratospheric wind reversal. The reversed mesospheric wind emerged as an extension of the westward wind typically found in the thermosphere. Concurrent with this mesospheric wind anomaly was the appearance of a slow eastward-propagating wavenumber-2 eddy pattern. At 0.01 hPa (~80 km) and 63.5°N, the Hovmöller diagram of height anomalies observed by Iida et al. (2014) clearly illustrate this pattern (see thick broken lines on the top panel of Figure 6.1) prior to SSW onset around 22 January 2009.
Figure 6.1. Hovmöller diagram of geopotential height anomalies for zonal wavenumbers 1-6 along 62.5°N at (top) 0.01 hPa and (bottom) 1 hPa. The contour interval is 200 m. The thick broken lines in the top panel denote an eastward traveling feature in the mesosphere before SSW onset. [Adapted from Iida et al., 2014]

The eastward-moving eddy pattern may be attributed to PWs. Figure 6.2 illustrates the vertical component of the wavenumber 2 EP flux at 60°N during the 2009 split SSW. Before onset, several upward bursts of PW2 activity occurred and dissipated in the stratosphere. As SSW onset drew closer, the wave activity penetrated higher across the stratopause, with incipient wave dissipation in the mesosphere ~6 days prior to the onset date noted by the vertical magenta line. This dissipation potentially induced sufficient westward acceleration to alter the upper mesospheric wind direction well before the stratospheric wind reversal. Consistent with this dissipation, Coy et al. (2011) identified a critical surface in the mesospheric wind that corresponded with a PW2 with eastward phase speed less than 20 m s⁻¹, henceforth referred to as EPWs. Although wave activities shown in Figure 6.2 can originate from other latitudes, these authors suggested that EPWs in the mesosphere originated from the troposphere.
The direct vertical propagation of EPWs is not as clear-cut as implied by Figure 6.2. The observed stratospheric height anomalies near 60°N (bottom panel of Figure 6.1) suggests the appearance of quasi-stationary PW signatures of wavenumbers 1-2 around the same timeframe as EPWs. The varying wave characteristics across the stratopause is highlighted by the altitude-longitude cross-sections shown in Figure 6.3. In particular, the vertical phase tilt changed slightly around 0.1 hPa, suggesting that the stratospheric PWs below the zero-wind line were distinct from the mesospheric EPWs. The similarity in wavenumber and slight phase offset of the perturbations implies that the waves were correlated. While the presence of both EPWs and quasi-stationary PWs are possible in eastward background flow, the descending westward wind seen in the vertical profile prevented the slow mesospheric EPW from being long-lived as evident in Figure 6.1 (top panel). The westward tilt of the stratospheric Rossby waves indicates an upward energy propagation. On 20 January, the EPW displayed a slightly eastward tilt suggesting the group velocity of the wave could be directed downward.
Figure 6.3. Geopotential height perturbations at 60°N contoured every 0.16 km during (a) 12, (b) 16, (c) 18, and (d) 20 January 2009. Black horizontal lines denote altitude level of the zero zonal-mean zonal wind. Corresponding vertical profiles of zonal-mean zonal wind (m s$^{-1}$) and stability (10$^{-4}$ s$^{-2}$) are shown to the right of each panel. [Coy et al. 2011]

The meridional structure of the EP flux and background wind leading up to SSW onset are shown in Figure 6.4. Consistent with the vertical wind profile at 60°N seen in Figure 6.2, the westward wind appeared first above 80 km and descended into the stratosphere while spreading toward the tropics. Regions of large EP flux convergence in the mid-latitudes near 50 km and 80 km decelerated the eastward wind. The associated wave forcing weakened the polar vortex, leading to SSW. Peaking on 16 January, the EP flux convergence region above 60 km decelerated the upper mesospheric wind and may be connected with the underlying PW activity, like EPWs noted in Figures 6.1 and 6.3. The flux divergence (orange shading) near the 40 m s$^{-1}$ isopleth suggested a wave source in the upper region of the polar stratospheric jet.
Figure 6.4. Meridional cross sections zonal-mean zonal winds (line contours), Eliassen-Palm (EP) flux (vectors), and EP flux divergence (color-filled contours) of zonal wavenumbers 1-3 in the Northern Hemisphere. The arrow scale denotes $2 \times 10^6$ and $2 \times 10^8 \text{kg s}^{-2}$ for vertical and horizontal components of the EP flux, respectively, in the region of 100-10hPa. Vectors are magnified by a factor of $5n$ for the $1^{\text{st}}$, $2^{\text{nd}}$, ..., $n^{\text{th}}$ layers between $10^{2-n}$ and $10^{1-n}$ hPa. The shading denotes zonal wind deceleration; see the tone bar where the units are m s$^{-1}$ day$^{-1}$. The contour interval of zonal-mean zonal winds is 10 ms$^{-1}$. [Adapted from Iida et al., 2014]

Iida et al. (2014) noted that the source for the mesospheric EPWs could be attributed to instability of the background wind as its zonal-mean structure developed strong shear. As noted in Section 3.6.1, the region of negative meridional gradient of QGPV ($\dot{q}_y < 0$) is a necessary condition for instability. The presence of the wave’s critical surface inside this region serves as a source from which wave activity emanates. Figure 6.5 illustrates the meridional cross-sections of $\dot{q}_y$. 
corresponding to the wind structure in Figure 6.4. The potential region of instability is shown in cyan.

**Figure 6.5.** QGPV meridional gradient prior to the January 2009 split SSW. The cyan areas (with dashed contours) highlight the negative values regions, fulfilling the necessary conditions for shear instability. [from Iida et. al, 2014]

Well before SSW onset (13-16 January), a pocket satisfying the necessary instability criterion appeared above the stratospheric jet core (compare Figure 6.4 and 3.5) as the eastward winds weakened rapidly with altitude. Seen in Equation 3.21, $\bar{q}_y$ is highly dependent on the zonal wind curvature terms. As such, the poleward side of the subtropical mesospheric jet and the top side of the stratospheric jet are areas susceptible to instability. The near collocation of this instability pocket with the enhanced EP flux divergence region pointed to the possible role of jet instability
as the EPW source. With the polar mesospheric wind reversal, the instability pocket descended into stratosphere along with the zero-wind line.

Before SSW onset, the juxtaposed eastward jet structure (with one core in the mid-latitude stratosphere and another in the subtropical mesosphere) was potentially setup by GWD. As discussed in Chapter 4, GW amplitude can become large at these altitudes, break, and impart westward momentum on the background flow. The resulting drag on the wind would typically close off the top side of the stratospheric jet and may create strong meridional and vertical wind shear as the drag occurs at varying latitudes and altitudes. The wind structure could then become barotropically and/or baroclinically unstable leading to a negative $\vec{q}_y$ region like that shown in Figure 6.5. Instability of similar jet structure had been shown to generate PWs such as the 4-day wave, prominent in the winter stratopause (e.g., Manney & Randel, 1993; Orsolini & Simon, 1995).

In addition to framing the background flow, GWs may also induce planetary-scale perturbations in the MLT region as they are preferentially filtered by stratospheric wind (Smith, 1997; 2003). Lieberman et al. (2013) showed that the wintertime characteristics of MLT perturbations were qualitatively consistent with a simple model of MLT wavenumber 1 PW generated by dissipating GWs that had been filtered through the underlying stratospheric PW perturbations. Figure 6.6 demonstrates this model.
Figure 6.6. (top) Ageostrophic wind in the latitude-longitude plane induced by breaking of mesospheric GWs whose vertical transmission has been filtered by the stratospheric circulation (bottom). Stratospheric wavenumber-1 PW geopotential height and gradient winds on latitude-longitude plane. [from Lieberman et al., 2013]

Using the primitive momentum equations (Equations 2.1a and 2.1b) under geostrophic assumption for a flow with zero Lagrangian acceleration, the ageostrophic wind components can be expressed in terms of the external forcing components ($F_x$ and $F_y$) due to GWD:

$$-fv_{ag} = F_x$$  \hspace{1cm} (6.1a)

$$fu_{ag} = F_y$$  \hspace{1cm} (6.1b)

Figure 6.6 shows the effect of GW forcing in the MLT (top row) due to filtering by the underlying stratospheric high-low pressure system (bottom row). For example, the eastward stratospheric flow would allow the vertical propagation of westward GWs. Upon breaking in the mesosphere, the resultant westward forcing ($F_x < 0$) on the background wind by westward GWs would drive a northward ageostrophic wind ($v_{ag} > 0$) like that seen in the top row of the diagram and described by Equation 6.1a. Ultimately, the GW-induced ageostrophic winds would result in a
divergence (convergence) pattern over stratospheric high-pressure (low-pressure) region. In the MLT, the divergence pattern would produce a low-pressure region and the convergence pattern a high-pressure region. Subsequently, PW perturbations would be imprinted in the MLT by flow divergence induced by GWs that survived the wind perturbations in the stratosphere.

EPWs in the mesosphere may likewise be tied to GW filtering by the underlying PW signatures. Interestingly, McLandress & McFarlane (1993) noted that filtering of orographic GWs may have a larger overall impact than non-orographic GWs on mesospheric winds. Consequently, they could play a larger role in inducing secondary PWs (McLandress et al., 2013).

Song et al. (2020) suggested that EPWs that occur prior to the January 2009 SSW could be a result of GWD in the mesosphere. GW-induced vorticity in the atmosphere can cause amplitude growth through the third term in Equations 3.30 and 3.32. Figure 6.7 shows EP flux vectors for wavenumber-2 EPWs on 17 January 2009. In-situ wave growth is suggested by a downward-propagating EPW from a region of flux divergence occurs between 45-55km, shown by the magnitudes and directions of black EP flux vectors. The region of suggested wave growth coincides with regions of large GWD. While Figure 6.7 suggests that GWD is the source of EPW wave growth, zero isopleths of $n^2$ and $\bar{q}_\phi$ are in the same vicinity. With the appropriate wave geometry, EPWs could also result from over-reflection, or a mixture of multiple source mechanisms.
To date, reasons for the appearance of EPWs before SSW onset are uncertain. In briefly describing the possible mechanisms that may lead to mesospheric EPWs, we highlight the complex interplay between quasi-stationary PWs propagating from the troposphere, waves generated by shear instability, and GWs beyond the dynamics already discussed in Chapter 6. The growth and dissipation of these waves can affect the circulation and momentum in the mesosphere as well as the stratosphere. Since PWs, GWs, and the background wind structure can modulate each other, all three must be considered to fully characterize mesospheric PWs prior to SSWs.
Given the three methods explaining PW appearance in the middle atmosphere (direct propagation, GW filtering, and instability), these concepts can be extended and applied to unique PWs in the middle atmosphere during and after SSW onset. Limpasuvan et al. (2016) addresses the role of westward-propagating PWs in the mesosphere after SSW onset. Here, WPWs are identified as instability waves that aid in the recovery of the stratopause.

Without loss of generality, we limit our scope to years with ES-SSWs. Since ES-SSWs are defined by how the vortex recovers, we do not expect precursory mechanisms to be different between SSWs without an elevated stratopause and ES-SSWs. However, if a difference is found in future studies, the present study will still be valid for ES-SSWs in particular. Additionally, composites are expected to be similar as only 2 SSWs identified in Table D1 lack an elevated stratopause.

Since SSW can potentially impact surface meteorological conditions, results from this study may help improve our climate predictive skills and provide new insights on the dynamical coupling between the stratosphere and mesosphere. In supporting this goal, our research objectives are to:

1. Identify the source of EPWs before SSW onset,
2. Assess the impact and uniqueness of mesospheric EPWs leading up to SSW onset, and
3. Explore the behaviors of EPWs along with other PWs surrounding SSW and determine their possible source mechanisms.

Questions relevant to these objectives include: Are EPWs regular and/or unique features of SSWs? What mechanism(s) may cause wave growth prior to SSW events? How do EPWs directly, and possibly indirectly via GW filtering, alter the momentum budget in the mesosphere? Are the appearances of EPWs mechanistically related to other mesospheric features such after SSW onset such as the mesospheric westward-propagating waves?

This thesis addresses these objectives using satellite observations, available model simulations nudged with observations, and numerical experiments. The observations and model simulations
will help identify the origin, characteristics, and forcings of EPWs prior to SSW onset to illuminate their role in SSW development and during wintertime climatology (devoid of SSWs). To fully understand the robust behavior of EPWs and the precursory mesospheric dynamics, a composite of the observed events and associated dynamics was performed. Since SSWs only occur roughly every other winter, there is a limited number of SSW events to use for a composite analysis. By performing an ensemble study, it was possible generate more samples of SSWs. The case studies, composite, and ensemble experiments were performed using a global chemistry-climate model with an extensive altitude range. Details of research methodology are provided in the next chapter.
Chapter 7

Data and Methods

This chapter provides an overview of satellite observations and models used to address our research objectives. It also details the proposed methodologies to digest the observations, to set up the model, and to identify, classify and analyze SSWs.

7.1 Satellite Data

Satellite data from NASA will be used to validate our model. Aboard the polar-orbiting Earth Observing System (EOS) Aura satellite launched in 2004, Microwave Limb Sounder (MLS) measures along-track atmospheric composition, e.g. radiances near $O_2$ spectral lines. Although the data is given on 55 vertical levels between $10^3$ hPa and $10^{-5}$ hPa, the useful range for the geopotential height and temperature data products is between 261 and $10^{-3}$ hPa (~9.4 km – 96.7 km). The EOS-Aura satellite orbits the earth ~15 times per day giving daily global coverage (Livesey et al., 2017). Geopotential height outputs are deduced from the 118-GHz (2.54 mm wavelength) and 234-GHz (1.28 mm wavelength) $O_2$ spectral lines while temperature outputs are deduced from 118-GHz and 239-GHz (1.23 mm wavelength) $O_2$ spectral lines (Livesey et al., 2017).

Along-track satellite profiles spanning 24-hour intervals were used to construct daily maps. The data was first gridded using a Delaunay triangulation method for a spherical surface onto a regular grid. Any satellite recordings at the same location during different times, possibly due to the crossing of the ascending and descending tracks, were averaged. A Kriging interpolation was performed on the satellite data (e.g., Reese, 2005; Coakley et al., 2008;), in which the distance-weighted mean-squared error was minimized among surrounding points (Krige, 1951). For interpolation in the present study, the mean-squared error was exponentially weighted by $e^{-3d/16}$ where $d$ is the distance between the point being estimated and surrounding points in units of the average spacing between the points on a sphere. For efficiency, the algorithm was restricted to a search radius of 32 times the average distance between points. The EOS Aura
satellite orbits the earth ~14.6 times a day which allowed the zonally averaged calculations to also average over the daily cycles in solar radiation that fluctuate the temperature and chemical species abundance. Since the satellite takes instantaneous measurements, values recorded along ascending and descending tracks may fluctuate with the daily cycle.

Each day, the satellite’s sinusoidal tracks (when mapped on a Mercator projection) over the earth’s surface gain a ~157° phase lead, resulting in the satellite repeating its cycle every 16 days. Ideally, a 16-day running average would have been used to account for daily variations, but this is larger than the time scale of EPWs. This large phase offset, however, allowed a 9-day running average to sufficiently account for variation in daily cycles while preserving the global-scale vortex perturbations. An example of the interpolated results is shown in Figure 7.1. The features compare favorably with output of model showing the correct magnitudes and positionings of regions of low and high amplitude. Additionally, the model accurately simulates the polar temperature increase associated with the splitting polar vortex. Notably, the 9-day averaged variables needed to interpolate observations result in a weaker temperature increase than shown in model data. Further comparisons between the model and MLS are discussed in Section 7.2 and Section 8.7.1.
**Figure 7.1.** Polar stereographic plots at 31 hPa of geopotential height in km (bold black contours) and temperature in Kelvin (color-filled contours) from the WACCM-SD model (top) and MLS (bottom) data. Geopotential height contours is incremented every 0.2 km. A 9-day running average was performed on the MLS data. Underlying landmasses are outlined by thin black lines. Dates for each plot are before (left) and after (right) SSW onset on 22 January 2009.

### 7.2 The Model

The study will utilize the Whole Atmosphere Community Climate Model, Version 4 (WACCM) developed at the National Center for Atmospheric Research (NCAR). As part of the Community Earth System Model Version 1.2 (CESM1), WACCM is an atmosphere-only global chemistry-climate model that extends up to ~145 km ($5.1 \times 10^{-6}$ hPa). Details of WACCM are provided by Marsh et al. (2013). More recent additions include convective and frontal non-orographic parameterizations of GWs as well as mountain stresses which have improved the frequency of SSWs in the Northern Hemisphere (Richter et al., 2010).
In developing a model dataset representative of 1978 to 2013, WACCM was run in the specified dynamics configuration. Referred to as WACCM-SD, this configuration has a horizontal resolution of 0.95° latitude by 1.25° longitude and 88 vertical levels and key dynamical variables output daily. Up to 50 km (~0.79 hPa), the model’s temperature and dynamics are constrained with six-hourly Modern-Era Retrospective Analysis for Research and Application (MERRA) Version 2 reanalysis (Rienecker et al., 2011). Above 60 km (~0.19 hPa), the model is fully interactive and free-running. From 50 km to 60 km (~0.79 hPa to ~0.19 hPa), a linear transition is applied between the nudged output below and the overlying free-running region.

As shown in other studies (e.g. Chandran et al., 2013; Tweedy et al., 2013; De Wit et al., 2014), WACCM-SD sufficiently models tracer responses above 0.79 hPa and mimics the observed dynamics in the MLT region reasonably well, especially during SSWs. Figure 7.2 further supports the realistic response of WACCM-SD when compared to satellite observations. The polar averaged zonal-mean temperature in model and observations show a similar stratopause descent and elevated stratopause behavior during the January 2013 SSW.
Discrepancies in trace species distributions between WACCM-SD and observations have been attributed to unrealistically weak GW driving (parameterized and resolved) and, consequently, weak mean meridional circulation in the model (Smith et al., 2011; Randall et al., 2015). Regardless, the dynamics in WACCM-SD should be sufficient to support our study objectives and can be compared with satellite data.

We will analyze the available WACCM-SD to examine the impacts of EPWs during SSW events between 1978 and 2013. However, over this period, only 15 displaced and 6 split SSWs were identified, at least according to Albers & Birner (2014). Having only a few SSWs may lead to results
with large uncertainties. To gather more robust results, the nudged WACCM-SD output may serve as the reference run for ensembles, further discussed in Section 7.4.

### 7.3 SSW Identification

In analyzing satellite and model data, we must identify and classify SSWs. SSW identification has been traditionally based on the World Meteorological Organization (WMO) definition, dating back to 1952 when the first SSW was initially observed (Scherhag, 1952). This definition identifies a (major) SSW as occurring when the 10-hPa zonal-mean zonal wind reverses as a result of the switched meridional temperature gradient in the polar region. A minor warming occurs if the temperature changes in the stratosphere by 25 degrees Celsius within a period of a week or less but does not cause a zonal-mean wind reversal at 10 hPa (Mchturff, 1978). However, over the years, variation of this definition has crept into practice and SSW identification is currently under debate (Butler et al., 2015).

Studies focused on the coupling between SSW and the MLT region may specialize to major SSW events in which the vortex recovery results in a stratopause elevated above its climatological altitude position. These specialized SSWs with elevated stratopause are referred to as ES-SSWs (Limpasuvan et al., 2016). Based on the zonal-mean zonal wind and temperature averaged between 70°N and 90°N during the extended winter, Limpasuvan et al. (Limpasuvan et al., 2016) identified an ES-SSW event if: (1) the temperature falls below 190 K between 80-100 km, (2) the zonal-mean zonal wind reverses from eastward to westward at 1 hPa and persists longer than 5 days, and (3) the stratopause altitude based on the zonal-mean temperature maximum between 20-100 km (e.g., Tweedy et al., 2013) exhibits a vertical discontinuity of at least 10 km. The criterion accounts for the occurrence of the mesospheric anomalies in conjunction with an elevated stratopause, ignored by the WMO definition.

Both the WMO definition of SSWs and the criteria for ES-SSWs by Limpasuvan et al. (Limpasuvan et al., 2016) will be applied to identify suitable warming cases. As our region of interest includes the mesosphere, we will define the onset date according to the polar averaged zonal-mean zonal
wind reversal at 1 hPa. Using 10 hPa would not be suitable as wind-reversals descend at different rates, which is evident in comparing the zero wind lines of Figures 7.3, 7.4 and 7.5. The tilt of the zero-wind line shows that the downward propagation of the 23 February 1999 wind-reversal is slower than that of the 6 December 1987 by \( \sim 2 \) days. The criteria of Limpasuvan et al. (Limpasuvan et al., 2016) are more rigorous. As such, we expect most ES-SSWs to be a subset of SSWs identified by the WMO criteria (WMO-SSWs). Pertaining to the dynamics prior to SSW onset, the difference between ES-SSWs and SSWs without an elevated stratopause is expected to be negligible since the elevated stratopause is a post-warming phenomenon. In the proposal for this study, results for ES-SSWs are assumed to be results for SSWs in general.

SSWs will be identified graphically with the aid of markers to indicate the fulfillment of the aforementioned three criteria. Figure 7.3 shows an example of how these markers can be applied to a dataset so that each case can be visually verified. The presence of the magenta box marking the onset date shows the fulfillment of all WMO-SSW and ES-SSW criteria. The fulfillment of individual WMO-SSW and ES-SSW criteria are also shown in Figure 7.3. Periods of persistent wind reversal (\( \geq 5 \) days) at 1 hPa are shown by black vertical lines with tick marks pointing toward eastward winds. Periods of persistent wind reversal at 10 hPa are shown by light grey vertical lines with tick marks pointing toward eastward winds. Time periods in which the temperature drops below 190 K between 80 km and 100 km is shown by magenta vertical lines with tick marks pointing toward warmer periods. Times when the discontinuity in the stratopause is greater than 10 km is denoted by black Xs at the bottom of the plot.
Figure 7.3. An example of SSW identification markers overlaid on a zonal-mean zonal wind plot averaged from 70° N to 90° N. Bold black line indicates the location of the stratopause. Thin black contour indicates the zero-wind line.

7.4 Ensemble Setup

With the limited number of SSW occurrences in observations (and WACCM-SD), it is not possible to develop a statistically robust picture of SSWs, especially if we seek to understand the dynamical differences leading up to displaced and split vortex events. More years of observational data are needed to increase the sampling number of SSWs. Alternatively, we can utilize models to generate more SSW events through long-term simulations (like a few hundred years) or short-term ensemble runs with many members. While letting the model run freely may be the easiest solution to gaining more SSW cases, we are dependent on the model's ability to produce SSWs at an appropriate frequency. Additionally, a free-run approach would produce a taxing amount of unusable data such as summers and winters with only minor warmings.

Instead, an ensemble approach is proposed for our study. While an ensemble would typically proceed in free-running mode, our ensemble maintains its higher vertical resolution, 88 versus 66 levels, by resuming in the specified dynamics configuration. As noted in the WACCM-SD run
description above, the model was nudged to observations up to 50 km and free-running above 60 km, with a linear transition in between those altitude levels. This will be called the root model run. In lowering the nudged altitude, the model (albeit in specified dynamics mode) becomes free-running across a greater altitude range. For our proposed ensemble setup, the model is nudged with a linear transition from 0 km to 0.4 km and initialized with WACCM-SD output from the root model run. Essentially, above 0.4 km, the model is free running.

An ensemble is created by randomly perturbing the initial atmospheric temperature condition by an amount below the model’s rounding error of $\sim 10^{-14} \text{K}$ (e.g., Kay et al., 2015). For example, to generate a 50-member ensemble, the initial temperature condition is randomly perturbed up to $n \times 10^{-14} \text{K}$, where $n = 1, 2, \ldots, 50$. Ensembles can be discarded as needed and can be run in parallel with other members.

An ensemble member will be considered “normal” if the criteria for SSW by Limpasuvan et al. (Limpasuvan et al., 2016) and WMO are not met within a month of the root onset date, the SSW onset date in the root run. Additionally, there should be no persistent (>5 days) temperature or wind reversals prior to the root onset date. While this definition does not explicitly remove winters with minor warmings, it allows variability in the normal members and primes the composite analysis for a binary argument; i.e. members chosen for the composites and anomalies either exhibited all features of an SSW or no features of an SSW. The time constraint of two months surrounding the root onset date was chosen since the EPWs are expected to occur 1 to 2 weeks prior to onset. Additionally, this implies that members with a final warming (when the polar regions transition to late springtime conditions) less than a month after the root onset date will not be considered a normal member.

To assess the appropriate initialization date for WACCM, a case study was performed on the 2009 split SSW case in which ensemble runs were initialized 30, 40, and 50 days prior to the root SSW onset date for this event (i.e., 22 January 2019). Figure 7.4 shows that the probability of producing a member with a normal scenarios or an SSW are most similar when the ensemble is initialized 40 days prior to the root onset date. For these 20 ensemble runs, 5 produced normal
scenarios and 4 produced an SSW. The ensemble runs produced SSWs well below the normal rate of SSW occurrence, roughly 60% of the time (Butler et al., 2015), suggesting that the production of the SSW is sufficiently random. In addition, this lead time well exceeds our current SSW predictability around 20 days (Domeisen et al., 2020; Karpechko, 2018), allowing for more randomized model outcomes. An ensemble run initialized 40 days prior to the selected SSW onset would allow processes persisting a month or less to evolve and react to varying initial conditions.

![Probability of SSW and Normal Occurrences](image)

**Figure 7.4.** Probability of a normal member vs. ensemble initialization time in days prior to SSW onset. 20 ensemble members were run for each initialization date.

For our ensemble runs, two split and two displaced SSW events with clear split/displaced vortex characteristics were selected from the WACCM-SD dataset between 1978 and 2013 (described in Section 7.2). The amount of SSW and normal members collected for each of the 4 sets, distinguished by initialization date, are recorded in Table 7.1.

The 4 selected SSW events had onsets of 12 February 1984, 9 January 2006, 22 January 2009, and 5 January 2013. The former two SSWs are displaced type (Charlton & Polvani, 2007; Kuttippurath & Nikulin, 2012) and latter two SSWs are split type (Coy & Pawson, 2015; Kuttippurath & Nikulin, 2012; Manney et al., 2009).
Additionally, these SSW events have a diverse range of quasi-biennial oscillation (QBO) phases, based on the equatorial stratospheric winds at 50 hPa. This may be of importance in a composite since the Northern Hemisphere (NH) polar vortex tends to be warmer (colder) and more disturbed (stronger) during a westward (eastward) QBO phase (Holton & Tan, 1980, 1982). Furthermore, the non-linear interaction between the QBO and the El Niño Southern Oscillation enhances the frequency of SSWs (Richter et al., 2011). The transition of QBO phases could also affect the timing of SSW development; Gray (2003) suggests in a mechanistic primitive-equation model a wind profile consistent with an eastward to westward QBO phase transition postpones warming events in the NH.

<table>
<thead>
<tr>
<th>Original SSW onset date (YYYY-MM-DD)</th>
<th>Number of normal members</th>
<th>Number of SSW members</th>
<th>QBO Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984-02-21</td>
<td>16</td>
<td>12</td>
<td>Weak (&lt; 5 m s⁻¹)</td>
</tr>
<tr>
<td>2006-01-09</td>
<td>12</td>
<td>33</td>
<td>Westward</td>
</tr>
<tr>
<td>2009-01-22</td>
<td>29</td>
<td>14</td>
<td>Eastward</td>
</tr>
<tr>
<td>2013-01-05</td>
<td>11</td>
<td>17</td>
<td>Westward</td>
</tr>
</tbody>
</table>

Table 7.1. Number of normal and SSW ensemble members generated with respect to reference SSW onset dates.

Composites are made with respect to SSW onset date. For a composite of normal winters with no SSW onset, the original SSW onset date listed in Table 7.1 was used. The risk of bias for a particular set of ensemble members (described by a row in Table 7.1) was eliminated by averaging the ensembles members in each set first, then averaging the sets together. Anomalies are calculated by subtracting the composited diagnostic across normal members from the composited diagnostic across SSW members.

The normalized Gaussian distributions of diagnostic values can act as their probability distribution curve. These diagnostic values were collected with respect to SSW onset date at a specified height, latitude, and longitude (or averaged over a range of heights, latitudes, or longitudes). As illustrated in the Figure 7.5 schematic, normal Gaussian probability distributions
of diagnostic values collected from SSW ensemble members and normal ensemble members (weighted such that each ensemble set has an equal contribution) were assumed. The area where these distributions overlap was calculated. This area is the probability that a value found at a location and reference day during an SSW would also be found during a normal winter. Subtracting the area from unity gives the area shaded in green in Figure 7.5. This area represents the probability that, given an SSW occurs, the diagnostic value will be different from that found during normal winters, henceforth abbreviated as P(\text{Ab|SSW}).

\textbf{Figure 7.5.} Schematic of normalized Gaussian distributions of diagnostic values during winters with SSWS (left curve) and normal winters (right curve). Values are collected with respect to SSW onset date at a specific height, longitude, and latitude. The green shaded region shows the probability of a diagnostic value being abnormal given that the winter contains an SSW or P(\text{Ab|SSW}).

Alternatively, the value describes the probability of an anomaly surrounding SSW being abnormal. This diagnostic is particularly useful in determining what anomalies exclusively occur during SSWs and do not occur during normal winters. For example, a low probability associated with a large anomaly suggests that, while the diagnostic value is typically larger than normal, the value is still found to be large in scenarios where no SSW occurs. Thus, a large anomaly at this
location and reference day may be associated with SSWs but is not a good indicator for the occurrence of an SSW.

7.5 Analyses

The linearized equation governing quasi-geostrophic potential vorticity (PV) or $q$ offers a method of separating contributions from instability and GWD (Andrews et al., 1987):

$$q_t' + \frac{(\bar{u} - c_x)}{R_e \cos \phi} q'_{x'} + \nu' \bar{q}_\phi' = X' + O(\alpha^2)$$

(7.1)

where $\phi$ is the latitude, $z$ the log-pressure height. Here, $R_e$ is the Earth’s radius, $\bar{u}$ is the zonal-mean zonal wind, and $\nu$ is the meridional wind velocity. Subscripts of $\lambda$, $\phi$, and $t$ indicate partial derivatives with respect to longitude, latitude, and time, respectively. From left to right, the first term describes the change in the PV perturbation ($q'$), the second term describes the zonal advection of PV, and the third term describes the meridional advection of PV, where $\nu'$ is the meridional velocity perturbation and $\bar{q}_\phi$ is the mean meridional PV gradient. $X'$ represents effects external forcings, including GWD. $O(\alpha^2)$ represents the error from linearization and may become large in regions of PW breaking.

The meridional quasi-geostrophic potential vorticity gradient $\bar{q}_\phi$ was used to examine the stability of the middle atmosphere based on 3-day average of dependent field variables (O’Neill & Youngblut, 1982):

$$\bar{q}_\phi = 2\Omega \cos \phi - R_e^{-1} \left( \frac{(\bar{u} \cos \phi)}{\cos \phi} \right)_\phi - \frac{\alpha f_0^2}{\rho_0} \left( \frac{\rho_0}{N_B^2} \frac{\bar{u}_z}{z} \right)_z$$

(7.2)

where $\phi$ is the latitude, $z$ the log-pressure height, $f_0$ is the reference Coriolis parameter, $\rho_0$ is the reference density, and $\Omega$ the Earth’s angular frequency. We refer to the positive definite first term on the right hand side (RHS) as the “beta term” associated with the gradient of $f_0$, the second term as the “barotropic term” associated with horizontal wind curvature, and the third term as the “baroclinic term” associated mainly with the vertical wind curvature. In figures, $\bar{q}_\phi$ is
nondimensionalized by $\Omega$. The $\tilde{q}_\phi$ formulation is based on 3-day averages of the dependent field variables which inherently filtered out waves with periods less than 3 days (or where $c_x > 77 \text{ m}\cdot\text{s}^{-1}$ at 60°N).

Eastward, westward, and stationary components of diagnostics were optionally obtained by first implementing a 31-day sliding Hanning window to dependent field variables (wind, temperature, etc.) and applying a Fourier transform. The Eliassen-Palm (EP) flux was computed using formulation associated with the transformed Eulerian-mean (TEM) equations given in Andrews et al., (1987). Section 3.4.2 discusses the EP flux in further detail. For these calculations, 5-day running averages were applied to the dependent field variables (wind, temperature, etc.) to remove perturbations with periods less than 5 days (or where $c_x > 46 \text{ m}\cdot\text{s}^{-1}$ at 60°N). The EP flux represents the product of the wave group velocity and the wave activity density (Andrews et al., 1987). In the NH, EP flux divergence (convergence) corresponds with wave energy growth (decay) and/or an eastward (westward) acceleration on the background wind.

The squared refractive index ($n^2$) can be used to better understand how PWs of certain zonal wavenumbers ($s$) and zonal phase speeds ($c_x$) propagate in $\bar{u}$ (Andrews et al., 1987):

$$n^2 = \frac{\bar{q}_\phi}{R_e (\bar{u} - c_x)} - \left( \frac{s}{R_e \cos \phi} \right)^2 - \left( \frac{f_0}{2 N_B H} \right)^2$$

(7.3)

where $H$ is the scale height. PWs tend to propagate towards a large positive squared refractive index and are unable to propagate in regions with a negative squared refractive index. The $n^2$ is difficult to composite as it varies widely, often approaching infinity. Therefore, we use $n^2$ as a reference to show how our diagnostics affect PW propagation instead of showing $n^2$ directly.
The $\vec{u}$ structure in the stratosphere and mesosphere can exhibit regions of unusually strong wind shear. There, the flow can become barotropically and/or baroclinically unstable, leading to the appearance of unstable PWs (e.g., Dickinson, 1973; Leovy & Webster, 1976; Matthias & Ern, 2018). For shear instability to occur, the generally positive meridional (quasi-geostrophic) potential vorticity gradient ($\tilde{q}_\phi$) associated with the wintertime circulation must become negative (e.g., Murray L. Salby, 1996). To serve as a source for an instability wave of a certain zonal phase speed ($c_x$), that region must also contain a critical layer, where the mean zonal flow matches $c_x$ (Dickinson, 1973). Hartmann (1983) used a linear barotropic model and a quasi-geostrophic baroclinic model to examine instabilities of the eastward stratospheric polar night jet. He found that, when the instability was seated on the poleward flank of the jet, the most unstable modes were wavenumber-1 and -2 waves with periods of a few days. When the instability was seated on the mid-latitude flank of the jet, the most unstable modes were wavenumber-1 to -3 waves with periods of a week or more. Manney et al. (1988) suggests that these periods are likely slightly longer in observations since the nonlinear effect of instability tends to weaken and broaden the jet (Pedlosky, 1987). Orsolini & Simon (1995) used a fully nonlinear nondivergent barotropic model to simulate the generation and life cycle of unstable PWs arising from the instability of the polar night jet, as well those arising from the instability of a double jet representing the mesospheric subtropical jet and the upward extension of the stratospheric polar night jet. These authors found similar low-wavenumber instability waves as Hartmann (1983) with periods on the order of days for the single jet case. Planetary-scale vortices developed to expel the low potential vorticity on the poleward flank of the jet into lower latitudes, acting to remove the sign reversal of the meridional potential vorticity gradient. Eddies of higher wavenumbers (3-4) were found in the double-jet case, accelerating and stabilizing the flow between the two zonal wind maxima. Although perturbations of higher wavenumbers have larger growth rates, Hartmann (1983) suggested that low-wavenumber instabilities (given the
predominance of low-wavenumber disturbance in the stratosphere) would be more likely to derive energy from an unstable flow than higher wavenumbers.

Prior to the split SSW of January 2009, a couple of studies noted the presence of slow eastward-propagating PWs, hereafter EPWs, in the mesosphere (L. Coy et al., 2011; Iida et al., 2014). Using a high-top forecast model with data assimilation, Coy et al. (2011) suggested that the mesospheric EPWs directly propagated from the troposphere with the underlying bursts of wavenumber-2 PW activity prior to the SSW onset and their eventual dissipation in the lower mesosphere. Using satellite observations, Iida et al. (2014) instead suggested that the EPW appearance before SSW onset could be generated in situ by shear instability of the polar night jet. Based on reanalyses, Song et al. (2020) demonstrated that the amplification of wavenumber-2 EPW before the 2009 SSW onset is likely attributed to GW forcing in the upper stratosphere and lower mesosphere.

Regardless of their source, EPWs may be a common feature leading up to SSW. In the composite study of Limpasuvan et al. (2016) based on 13 SSW events, a robust signature of zonal wavenumber-1 EPW with an eastward period of around 10 days was clearly evident between 40-60 km and over the polar region, intensifying roughly 10 days before SSW onset (see their Figure 10). However, these authors did not discuss the cause of the wave presence and focused only on the wavenumber-1 westward-travelling wave, developing after the SSW onset. Hence, the exact nature of EPWs and why they occur before SSW onset remain unclear. Using a high-resolution global circulation model (GCM) integrated over three years, Sato and Nomoto (2015) suggested that EPWs may be generated by baroclinic instability in the mesosphere.

To provide a detailed account of the dynamics leading up to SSW onset, Chapter 8 examines mesospheric EPWs prior to the 2009 SSW using a high-top GCM constrained by reanalysis below 50 km. The evolution of the background flow conditions supporting EPW is explored through the interplay of GWs and PWs. We identify a similar double $\bar{u}$ maxima in the meridional direction, as noted by Orsolini & Simon (1995). In fact, this wind configuration consists of well-separated and strengthened subtropical mesospheric and polar night jets. The strong shear instability between
the two jet cores promotes the growth of unstable EPWs whose characteristics are consistent with wave over-reflection. To our knowledge, this is the first study to examine the mesospheric instabilities prior to an SSW event from the perspective of over-reflection. Our study may lead to a better understanding of mesosphere-stratosphere coupling and help assess the role of the mesosphere in SSW predictability.

SSW events were identified using the criteria of Limpasuvan et al. (2016) discussed in Section 7.3. These SSWs capture the strong coupling between the upper stratosphere and the mesosphere-lower thermosphere during SSWs which is pertinent in this study. We identified 13 ES-SSW events in our simulations between 1980-2013. In YYYYMMDD format, the onset dates of these ES-SSWs are 19840221, 19841230, 19870122, 19890219, 19950127, 19971223, 20020213, 20031220, 20060109, 20090122, 20100124, 20120113, and 20130105. We emphasize that the SSW onset is defined based on $\bar{u}$ at 1 hPa reversing direction. While the current study focuses on the 2009 event, winters when other SSW events occurred are removed from the climatology of quiet, non-SSW winters.

8.1 Interplay of PWs and GWs

The $\bar{u}$ meridional cross-sections prior to the onset are compared to the December-February (DJF) climatology in Figure 8.1. Here, the wind climatology excludes years with ES-SSWs with onset dates listed in the Chapter 8 introduction. On 25 December (corresponding to the gold vertical line in Figure 8.1a), the wind structure is similar to climatology (Figure 8.1d), with a single maximum near the mid-latitude stratopause. The eastward wind magnitude is however stronger than climatology. A few days later (on 31 December corresponding to the brown vertical line in Figure 8.1a), the wind structure departs significantly from climatology, consisting now of two local maxima of comparable strength (Figure 8.1c). One maximum appears in the subtropical upper mesosphere and the other corresponds to a strengthened polar night jet. Iida et al. (2014) reported a similar double-maxima wind configuration before the 2009 SSW onset in MLS observations.
The rapid wind evolution between 25 and 31 December was investigated with respect to PW forcing (corresponding to disturbances of zonal wavenumbers 1-6) and total GWD (described in Section 2.3) on the background wind (Figure 8.2). On 25 December, PW activity is weak with the expected upward and equatorward propagation through the region of eastward wind from the mid-latitude lower stratosphere to low-latitude upper mesosphere (Figure 8.2a). Nevertheless, upward PW activity appears in the upper region of the polar night jet and its EP flux convergence...
leads to a westward forcing (blue contour) of around 10 m·s⁻¹·day⁻¹. By 31 December, PW activity greatly intensifies, penetrating well into the subtropical mesosphere before damping. Centered near 35°N and 80 km, the EP flux convergence exerts strong westward forcing in excess of 60 m·s⁻¹·day⁻¹ between the subtropical and polar jet cores.

**Figure 8.2.** Altitude-latitude sections of (a, b) PW EP flux and its divergence and (c, d) total GWD (resolved and parameterized) during 25 December and 31 December 2008. Westward (dotted black contour) and eastward (solid thin black contour) winds increment by 10 m·s⁻¹, with the zero-wind line thickened. Incremented by 20 m·s⁻¹·day⁻¹, GWD and PW EP flux divergence are contoured in blue for westward forcing and red for eastward forcing. For (a) and (b), the 10 m·s⁻¹·day⁻¹ isopleth is indicated by the thin blue contour to illustrate the broad extent of PW forcing. The meridional EP flux vector component was scaled by $(100\pi R e \rho_0)^{-1} \cos \phi$ and the vertical component by $(R e \rho_0)^{-1} \cos \phi$. Grey shading indicates regions of negative $\tilde{q}_\phi$.

Figures 8.2c,d illustrate the corresponding GWD. On 25 December, when $\tilde{u}$ is similar to climatology, strong westward GWD (blue contours) caps the top of the eastward jet below the
The forcing is most prevalent around 80 km between 30-50°N. Such GWD pattern is due to westward GWs that are allowed to reach the mesosphere by the strong eastward wind below 60 km. With the development of the double-maxima wind configuration on 31 December, eastward GWD (red contours) becomes more apparent. Wintertime PWs propagating upward from the troposphere can break along the edge of the polar jet, as suggested by the EP flux pattern in Figure 8.2b, and exert a westward acceleration along the equatorward flank of the polar vortex below 70 km. The resulting weakened eastward wind allows eastward GWs to reach the mesosphere and impose an eastward $\vec{u}$ tendency near the subtropical mesospheric jet core, evident in Figure 8.2d. GWD also becomes increasingly westward in the mid-latitudes (with values exceeding 60 m·s$^{-1}$·day$^{-1}$) and concentrated near the zero-wind line. Such strong decelerative effects increase the wind shear as indicated by the constricted isotachs.

**Figure 8.3** shows the latitude-time evolution of PW forcing and total GWD (both as filled contours) averaged between 0.2 hPa and 0.02 hPa. This pressure range is where strong GWD appears in Figures 8.2c,d. We overlay the eastward $\vec{u}$ averaged between 1.0 and 0.1 hPa as dotted line contours to capture the evolution of the polar night jet in the upper stratosphere (e.g., see Figures 8.1b,c). To illustrate the evolution of subtropical mesospheric jet, we also superimpose eastward $\vec{u}$ averaged between 0.1 and 0.01 hPa in solid line contours. For clarity, only eastward wind values of 30 m·s$^{-1}$ and greater are shown.
Figure 8.3. Latitude-time evolution (averaged between 0.2 hPa and 0.02 hPa) showing (a) PW forcing and (b) total GWD, both in m s\(^{-1}\) day\(^{-1}\). The eastward \(\bar{u}\) (m s\(^{-1}\)) values averaged from 1.0 to 0.1 hPa indicates the polar jet (dotted contours). The eastward \(\bar{u}\) values averaged from 0.1 to 0.01 hPa indicates the subtropical jet (solid contours). Vertical gold and brown lines mark 25 and 31 December 2008, respectively. The latter date corresponds to the formation of a double-maxima wind structure. The dashed vertical line indicates SSW onset on 22 January 2009.
At this altitude range, we clearly see that, equatorward of 50°N, westward PW forcing maximizes after 25 December (indicated by the gold vertical line), consistent with the transition between Figures 8.2a,b. Strong westward PW forcing then appears over the entire NH just before SSW onset (indicated by the vertical dashed line). Seen in Figure 8.3b, strong westward GWD (blue regions) persists over the polar jet and followed its migration northward. With the SSW onset, eastward GWD eventually dominates throughout NH due to the underlying stratospheric wind reversal, as observed by De Wit et al. (2014).

By 31 December (brown vertical line), the formation of the double-maxima wind configuration noted in Figures 8.1 and 8.2 commences as the polar jet migrates poleward. Concurrently, the westward PW forcing and eastward GWD both peak around 30°N between the subtropical jet and the polar jet. Comparing Figures 8.2b,d, we see that near 80 km and equatorward of 40°N, the eastward GWD (red contours) exceeds the westward PW forcing (blue contour). The resulting net eastward forcing (from GWs and PWs) would help maintain the subtropical mesospheric jet core. The nearby westward PW forcing maxima between the subtropical and polar jets would slow the local eastward wind between the jet cores. To this end, GWD and PW forcing conspire to help form the double-maxima wind configuration. In their model simulation, Sato and Nomoto (2015) identified similar interplay between GWD and PW forcing in the formation of the double-maxima wind configuration (see their Figures 8.6 and 8.7).

8.2 Static and Shear Instabilities

In Figures 8.2a,b on 25 and 31 December, a region of negative $q\phi$ exists below the zero-wind line suggesting a configuration in which the turning level lies below the critical layer, similar to the idealized schematic in Figure 3.6. Therefore, a pre-existing configuration is available to support over-reflection given the introduction of sufficient incident PW activity and/or instability.

From the TEM perspective, net wave forcing drives a (residual) mean meridional circulation, ($\tilde{v}^*$, $\tilde{w}^*$), to help maintain thermal wind balance. Here, $\tilde{v}^*$ and $\tilde{w}^*$ represent the meridional and vertical motion, respectively. Figures 8.4a,b show the mean meridional circulation as vectors for
25 and 31 December. On 25 December, the poleward motion corresponds to the westward GWD (seen in Figure 8.2c) and, by continuity, results in strong downward motion over the polar region (Figure 8.4a). As highlighted by the negative $\vec{w}^*$ region (cool-colored shading), such downwelling extends largely across NH between 60-90 km and corresponds to a well-documented wintertime phenomena driven by GWD that helps maintain the stratopause (e.g., Hitchman et al., 1989).

With the formation of the double-maxima wind configuration (Figure 8.4b), upwelling (warm-colored shading) replaces downwelling around 30°-50°N and 60-80 km as eastward GWD becomes dominant near the subtropical mesospheric jet core.

**Figure 8.4.** Altitude-latitude sections of (a, b) $\vec{w}^*$ and (c, d) $N_B^2$ as shading. (a, b) The residual circulation vectors ($\vec{v}^*$, $\vec{w}^*$), multiplied by (1, 1000), are overlaid. Only eastward $\vec{u}$ is shown and contoured every 10 m·s$^{-1}$ as thin black contours. The zero-wind line is thickened.
Anomalous upwelling can lead to adiabatic cooling and affects (vertical) static stability by altering the vertical temperature gradient. Illustrating $N_B^2$, Figure 8.4c reveals that static stability between 60-90 km tends to be weak (blue shading) equatorward of 40°N. By 31 December (Figure 8.4d), the strong upwelling identified in Figure 8.4b has led to extensive cooling and, consequently, widespread areas of weakened static stability between the subtropical mesospheric and polar jets of the double-maxima wind configuration.

The baroclinic term of $\bar{q}_\phi$, defined in Equation (1), depends on $N_B^2$ and the vertical wind shear. The diminished static stability and enhanced vertical wind shear would increase the contribution from the baroclinic term and reduce the overall $\bar{q}_\phi$, assuming other terms in Equation (1) are fixed. Comparing Figures 8.4c,d, we note that strong vertical wind shear (evidenced by the tightened isotachs) coincides with area of drastic $N_B^2$ decline. With development of a double-maxima wind structure, we expect $\bar{q}_\phi$ to decrease.

Figure 8.5 illustrates the relative contribution of various terms to $\bar{q}_\phi$ in Equation (1). Positive and negative (nondimensionalized) $\bar{q}_\phi$ values are shown in Figures 8.5a-c. The negative $\bar{q}_\phi$ region is bounded by grey shading and corresponds to the grey shading in Figures 8.2a,b. In Figures 8.5d-f, regions where the magnitude of the beta, barotropic, or baroclinic terms in Equation (2) dominates over other terms are shaded in blue, grey, or red, respectively. Light and dark tints of each color indicate whether the term is contributing positively (light) or negatively (dark) to $\bar{q}_\phi$.

For example, since the planetary vorticity term is always positive, the dominance of the beta term is always represented as a light blue.
Figure 8.5. Altitude-latitude sections of (a-c) non-dimensionalized $\bar{q}_\phi$ and (d-f) the dominant terms contributing to $\bar{q}_\phi$. Dates are shown at the upper right of each row. (d-f) In reference to the right-hand side of Equation 1, blue-, grey-, and red-shaded regions indicate the dominance of the first, second, and third term, respectively. Dark (light) colors represent a negative (positive) contribution to $\bar{q}_\phi$. Eastward $\bar{u}$ is contoured every 10 m·s$^{-1}$ as thin black contours, with the zero-wind line thickened.

With the development of the double-maxima wind configuration in Figures 8.5b,c, increased positive $\bar{q}_\phi$ values developed into well-organized cores (brown-shaded contours), nearly
collocated with the local wind maxima. Figures 8.5e,f show that the enhanced values are dominated by the barotropic term (grey region), indicating the strong horizontal wind curvature. The increased curvature is attributed to the strong westward GWD near the zero-wind line and the adjacent eastward GWD (see Figure 8.2d), as well as the pronounced westward PW forcing in the subtropics. A valley of low $\bar{q}_\phi$ developed between the local maxima extending diagonally from the poleward side of subtropical jet near 80 km. Regions of negative $\bar{q}_\phi$ (grey- and purple-shaded contours) increased in magnitude on 31 December and, approaching onset, peaked above the zero-wind line near 80 km and 50°N. Sato and Nomoto (2015) found a decrease in their modified PV gradient that coincides with $N^2$ decline (attributed also to upwelling and adiabatic cooling) in a similar location between jet cores.

Given the split nature of the 2009 SSW, we investigate the evolution of the wavenumber-2 geopotential height amplitude, $\bar{Z}_g$, averaged between 45°N and 55°N in Figure 8.6. Following the formation of the double-maxima wind configuration on 31 December (marked by the brown vertical line), the wavenumber-2 $\bar{Z}_g$ with a mean eastward phase speed appears over a deep layer between 1 hPa and 0.1 hPa. About 10 days before SSW onset, the wavenumber-2 peaked near 5 m·s$^{-1}$. Iida et al. (2014) observed very similar slow EPWs associated with the double-maxima wind configuration in observations. A similar shift toward the dominance of slow eastward phase speed for PWs was also noted in Sato and Nomoto (2015).
Figure 8.6. Zonal phase speed vs. time plot of wavenumber-2 geopotential height amplitude or $Z_g$ (m) at 33, 51 and 67 km averaged between 45°N and 55°N. White contours increment by 5 m. Positive phase speed indicates eastward movement. The vertical brown line corresponds with the formation of a double-maxima wind structure on 31 December 2008. The vertical dashed line indicates SSW onset on 22 January 2009.

The presence of unstable EPW arising from the reversal of PV is further suggested by the non-zonally averaged view of the circulation. Figure 8.7 illustrates the carbon monoxide (CO) distribution (as filled contours) at 0.1 hPa. At this altitude, CO serves as a nearly conservative tracer that mimics PV (e.g., Solomon et al., 1985). Overlaid on this figure are the geopotential height (black) contours, outlining the polar vortex. On 31 December (Figure 8.7a), a wavenumber-1 perturbation appeared as polar vortex shifted off the pole, increasing the local zonal wind near the International Date Line. By 6 January (Figure 8.7b), the vortex further deformed with features indicative of PW breaking, and an irreversible mixing, along the vortex’s edge. A filament of low-CO bluish air (and high PV, not shown) was advected equatorward around 30°N, just poleward of the subtropical jet. This filament structure illustrates the local meridional gradient reversal of PV that destabilized the flow. Finally, by 16 January, the CO distribution is dominated by a wavenumber-2 pattern as a result of two partially separated low-pressure systems. This pattern migrates slowly eastward thereafter, as suggested by Figure 8.6c.
As seen in Figure 8.6, the presence of wavenumber-2 EPWs persisted up through SSW onset. This persistence coincides well with the lingering presence of the double-maxima wind configuration shown in Figure 8.3. Between 25 December and 22 January, we see the intensification of the subtropical mesospheric core (solid contour in Figure 8.3) and the polar jet core (dotted contours) as the latter continues to migrate poleward. We identified an increase in the magnitude of negative $\bar{q}_\phi$ with the development of strong wind shear and diminished static stability near the subtropical mesospheric and polar jet cores (see Figure 8.5b). This development suggests a strong destabilization of the background wind near the zero-wind line. Figure 8.7c shows that this growing instability has a wavenumber-2 pattern. Hence, the development and persistence of the double-maxima wind structure prior to SSW encouraged the in-situ generation of a wavenumber-2 EPW leading up to SSW onset via shear instability.

8.3 Generation of EPWs from Shear Instability

Figure 8.8a show the EP flux for wavenumber-2 PWs (hereafter PW2s) of all phase speeds just prior to SSW onset (16 January 2009). Here, the double-maxima wind structure persisted from its formation around 31 December 2008. An extensive region of negative $\bar{q}_\phi$ (grey-shaded area)
remained in a meridional local minimum of $\bar{u}$ near 0.1 hPa. Consistent with the split polar vortex, PW2s dominated the wave field. Strong PW2 activity emanated upward from the troposphere and was refracted equatorward upon reaching the stratopause and above. Westward forcing (blue contours) associated with PW2s appear mainly above and on the equatorward side of the polar jet. Its strong decelerative effects on $\bar{u}$ further allows eastward GWD to reach the upper mesosphere as noted in Figure 8.3b.

**Figure 8.8.** Altitude-latitude sections of PW2 EP flux and EP flux divergence for (a) all phase speeds and (b) eastward phase speeds. Negative (blue) and positive (red) EP Flux divergence is contoured every (a) 5 m·s⁻¹·day⁻¹ and (b) 2 m·s⁻¹·day⁻¹. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s⁻¹, with the zero-wind line thickened. The negative $\bar{q}_\phi$ region is grey-shaded. The presence of a critical layer inside the negative $\bar{q}_\phi$ regions is shaded in green. The EP Flux vector components are scaled as done in Figure 8.2, but the EPW2 vector magnitudes are multiplied by 10.0.

We elucidate EPWs by band-pass filtering for wavenumber-2 disturbances with eastward phase speeds of 5 m·s⁻¹ and greater. We refer to the filtered result as EPW2. The selected phase speed range for filtering was based on the identified PW2 eastward peak in Figure 8.6. Considering the broad phase speed distribution shown in Figure 8.6, the band-passed EPW2 is a small contribution to the total PW2. Nevertheless, filtering should better illustrate the EPW
characteristics by minimizing the influence from strong quasi-stationary PW activity, suggested by the phase speed distribution in Figure 8.6.

The resulting EPW2 EP flux is shown in Figure 8.8b. As noted by Coy et al. (2011), we see the upward EPW2 from the troposphere. Upon reaching the stratopause, its flux convergence imposes westward forcing on the equatorward side of the polar jet. EPW2 activity also emanates from the bottom edge of the negative (grey) $\bar{q}_\phi$ region with an overlying critical layer (green shading). This emanation results in a strong EPW2 EP flux divergence (red contours) and overlaps with the mid-latitude region of PW2 EP flux divergence in Figure 8.8a. Such overlapping highlights the dominance of EPW2 to the overall PW2 activity in that region. The collocation of a negative $\bar{q}_\phi$ region, a critical layer, and emerging EPW2 activity strongly suggests in-situ EPW2 growth from shear instability. Occurring in the eastward flow regime, such growth indicates that these unstable waves have an eastward phase speed, as supported by the eastward shift in the phase speed distribution in Figure 8.6 around 16 January. Comparing Figures 8.5c,f, the negative $\bar{q}_\phi$ region is dominated by the barotropic term (dark grey region) and baroclinic term (dark red region) on the equatorward and poleward side, respectively. Thus, the emanating EPW2 flux activity pointed equatorward as well as downward toward a flux convergence region in the subtropics, where it merged with the EPW2 activity from below. Overall, the characteristics of these unstable waves are consistent with EPW2 identified by Coy et al. (2011) and Iida et al. (2014).

8.4 Propagation of EPW2

We explore the linear propagation of EPW2 in Figure 8.9 based on the (nondimensionalized) squared refractive index ($n^2$). As defined in Equation (2), $n^2$ depends on $N_B^2$, $\bar{q}_\phi$, and the critical layer (via the quantity $\bar{u} - c_x$). In particular, $n^2$ becomes negative when $\bar{q}_\phi$ is negative. It decreases with either decreasing $N_B^2$ or $\bar{q}_\phi$ and becomes infinite at the critical layer.
In Figure 8.9, we compute \( n^2 \) for a stationary PW2 \((c_x = 0)\) and for a PW2 with eastward phase speed of 10 m·s\(^{-1}\) at 50°N. As noted in Figure 8.6, the phase speed of the emergent EPW2, after the formation of the double-maxima wind configuration, tends to focus around 5 m·s\(^{-1}\). However, given the broad phase speed distribution of PW2 perturbations, we used the phase speed of 10 m·s\(^{-1}\) (at 50°N) in computing \( n^2 \) to illustrate the upper bound characteristics of EPW2 propagation. The white areas in Figure 8.9 represent negative \( n^2 \) values. Red regions correspond to extremely large \( n^2 \) values (>100), often occurring near critical layers.

On 25 December, a broad region of low (bluish) \( n^2 \) tilts along with the eastward wind structure for both EPW2 (Figure 8.9a) and stationary PW2 (Figure 8.9d). Although EPW2 is expected to be
weak above 10 hPa at this time (based on Figure 8.6), any EPW2 propagation would tend to be refracted toward larger $n^2$ values on the equatorward flank of the jet and dissipate near its critical layer. This propagational pattern would be similar for stationary PW2. With the formation of the double-maxima wind configuration on 31 December, PW2 started to grow (as seen near the vertical brown line in Figure 8.6) throughout the stratosphere as $n^2$ drastically changed. The aforementioned valley of low $\bar{q}_\phi$ values (found in Figure 8.5b between the subtropical mesospheric and polar night jets) reduced $n^2$ around 30°-40°N and 60-80 km, labeled as feature A. This tends to limit PW2 from propagating further upward and equatorward toward the subtropical mesospheric zero-wind line. Moreover, an enhanced positive $\bar{q}_\phi$ in Figure 8.5b corresponds to a localized region of large $n^2$ around at 44°N and 60 km, labeled as feature B. Taken together, changes in $n^2$ at features A and B established a waveguide that encouraged PW2s to propagate toward, but not beyond, the intervening region between the subtropical and polar jets. Correspondingly, strong PW EP flux convergence occurred (as evident in Figure 8.2b), indicating the dissipation of stationary PW2 and EPW2 at that location.

By 16 January, EPW2 became very strong near the stratopause (see Figure 8.6). The northward migration of the polar jet increased the $\bar{u}$ curvature and increased $\bar{q}_\phi$ as evident by the maxima in Figure 8.5c at 10 hPa and 70°N. This increased curvature was expected to also enhance $n^2$ values since $n^2$ depends on $\bar{q}_\phi$ through Equation (3). Consequently, comparing Figures 8.9a,c as well as Figures 8.9d,f, we see that regions of positive $n^2$ spread toward higher latitudes between 10 and 0.1 hPa, allowing EPW2 and PW2 (originating in the mid-latitude troposphere) to propagate more vertically along the edge of the polar vortex, as shown in Figure 8.8b. The predominantly vertical propagation of PWs prior to SSW onset is common (e.g., Limpasuvan et al., 2012).

### 8.5 Unstable EPW from an Over-reflection Perspective

Unstable waves can manifest as over-reflection and transmission as suggested by studies like Lindzen (1980) and Harnik & Heifetz (2007). Illustrated in Figure 3.6, such a perspective links the
influence of upward propagating tropospheric EPW2 disturbances reported by Coy et al (2011), to the production of mesospheric unstable waves during the 2009 SSW suggested by Iida et al. (2014). Since the vertical geometry for instability idealized in Figure 3.6 is homomorphic with the meridional geometry (Lindzen, 1988), a similar meridionally-oriented geometry could also encourage over-reflection. Hence, unlike the idealized scenario, the over-reflected waves in reality can have both vertical and meridional components in their group velocity.

The illustration in Figure 3.6 bears a strong resemblance to Figure 8.8b. In particular, we see the upward propagation of EPW2 impinging on the bottom portion of the unstable (grey) region of negative $q_\phi$. As diagnosed in Figure 8.9c, the waveguide (by 16 January) readily allowed upward EPW2 activity to reach the unstable region, where the EPW2 critical layer (green region) resided. Downward EPW2 EP flux vectors point away from the unstable region in Figure 8.8b, leaving behind a strong region of EP flux divergence. The emergent EPW2 activity is below the wave evanescent region of negative $n^2$ values, as shown in the white region (labeled as feature C) in Figure 8.9c. As such, these vectors can be interpreted as the over-reflection of upward propagating EPW2. Overall, we expect the downward energy propagation to have been negated or masked by the persistent upward EPW2 activity from below. As noted above, given the complex nature of $q_\phi$, the EPW2 activity emerging from the unstable region can also point equatorward.

Comparing Figures 8.9c,f, the evanescent region (feature C) was thinner for EPW2 than for PW2, suggesting that EPW2 was more conducive to over-reflection since it would need to tunnel a shorter vertical distance to its critical layer (as suggested in Figure 3.6). Thus, there was a bias to produce unstable EPWs. After the formation of a double-maxima wind configuration, the poleward movement of the polar jet allowed PW2 and EPW2 to propagate vertically (discussed in Section 8.4), further thinning the evanescent region. The eastward bias in the resultant unstable wave production is shown by the eastward shift in $Z_y$ phase speed distribution in Figure 8.6.
Figure 8.10a illustrates the altitude-time evolution of the vertical component of EPW2 EP flux, averaged from 45°N-55°N. The upward flux is shown in brownish filled contours and downward in bluish filled contours. Following the formation of a double-maxima wind configuration on 31 December (brown vertical line), EPW2 propagated from the surface up to the stratopause. Upon reaching the bottom of the unstable region (black stipples) in which a critical layer resides, over-reflection occurred as evidenced by the negative (bluish areas) EP flux near 60 km and around 3 January. There, EP flux divergence (red contours) imposes an eastward forcing of 4 m·s⁻¹·day⁻¹.
Figure 8.10. Altitude-time section of PW2 vertical EP flux (filled-contours) for waves of (a) eastward phase speeds greater than 5 m·s⁻¹ and (b) all phase speeds. Regular-sized (thin) black contours of +(-) 0.2 × 2ᵐ kg·s⁻², m ∈ [1,2,3,4,5,6,7,8] outline upward (downward) fluxes. Red contours show EP flux divergence and are incremented by (a) 2 m·s⁻¹·day⁻¹ and (b) 5 m·s⁻¹·day⁻¹. Thick black contour depicts the zero-wind line. Locations with negative ｑ_φ are marked by black stipples. Locations where a critical layer exists inside a region of negative ｑ_φ are marked by green stipples. Stippled regions shown in (a) also apply to (b). Latitudinal averaging is between 45°N and 55°N.

Around 10 January, another stronger episode of upward EPW2 emerged from the surface. Like the earlier episode, we see over-reflection from the unstable region and strong EP flux divergence just before the wind reversal as marked by the zero-wind line (thickened black contour). Coy et al. (2011) reported similar bursts of PW2 signal originating near the surface that reached 1 hPa
within days to weeks. Overall, these over-reflection characteristics are consistent with features in 5.12b.

The evolution of EP Flux for all PW2 is shown in Figure 8.10b. In comparison to Figure 8.10a, the EPW2 flux comprised some of the persistent upward PW2 bursts, especially after 31 December. In reference to Figure 3.6, wave transmission could appear in the positive $n^2$ region above the evanescent region. From Figure 8.9c, this evanescent region (near feature C) is roughly 20 km thick. Transmitted waves propagating from the critical layer can readily deposit their momentum creating a region of EP flux convergence (westward acceleration). This pattern is evident in Figure 8.8a as a region of PW2 EP flux convergence (blue contours) that sits atop a region of EP flux divergence (red contours) in the unstable (grey) region. Additional evidence of transmission also appears in Figure 8.10b with upward PW2 flux (brown region) above the region of downward flux (blue shading) along with EP flux divergence (red contours).

8.6 Summary of the January 2009 SSW Case Study

The anomalous growth of the mesospheric EPWs with a zonal phase speed of ~10 m·s$^{-1}$ prior to the 2009 split SSW was identified in this paper, in agreement with past studies (e.g., L. Coy et al., 2011; Iida et al., 2014; Song et al., 2020). Our diagnoses reveal new key insights about these precursory EPWs:

1. They arise from shear instability in the mesosphere via wave over-reflection.

2. Created by GW and PW forcing ~20 days before SSW onset, the unstable flow is characterized by a double-maxima wind structure with a subtropical mesospheric core and polar stratospheric core.

3. This wind configuration sets up a unique wave geometry that, from an over-reflection perspective, favors the production of eastward-propagating PWs.

Preceding the formation of the double-maxima wind configuration was the presence of subtropical eastward GWD adjacent to the mid- to high-latitude westward GWD in the
mesosphere. This distinctive GWD pattern induced a subtropical upwelling that locally lowered static stability and, consequently, altered the refractive index. Changes in wave propagation led to enhanced PW damping near the intervening region between the wind maxima, further promoting the jet separation. Thus, a positive feedback loop was created in which the double-maxima wind configuration was sustained, while the mesospheric flow became more susceptible to shear instability. With the formation of a double-maxima wind structure, the polar jet core strengthened and migrated poleward as the wave evanescent layer in the unstable region became thinner. The northward-shifted polar jet also guided the upward-propagating PWs more vertically toward the thinning evanescent region (see Figures 6 and 10). The vertical orientation increased the likelihood of over-reflection, particularly for EPWs. The poleward migration of the polar night jet may likewise provide the background conditions that favor SSW onset through resonance (e.g., Albers & Birner, 2014).

The background flow evolution leading into the January 2009 SSW supported a wave geometry suitable for wave over-reflection (c.f., Figure 3.6). As a result, PWs with eastward phase speeds were generated near the stratopause from instability. A composite study by Domeisen et al. (2018) found a tendency for an eastward shift in the PW zonal phase speed distribution with altitude prior to SSW onset. These authors also noted that PWs propagating upwards into the stratosphere are limited to low wavenumbers by the strong wintertime stratospheric background wind exceeding a critical speed. Since the critical speed of the background wind is relative to the wave, eastward-propagating (westward-propagating) waves experience a higher (lower) critical speed. Thus, eastward-propagating waves would be able to propagate into stronger stratospheric winds, with a larger effect at higher wavenumbers. However, over-reflection also generates PWs with a bias toward eastward phase speeds. This suggests a possible compounding effect that would shift PWs in the stratosphere toward eastward phase speeds, particularly for split SSWs.

In examining other SSW events (not shown), we found that the generation of EPWs were common prior to SSW onset and were commonly associated with a double-maxima wind
configuration, formed at different times with respect to SSW onset. Overall, EPWs could significantly impact the mesospheric wind structure and play a key role in the nature and timing of SSW events. In particular, EPWs help reduce the preexisting wind shear and, thereby, stabilize the polar vortex. GWD may likewise generate a non-conservative wave source for EPWs (Song et al., 2020) and work in tandem with $\tilde{q}_\phi$ reversal in generating in-situ EPWs. Further research is needed to understand the source mechanisms of in-situ wave generation in the mesosphere. The over-reflection perspective offers a framework to connect variability in the stratosphere to stability in the mesosphere.

8.7 Extended Research

8.7.1 MLS Validation

MLS data was used to validate our findings from the January 2009 SSW case study. In this section, we show some examples of MLS observations capturing key features in our case study. Comparable to Figures 8.4c,d, Figure 8.11 shows $N_B^2$ for 25 and 31 December 2008. The blue shaded region representing low $N_B^2$ at 65 km extends to 35°N in Figure 8.11a. In Figure 8.11b, this region extends above 50°N. This decrease in $N_B^2$ coincides with the development of a separation of the subtropical (85 km) and polar (55 km) jet. Here, the subtropical jet is not as well defined as in Figure 8.6. The geostrophic wind approximation used to calculate MLS winds shown by Equations 2.7a,b become less accurate approaching the Equator where $f$ approaches zero. This results in an over-estimation of wind strength. As a result, the subtropical jet in our MLS data appears to extend farther south than in the model winds.
Regardless, Figure 8.11 shows an adiabatic cooling that results in a reduction of static stability and $\bar{q}_\phi$, like in Figure 8.4. This makes the region between the subtropical and polar jets more susceptible PW dissipation and a further separation of the jets. The values of $N^2_B$ between 5.15 and Figure 8.4 are sufficiently similar and suggests that the model accurately represents the zonal-mean thermal structure during the formation of a double-maxima.

Figure 8.12, comparable with Figure 8.6, shows the phase speed distributions of $\bar{Z}_g$ at 10, 1, and 0.1 hPa. After the formation of a double-maxima configuration on 31 December 2008 (brown vertical line), there is a shift in phase speed towards eastward waves. Similar to the signal shown by WACCM-SD, the eastward shift in phase speed distribution persists until SSW onset (dashed vertical line). The poleward shift of the poleward jet (which vertically-orient PWs) and the diminished $\bar{q}_\phi$ values between the subtropical and polar jets suggest that over-reflection is more likely.
Figure 8.12. Zonal phase speed vs. time plot of wavenumber-2 GHP amplitude (m) at 33, 51 and 67 km averaged between 45°N and 55°N. White contours increment by 5 m. Positive phase speed indicates eastward movement. The vertical brown line corresponds with the formation of a double-maxima wind structure on 31 December 2008. The vertical dashed line indicates SSW onset on 22 January 2009. Observations were retrieved from MLS.

Figure 8.13 shows the WN2 EP flux on 16 January 2009. Similar to Figure 8.8, a double-maxima zonal wind configuration is present prior to SSW onset: a polar jet around 60°N and 50 km and a subtropical jet around 30°N and 75 km. Observations agree that a critical layer is present inside a layer of negative $\bar{q}_\phi$ from 45-70°N above 60 km. Figure 8.13a shows a large divergence emerging below the negative $\bar{q}_\phi$ region. This feature is also found in our model (Figure 8.8a) except the divergence in Figure 8.13 occurs at higher latitudes. Here, there are no downwards EP flux vectors that would be expected from the occurrence of over-reflection. When filtered for eastward waves with phase speeds greater than 5 m s$^{-1}$ (Figure 8.13b), the EP flux vectors show that the direction of energy propagation changes abruptly due to the region of EP flux divergence, shifting from a mostly upward to a purely southward orientation. Filtering by phase speed provides the ability to better visualize the change in the direction of PW propagation due to over-reflection (also the case in Figure 8.8). However, model data reveals downward propagation from the EP flux divergence region, expected from over-reflection, albeit over a small vertical range (<
10 km). The coarser vertical resolution of MLS could explain why this downward EP flux is not captured.

**Figure 8.13.** Altitude-latitude sections of PW2 EP flux and EP flux divergence for (a) all phase speeds and (b) eastward phase speeds. Negative (blue) and positive (red) EP Flux divergence is contoured every (a) 5 m·s\(^{-1}\)·day\(^{-1}\) and (b) 2 m·s\(^{-1}\)·day\(^{-1}\). Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s\(^{-1}\), with the zero-wind line thickened. The negative \(\vec{q}_\phi\) region is grey-shaded. The presence of a critical layer inside the negative \(\vec{q}_\phi\) regions is shaded in green. The EP Flux vector components are scaled as done in Figure 7, but the EPW2 vector magnitudes are multiplied by 10.0. Data retrieved from MLS.

### 8.7.2 Other Case Studies

Other SSWs were examined to investigate the source, uniqueness, characteristics, and impacts of EPWs. The January 2009 SSW discussed previously in the main text was only associated with dominant eastward-propagating unstable wavenumber-2 PWs prior to a split SSW. Here, we demonstrate the origin of the double-maxima wind followed by the appearance of EPWs for four other SSW events with onset dates of 9 January 2006 (displaced vortex), 24 January 2010 (displaced vortex), 13 January 2012 (displaced vortex), and 5 January 2013 (split vortex). By performing these case studies, the present study shows that the instability features discussed in Section 8.2 are not exclusive to the 2009 split SSW.
**Figure 8.14** shows the GWD, PW forcing, and $N_s^2$ over four winters containing SSW events. While a double-maxima zonal wind configuration forms sporadically throughout the winter, all four SSWs exhibited the persistence of the double-maxima leading up to SSW onset. The separation between the polar jet (dotted contours) and the subtropical jet (solid contours) is shown in Figure 8.14 by the poleward movement of the polar jet. During the 2005-2006 winter, the double-maxima formed in November (not shown) and persists until onset. As a result, the polar jet and subtropical jets are already separated at the beginning of December.

In all cases, the appearance of the double-maxima wind configuration was simultaneous with an eastward GWD in the subtropics between 0.2 hPa and 0.02 hPa, shown in Figures 8.14(a-d)b. This eastward GWD migrated north approaching SSW onset, which coincided with the poleward march of the polar jet. Ultimately, at SSW onset, eastward GWD in the mesosphere extended throughout the northern hemisphere as the stratospheric zonal wind reversed (dashed vertical line).
The Formation of a Double-Maxima in Zonal-Mean Zonal Wind Prior to SSW Onset
Averaged Between 0.2 – 0.02 hPa

a) 2005-2006

aa) PW Forcing
WACCM-SD

ac) $N_B^2$
WACCM-SD

ab) GW Drag
WACCM-SD

ad) $N_B^2$
MLS

b) 2009-2010

ba) PW Forcing
WACCM-SD

bc) $N_B^2$
WACCM-SD

bb) GW Drag
WACCM-SD

bd) $N_B^2$
MLS
Figure 8.14 Latitude-time sections averaged between 0.1 hPa and 0.01 hPa of ((a-d)a) PW forcing from WACCM-SD, ((a-d)b) GWD from WACCM-SD, ((a-d)c) upwelling or positive $\vec{w}^*$ from WACCM-SD, ((a-d)d) upwelling from MLS, ((a-d)e) $N^2_B$ from WACCM-SD, and ((a-d)f) $N^2_B$ from MLS. GWD is calculated by totaling parameterized the GWD and the EP flux of wavenumbers greater than 6. Black contours indicate eastward winds incremented by 10 m·s$^{-1}$. The zero-wind line is indicated by a bold black contour. Solid vertical line marks SSW onset. Dashed vertical line marks the formation of a double-maxima configuration. Dashed line is missing for the 2005-2006 winter because the double-maxima wind structure forms prior to December.

For most cases, the poleward movement of the polar jet and low $N^2_B$ values occurred in tandem. The magnitude and timing of decreased $N^2_B$ could be slightly different between model and observations, possibly due to differences in the actual and parameterized GWD. However, the fluctuations of $N^2_B$ found in the model broadly agrees with observations.

As in the January 2009 case study, the modification of $N^2_B$ would alter PW propagation through $n^2$. However, the contribution of PW forcing and GWD to the double-maxima wind formation varies as seen in Figure 8.15. Since the double-maxima structure persisted throughout much of the winter prior to the January 2006 SSW, the formation of the double-maxima structure is not shown in Figure 8.15. All cases presented in Figure 8.15 show a similar juxtaposition of subtropical eastward GWD and mid- to high-latitude westward GWD in the mesosphere. This was also noted in Section 8.2 in reference to the double-maxima formation during the 2008-2009 winter. Additionally, upward-propagating PWs dissipated on the equatorward side of the polar jet and the poleward side of the subtropical jet, aiding in the formation of a double-maxima. The correlation of eastward and westward GWDs, westward PW forcing, upwelling, and a decrease in $N^2_B$ suggests that the mechanism for the formation of a double-maxima wind structure proposed in Section 8.2 is common leading up to SSW onset, irrespective of the polar vortex being split or displaced.
Figure 8.15. GW forcing (m·s⁻¹·day⁻¹) and wavenumber 1-6 PW forcing (m·s⁻¹·day⁻¹). Eastward (red contours) and westward (blue contours) accelerations are incremented by 20 m·s⁻¹·day⁻¹. Thin black contours indicate eastward zonal-mean zonal winds. EP flux vectors (m²·s⁻²) illustrate the direction of the waves’ group velocity. The meridional EP Flux vector component was scaled by \((100\pi R_\text{e}\rho)^{-1}\cos \phi\) and the vertical EP flux vector component was scaled by \((R_\text{e}\rho)^{-1}\cos \phi\). Grey regions indicate negative \(\bar{q}_\phi\).

Discussed in Section 8.2, the double-maxima wind configuration was conducive to the production of instability waves. In Section 8.7.3, a composite study summarizes the robust instability wave signal prior to SSW. To avoid redundancy, the present section will discuss select cases to illuminate specific characteristics of instability waves.

During the January 2013 and December 1984 SSW, EPW2s were particularly strong (see Figure 8.16). In fact, the EP flux divergence associated with EPW production from instability provided a strong eastward acceleration that supported a tertiary zonal-mean zonal wind maxima between
the subtropical and polar jet. For WACCM-SD, EP flux convergence zones to the north and south of the divergence region suggest that eastward momentum was transferred from the equator to the pole. MLS observations are not available before 2004 and therefore cannot be compared with Figure 8.16a. For 4 January 2013, MLS observations are shown in Figure 8.16c. While the structure of the instability agrees between the model and observations, the winds and the EP flux divergence are different. In Figure 8.16c, the tertiary zonal wind maxima is not as noticeable and appears as a poleward extension of the subtropical jet. This could be a result of the geostrophic wind approximation. Furthermore, geostrophic approximations of the wind can result in large discrepancies in the EP flux divergence (Boville, 1987).
Figure 8.16. Height vs. latitude sections from negative (blue contours) and positive (red contours) EP Flux divergence (m·s⁻¹·day⁻¹) for EPW2s moving faster than 5 m·s⁻¹; EP Flux divergence/convergence contours increment by 2 m·s⁻¹·day⁻¹. EP flux vectors (m²·s⁻²) illustrate the direction of the waves’ group velocity. The meridional EP Flux vector component was scaled by $10 \times (100 \pi R_e \rho)^{-1} \cos \phi$ and the vertical EP Flux vector component was scaled by $10 \times (R_e \rho)^{-1} \cos \phi$. The EP Flux divergence was scaled by $(\rho R_e \cos \phi)^{-1}$. Eastward zonal-mean zonal winds are incremented by 10 m·s⁻¹ (thin black contours) with the zero-wind line in bold. A region of negative $\bar{q}_\phi$ (grey-shaded region) fulfills the necessary condition for instability. The presence of a critical layer inside a region of negative $\bar{q}_\phi$ is indicated by green shading. Data was retrieved from (a,b) WACCM-SD and (c) MLS.

In all cases, EP flux divergence occurred in eastward winds, implying that the instability wave produced would have a mean eastward phase speed. Finally, instability waves occurring during
the 2005-2006 winter are shown in Figure 8.17 to demonstrate that eastward mesospheric instability PWs occur as wavenumber-1 and -2 perturbations. Here, the model and observations are in good agreement.

**Figure 8.17.** Height vs. latitude sections from (a, c) WACCM-SD and (b, d) MLS of negative (blue contours) and positive (red contours) EP Flux divergence (m·s⁻¹·day⁻¹) for EPW1s (top row) and EPW2s (bottom row) moving faster than 5 m·s⁻¹; EP Flux divergence/convergence contours increment by 2 m·s⁻¹·day⁻¹. EP flux vectors (m²·s⁻²) illustrate the direction of the waves’ group velocity. The meridional EP Flux vector component was scaled by $10 \times \left( \frac{100\pi R_e \rho}{\cos \phi} \right)^{-1}$ and the vertical EP Flux vector component was scaled by $10 \times \left( \frac{R_e \rho}{\cos \phi} \right)^{-1} \cos \phi$. The EP Flux divergence was scaled by $(p R_e \cos \phi)^{-1}$. Eastward zonal-mean zonal winds are incremented by 10 m·s⁻¹ (thin black contours). The zero-wind line is represented by a thick black contour. A region of negative $\bar{q}_\phi$ (grey-shaded region) fulfills a necessary condition for instability. The presence of a critical layer inside a region of negative $\bar{q}_\phi$ is indicated in green shading.
In this section, we have shown that the interplay of PW and GWD in creating a double-maxima configuration are not unique to the often-studied January 2009 split SSW. Furthermore, after this double-maxima wind formation, instability EPWs were found to be associated with wavnumber-1 and wavenumber-2 perturbations prior to both split and displaced SSWs. Therefore PWs sourced from instability are not unique to the type of SSW. Accordingly, the rest of the study will focus on the nature of PWs sourced from instability surrounding SSWs in general.

While instability waves are not necessarily dependent on a double-maxima wind configuration, strong and persistent occurrences of these instability waves were found after the formation of a double-maxima and prior to SSW. In all cases, these EPWs significantly impact the wind structure of the middle atmosphere. In some cases, EPWs can produce enough local eastward acceleration to generate a tertiary local maximum between the polar and subtropical jets.

8.7.3 Small Composite Study

A composite of SSW events listed in the Chapter 8 introduction was developed to elucidate the robust behavior of EPWs in the mesosphere prior to SSW onset. Following Limpasuvan et al. (2016), the composite is based on aligning all SSW events with respect to their onset dates. By convention, Day 0 is the onset date of the composite.

Analogous to Figure 8.1, Figure 8.18 shows the evolution of the composite \( \bar{u} \) at 60°N. Again, the wind reversal above 60 km occurs a few days before the stratospheric wind reversal and the wind strengthens prior to onset (vertical dashed line). Like Figure 8.1b, the composite wind meridional structure well before onset (e.g., Day -35 shown in Figure 8.18b) is similar to the DJF climatology illustrated in Figure 8.1d. By Day -10 (Figure 8.18c), the double-maxima wind configuration appears as in the 2009 SSW case (Figure 8.1c). Thus, the double-maxima wind configuration appears to be a robust characteristic leading toward SSW onset.
Figure 8.18. (a) Altitude-time of $\bar{u}$ at 60°N composited for all SSWs in Table 1. SSW onset at day 0 is indicated by a vertical dashed line. Westward (dotted black contour) and eastward (solid thin black contour) wind increment by 10 m·s$^{-1}$, with the zero-wind line thickened. (b, c) Altitude-latitude composites of $\bar{u}$ (b) 35 days and (c) 10 days before SSW onset.

Figures 8.19a,b illustrate the composite evolution of the jet cores in the double-maxima wind configuration along with the interplay between PW and GW forcing. Like in the 2009 SSW event, the polar jet core (dotted black contours) deviates farther poleward from the subtropical jet core (solid black contours) approaching Day 0 (black vertical line). Westward PW forcing becomes stronger between the polar and subtropical jets as the wind maxima separate (Figure 8.19a). Here, the PW forcing is averaged between 1.0 and 0.01 hPa to emphasize its effect in separating the polar jet (~1.0-0.1 hPa) from the subtropical jet (~0.1-0.01 hPa). Westward
acceleration from PW forcing maximizes and spreads throughout the winter hemisphere at SSW onset, eroding the polar jet. Illustrated in Figure 8.19b, the GWD pattern is similar to the 2009 SSW event (Figure 8.3b). Westward GWD accompanies the underlying polar jet’s migration into high latitudes while the eastward GWD becomes dominant at lower latitudes. Akin to the 2009 SSW case, the net eastward forcing in the subtropical regions would lead to localized upwelling and adiabatic cooling. These conditions lead to a decrease in $N^2_B$ at mid-latitudes (Figure 8.19c). Upon onset, strong eastward GWD dominates the northern hemisphere and induces an adiabatic mesospheric cooling above the descending westward wind (Limpasuvan et al., 2012). While the growth of EPW with zonal wavenumber 1 (i.e. EPW1) shows up well in a composite of all SSWs, the growth of EPW2 is not as evident. This disparity is attributed to the disproportionately large number of displacement SSW events. PW1s enhance prior to displacement and some split SSWs while PW2s only notably amplify prior to split SSWs (e.g., Bancalá et al., 2012).
Figure 8.19. Latitude-time of (a) PW forcing, (b) total GWD, and (c) $N_B^2$ composited of all SSWs in Table 1. Vertical averaging are indicated to the left of each plot. Zonal-mean zonal wind averaged from 1.0 to 0.1 hPa indicates the position and strength of the polar jet (dotted contours) and the zonal-mean zonal wind averaged from 0.1 to 0.01 hPa indicates the position and strength of the subtropical jet (solid contours).
To better elucidate EPWs, Figure 8.20 shows separately the composite EPW1 and EPW2 activity for a 5-day averaged period (days -10 to -5) before displacement and split SSWs. In the composite, four split SSWs were used with onset dates of 19841230, 19890219, 20090122, 20130105 and nine displaced SSWs were used with onset dates of 19840221, 19870122, 19950127, 19971223, 20020213, 20031220, 20060109, 20100124, 20120113 (in YYYYMMDD format).

As shown in Figure 8.20, a double-maxima configuration exists prior to both split and displacement SSWs. We caution that the composite of split SSWs is comprised of only 4 events. Prior to both split and displacement SSWs, the meridional structure of $\vec{u}$ prominently features the double-maxima wind configuration. Strong wind shear above the polar night jet and poleward of the subtropical jet foster an unstable (grey-shaded) region of negative $\vec{q}_\phi$. EPW1 and EPW2 activity emerge from the unstable region with downward and equatorward fluxes. The composite EPW1 activity is stronger and directed more downward compared to EPW2. The
resulting EP flux divergence imposes eastward acceleration above the polar jet. The associated EP flux convergence imposes westward acceleration on the flanks of the subtropical and polar jets around 50°N.

Overall, the unstable EPW activity in the composite suggests the possible occurrence of over-reflection similar to the 2009 split SSW case. Strong upward PW activity from the lower stratosphere impinges on the unstable region. The emerging over-reflected EPWs appear to merge with the underlying EPWs in producing strong westward forcing on the equatorward flow of the polar night jet. A composite study by Domeisen et al. (2018) finds a tendency for an eastward shift in zonal phase speed for PWs with altitude prior to SSW onset. These authors also note that PWs propagating upwards into the stratosphere are limited to low wavenumbers by the strong wintertime stratospheric background wind exceeding the critical velocity. Since the critical velocity of the background wind is relative to the wave, eastward-propagating (westward-propagating) waves experience a higher (lower) critical velocity. Thus, eastward-propagating waves would be able to propagate into stronger stratospheric winds, with a larger effect at higher wavenumbers. However, over-reflection also generates PWs with a bias toward eastward phase speeds (as discussed previously). This suggests a compounding effect that would shift PWs in the stratosphere toward eastward phase speeds, particularly for split SSWs that are associated with a higher wavenumber disturbance (wavenumber 2).
Chapter 9

Ensemble Composite Study

The zonal-mean zonal wind during normal winters and winters with SSWs are shown in Figure 9.1. By definition, the stratosphere maintains an eastward flow throughout a normal winter and experiences an eastward-to-westward wind reversal during winters with SSWs. For normal winters, the mesospheric zero-wind line is maintained at 82.6 km with a standard deviation of 12.4 km based on the altitude of the mesospheric zero-wind line 41 days surrounding the central date of each normal ensemble member. The stark contrast between a steady zero-wind line during normal winters and the dramatic descension of the zero-wind line during ES-SSW winters depends on the stability of the mesospheric flow.
Figure 9.1. Height-latitude composites of $\bar{u}$ averaged from 60-70°N during (a) normal and (b) SSW winters. $\bar{u}$ is shown by thin black contours and is incremented by 10 m s$^{-1}$. Blue and green lines emphasize the location of the -5 and 5 m s$^{-1}$ isotachs, respectively. Grey shading shows regions of negative $q_\phi$.

The stability of the flow is diagnosed by $q_\phi$, such that negative $q_\phi$ (shaded regions in Figure 9.1) in an eastward zonal-mean flow suggests a region instability. Likewise, positive $q_\phi$ in a westward zonal-mean flow also suggests instability (Dickinson, 1973). Since this is a composite study, the exact position of critical levels for EPWs, QSPWs, and WPWs cannot be explicitly shown. Blue (green) contours emphasize the -5 (5) m s$^{-1}$ isotachs that separate the wind regimes containing critical levels for WPWs, QSPWs, and EPWs. Since PWs propagate westward relative to the zonal wind, EPWs would find their critical level beyond green contours (winds greater than 5 m s$^{-1}$),
QSPWs between green and blue contours, and WPWs within blue contours (winds less than -5 m s\(^{-1}\)).

Depending on the relative positioning of the source level, turning level, and critical level, several wave geometries are possible as shown in Figure 9.2.
Figure 9.2. Wave geometries of vertical PW propagation between 20 km and 90 km. In our composite, these scenarios roughly occur (a) before day 0, (b, c) days 0 - 10, (d) and days 10 - 20.
Red, yellow, and blue colors represent eastward, stationary, and westward zonal phase speeds, respectively. Figure 9.2a shows the same background wind configuration as Figure 3.6 such that eastward waves would experience a thinner evanescent region, allowing them to tunnel more easily to the critical level and stimulate PW growth at the critical level. The modulation of phase speed by the background geometry of the instability layer will alter the unstable PW’s future propagation paths and its eventual dissipation. While additional non-conservative considerations (such as asymmetric GW drag) are necessary, the over-reflection perspective is powerful as it connects incident, transmitted, and over-reflected waves to one another.

In Figure 9.3a, positive (red-shaded contours) and negative (grey- and purple-shaded contours) \( q_\phi \) values are an indicator of the stability of the background flow. In Figure 9.3b, the dominance of beta, barotropic, and baroclinic terms of \( \tilde{q}_\phi \) (see Equation 7.2) are shown by blue, grey, and red shading, respectively. A negative (positive) contribution of the dominant term is indicated by a darker (lighter) shading. For example, since the planetary vorticity is always positive in the Northern Hemisphere, the beta term is always shown by light blue shading. The zero-wind line (thick black contour) separates the regions of mean eastward and westward flow.
Figure 9.3. Height-time (a,b) composites and (b) anomalies of $\bar{q}_\phi$ during SSW winters. (a) Positive (negative) and negative $\bar{q}_\phi$ values are shown by red (grey and purple) shaded contours. (b) Dominance of the beta, barotropic, and baroclinic terms are indicated by blue, grey and red contours, respectively. Lighter/darker shading indicates a positive/negative contribution of the dominant term. (c) Positive (negative) anomalies of $\bar{q}_\phi$ are indicated by medium (thin) black contours. The probability that the anomaly is abnormal is given by orange-shaded contours. The zero-wind line is indicated by a thick bold contour.
Prior to SSW events, the background flow in the middle atmosphere only becomes baroclinically unstable (indicated by dark red shading in a negative $\bar{q}_\phi$ region; c.f., Figures 2.26a,b) due to curvature of the shear as the wind transitions from eastward in the mesosphere to westward in the lower thermosphere. Figure 9.3c shows positive (negative) $\bar{q}_\phi$ anomalies in solid (dashed) contours with orange-shaded contours indicating $P(\text{Ab}\mid\text{SSW})$ (see Section 7.4 for explanation). Small $\bar{q}_\phi$ anomalies and low $P(\text{Ab}\mid\text{SSW})$ values before day -10 indicate that negative $\bar{q}_\phi$ in the MLT is common during normal winters too, also shown in Figure 9.1a.

In Figure 9.1a the blue, green, and zero-wind isotachs are embedded in the $\bar{q}_\phi < 0$ (shaded) region associated with the normal wintertime MLT. Here, the critical levels for WPWs, QSPWs, and EPWs exist past a turning level. Similar to the schematic in Figure 9.2a, eastward-propagating waves (EPWs) would experience a critical level lower to the ground than QSPWs or WPWs resulting in a thinner evanescent region. Therefore, over-reflection would be more favorable towards EPW production during a normal winter. Wave activity that is not over-reflected is refracted meridionally due to the turning level. As a result, the height variability of the mesospheric zero-wind line due to the dissipation of upward-propagating PWs is curtailed. A similar wave geometry is seen before day 0 in Figure 9.1b. Rhodes et al. (2021) discusses how the zonal wind configuration leading up to SSW onset becomes conducive for the over-reflection of EPWs.

The zonal-mean zonal wind is closely tied to $\bar{q}_\phi$ in the middle atmosphere with a correlation coefficient of 0.98. This coefficient was calculated from the averaged values of the zonal-mean zonal wind and $\bar{q}_\phi$ between 10 hPa and 1 hPa and between 60°N and 70°N. These averaged values were correlated across days -20 and 30 for all SSW ensemble members. Given this high correlation, $\bar{q}_\phi$ can be used as a proxy diagnostic for the zonal-mean zonal wind in the middle atmosphere. For example, as the westward winds slow their descent in Figure 9.3a after SSW onset, they become weaker and correspond to diminishing magnitudes of negative $\bar{q}_\phi$. 
Therefore, the stabilization of the stratospheric winds can be gauged by increased $\bar{q}_\phi$, in which positive $\bar{q}_\phi$ and eastward zonal-mean winds are restored upon SSW recovery. Approaching onset in Figures 9.3a,b, the normally unstable baroclinic flow in the mesosphere starts to stabilize, indicated by the increase in $\bar{q}_\phi$. After SSW onset, the mesosphere shows positive $\bar{q}_\phi$ and eastward zonal-mean winds in the mesosphere indicating a stable configuration.

9.1 Before Day 0

9.1.1 WPWs and QSPWs

Figure 9.4a-c depicts the composited EP fluxes for (a) WPWs, (b) QSPWs, and (c) EPWs such that the shaded contours indicate vertical flux while the blue (red) contours show EP flux convergence (divergence). Even though they have critical levels aloft, the upward flux does not result in flux convergence in the mesosphere. Figure 9.1b shows that prior to day 0 a thick layer of negative $\bar{q}_\phi$ encompasses the critical levels of WPWs and QSPWs. The same wave geometry is true for winters devoid of SSWs, shown in Figure 9.1a. According to the first term in Equation 7.3, the decrease in $\bar{q}_\phi$ would result in a decrease in $n^2$ such that $n^2$ is negative when $\bar{q}_\phi$ is zero. As a result, PW group velocities become less vertically oriented as they are refracted away from the turning level. Resultantly, PWs are unable to propagate to their critical level aloft, which exist beyond the turning level. Figure 9.2a suggests that WPWs and QSPWs could tunnel past the turning level and over-reflect. However, this would be more likely to occur for EPWs since their critical level is closer to the turning level, i.e., they experience a thinner evanescent region.
Figure 9.4. Height-time plot averaged between 60°N and 70°N. (a-c) Upward (downward) vertical EP fluxes are outlined by black regular-sized (thin) contours incremented by $+(-) 0.2 \times 2^m \text{ kg s}^{-2}$, $m \in [1,2,3,4,5,6,7,8]$ and filled with tan (blue) shading. EP flux convergence (blue contours) and divergence (red contours) are incremented by $2 \text{ m s}^{-1} \text{ day}^{-1}$. (d-f) Positive (negative) anomalies of EP flux divergence are indicated by solid (dashed) contours and the probability that the anomaly is abnormal is given by orange-shaded contours. The zero-wind line is indicated by a thick bold contour.

Figures 9.4d-f shows positive (negative) EP flux divergence anomalies in solid (dashed) contours. Orange-shaded contours show $P(\text{Ab} | \text{SSW})$. Prior to SSW, this probability is low, suggesting that
EP flux prior to onset is not a good indicator of SSW occurrence. Albeit from day -5 to day -1, flux convergence anomalies in the stratosphere for WPWs and QSPWs get larger along with the $P(\text{Ab}|\text{SSW})$ approaching onset.

### 9.1.2 EPWs

As during a normal winter, discussed in the Chapter 9 introduction, the wave geometry at 70 km before SSW onset in Figure 9.1b favors the over-reflection of EPWs. This bias is exemplified in Figure 9.4 since divergence near the stratopause prior to SSW is seen primarily in the eastward component of the flux. EPWs produced from this divergence region propagate downward as suggested by the coinciding negative vertical EP flux (blue shaded region). Approaching onset, Figure 9.4c shows a persistent upward flux doesn’t only sustain a region of EP flux divergence, but results in an increasing amount of EP flux divergence at 70 km approaching SSW onset. This is consistent with the over-reflection mechanism; as the evanescent region becomes thinner, incident waves more easily over-reflect. In Figure 9.4c, the EPW EP flux divergence increases in tandem with $\tilde{q}_\phi$ (shown in Figure 9.3a). Over-reflection acts to relieve instability, but by doing so it reduces negative $\tilde{q}_\phi$ that separates upward-propagating PWs from the critical level. A positive feedback loop is established in which upward-propagating PWs become increasingly effective at stabilizing the stratopause the longer they persist.

Figure 9.4f at 70 km shows small average EP flux divergence anomalies with $P(\text{Ab}|\text{SSW}) < 0.1$. While EP flux divergence for EPWs is a robust feature prior to SSW growth, the flux divergence at a specific height and reference day (with respect to SSW onset) is not a good indicator of SSW. Additionally, the low $P(\text{Ab}|\text{SSW})$ for all PWs (cf., Figures 9.4a-c) suggests that the strength of the tropospheric forcing may not be the only factor in producing an SSW. As mentioned previously, the persistence may be just as important, if not more important, than the strength of tropospheric forcing in initiating an SSW event.
9.2 Day 0 to Day 10

9.2.1 WPWs

Days 0-10 are marked by descended westward winds in the stratosphere and eastward winds in the MLT. This creates a layer of reversed stratospheric winds (RSW) that has a zonal-mean westward flow and persists for over 2 weeks. WPWs with faster westward phase speeds than the westward wind velocity should be able to propagate past the RSW unencumbered since they would not experience a critical level. Alternatively, WPWs with phase speeds slower than the background wind will interact with the RSW. Since EP fluxes for WPWs with phase speeds less than \(-20 \text{ m s}^{-1}\) are negligible compared to those in Figure 9.4 (not shown), we focus on the interaction of WPWs with the RSW to explain their propagation through the middle atmosphere.

Wave geometries experienced at the bottom boundary of the RSW are illustrated in Figure 9.2b. Slower WPWs (right green arrow) may find their critical level and be absorbed while faster WPWs (left green arrow) may find a turning level before their critical level and over-reflect. In Figure 9.1b, the turning level (edge of the grey shading) and the zero-wind line remain close together at the bottom boundary of the RSW; even WPWs with phase speeds between \(-5 \text{ m s}^{-1}\) (blue line) and \(-20 \text{ m s}^{-1}\) would experience their critical level past the turning level and be prone to over-reflection. At the upper boundary of the RSW, the turning level is well-separated from the zero-wind line and is generally maintained within the RSW. Around day 4, the 5 and 10 m s\(^{-1}\) isotachs rest above the turning level at the upper boundary of the RSW. Therefore, WPWs roughly between 5 and 10 m s\(^{-1}\) would be prone to over-reflecting at both the bottom and top boundaries, each time extracting energy from the background flow and transmitting it aloft. As a composite, these wave geometries are approximated by the average \(\bar{q}_\phi\) and \(\bar{u}\). In individual case studies, a wider range of phase speeds may be able to over-reflect at both boundaries.

Figure 9.4a shows that upward EP flux (brown shading) associated with WPWs appears above 80 km. Compared to EPWs and QSPWs, WPWs have a relatively strong presence above 80 km with a larger upward EP flux indicated by brown shading (cf., Figures 9.4a-c). Additionally, an EP flux
convergence region exceeding 20 m s\(^{-1}\) day\(^{-1}\) in Figure 9.4a at ~95 km shows that the resulting dissipation of WPWs have a significant impact on the background wind. This abnormal EP flux convergence is unique to SSWs with P(\text{Ab}|\text{SSW}) > 0.6.

**Figure 9.5** shows the geopotential height perturbation amplitude, \(Z_g\), for PWs with varying ranges of phase speeds. While WPW \(Z_g\) decreases around 55 km after SSW onset in Figure 9.5a, WPWs still maintain a presence within the RSW. This is reflective of the limited phase speed range of WPWs allowed to be transmitted into the RSW. Since WPWs exist in the stratosphere during the normal winter, their \(Z_g\) values within the RSW are not shown to be an abnormal anomaly associated with SSWs (Figure 9.5d). However, enhanced WPW \(Z_g\) in the MLT are unique to SSWs with P(\text{Ab}|\text{SSW}) > 0.6.
Figure 9.5. Height-time plot averaged between 60°N and 70°N. (a-c) Geopotential height perturbation amplitudes, $Z_g$, are incremented by 100 m (d-f) Positive (negative) $Z_g$ anomalies are indicated by solid (dashed) black contours incremented by 100 m s$^{-1}$. The probability that the anomaly is abnormal is given by orange-shaded contours. In all sub-plots, the zero-wind line is indicated by a thick bold contour.

9.2.2 QSPWs

At day 0, the critical level becomes exposed outside of a region of reversed $\bar{q}_\phi$. The exposure of the critical level is not as clear from a height-latitude composite (Figure 9.1b) but is suggested to
occur as $q_\phi$ increases while the positioning of the zero-wind line remains steady approaching onset. The zero-wind line is the critical level for stationary PWs which contain much of the upward flux relative to traveling PWs (cf., Figures 9.4a-c). The height-latitude plot in Figure 9.6 better illustrates the exposure of the QSPW critical level. The EP flux vectors suggest the group velocity of PWs of all phase speeds with their divergence (convergence) in red (blue) contours suggesting the induced accelerations associated with their growth or dissipation. On day -5 (Figure 9.6a), the zero-wind line is embedded in a region of negative $q_\phi$ (grey shading) resulting in upward propagating PWs dissipating away from the zero-wind line. The exposure of the zero-wind line, the critical level for QSPWs, at day 0 (Figure 9.6b) greatly enhances the first term in the $n^2$ equation (Equation 7.3). As a result, PW group velocities are directed toward the mesospheric zero-wind line. Their subsequent convergence at the zero-wind line causes the isotach to drastically descend by more than 40 km. This is largely due to QSPWs, as they have the largest flux convergence in this region by a magnitude of 10 (c.f., Figures 9.4a-c). From 40-65km in Figure 9.4e, there is a large flux convergence anomaly associated with the descent of the zero-wind line with $P(\text{Ab}|\text{SSW}) > 0.5$. This is a substantially large probability given the varying position of the zero-wind line between SSWs.
As expected, Figure 9.5b shows a QSPW $\overline{Z}_g$ upon SSW onset exceeding 800 m. Interestingly, low values of $P(\text{Ab}|\text{SSW})$ and $\overline{Z}_g$ in Figure 9.5e suggests that these QSPWs amplitudes are not abnormal. Therefore, the amplitude of QSPWs alone is not a good indicator for SSW occurrence.

### 9.2.3 EPWs

Interestingly, flux divergence for EPWs at 75 km continues to increase after SSW onset (Figure 9.4c). This continued flux divergence is surprising from the over-reflection perspective since an upward-propagating EPW should encounter a critical level before the turning level, resulting in flux convergence or wave absorption (suggested in Figure 9.1b by the positioning of the green contour well outside of the grey-shaded region). Furthermore, a new region of EPW flux divergence emerges at 50 km. These two layers of flux divergence forming at the top and the bottom of the wind reversal are a common feature, appearing in most of our ensembles at varying NH latitudes and heights. Both at 75 and 50 km, a flux convergence anomaly exceeding 7
m s\(^{-1}\) day\(^{-1}\) with \(P(\text{Ab}|\text{SSW}) > 0.5\) suggests that this is a robust feature associated with SSWs (see Figure 9.4f).

The presence of EPWs in the westward wind should not be possible by upward-propagating EPWs since they would experience a critical level and dissipate in the lower stratosphere (illustrated in Figure 9.2b). As a common feature after SSW onset, the happenstance of nonlinear interaction is an insufficient explanation. However, another source mechanism could be at play such as the asymmetric dissipation of GWD (discussed in Chapter 6).

During a normal winter, westward GWD caps the polar jet and maintains the positioning of the stratopause (e.g., Limpasuvan et al., 2012). This is demonstrated in Figure 9.7 which examines GWD (shaded contours) with respect to \(\bar{u}\) (line contours). Figure 9.7 only looks at features of wavenumbers 0 and 1 to focus on the zonal-mean diagnostic and the relative effect of wavenumber-1 perturbations. While patterns of wavenumbers greater than 1 may be relevant for specific cases, wavenumber-1 perturbations are common surrounding SSWs resulting from both split and displaced polar vortices (Bancalá et al., 2012). When comparing several different SSWs, the vortex may be displaced or split over different longitudes. To help remove phase variability during vortex breakdown, values are composited with respect to the average phase of the stratospheric wavenumber-1 geopotential between 10 and 5 hPa such that the ridge is always centered at 0° longitude.
Figure 9.7. Altitude vs. relative phase shift of zonal GWD averaged between 60°N and 70°N. Before compositing, the average phase between 10 and 5 hPa of the wavenumber-1 geopotential higher perturbation are aligned such that the PW trough is centered at 180° longitude. (a,b) GWD is in filled contours and overlaid by the westward (eastward) wind velocity in thin (bold) contours. (b,c) Orange-shaded contours show the probability of an anomaly being significant (i.e., the value found at a coordinate and time with respect to SSW onset is unique to normal or SSW winters) and overlaid by positive (negative) anomalies in solid (dashed) black contours.

On day -10 near 75 km (Figure 9.7a), westward GWD caps the strong eastward winds. Expectedly, this GWD is not significantly different than during normal winters, illustrated by low probability values in Figure 9.7b. After SSW between 40 and 80 km, the stratospheric low-pressure system with strong eastward wind (thick black contours in Figure 9.7a) is replaced by a stratospheric high-pressure system with strong westward winds (thin black contours in Figure 9.7c) that constitute the RSW region. In Figure 9.7c, the westward wind favors eastward-propagating GWs,
which impose an eastward GWD above the RSW. Resultantly, the eastward GWD imposes an eastward wind in the mesosphere. This eastward GWD has a large wavenumber-1 component with the greatest GWD occurring between a 0° and 100° phase shift from the stratospheric PW ridge. The GWD 4 days after SSW near 75 km has an anomaly of > 60 m s\(^{-1}\) day\(^{-1}\) and a \(P(\text{Ab}|\text{SSW}) > 0.7\) (Figure 9.7d). As such, this feature is a reliable consequence of SSWs. The capping of the RSW by eastward GWD was also found by Limpasuvan et al. (2012).

A Hovmöller diagram of GWD overlaid with geopotential height, \(Z_g\), contours averaged from 0.1-0.01 hPa and 60-70°N (Figure 9.8a) shows the interaction of GWD with regions of low and high pressure (indicated by regions of low and high geopotential height respectively). As in Figure 9.7, diagnostics show wavenumber-0 and -1 features and are composited with respect to the stratospheric wavenumber-1 geopotential. Longitudes are repeated to better illustrate patterns across the periodic boundaries. At day 0, a region of high geopotential height becomes dominant. This coincides with the formation of the RSW layer associated with the movement of a high-pressure system over the pole (reflected by regions of negative \(\bar{q}_\phi\) in Figure 9.3a). Resultantly, as discussed for Figure 9.7, GWD switches from a westward forcing to an eastward forcing with a large wavenumber-1 component.
Figure 9.8. Altitude vs. relative phase shift of zonal GWD averaged from 60-70°N and 0.1-0.01 hPa. Before compositing, the average phase between 10 and 5 hPa of the wavenumber-1 geopotential higher perturbation are aligned such that the PW trough is centered at 180° longitude. (a) GWD is in filled contours and overlaid by geopotential height in black contours. (b) Orange-shaded contours show the probability of an anomaly being significant (i.e., the value found at a coordinate and time with respect to SSW onset is unique to normal or SSW winters) and overlaid by positive (negative) anomalies in solid (dashed) black contours.

GWs propagate through the stratospheric wavenumber-1 $Z_g$ ridge shown by the phase shift, and dissipate on the west (east) side of the mesospheric wavnumber-1 $Z_g$ ridge (trough) shown in reference to $Z_g$ contours. GWD counteracts the underlying westward winds (shown in Figure 9.7b) and thereby destroys the high-pressure system from west to east. This results in the
manifestation of an EPW, shown in Figure 9.8a by the eastward shift in the high-pressure system over time between day 0 and day 5. Resultantly, EPW growth can be seen in Figure 9.4c near 80 km. After day 5, eastward GWD subsides as a low-pressure system indicated by low values of geopotential height reforms in this region.

Interestingly, this interaction manifests an EPW that has a destructive interference with the high-pressure system resulting in an overall loss of geopotential height amplitude at 80 km. The growth of this wave restores a low-pressure system to the stratosphere. This restoration is particularly evident in Figure 9.3a as \( \tilde{q}_\phi \) becomes positive at this level. An abnormally large positive \( \tilde{q}_\phi \) (red-shaded contours) due to an enhanced vertical wind curvature (indicated Figure 9.3b by the light red shading) appears around day 5 and 85 km with \( P(\text{Ab} | \text{SSW}) > 0.6 \). The decrease in wave activity in this region due to GWD could produce a dynamically quiet region and support an enhancement of \( \tilde{q}_\phi \) due to relaxation of the atmosphere towards thermal wind balance. The radiative relaxation rate near the stratopause is roughly 5-7 days (Gille & Lyjak, 1986) which agrees with the time scale of the \( \tilde{q}_\phi \) enhancement.

In Figure 9.8, eastward GWD forcing is seen between 0° and 100° longitude a few days before onset. In Figure 9.4c, EPW growth also begins a few days before onset. Therefore, the EPWs generated from instability prior to SSWs and may be one contiguous feature ultimately enhancing the stability of the mesosphere. As pointed out in Section 9.2.2, the over-reflection of upward-propagating PWs induces wave growth from instability. These two source mechanisms are interconnected through the PV equation (Equation 7.1). Further research is needed to investigate the dominance of these source mechanisms in generating instability waves.

However, this is not the only evidence of EPW wave growth after SSW onset. In Figure 9.4c, EPW flux divergence occurs at the bottom of the RSW as well. The wind structure and \( \tilde{q}_\phi \) between days 0 and 5 in Figure 9.1b is similar to the schematic in Figure 9.2c. For this wind configuration, upward-propagating EPWs would be absorbed by an exposed critical level. However, Figure 9.3c suggests that wave growth occurs around 45 km. This region of flux divergence can be explained
when wave growth from asymmetric GWD is incorporated into the over-reflection perspective. Figure 9.2c shows EPW generated at the upper level of the reversed RSW. These EPWs become trapped with the upper and lower levels of the RSWs conducive to over-reflection. As at the upper boundary of the RSW, over-reflection at the lower boundary of the RSW would work to eliminate the high-pressure system and apply an eastward acceleration to the background winds. The trapping of these EPWs is supported by Figure 9.5d as $\overline{Z_g}$ for EPWs remains enhanced inside the RSW after onset. The eastward accelerations at the bottom and top of the RSW would modulate the thickness and duration of the RSW, albeit convergence from upward-propagating PWs would also play a role.

### 9.3 After Day 10

#### 9.3.1 WPWs and QSPWs

Like EPWs, the presence of WPWs inside the RSW decreases. As the westward winds become weaker, WPWs of slower phase speeds can propagate past the RSW unencumbered. The WPW $\overline{Z_g}$ above the RSW are maintained (see Figure 9.5a) and EP flux convergence in the mesosphere continues (see Figure 9.4a). WPW flux convergence in the mesosphere were found to contribute to SSW recovery by promoting the reformation and descent of the stratopause around 80 km (Limpasuvan et al., 2016). Figure 9.3a shows that the reformation of the stratopause begins around day 10 as the mesospheric winds become baroclinically unstable. The instability is indicated by a region of negative $\overline{\eta \phi}$ (yellow contours) in an eastward flow. The dominance and negative contribution of the baroclinic component is indicated by the dark red shading.

Interestingly, Figure 9.4b shows that the flux convergence of QSPWs in the MLT increases after day 5. Before day 5, QSPWs were absorbed by the RSW. Figure 9.4b shows that flux convergence from QSPWs still maintains the RSW after day 5 but decreases in strength. In Figure 9.5c, the region where QSPW $\overline{Z_g}$ is greater than 500 m extends between 30 and 70 km before and just after SSW onset. After day 5, this region has reduced in size and descended to the lower boundary of the RSW. While enhanced QSPWs at the lower boundary of the RSW, the value of $\overline{Z_g}$ is not
abnormal as larger QSPWs can also exist in the normal winter stratosphere; \( P(\text{Ab}|\text{SSW}) < 0.1 \) around day 10 at the lower boundary of the RSW in Figure 9.5e. However, negative \( \overline{Z_g} \) at 50 km in Figure 9.5e reflects the blocking of upward-propagating QSPWs by the bottom RSW boundary.

In Figure 9.3a, the composited \( \overline{q\phi} \) shows the turning level descends below the zero-wind line after day 10 which corresponds with the increase of QSPW flux convergence in the MLT in Figure 9.4b. Figure 9.4e shows the QSPW flux convergence increase in \( P(\text{Ab}|\text{SSW}) \), albeit not to a large value. However, this low \( P(\text{Ab}|\text{SSW}) \) value could result from a large variation in the positioning of the stratopause relative to SSW onset date. In Figure 9.5e, positive \( \overline{Z_g} \) anomaly with \( P(\text{Ab}|\text{SSW}) > 0.6 \) confirms that the presence of QSPWs in the mesosphere are associated with SSW recovery.

Therefore, the wave geometry presented in Figure 9.2d for QSPWs is possible and generally occurs late in SSW recovery as the RSW becomes weaker. As a result of QSPWs experiencing an evanescent region instead of a critical level, WPWs would experience a thicker evanescent region and become less likely to over-reflect. Additionally the max winds become weaker than -5 m s\(^{-1}\), WPWs faster than this phase speed will propagate past the RSW unencumbered, but not extract any energy from the background flow from over-reflection. Resultantly, WPW flux convergence in the MLT decreases while QSPW flux convergence increases after day 5. Limpasuvan et al. (2016) has shown that WPWs aid in the recovery of the stratopause and suggests that WPWs could be responsible for developing an elevated stratopause. Interestingly, the as QSPW flux convergence increases (and WPW flux convergence decreases) in the MLT, the zero-wind line associated with the stratopause descends back to its climatological positioning. The influence of PWs with different phase speeds on the positioning of the stratopause is still unknown, but currently extends above our model boundary. While further research is needed to understand stratopause recovery, our study offers a mechanism by which the RSW can modulate PW phase speeds present in the MLT during SSW recovery.
9.3.2 EPWs

In the scenarios shown by Figures 9.2b-d, EPWs are unable to propagate past the RSW and are trapped in the troposphere. This effect can be seen after day 5 in Figure 9.5d as $\bar{Z}_g$ is relatively low in the RSW. However, there is still a significant $\bar{Z}_g$ for EPWs in the mesosphere. This is likely due to eastward GWD. Unlike the scenario presented in Figure 9.2c, the upper RSW boundary is much lower and is unable to seed the region where EPWs can over-reflect. Without the over-reflection of EPWs to rapidly restore the stratospheric polar vortex (discussed in Section 9.2.3), the RSW can remain for prolonged periods of time.

After day 5, EP flux divergence associated with eastward GWD diminishes as the upper boundary of the RSW descends below 60 km. The larger density below 60 km would reduce the amount of drag produced from GW momentum deposition and therefore reduce the effectiveness of GWD as a source for EPW growth. In Figure 9.5d, a negative $\bar{Z}_g$ anomaly near 50 km with $P(\text{Ab}|\text{SSW}) > 0.6$ shows that the stratosphere is abnormally devoid of EPWs. From the over-reflection perspective, this illustrates the inability for EPWs to propagate directly into the RSW region unless asymmetric GWD generates EPWs from the upper boundary of the RSW.

The descended region of negative $\bar{a}_\phi$ remains closer to the lower boundary than the upper boundary of the RSW. These features suggest that the prolonged RSW in the stratosphere is being maintained by PW dissipation at the bottom of the RSW, while dynamics at top of the RSW act to restore the eastward flow. This is supported by schematics in Figure 9.2b-d where the bottom boundary of the RSW supports PW absorption from QSPWs and EPWs due to an exposed critical level. On the other hand, the top boundary of the RSW can relieve stability through the over-reflection of PWs inside the RSW region, eastward GWD, or a relaxation to thermal wind conditions due to a lack of wave activity. In Figure 9.3c, a significant ($P(\text{Ab}|\text{SSW}) > 0.8$) negative anomaly in $\bar{a}_\phi$ has its maxima located on the bottom RSW boundary around 35 km. Expectantly, instability and maintenance of the prolonged RSW is maintained by flux convergence at the bottom RSW boundary.
9.4 Summary and Discussion

In this study, used the over-reflection perspective can be applied to explain PW behaviors with wave geometries during SSWs. In doing so, we revealed several unique features related to these processes:

(1) Prior to SSWs, there is a tendency for over-reflection to produce waves with eastward phase speeds. This is illustrated by the persistent EP flux divergence around 70 km in Figure 9.4c.

(2) An unstable mesosphere inhibits the zero-wind line from PW absorption. On day 0 the mesosphere stabilizes ($\tilde{q}_\phi$ increases). The rapid descent of the zero-wind line is a result of the exposure of the critical level to upward-propagating planetary waves. A positive feedback loop is created by PWs interacting with a thinning evanescent region. This suggests that the persistence of tropospheric forcing may be just as important as the strength of tropospheric forcing in inducing SSWs.

(3) After day 0, GWD acts as a source mechanism on the upper boundary of the RSW. EPW can become trapped resulting in two layers of EP flux divergence at the top and bottom boundaries of the RSW. To our knowledge this is the first time this feature has ever been discussed.

(4) Over-reflection allows WPWs to tunnel into the RSW region, over-reflect and dissipate in the MLT. The production of WPWs from instability was also described by Limpasuvan et al. (2016).

(5) The wave geometry present at the bottom boundary of the RSW region may relate to the phase speed range of PWs in the MLT.

(6) The maintenance of RSW in the stratosphere is due to tropospheric forcing suggested by the concentration of negative $\tilde{q}_\phi$ in the lower portion of the RSW.

Kodera et al. (2016) categorizes SSWs as reflective or absorptive coinciding with how upward-propagating PWs interact with the lower critical level. The present study does not distinguish between these two types but shows that SSWs can have absorptive or reflective characteristics depending on its wave geometry during SSW recovery. Since this wave geometry can shift over time, a single SSW event could exhibit properties of both a reflective and absorptive SSW. Our study also explains why reflective SSWs, in Kodera et al. (2016), experienced a quicker recovery. A period of stronger over-reflection would have the properties of a reflective SSW, in which
resultant instability wave would act to restore the polar vortex by inducing an eastward acceleration.

From Equation 7.1, two mechanisms can support the growth of a PW ($q'_ω$): the background wind curvature indicated by $\bar{q}_φ$ and the nonconservative effects $X'$ resulting from forcings like unresolved GWD. Song et al. (2020) discussed wavenumber-2 EPWs prior to the January 2009 SSW and found a correlation between unresolved GWD and an increase in $q'_ω$. The study suggests that GWD drives the growth of wavenumber-2 EPWs (as opposed to instability) since its growth occurs away from the turning level. This study stands juxtaposed to Rhodes et al. (2021) in which shows wavenumber-2 EPW flux divergence from regions where the necessary conditions for instability are fulfilled. Upon further examination of Equation 7.1, nonconservative forcings can affect the wave amplitude and/or the stability depending on the background wind conditions. Near a wave’s critical level for example, the refractive index approaches infinity (Equation 7.3) forcing the group velocity of and the amplitude of the wave to approach zero. Instead of amplifying a wave, nonconservative forcing would act to enhance instability through $\bar{q}_φ$. In turn, this instability can generate PWs through over-reflection. Following Song et al. (2020), further evaluation of the source mechanisms that generate PWs in situ in the MLT, through the application of the potential vorticity equation, would be valuable in understanding how stability in this region is modulated.

In this study, we have noted the significant presence of WPWs in the MLT after SSW onset. Our study, like Limpasuvan et al. (2016), identifies instability as their source mechanism. Sassi et al. (2016) also identifies WPWs in the MLT and shows that they can significantly impact the meridional circulation, enhancing upwelling in the tropics and downwelling at the pole. This downwelling can result in a descent of nitric oxides produced from energetic particle precipitation over the pole, as shown in a January 2013 case study by Orsolini et al., (2017). In a case study of the January 2009 split SSW, Harvey et al. (2021) has reported that this enhanced descent of nitric oxides occurs in the troughs of WPWs in the MLT.
More research is needed on the application of critical-layer theory to the MLT and PW interactions with atmospheric boundaries like the stratopause. Our study attempts to implement critical-layer theory to the middle atmosphere by applying the wave geometries associated with the over-reflection perspective. This framework enabled us to discuss and explain PW interactions with various boundaries surrounding SSWs.
Chapter 10

Conclusion

Three research objectives were laid out in this dissertation: To (1) identify the source of EPWs before SSW onset, (2) assess the impact and uniqueness of mesospheric EPWs leading up to SSW onset, and (3) explore the behaviors of EPWs along with other PWs surrounding SSW and determine their possible source mechanisms.

The first objective was addressed in the January 2009 SSW case study (Chapter 8). Here, the source of EPWs occurring prior to SSW onset was identified to be mesospheric instability. To better understand how upward-propagating PWs (acting as incident waves) affected mesospheric stability, the over-reflection perspective was applied. A detailed discussion addressed how the formation of a double-maxima wind configuration prior to SSW onset enhanced the ability for PWs to over-reflect, resulting in unstable PW growth. The double-maxima configuration was formed by the interplay between GWD and PW forcing. Additionally, we showed that the association of EPWs with the double-maxima configuration was common by examining other case studies (Section 8.7.2). These EPWs occurred as both wavenumber-1 and -2 perturbations and were not unique to the type of SSW. As a result, we focused our analysis on the sources and impacts of unstable PWs surrounding all types of SSWs.

The second objective was addressed through other case studies (Section 8.7.2), a composite study (Section 8.7.3), and an ensemble study (Chapter 9). Through case studies, we found that EPW growth had a significant impact the wintertime $\vec{u}$, sometimes creating a tertiary $\vec{u}$ maxima. The composite study addressed the general propagation pattern of wavenumber-1 and -2 unstable EPWs. EPWs generally extracted westward momentum from the region of shear instability above the polar jet. The westward momentum was subsequently transported to regions of eastward flow, resulting in westward accelerations along the peripheries of the polar and subtropical jets. The ensemble study further illuminated the effect that EPWs had on mesospheric stability. From the over-reflection perspective, the EPW wave growth increases
mesospheric $\bar{q}_\phi$ and thus promotes stability in the mesosphere. The increase in $\bar{q}_\phi$ lead to an exposed critical level, resulting in a rapid descent of zonal-mean westward winds.

Finally, the third objective was addressed in our ensemble study (Chapter 9) as the over-reflection perspective was applied to better understand in-situ wave growth surrounding SSWs. The application of this perspective offered explanations of the presence of QSPWs and WPWs in the mesosphere after SSW onset. Additionally, the ensemble study identifies and addresses the source mechanisms of a unique EPW feature after SSW onset. GW filtering creates EPWs inside the westward stratospheric wind. The unique wind geometry results in EPWs becoming trapped within the RSW layer. To our knowledge, this is the first time that this feature has ever been analyzed.

To this end, this dissertation has advanced our knowledge of EPWs by investigating its source mechanisms and characteristics in case studies, a composite study, and an ensemble study. The over-reflection framework was applied to better illustrate how wave growth from instability was connected to the surrounding atmospheric variability. Ultimately, our analysis not only enhanced our understanding of EPWs prior to SSWs, but unstable waves surrounding SSWs in general. In doing so, we have made novel advancements on the application of critical layer theory to the middle atmosphere. This research will serve as the impetus for future research to better understand the source mechanisms and behavior of unstable PWs in the MLT.
Appendices

Appendix A: Variable Glossary

$A_ψ$: real amplitude of the streamfunction perturbation

$A$: wave-activity density

$\beta \equiv \frac{\partial f}{\partial y}$: coefficient used in $\beta$-plane approximation

$\beta_e$: effective beta; incorporates the meridional and vertical curvature of the background wind in addition to $\beta$.

c: wave phase speed (aka trace speed)

$[c_x, c_y, c_z]$: wave phase speeds in the x-, y-, and z-directions, respectively

$c_g$: wave group velocity

$[c_{gx}, c_{gy}, c_{gz}]$: wave group velocities in the x-, y-, and z-directions, respectively

$c_p$: specific heat capacity at constant pressure

$D$: frictional and diabatic acceleration on the wind

$\varepsilon$: the square of the ratio of the Coriolis parameter and the buoyancy frequency, $f_0^2 / N^2(z)$

$\varepsilon(t)$: phase shift as a function of time

$\zeta$: absolute vorticity

$[\phi, \theta, z]$ : longitude, latitude, and radial directions such that $\phi$ indicates the angle eastward from the prime meridian, $\theta$ is the angle north of the equator, and $z$ is the height above the surface of the sphere

$\theta_T$: potential temperature

$\rho$: density
\( \psi \): horizontal stream function such that
\[
u' = -\frac{\partial \psi}{\partial y} \quad \text{and} \quad v' = \frac{\partial \psi}{\partial x}
\]

\( \Psi \): amplitude of the streamfunction

\( \omega \): frequency

\( \Omega \): Earth’s rotational frequency

\( \tilde{F} \equiv [F_x, F_y, F_z] \): Eliassen-Palm flux on a tangent plane

\( f \): Coriolis parameter

\( g \): gravitational acceleration constant \( \approx 9.81 \frac{m}{s^2} \)

\([k, l, m]\): spectral wavenumbers in the \( x_T, y_T, \) and \( z_T \) directions respectively

\( N_B^2 \): Brunt–Väisälä frequency, or the buoyancy frequency, which is the angular frequency of an air parcel oscillating in a statically stable atmospheric background

\( n^2 \): index of refraction

\( O_3 \): high order / non-linear wave accelerations

\( p \): pressure

\( Q \): diabatic heating rate per unit mass

\( q \): potential vorticity

\( R_e \): the radius of the Earth \( \approx 6,317,000 \) m

\( T \): Temperature

\( \langle T \rangle \): Mean temperature of air column. For a scale height of 7 km, this temperature is approximated to be 240 K. See Andrews et al. (1987) for full definition.

\([u, v, w] = \tilde{v} \): wind velocity vector in the longitudinal, latitudinal, and radial directions

\([u_g, v_g] = \tilde{v}_{\text{g}} \): horizontal geostrophic wind velocity vector
Φ: geopotential

\[ [x_T, y_T, z_T] : \text{orthogonal directions on the tangent plane of a sphere such that } x_T \text{ is pointing in the eastward direction, } y_T \text{ is pointing in the northward direction, and } z_T \text{ is pointing in the } \hat{z} \text{ direction. The } T \text{ subscripts stands for tangent-plane. Later in the variables are simplified to } [x, y, z]. \]

\( Z_g \): geopotential height

\( \bar{Z}_g \): geopotential height amplitude

**Appendix B: Potential Vorticity**

A vorticity equation can be derived through the primitive equations by taking the partial derivative of the x-direction momentum equation with respect to y and the partial derivative of the y-direction momentum equation with respect to x. Geostrophy may also be assumed such that the geostrophic winds follow along the stream function \( \psi = f_0^{-1} \Phi \). This process is explained in detail using quasi-geostrophic primitive equations to derive the quasi-geostrophic absolute vorticity equation by Andrews et al. (1987, p. 121).

\[
q_g = 2\Omega \sin(\phi_0) + 2\Omega a^{-1} \cos(\phi_0)y + \psi_{xx} + \psi_{yy} + \rho_0^{-1}(\rho_0 \varepsilon \psi_z)_z \tag{B1}
\]

where subscripts of a coordinate variable indicate the partial derivative of the subscripted variable with respect to the subscript. The summed term \( 2\Omega \sin(\phi_0) + 2\Omega a^{-1} \cos(\phi_0)y \) is the beta-plane approximation for planetary vorticity, \( \varepsilon = f_0^2 / N^2(z) \) is the square of the ratio of the Coriolis parameter and the buoyancy frequency, and \( \psi \) is the stream function. The third and fourth components on the right-hand side represent the horizontal curl of the wind vector and the last component on the right-hand side represents the dynamical contribution of “stretching” to the vorticity.

Finally, we find the basic northward potential vorticity gradient by taking a partial derivative of the mean of the geostrophic potential vorticity with respect to y.
\[
\bar{q}_y = 2\Omega \alpha^{-1} \cos(\phi_0) - \bar{u}_{yy} - \rho_0^{-1}(\rho_0 \epsilon \bar{u}_z)_z
\]  

(B2)

A couple terms are eliminated since the mean meridional wind is zero \((\bar{v}_{xy} = 0)\) and the meridional gradient of the constant \(f_0\) becomes zero. A similar process can be done with spherical coordinates to get the meridional gradient of the mean quasi-geostrophic potential vorticity.

\[
\bar{q}_\phi = 2\Omega \cos(\phi_0) - \left(\frac{1}{a \cos(\phi_0)} \left(\bar{u}_y \cos(\phi_0)\right)\phi\right) - \frac{a}{\rho_0} \left(\frac{\rho_0 2\Omega \sin(\phi_0)}{N^2} \bar{u}_z\right)_z
\]  

(B3)

The first term accounts for the rotational effects of the earth, the second component accounts for the horizontal relative vorticity due to horizontal motions, and the third component accounts for the horizontal vorticity due to vertical motions (or the stretching/compressing of an air column).

When considering shear instability, it is often convenient to consider the streamfunction as a wave that is able to grow in time. Both the wave and its phase speed are assumed to have imaginary components. Under this consideration, the following equation must be satisfied (the full derivation is explained by Houghton (1977, pp. 122–123).

\[
\int_A^B \int_A^B \bar{q}_y \frac{|\psi|^2}{|\bar{u} - c|^2} dy = 0
\]  

(B4)

where \(c_i\) is the imaginary part of the phase speed which must be positive for an instable wave.

Since the integral is equal to zero, \(\bar{q}_y\) must change sign at some point in the interior of the flow for an unstable wave to satisfy the conservation of potential vorticity. However, although this is necessary condition, a change of sign does not indicate a presence of PW wave breaking, but the potential.

**Appendix C: PW Generation Through Instability: Counter-propagating Rossby Waves**

First, consider the linearized PV equation for a conservative flow.

\[
q_t + U q_x + v \bar{q}_y = 0
\]  

(C1)
Here, subscripts of directional unit vectors imply partial derivatives, $q$ is the PV, and $\bar{q}_y$ is equal to the $\beta_e$ in Equation 3.21. The basic state zonal velocity is represented by $U$.

From the CRW perspective, a wave is described as a perturbation in PV that lags a perturbation in meridional velocity by a phase of $\frac{\pi}{2}$ such that:

$$q = |q|(y, z, t)e^{i(kx + \epsilon(t))}$$  \hspace{1cm} (C2a)

and

$$v = |v|(y, z, t)e^{i(kx + \epsilon(t) - \pi/2)}$$  \hspace{1cm} (C2b)

Here, $q$ and $v$ are represented as zonal perturbations with amplitudes $|q|$ and $|v|$. The phase shift with time is represented by $\epsilon(t)$. If Equations C2a, b are substituted into Equation C1, then a relationship can be developed between the relative phase speed $(U - c_x)$ and $\bar{q}_y$.

$$U - c_x = \frac{\bar{q}_y}{k} \left( \frac{|v|}{|q|} \right)$$  \hspace{1cm} (C3)

Here, $c_x = k^{-1} \frac{\partial \epsilon(t)}{\partial t}$. In this case, a wave with a real-valued relative phase speed (a propagating wave) must have the same sign as $\bar{q}_y$. This relationship was also realized from the over-reflection perspective through the first term of the refractive index (Equation 3.27b).

The CRW considers the interaction of two waves, each with unique perturbations in $q$ and $v$, across a critical level. These waves can be visualized through Figure C1 where each wave has perturbations in potential vorticity along the zonal axis. Solid arrows represent the meridional velocity at their local maximums and minimums while dashed arrows represent the shear induced on one wave by the meridional velocity of another. While the background conditions, $\bar{q}_y$ and $U$, in Figure C1 are specified, the interaction of waves regardless of background conditions can be first generally considered by assuming the linearized PV is conservative (as in Equation 3.27).
First, Equations C2a,b can be substituted into Equation C1. Then the real and complex components can be separated into equations that describe the change in amplitude and phase of the PV perturbations induced on one wave by another wave. While this is described in more detail by Harnik and Heifetz (2007), the result becomes:

\[
\Delta |q_1| = -\bar{q}_y |v_2| |q_1| |q_2| \sin(\epsilon) \tag{C4a}
\]

\[
\Delta |q_2| = \bar{q}_y |v_1| |q_1| |q_2| \sin(\epsilon) \tag{C4b}
\]

and

\[
\Delta \epsilon_1 = -kU_1 + \bar{q}_y |v_1| |q_1| + \bar{q}_y |v_2| |q_2| |q_1| |q_2| \cos(\epsilon) \tag{C5a}
\]

\[
\Delta \epsilon_2 = -kU_2 + \bar{q}_y |v_2| |q_2| + \bar{q}_y |v_1| |q_1| |q_2| |q_1| \cos(\epsilon) \tag{C5b}
\]

Here, \(\Delta\) indicates the change induced on a variable and subscripts indicate which wave the variable is referencing. By itself \(\epsilon\) represents the phase difference between the two waves \(\epsilon_2 - \epsilon_1\).
Equations C4a and C4b show that if \( \bar{q}_{y_1} \) and \( \bar{q}_{y_2} \) are of opposite sign, then mutual wave growth can occur. For this special case of wave growth, the third terms in Equation C5a and C5b reveal a phase-locking mechanism between the two waves: the cosine in the third term changes the sign and magnitude of the third term depending on the direction and magnitude of the phase difference. Thus, the CRW perspective provides a mathematical conceptualization of wave growth across a critical level.

**Appendix D: Identification and Classification of SSWs**

Other polar vortex characteristics are used to classify SSWs beyond the presence of an elevated stratopause. Typically, without PW presence in the wintertime stratosphere, the polar vortex is expected to be nearly zonally symmetric, with eastward circumpolar flow established by the temperature contrast between the cold polar region and warm low latitudes. With SSW onset, the vortex can become displaced off the pole, manifesting in the appearance of a zonal wavenumber-1 eddy pattern or become split with a zonal wavenumber-2 eddy pattern. Spectral analysis is typically used to characterize the eddy amplitude and subsequently classify the SSW type (i.e., split or displaced SSWs). Ideally, a displaced SSW would have a spike in wavenumber-1 amplitude and split SSW a peak in wavenumber-2 amplitude.

**D.1 Computation of Circulation on a Spherical Grid**

The circulation can be computed through the application of Stokes theorem in spherical coordinates.

\[
\oint_C \mathbf{v} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{v}) \cdot d\mathbf{S} \\
(D1)
\]

Stokes theorem states that the circulation of an enclosed path, which is the integral of the velocity component in the direction of the path along a surface, is equal to the areal sum of vorticity in the direction of the area vector on a surface bounded by the enclosed path. In the first definition, the direction and location of each point along the enclosing path must be known. For the second definition, only the coordinates that define the enclosed surface need to be
known. Because our data is spherical and discrete, the latter definition of circulation is more readily computed.

The curl in spherical coordinates is calculated in reference to spherical coordinates \((\phi, \theta, \text{ and } z)\). Before taking the curl, the scale factors must be considered.

\[
h_\phi = R_e \cos \theta \\
h_\theta = R_e \\
h_z = 1
\]

On an isobaric spherical surface, the area vector is approximately the unit vector in the radial direction, \(\hat{z}\). Therefore, only the radial component of the curl is needed. Additionally, each discrete point \(|d\vec{S}|\) on the surface area represents varies depending on latitude such that

\[
|d\vec{S}| = R_e^2 \cos \theta \, d\phi d\theta \rightarrow \Delta S = R_e^2 \cos \theta \, \Delta \phi \Delta \theta 
\]

Therefore, the circulation surrounding an area of a discrete set of data of size \((n, m)\) can be defined as:

\[
\text{Circ.} = \sum_{N=0}^{n-1} \sum_{M=0}^{m-1} \frac{1}{R_e \cos \theta_M} \left( \frac{\Delta v}{\Delta \phi_N} - \frac{\Delta (\cos \theta_M u)}{\Delta \theta_M} \right) R_e^2 \cos \theta \, \Delta \phi_N \Delta \theta_M 
\]  

Simplifying, the scaling components from the curl and differential area cancel.

\[
\text{Circ.} = R_e \sum_{N=0}^{n-1} \sum_{M=0}^{m-1} \left( \frac{\Delta v}{\Delta \phi_N} - \frac{\Delta (\cos \theta_M u)}{\Delta \theta_M} \right) \Delta \phi_N \Delta \theta_M 
\]

Theoretically, calculating circulation through the Stokes theorem should give the same result as calculating the circulation directly. To directly calculate the circulation, the path of the maximum flow has to be known beforehand. Using the Stokes Theorem, the shape of the flow can be changed such that all the points within the maximum flow are known. The latter technique is particularly useful for calculating the circulation of a vortex that is constantly changing shape.
D.2 SSW Classification Algorithm

A new method of classification is proposed to show the state of the vortex at all levels and times during the SSW. Unlike other classification systems, geostrophy is assumed such that the vortices can be identified using only the geopotential variable.

The data is smoothed 10° longitude by 5° latitude and the local minimums are defined at points with the lowest geopotential within 50° longitude and 28° latitude.

Only local minimums above 45°N are considered to avoid the equatorial region where geostrophy is no longer a sufficient approximation. While it may be more accurate to search for geopotential minimums within a specified distance across the surface of a sphere, we found that it is sufficient to simply search by degrees using the spherical data. It is only necessary to store up to 4 local minimums although more may be stored if desired.

![Figure D1: Model vortices with a contour of the geostrophic wind components (m s⁻¹). Arbitrary X and Y directions are given in meters. Bolded white stars show local minimums in geopotential and dotted white lines follow the region of maximum flow surrounding the local minimums. Model vortices are defined as two upside-down Gaussian curves with a standard deviation of 250 km and a flat circular bottom inserted at the bottom of the Gaussian curve. The flat bottoms have radii of 500 km and 250 km. The depth of each vortex is 100 km. The progression from the left image to the right images show the two vortices colliding.](image-url)
An artificial geopotential field was created to test the classification algorithm. In Figure D1, local minimums are indicated by asterisks, illustrating that the current method can distinguish between one and two local geopotential minimums. Now that the local minimums of the vortices are identified, the points inside the vortices must be located.

From the local minimum, point locations inside the vortex are gathered by searching outwards along the isobaric surface for maxima in geostrophic wind. A tolerance of 0.25 km is used to include maximum points and exclude small variations in geostrophic wind. This process is repeated for any additional local minimums.

From Step 2, the points inside the vortices are located, including the points at the edge of the enclosed region (indicated by the white dotted line in Figure D1). When the vortices are well separated, all points inside the vortex are included. When the vortex is not well separated, some points may be excluded due to a local minimum inside the vortex larger than the tolerance (0.25 km). However, these areas should not have a large contribution to the circulation.
Figure D2. Colored contours represent the geostrophic wind (m/s). M1 represents “method 1”: First assume a circle, calculate the average radius and the maximum velocity, and finally approximate the circulation directly by assuming the maximum velocity is constant along the circumference of the circle. M2 represents “method 2” in which the circulation is calculated through the circulation approximation via Stokes Theorem discussed in Appendix C. Although arbitrary, if X and Y distances are in meters, circulation units are in m$^2$ s$^{-1}$. Black filling represents the region in which material is geostrophically conserved inside the vortex, contained by a maxima in geostrophic wind.

Figure D2 shows all points located within the vortices in black. When the vortices are well separated, the black points fill in the entire inner region of the vortices. In Figure D2c, the vortices become too close such that there can only be one geopotential minimum in the area. Therefore,
our algorithm only locates points spreading out from the local minimum and excludes some points related to the dip in geopotential in the top right portion of the vortex. However, these points are not in regions of high vorticity and should not significantly alter the circulation.

Two methods for calculating circulation are used to compare. The first method (M1) assumes a circular vortex. The average radius of the vortex, i.e. average distance the edge points are away from the local minimum, and the maximum velocity are used to infer the circulation around the vortex, $\mathcal{C}_{circ} = 2\pi r v_{max}$. The second method (M2) uses the Stokes theorem as described in Appendix C. Since the individual vortices are circular, Method 1 is accurate while the vortices are well-separated or superimposed (Figures D2a, D2d, D2f). In these scenarios, Method 1 is approximately the same as Method 2, although Method 2 tends to slightly underestimate the circulation. In Figure D2b, Method 1 over-estimates the circulation as the vortex interaction extends the average radius of the vortex. Using Method 2, the circulations calculated in Figure D2b are slightly less than the circulations in Figure D2a; this is expected as the vortices in Figure D2b are destructively interfering. In Figure D2c, both methods produce a similar circulation approximation. These values are reasonable considering the constructive interference of the two vortices. Finally, both Methods capture the circulations added together Figure D2d as the vortices become superimposed. Overall, Method 2 proves to be robust at capturing the circulations of the two vortices as they are interacting.

If any point locations overlap, these vortices are the same and the vortex with the lowest geopotential should be chosen as the dominant vortex. Therefore, the classification system would not classify Figure D2b as two distinct vortices.

The vortices are determined to be separate if they are separated by two maxima in geostrophic wind. This separation indicates that material cannot freely escape at the boundaries of the vortex or be exchanged between two vortices without an ageostrophic forcing. If any point locations selected from Step 2 overlap, these vortices are the same and the vortex with the lowest geopotential should be chosen as the dominant vortex. For example, the classification system would not classify Figure D2b as two distinct vortices. Once the number of vortices is
distinguished, the circulation of the vortex/vortices can be determined. Circulation within the polar vortex can be calculated using the Stokes Theorem (discussed Appendix C).

If multiple vortices are present, compare the circulation of each to the vortex with the largest circulation, i.e. the circulation of the dominant vortex. If the circulation of the vortex is less than half the circulation of the dominant vortex, then this vortex should be excluded.

Finally, the number of vortices present and the circulation of those vortices can be recorded for a specific height and time. This process can be repeated for other heights and times.

![Figure D3](image1.png)

**Figure D3.** Polar plots of low-pressure system detection algorithm at approximately 10 hPa. Overlying continent boundaries, the thin black lines are geopotential height contours (km). The bold black lines outline the detected regions of influence of the low-pressure system(s), i.e. the regions where there is an increasing geopotential gradient and thus a sufficient force to retain material within the regions of maximum wind outlining the vortex. If the regions connect, the vortices are not independent of one another. Bold black stars indicate points of low-pressure.

For real data, the algorithm is sufficient to detect how many low pressure systems occur over the pole. For the 22nd and 23rd of January, while two low-pressure systems are present above the pole, the regions outlined by their maximum winds are connected. This means that material is able to follow geostrophic flow between the two low-pressure systems and their circulations are dependent on each other. On the 24th of January, the circulations of the vortices are independent.
and the onset of the vortex split occurs. Note that this occurs about 2 days after the onset date as the onset date is defined by wind reversal aloft at 1.0 hPa.

D.3 Development of an SSW Classification Method

Ideally, spectral analysis done at specific latitudes will reveal a clear dominance in wavenumber 2 for a split SSW, especially like the January 2009 event. However, the optimal latitude band to diagnose the contribution of each wavenumber will vary from case to case as the vortex may be more or less constrained about the pole. This method assumes that a vortex split would be symmetric about the pole, which is not necessarily the case leading up to SSW onset as the polar vortex tends to wobble. Additionally, the time when a vortex splits is difficult to discern only by looking at wavenumber amplitudes. As noted below, the vortex evolution around SSW onset can be complex with the vortex displacing and splitting at different altitudes and changing from one type to another.
**Figure D4.** Amplitude of zonal wavenumber 1 (top left) and wavenumber 2 (bottom left) geopotential height perturbations (km) averaged from 60°N to 90°N. Thin black contours in wave amplitude plots increment by 0.01 km. On wave amplitude plots, the thick solid black line indicates the zero-wind line and the dashed horizontal line highlights the 10 hPa level. Height and time of polar plots (right) in descending chronological order are specified by magenta stars on the wave amplitude plots. Solid black contours in polar plots indicate geopotential height (km) incremented by 0.2 km. Blue and red regions show the relative low-pressure and high-pressure regions, respectively.
Consequently, wavenumber-1 and wavenumber-2 amplitudes of geopotential or meridional wind provide evidence that only suggests a split or displaced SSW occurrence. This ambiguity can be illustrated in Figures 8.4 and 8.5 which show the zonal wavenumber-1 and -2 geopotential height amplitudes (left plots) as well as a top-down view of the polar vortex structure (right plots). Zero-wind lines (bold lines on the left plots) indicate wind reversal while the amplitudes suggest the state of the polar vortex. Both Figures 8.4 and 8.5 have been categorized as split SSWs by Albers & Birner (2014). Classification of SSW was difficult by looking at the wavenumber amplitudes alone. Instead, it was necessary to use corresponding polar plots to determine if any splitting occurs. In both examples, there was a large wavenumber-1 component relative to the
wavenumber-2 component during onset. However, both vortices eventually split with the wavenumber-2 amplitude remaining relatively small. In December 1987 (Figure D4), the vortex did not split until well after wind reversal as the weakened low-pressure system broke down around the dominant high-pressure system (highlighted in red in the geopotential height contour). Therefore, the SSW during the winter of 1987 could be categorized as displaced.

Similarly, in Figure D5, zonal-mean zonal wind reversal occurred as a result of the low-pressure system being displaced. In this scenario, after displacement, the low-pressure system split into two vortices as the vortex was caught between a dominant high-pressure system moving poleward and a persistent high-pressure system over Western Asia. This made classification more difficult since the wind-reversal was not necessarily due to the vortex splitting, but the splitting vortex was associated with the initial wind-reversal. For example, a study may categorize this SSW as a displacement because the cause of wind-reversal at 10 hPa was due to the polar vortex moving off the pole. However, the background wind configuration, two high-pressure systems and two low-pressure systems, was akin to that of a split case; therefore, precursory conditions related to the structure of the vortex and the recovery of the vortex could resemble that of a split case.

Several alternative methods for SSW classification have emerged over the years. Once a SSW event is established, Charlton and Polvani (2007) identified various vortices by locating their edges using the QGPV gradient. From the closed path around each vortex edge, they computed the circulation strength. If two vortices are present and the circulation of one vortex is greater than half that of the other vortex, then the SSW event is classified as a split SSW. Here, the vortex with the larger circulation is assumed to be associated with the main wintertime polar vortex. Otherwise, the SSW event is classified as a displaced SSW. Mitchell et al. (2013) proposed another notable classification algorithm. Assuming an elliptical shape to the polar vortex, these authors can delineate the vortex as split, displaced, or stable while allowing for the classification of “mixed” events in which the vortex is neither classically displaced nor split. Recently, Lawrence and Manney (2018) employed a visualization algorithm to distinguish vortices based on potential
vorticity. In all, these methods require the full horizontal wind components when determining the vortex characteristic.

Here, a new method is proposed for SSW classification by identifying cyclonic vortices characterized by geostrophic winds. Based on their horizontal spatial scales (> 5,000 km), the dynamics around the polar vortex before SSW and potential vortices formed with SSW onset are expected to be governed largely by geostrophy. Using the geopotential field (smoothed with a 10° x 5° longitude-latitude running mean) at each isobaric surface, a vortex’s geographical center was initially found by its local geopotential minimum within a range of 50° longitude and 28° latitude. From each local geopotential minimum, we searched outward along the isobaric surface for the surrounding region of maximum geostrophic wind speed. A tolerance of 0.25 m s⁻¹ was applied such that variations in geostrophic wind smaller than the tolerance would be incorporated into the vortex region. If the loci of maximum wind speed of one vortex overlapped those of another vortex, then the circulations of the vortices were deemed dependent on one another and the vortex with the lowest geopotential is considered the dominant one. More information on the classification algorithm can be found in Appendix D.

The closed path tracing the wind speed maxima and the wind vector along that path constituted the circulation of the identified vortex at each altitude. From the Stokes Theorem, we calculated the circulation as an area integral of the absolute vorticity within the region enclosed by the path. If multiple vortices are present, the one with the largest circulation was identified as the dominant vortex which is presumably the polar vortex. The other identifiable vortices (if present) needed a circulation value of at least half that of the dominant one. More information on the calculation of vortex circulation can be found in Appendix C. At each altitude and time, the number of vortices (and their corresponding circulations) were determined, as illustrated in Figure D6.
Figure D6. Preliminary SSW classification plots for the winter of (a) 1987-1988 and (b) 1998-1999. Blue color-filled contours depict geopotential height amplitude (km) averaged between 60° N and 90° N and indicates the extent of vortex displacement; note these contours drawn where only one dominant vortex in the polar region exists. The occurrence of 2, 3, or 4 vortices are expressed by orange, green, and red points respectively. Onset dates are defined as the day of wind-reversal at 1 hPa.

The proposed method gave similar results to Mitchell et al. (2013) and Charlton and Polvani (2007) with some exceptions. Figure D6 demonstrates the altitude-time distribution of the number of vortices (between 1-4) during the same winters shown in Figures 8.1 and 8.2. Consistent with Figures 8.1 and 8.2, Figure D6a illustrates that the wind reversal during SSW was closely linked with the displaced vortex (i.e., large wavenumber-1 amplitude) in the upper stratosphere. However, Figure D6a reveals more information of the vortex evolution as SSW occurs. In December 1987, the displaced polar vortex led to zonal-mean wind reversal that propagated down into the lower stratosphere. As SSW occurred, the displaced vortex split into two (or more) vortices in lower stratosphere and near the stratopause, while, in the mid-stratosphere, the vortex shifted back over the pole. Charlton and Polvani (2007) categorized the December 1987 case as a split SSW and Mitchell et al. (2013) categorized the case as a mixed event.
Likewise, the February 1999 SSW (Figure D6b) began with a displaced vortex that quickly split almost immediately after the SSW onset. The split vortex occurred over a broad altitude range in the westward wind regime (bounded by the black zero-wind line). This case was classified as a split SSW by Charlton and Polvani (2007) and a displaced SSW by Mitchell et al. (2013). Figure D6 shows that vortex evolution can be complex. Depending on the timing of the zonal-mean zonal wind reversal (which is latitude dependent) as shown in Figures 8.4 and 8.5, one could argue that the February 1999 SSW was split or displaced.

Our method allows us to classify the polar vortex based on the overall structure. For December 1987, vortex splitting occurred many days after the wind reversal and the splitting did not occur throughout the stratosphere. This clearly shows that the structure of the vortex during SSW was more akin to displaced SSWs. In contrast, the February 1999 SSW exhibited vortex splitting immediately after wind-reversal at all altitudes in the stratosphere, suggesting that the polar vortex had an inherent split structure and should be classified as such.

As an additional example of our classification method, Figure D7 shows vortex evolution during the January 2006 and 2009 warmings which have been consistently classified as displaced SSW and split SSW, respectively. Between 40 km and 60 km, Figure D7a clearly demonstrates that the initial wind reversal in 2006 occurred as the vortex displaced off the pole. We also see a tendency for the vortex to split into multiple vortices near the stratopause shortly after SSW onset as the wind anomalies descended toward the tropopause. However, this split feature did not extend all the way through the stratosphere. For the January 2009 case, Figure D7b clearly reveals vortex splitting in tandem with SSW onset. The vortex splitting occurred at nearly all altitudes in the stratosphere. The vortex structures of December 1987 and January 2006 were similar and distinct from the vortex structures of February 1999 and January 2009.
Using these methods of identification and classification (reviewed above), composites of displaced and split SSWs can be constructed. Albers and Birner (2014) identified at least 15 displaced SSWs and 6 split SSWs between 1979 and 2013.

**D.4 Classification Results in the WACCM-SD Run**

Major SSW events from WACCM-SD between 1979 and 2013 were identified based on the criteria noted in Section 7.3. Listed in Table D1, the identified major SSWs fit the WMO criteria. However, not all of them had an elevated stratopause in their recovery phase. Mostly, ES-SSWs are a subset of the WMO-defined SSWs (hereafter, WMO-SSW). Each major SSW event was further classified based on the number of developed vortices during SSW onset at 1 hPa. In total, 4 split WMO-
SSW events and 14 displaced WMO-SSW events were found between 1979 and 2013. Of these, there are 4 split ES-SSWs and 12 displaced ES-SSWs.

<table>
<thead>
<tr>
<th>Onset Year (YYYY)</th>
<th>Onset Date (MM-DD)</th>
<th>ES-SSW?</th>
<th>WMO Major SSW?</th>
<th>Classification by Present Study</th>
<th>Classification from Albers et al. (2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>02-21</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>1984</td>
<td>12-30</td>
<td>Yes</td>
<td>Yes</td>
<td>Split</td>
<td>Split</td>
</tr>
<tr>
<td>1987</td>
<td>01-22</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>1987</td>
<td>12-06</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Split</td>
</tr>
<tr>
<td>1989</td>
<td>02-19</td>
<td>Yes</td>
<td>Yes</td>
<td>Split</td>
<td>Split</td>
</tr>
<tr>
<td>1995</td>
<td>01-27</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Not Present</td>
</tr>
<tr>
<td>1997</td>
<td>12-23</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Not Present</td>
</tr>
<tr>
<td>1998</td>
<td>12-14</td>
<td>No</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>1999</td>
<td>02-23</td>
<td>No</td>
<td>Yes</td>
<td>Displacement</td>
<td>Split</td>
</tr>
<tr>
<td>2001</td>
<td>01-24</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>2001</td>
<td>12-23</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>2002</td>
<td>02-13</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>2003</td>
<td>12-20</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>2006</td>
<td>01-09</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>2009</td>
<td>01-22</td>
<td>Yes</td>
<td>Yes</td>
<td>Split</td>
<td>Split</td>
</tr>
<tr>
<td>2010</td>
<td>01-24</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Displacement</td>
</tr>
<tr>
<td>2012</td>
<td>01-13</td>
<td>Yes</td>
<td>Yes</td>
<td>Displacement</td>
<td>Not Present</td>
</tr>
<tr>
<td>2013</td>
<td>01-05</td>
<td>Yes</td>
<td>Yes</td>
<td>Split</td>
<td>Not Present</td>
</tr>
</tbody>
</table>

Table D1. Identification and classification of SSWs from 1979 to 2013. SSW onset dates were defined by the point of wind reversal at 1 hPa (Limpasuvan et al., 2016). Each event was assessed to be a WMO-SSW and ES-SSW using criteria set by the World Meteorological Organization (WMO) and Limpasuvan et al. (Limpasuvan et al., 2016) respectively. The SSW type was classified by the shape of the vortex at onset and compared to the type of SSWs collected from multiple studies by Albers & Birner (2014).

The majority of SSW classifications agree with those of Albers & Birner (2014). As noted in Section D.3, discrepancy can arise due to the complexity of the vortex; there is disagreement in the community about the SSW type that occurred on December 1987 and February 1999 (red text in
Table D1). The well-known split SSW of January 1979 was also considered, but the zonal-mean zonal wind reversal, averaged between 70°N and 90°N, did not propagate down to 10 hPa. Therefore, it would not be an ideal reference case from which to base our ensembles.
List of References


Charney, J. G., & Drazin, P. G. (1961). Propagation of planetary-scale disturbances from the lower into the


Holton, J. R. (1982). The Role of Gravity Wave Induced Drag and Diffusion in the Momentum Budget of


from 1850 to 2005 simulated in CESM1(WACCM). Journal of Climate, 26, 7372–7391. https://doi.org/10.1175/JCLI-D-12-00558.1


