THE ARCTIC POLAR-NIGHT JET OSCILLATION

by

Peter Hitchcock

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Copyright © 2012 by Peter Hitchcock
The eastward winds that form each winter in the Arctic stratosphere are intermittently disrupted by planetary-scale waves propagating up from the surface in events known as stratospheric sudden warmings. It is shown here that following roughly half of these sudden warmings, the winds take as long as three months to recover, during which time the polar stratosphere evolves in a robust and predictable fashion. These extended recoveries, termed here Polar-night Jet Oscillation (PJO) events, are relevant to understanding the response of the extratropical troposphere to forcings such as solar variability and climate change. They also represent a possible source of improvement in our ability to predict weather regimes at seasonal timescales.

Four projects are reported on here. In the first, the approximation of stratospheric radiative cooling by a linear relaxation is tested and found to hold well enough to diagnose effective damping rates. In the polar night, the rates found are weaker than those typically assumed by simplified modelling studies of the extratropical stratosphere and troposphere. In the second, PJO events are identified and characterized in observations, reanalyses, and a comprehensive chemistry-climate model. Their observed behaviour is reproduced well in the model. Their duration correlates with the depth in the stratosphere to which the disruption descends, and is associated with the strong suppression of further planetary wave propagation into the vortex. In the third, the response of the zonal mean winds and temperatures to the eddy-driven torques that occur during PJO events is studied. The collapse of planetary waves following the initial warming permits
radiative processes to dominate. The weak radiative damping rates diagnosed in the first project are required to capture the redistribution of angular momentum responsible for the circulation anomalies. In the final project, these damping rates are imposed in a simplified model of the coupled stratosphere and troposphere. The weaker damping is found to change the warmings generated by the model to be more PJO-like in character. Planetary waves in this case collapse following the warmings, confirming the dual role of the suppression of wave driving and extended radiative timescales in determining the behaviour of PJO events.
For Lucy Pickard, my Grandma Africa.
Acknowledgements

Drink. It
is what you will have
to remember:

rain’s vowelless syntax,
how mathematics was an elegy,
the slenderness of trees.

Jan Zwicky – K. 219, Adagio

To my supervisor, Ted Shepherd, for the depth of his insight, for the intellectual freedom he afforded me, for the strong sense of community he has cultivated in his group, and for all of his support.

To Charles McLandress, Isla Simpson, Tiffany Shaw, James Anstey, Martin Keller, and Karen Smith for all of their answers to my questions, and for the many, many discussions.

To Shigeo Yoden for his hospitality and the remarkable generosity he showed in sharing his time with me during my visits to Kyoto. To Gloria Manney for her encouragement and interest in my work, and for the colour schemes.

To Charles McLandress, Michael Neish, Isla Simpson, and Michael Sigmond for technical support here in Toronto; to John Scinocca and Norm McFarlane for sharing the depth of their familiarity with CMAM and climate modeling in general; and to Shunsuke Noguchi and Masakazu Taguchi for their collaboration and support in working with the mechanistic model.

To my committee members, Paul Kushner and Kaley Walker, for their guidance throughout my time in Toronto. To my external examiner, Alan Plumb, for sharing his insight and his perspective on the relevance of these results. To Krystyna Biel and Ana Sousa for their administrative support. To NSERC, CFCAS, the Walter Sumner Foundation, CGCS, and JSPS for financial support.

To Patricia for all the rolled eyes, and to Heather for never touching her nose. To Bel Helen for the poetry, and to Brad and the Chorus for the songs that needed patience to fall in love with.

Finally, to my parents for their love, and to my father in particular for showing me just how much fun science can be.
Contents

1 Introduction 1
  1.1 Dynamical ingredients .............................................. 3
    1.1.1 Eliassen adjustment ........................................ 3
    1.1.2 Radiative heating ........................................... 5
    1.1.3 Wave, mean-flow interaction ................................. 6
  1.2 Variability of the Arctic Polar-night Jet ...................... 10
    1.2.1 Stratospheric sudden warmings .............................. 14
    1.2.2 The Polar-night Jet Oscillation ............................ 15
  1.3 Outline and contributions ....................................... 19

2 Data 22
  2.1 Model simulations .................................................. 22
    2.1.1 Canadian Middle Atmosphere Model .......................... 22
    2.1.2 Mechanistic Circulation Model ............................... 24
  2.2 Reanalyses .......................................................... 25
    2.2.1 ERA40 .......................................................... 25
    2.2.2 MERRA ........................................................ 25
    2.2.3 ERA Interim .................................................... 25
  2.3 Satellite observations ............................................. 26
    2.3.1 Microwave Limb Sounder .................................... 26

3 Radiative Damping 27
  3.1 Introduction ....................................................... 27
  3.2 Longwave relaxation rates ...................................... 30
    3.2.1 Theory .......................................................... 30
    3.2.2 Linear regression estimates ................................ 33
3.2.3 Effective longwave damping rates ........................................ 36
3.3 Shortwave relaxation rates ..................................................... 42
3.4 Radiative-photochemical equilibrium temperature ....................... 45
3.5 Discussion and conclusions ..................................................... 46

4 Statistical characterization .................................................... 49
4.1 Introduction ............................................................................. 49
4.2 Methods .................................................................................. 53
  4.2.1 Sudden warmings ............................................................... 54
  4.2.2 Annular modes and weak vortex events ............................... 57
  4.2.3 Polar-night Jet Oscillation events ....................................... 57
4.3 Results .................................................................................... 63
  4.3.1 Relationship to sudden warmings ....................................... 71
  4.3.2 Relationship to weak vortex events .................................... 76
4.4 Discussion ................................................................................. 79
4.5 Conclusions ............................................................................. 82

5 Zonal mean dynamics .................................................................. 84
5.1 Introduction ............................................................................. 84
5.2 Data .......................................................................................... 87
5.3 Composites ............................................................................... 88
5.4 Residual circulation ................................................................. 93
5.5 Temperature anomalies ............................................................ 99
  5.5.1 Radiative relaxation .......................................................... 100
  5.5.2 Radiative relaxation with Eliassen adjustment .................... 102
  5.5.3 Transient adjustment ........................................................ 104
5.6 Stratopause descent ................................................................. 108
5.7 Conclusions ............................................................................. 110
5.A Zonal-mean quasi-geostrophy on the sphere ............................... 111

6 Mechanistic circulation modelling ............................................... 114
6.1 Introduction ............................................................................. 114
6.2 Model setup ............................................................................. 116
6.3 Time-mean response ............................................................... 118
  6.3.1 Stratospheric changes ....................................................... 118
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.2</td>
<td>Tropospheric changes</td>
<td>123</td>
</tr>
<tr>
<td>6.4</td>
<td>Variability</td>
<td>126</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Abacus plots</td>
<td>126</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Stratospheric composites</td>
<td>130</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Tropospheric response</td>
<td>132</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Summary of transient response</td>
<td>136</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusions</td>
<td>137</td>
</tr>
<tr>
<td>6.A</td>
<td>Uncertainty estimation</td>
<td>139</td>
</tr>
<tr>
<td>6.A.1</td>
<td>Time-averaged quantities</td>
<td>139</td>
</tr>
<tr>
<td>6.A.2</td>
<td>Annular mode timescales</td>
<td>139</td>
</tr>
<tr>
<td>7</td>
<td>Conclusion</td>
<td>141</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary</td>
<td>141</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Statistics of the PJO</td>
<td>141</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Dynamics of the PJO</td>
<td>142</td>
</tr>
<tr>
<td>7.2</td>
<td>Discussion</td>
<td>145</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Detecting and attributing changes in the Arctic vortex</td>
<td>146</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Seasonal forecasting</td>
<td>147</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Surface influence of stratospheric forcings</td>
<td>148</td>
</tr>
<tr>
<td>7.3</td>
<td>Future work</td>
<td>149</td>
</tr>
</tbody>
</table>

**Bibliography** | 151 |
List of Tables

4.1 SSW classification in MERRA ............................................. 55
4.2 Percentage of variance explained by EOFs of polar-cap averaged temperatures ............................................. 58
4.3 Probability of PJO occurrence following sudden warmings ............................................. 73
4.4 Probability of PJO occurrence following weak vortex events ............................................. 77
List of Figures

1.1 EOFs of zonal mean zonal wind in the Arctic stratosphere during the winter showing the poleward and downward migration of anomalies associated with the PJO ................................................................. 16
1.2 Temperature anomalies observed by the MLS instrument during the winter of 2008-2009, showing a clear example of a PJO event. ..................... 17

3.1 Linear regression of longwave heating rates against local temperature anomalies ......................................................... 34
3.2 Regression of longwave heating rates including a term quadratic in the local temperature anomalies ........................................ 35
3.3 Vertical profile of effective longwave damping rates by latitude band, compared against parameterization of Fels (1982) .......................... 37
3.4 Effective longwave damping rates in the meridional plane by season ... 38
3.5 Effective longwave damping rates in the tropics as a function of zonal wave number ................................................................. 39
3.6 Effective longwave damping rates above the poles; failure of regression in lower Antarctic stratosphere during austral spring ....................... 40
3.7 Influence of non-local temperature anomalies on radiative heating rates during the breakdown of the Antarctic polar vortex ....................... 41
3.8 Linear regressions of shortwave heating rates against local temperature anomalies as a function of local time ............................... 42
3.9 Effective shortwave damping rates in the meridional plane for December and June ................................................................. 43
3.10 Effective total radiative damping rates in the meridional plane as a function of season ................................................................. 44
3.11 Linear estimate of radiative-photochemical equilibrium temperature, compared with temperatures computed by a radiative-photochemical model (Fels 1985) .................................................. 46

4.1 Sudden warming occurrence frequency as a function of month in the reanalyses and in the CMAM ensemble ................................................................. 56

4.2 First and second EOF of polar-cap averaged temperatures ................. 58

4.3 Polar-cap averaged temperature anomalies during the winters of 2007–08 and 2008–09 as depicted by the phase-space trajectories of the projection onto the first two EOFs and by ‘abacus’ plots ........................................... 60

4.4 Abacus plots of Arctic variability in ERA40, MERRA, MLS and one member of the CMAM ensemble ................................................................. 64

4.5 Histograms of the rate of rotation and of the phase during the most rapid periods of amplification in the phase space of the first two EOFs ........ 65

4.6 PJO event occurrence frequency and duration as a function of the reference phase $\theta_c$ .................................................................................. 67

4.7 PJO event occurrence frequency and duration as a function of decade in the CMAM ensemble ................................................................. 68

4.8 PJO event occurrence frequency and duration in the reanalyses and the CMAM ensemble as a function of season .................................................. 69

4.9 Decorrelation timescales of the NAM in the ERA Interim reanalysis and each member of the CMAM ensemble .................................................. 70

4.10 Abacus plots showing vortex splits and displacements in the reanalyses and in one member of the CMAM ensemble ........................................... 72

4.11 Composites of zonal wind and vertical EP flux anomalies during PJO and non-PJO sudden warmings .................................................. 75

4.12 Abacus plots showing weak vortex events in the reanalyses and in one member of the CMAM ensemble .................................................. 76

4.13 Composites of the NAM index during PJO and non-PJO weak vortex events .................................................. 78

4.14 Abacus plot of the merged reanalyses showing some commonly applied definitions of anomalously strong or cold events in the vortex ........ 80

4.15 Difference in climatologies of polar-cap averaged temperatures computed for selected decades in the CMAM ensemble ........................................... 81

5.1 Abacus plot of a 96-year CMAM time-slice run ........................................... 87
5.2 Composites of absolute wind and temperature anomalies during PJO events in the MERRA reanalysis and the CMAM time-slice run ................. 89
5.3 Composites of zonal wind tendencies and EP flux divergences during PJO events in the CMAM time-slice run ................................. 90
5.4 Composites of parameterized gravity wave drag and vertical momentum fluxes during PJO events .................................................. 91
5.5 Composites of anomalous vertical residual velocity, anomalous squared Brunt-Väisälä frequency, and anomalous adiabatic heating rates during PJO events .............................................. 91
5.6 Composites of temperature tendencies and diabatic heating rates during PJO events ................................................................. 92
5.7 Temperature anomalies, zonal wind, EP flux divergence, and vertical residual circulation during PJO event case study ...................... 94
5.8 Transient and downward control decompositions of the vertical residual velocity during two phases of the PJO event case study .......... 97
5.9 Profiles of the effective radiative damping timescales and of squared Brunt-Väisälä frequency during the PJO event case study ........... 99
5.10 Temperature anomalies predicted during the PJO event case study by radiative relaxation alone .................................................. 101
5.11 Temperature anomalies predicted during the PJO event case study by radiative relaxation with Eliassen adjustment .................. 103
5.12 Temperature anomalies predicted by the transient Eliassen adjustment to the eddy-driven torques during the PJO event case study ....... 105
5.13 Sensitivity of the temperature anomalies predicted by transient Eliassen adjustment to the profile of squared Brunt-Väisälä frequency and to the profile of radiative damping timescales ...................... 107
5.14 Temperature anomalies predicted by transient Eliassen adjustment during a second PJO event case study ............................... 108
5.15 Zonal wind, parameterized orographic gravity wave momentum fluxes and wave drag during the PJO event case study .................. 109

6.1 Vertical profiles of the prescribed radiative damping timescales ........ 117
6.2 Time-mean changes in zonal mean temperatures, winds, residual circulation, and planetary-scale EP fluxes in the wave-one weakened-damping run ................................................. 119
6.3 Time-mean changes in zonal mean temperatures, winds, residual circulation, and planetary-scale EP fluxes in the wave-two weakened-damping run ................................................. 120
6.4 Changes in lower stratospheric temperatures, residual circulation, and upwards EP fluxes as a function of the lower stratospheric radiative damping rate .................................................. 122
6.5 Vertically integrated zonal mean angular momentum budget in the control and the weakened-damping runs .......................................................... 124
6.6 EOFs of polar-cap averaged temperatures in the control and the weakened-damping runs, and the fraction of variance explained by the first two EOFs as a function of the lower stratospheric radiative damping rate .................. 127
6.7 Abacus plots for the wave-one control and weakened-damping runs .................. 129
6.8 Abacus plots for the wave-two control and weakened-damping runs ............ 130
6.9 Composites of the zonal wind at 60° N and vertical EP flux over the polar cap during weak vortex events .......................................................... 131
6.10 Composites of the upper tropospheric zonal wind and the temperature difference between 300 hPa and 100 hPa during weak vortex events .......... 132
6.11 Change in the vertically integrated momentum budget during weak vortex events in the weakened-damping runs .................................................. 134
6.12 Vertical profile of NAM timescales in the wave-one and wave-two runs .... 135
6.13 Summary of composited metrics during weak vortex events as a function of the lower stratospheric radiative damping rate .................. 136
Chapter 1

Introduction

Where there were unnumbered paths of air, now
the one shaft of your plunge, whose walls
are the shrieks of your old nemesis, gravity

*Don McKay – Plummet*

As a result of the mis-alignment between the spin axis of the Earth and its orbital plane, for several months of the year the air above each pole lies in perpetual night. Above the relatively well-mixed, neutrally-stable troposphere, the contrast in the radiative heating of the mid-latitude and polar stratosphere generates a strong meridional thermal gradient. In turn, strong westerly (eastward) winds spin up, which produce a Coriolis force in balance with the resulting pressure gradient. These circumpolar winds, alternatively called the polar-night jet or the stratospheric polar vortex, form over both poles during their respective winters.

The winds are disturbed intermittently by undulations in the winds and temperatures, with horizontal wavelengths of the order of the Earth’s circumference. These undulations, known as planetary-scale Rossby waves, are generated by large scale topographical features such as the Tibetan plateau and the Rocky mountains, and by the contrast in heating rates above the continents and oceans. That the features which generate the waves are stronger north of the equator than south of it distinguishes the dynamics of the two polar vortices. The waves can propagate upwards only in westerly winds (Charney and Drazin, 1961), and can be thought of as carrying with them something equivalent to easterly angular momentum (Shepherd, 1990). When the waves propagate into a region, therefore, they decelerate the zonally symmetric winds. If they subsequently propagate
away, the background winds accelerate once more; if, on the other hand, they dissipate by some process, the background winds remain weakened.

These waves are strongly intermittent (Holton and Mass, 1976) and can be quite nonlinear in the stratosphere where they often break as they encounter critical lines (McIntyre and Palmer, 1983). Understanding in detail just how these waves organize themselves, temporally and spatially, remains a major theoretical challenge. Their effects are most spectacularly manifest in stratospheric sudden warming events, which occur in roughly two out of every three winters, and can begin at any point in the season. Over the course of several days, the waves amplify rapidly and their propagation focuses on the pole, depositing sufficient momentum to completely reverse the winds in the polar-night jet. For reasons explained below, this can warm the mid-stratosphere by upwards of 50 K.

The central goal of this thesis is to demonstrate that, in contrast to the suddenness of the onset of the warming events, the recovery of the polar vortex following the warming can in some cases take as long as three months. Over this extended recovery period the anomalous circulation follows a very characteristic, robust pattern. These long timescale recoveries are referred to here as Polar-night Jet Oscillation (PJO) events, adopting the nomenclature of Kuroda and Kodera (2001). While they are of intrinsic dynamical interest, since they represent a significant fraction of the variability of the Arctic vortex, it is also important to properly characterize them in order to detect or attribute changes in the polar region in response to any external forcings. Moreover, they play a significant role in the coupling between the extratropical stratosphere and troposphere. As such they represent a potential source of skill in seasonal forecasting, and may play a role in transmitting the stratospheric effects of a variety of external forcings to the surface.

The remainder of this chapter gives a more thorough background to the thesis, focusing in Section 1.1 on the dynamical ingredients relevant to this phenomenon, including radiative transfer, wave-mean flow interaction, and the Eliassen adjustment process. Section 1.2 reviews common paradigms of Arctic variability, into the context of which these events must be placed. The PJO and the novel perspective advanced by this thesis are introduced towards the end of this section, after the required background concepts have been outlined. Finally, the chapter closes with an outline of the projects carried out which make up the thesis.
1.1 Dynamical ingredients

1.1.1 Eliassen adjustment

The atmosphere, for most of the domain under present consideration, is in local thermodynamic equilibrium (LTE). As a macroscopic fluid, it is fully described by seven fields: the three components of velocity, and the thermodynamic quantities of internal energy, density, temperature, and pressure. The internal energy is given, to a good approximation in the dry middle atmosphere, by the sensible heat content, and the latter three quantities are diagnostically related by the equation of state for an ideal gas. This leaves five first order prognostic equations and thus five types of linear modes. Scaling arguments relevant for the large scale flow of the atmosphere, namely that (a) flow velocities in the atmosphere are small compared to the speed of sound, and (b) the thickness of the atmosphere is small compared to its horizontal extent, permit two of these modes to be filtered out, reducing the prognostic continuity and vertical momentum equations to diagnostic ones. The three remaining modes—two buoyancy-driven modes, referred to as gravity waves, and one vorticity-driven mode, referred to as Rossby waves—are the most relevant to the large scale circulation.

In the extratropics, where the spin axis of the Earth projects significantly on to the local vertical, there is a strong separation of timescales in the large scale circulation between the gravity-wave modes (with timescales less than the inverse of the Coriolis parameter $1/f$) and the Rossby-wave mode (with non-linear timescale\(^1\) of order $L/U$, where $L$ and $U$ are characteristic length and velocity scales, respectively). To leading order in the Rossby number, defined as the ratio between these timescales $Ro = U/Lf$, the gravity-wave modes too can be filtered out. This leaves a single prognostic equation, and thus a single field\(^2\) to describe this leading-order, balanced state. Neglecting diabatic and dissipative processes, this yields the material conservation of quasi-geostrophic potential vorticity $q$,

$$\frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x} + v \frac{\partial q}{\partial y} = 0 \quad (1.1)$$

as the equation of motion (spherical geometry is neglected here for simplicity). An expression for the potential vorticity is given below. The horizontal winds $u$ and $v$ are,

---

\(^1\)The timescale $(L\beta)^{-1}$ associated with the meridional gradient $\beta$ of the Coriolis parameter is assumed here to be of the same order as the non-linear timescale.

\(^2\)There are, in addition, non-trivial dynamics at the lower boundary.
in this limit, diagnostically related to the pressure $p$ by what is known as geostrophic balance: the balance of horizontal pressure gradients by the Coriolis force

$$fu = -\frac{\partial p}{\partial y}, \quad fv = \frac{\partial p}{\partial x}.$$  \hspace{1cm} (1.2)

The Coriolis parameter $f = 2\Omega \sin \phi$ is twice the projection of Earth’s angular velocity $\Omega$ onto the local vertical at the latitude $\phi$. Combining this with hydrostatic balance (the balance of the weight of the atmosphere above by the pressure) yields thermal wind balance, which relates the vertical wind shear to horizontal gradients in temperature. Taking the zonal mean (around latitude circles, indicated by an overline) yields the expression of this balance most relevant to the current exposition,

$$f \frac{\partial \bar{u}}{\partial z} = -\frac{R}{H} \frac{\partial \bar{T}}{\partial y},$$  \hspace{1cm} (1.3)

where $R = 287 \text{ J K}^{-1} \text{ kg}^{-1}$ is the dry gas constant, $H$ is the vertical e-folding length scale over which the background density profile $\rho_0 = \rho_s e^{-z/H}$ decays from its surface value $\rho_s$, and $y$ is the latitudinal coordinate, increasing towards the North Pole.

Under these scaling assumptions, the flow will maintain thermal wind balance at every instant. In order to maintain this balance, the effects of any applied force (which would otherwise simply accelerate the winds) must be distributed to the temperature field as well. Similarly, any diabatic heating must also influence the wind field. This is achieved through an overturning circulation in the meridional plane, via a process known as Eliassen adjustment (Eliassen, 1951). In addition to distributing the effects of the forcing between the wind and temperature fields, the induced overturning circulation can also distribute the effects spatially. The transient adjustment to a time varying force local to the Arctic polar vortex can therefore have a global signature (Plumb, 1982). The steady state circulation achieved in response to a constant forcing, however, can be shown to close exclusively downward (resulting in a vertically integrated balance between the applied force aloft and frictional processes at the surface), giving rise to the notion of ‘downward control’ (Haynes et al., 1991). This adjustment plays a central role in the dynamics of the polar vortex, and is discussed more thoroughly in Chapter 5.
1.1.2 Radiative heating

In order to describe the Eliassen adjustment process discussed above, one needs to understand both the zonal mean torques and diabatic heating which drive the adjustment. The diabatic processes in the middle atmosphere, where moist processes can be safely neglected, are primarily radiative in origin. In general the net radiative term is given by the residual between heating from the absorption of solar radiation by the ozone layer (particularly in the ultraviolet region of the spectrum), and infrared cooling from greenhouse gases. Since the stratosphere lies above the effective radiating level of the Earth (i.e. the level at which the bulk of the thermal radiation lost to space is emitted, and therefore the level at which temperatures must adjust in order to balance the incoming solar radiation), gases with spectral features in the infrared tend to have a net cooling effect at these altitudes. Through most of the stratosphere the dominant contributor to this cooling is the collection of vibrational–rotational transitions of carbon dioxide, which emit light of wavelengths near 15 $\mu$m, though ozone plays a quantitatively important role, and in the lower stratosphere other minor constituents including water vapour also contribute. In the absence of the shortwave heating during the polar night, the infrared cooling dominates; rates in the stratosphere are of the order of several degrees Kelvin per day.

Due once again to the aspect ratio of the atmosphere (relevant horizontal length scales are two orders of magnitude larger than relevant vertical length scales), to an approximation good enough for the purposes of this thesis, the radiative transfer can be thought of as occurring vertically in a column of air above each point on the Earth’s surface. The net radiative heating rates at a given height can then be written as a sum of the radiation emitted by the rest of the column and absorbed locally, and that emitted locally. In general, the heating rates are therefore determined non-locally, as heat is transferred to and from remote levels of the atmosphere. They are also non-linearly related to the temperatures, mostly as a result of the curvature of the Planck function, which determines the dependence of radiated power on temperature. A realistic representation of radiative heating in the middle atmosphere therefore requires detailed consideration of these effects.

However, the residual between the local emission and absorption is often well approximated in the middle atmosphere by the component of the locally emitted radiation that escapes to space (Rodgers and Walshaw, 1966). Moreover, since the dominant emitter
of longwave radiation in the middle atmosphere, CO$_2$, is well-mixed, the local temperature is a good predictor of heating rates. This is not the case in the troposphere, where H$_2$O in its vaporous and condensed states plays a much more important role. So long as the temperature perturbations remain small, linearization of the Planck function is also justified. As a result, the longwave cooling can be well approximated by a local, linear damping (Dickinson, 1973), albeit with a coefficient that varies with height, latitude and season.

The effects of shortwave radiation provide another source of radiative damping due to the dependence of ozone mixing ratios on temperature (Craig and Ohring, 1958), although the linear and local approximations are less accurate than for longwave radiation. In photochemical equilibrium, higher temperatures give rise to reduced ozone abundances. This reduces shortwave absorption, leading to the damping of temperature perturbations where ozone is under photochemical control. On the other hand, when the effects of transport dominate the local photochemical processes, the dynamics tend to lead to positive correlations between ozone and temperature (Hartmann, 1981). However, ozone perturbations also have a significant, non-local ‘opacity’ effect as they allow more or less shortwave radiation to penetrate to lower levels. In general this leads to weaker correlations between temperatures and shortwave heating rates (Haigh, 1985). The local effects, in particular regions and seasons, are nonetheless sufficiently dominant that an effective linear damping can be defined.

This linear approximation to radiative cooling rates is sometimes known as a ‘Newtonian’ cooling approximation, and is of great use to theoretical and simplified modelling studies. It is explored in greater detail in Chapter 3.

### 1.1.3 Wave, mean-flow interaction

One of the most important (near-)symmetries of the atmosphere is the rotational symmetry about the rotational axis of the Earth. Although this symmetry is broken in essential ways by surface features, the decomposition of the flow into a zonal mean and an eddy component is an extremely fruitful one, and is relied upon heavily in this thesis. In particular, one can describe the effects of the eddies upon the mean as a zonal mean torque, relevant to the Eliassen adjustment of the vortex. The theory of wave, mean-flow interaction outlined here is classical (many of these details are found in Andrews et al. (1987) or Bühler (2009)). It is not used directly by much of the work that follows; how-
ever, the implications are quite foundational, so it is therefore important to review the relevant details.

For simplicity, as above, we will consider a Cartesian coordinate system in which the Coriolis parameter is taken to vary linearly in the meridional direction, \( f = f_0 + \beta y \), and the vertical direction is taken to be a log-pressure coordinate. On such a ‘beta-plane’, the quasi-geostrophic potential vorticity is given by

\[
q = f_0 + \beta y + \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{1}{\rho_0} \frac{\partial}{\partial z} \left( \rho_0 \varepsilon \frac{\partial \psi}{\partial z} \right), \quad \varepsilon = \frac{f_0^2}{N^2}.
\] (1.4)

The horizontal winds are related to the stream function \( \psi \) by \((u, v) = (-\partial_y \psi, \partial_x \psi)\).

One can always decompose any given field into its zonal mean and an anomaly:

\[
q = q + q'
\] (1.5)

If we assume, however, that \( q' \) is small, say of the order of some small parameter \( a \), linear theory permits much to be said. For instance, linearizing (1.1) about a state with constant zonal mean zonal wind \( \bar{u} = U \),

\[
\frac{\partial q'}{\partial t} + U \frac{\partial q'}{\partial x} + v' \beta = 0,
\] (1.6)

and substituting in a plane wave solution with phase speed \( c \) relative to the ground and wave numbers \( k, \ell, m \) in the zonal, meridional, and vertical directions, respectively, gives a dispersion relation for Rossby waves. This can be written as a condition for when such waves can propagate vertically (Charney and Drazin, 1961):

\[
0 < U - c < U_c, \quad U_c = \frac{\beta}{k^2 + \ell^2 + 4\varepsilon/H^2}.
\] (1.7)

The stationary waves \((c = 0)\) most relevant to the polar vortex, therefore, can propagate upwards only through westerly winds that are weaker than a critical velocity \( U_c \) which depends directly on the meridional gradient of potential vorticity. Here this gradient is simply that of the Coriolis parameter; for more general flows \( \beta \) would be replaced by the full meridional gradient of \( q \). This condition can of course be refined for more realistic background states (Matsuno, 1970), though the simple criterion given can already provide a paradigm for explaining many aspects of observed stratospheric variability. For instance, since \( U_c \) is larger for waves of larger horizontal extent, the background state
is quite effective at filtering out all but the longest waves; and indeed eddies of zonal wave number 1 and 2 are observed to dominate the winter stratosphere. Moreover, in the summer hemisphere, the easterly winds aloft prohibit these stationary waves from propagating into the stratosphere at all.

These waves produce fluxes of momentum and heat. Where these fluxes converge, they impart a force on the mean flow, in the form of a generalized Reynolds stress to account for thermal-wind balance (1.3). For concreteness, the zonal mean zonal momentum equation consistent with (1.1) can be written as

\[ \bar{u}_t - f_0 \bar{v} = -\frac{\partial}{\partial y} \bar{u}' \bar{v} + \frac{1}{\rho_0} \frac{\partial}{\partial z} \left( \rho_0 f_0 \frac{R \bar{v}' T}{H N^2} \right) = \frac{1}{\rho_0} \tilde{\nabla} \cdot \tilde{F}. \] (1.8)

Note that this holds for finite amplitude disturbances under quasi-geostrophic scaling. The presence of the heat-flux term in this expression is a result of a redefinition of the overturning meridional circulation:

\[ \bar{v}' = \bar{v} - \frac{1}{\rho_0} \frac{\partial}{\partial z} \left( \rho_0 \frac{R \bar{v}' T}{H N^2} \right), \quad \bar{w}' = \bar{w} + \frac{\partial}{\partial y} \left( \frac{R \bar{v}' T}{H N^2} \right). \] (1.9)

This framework is known as the Transformed Eulerian Mean (TEM) and has several advantages, one of which is related to an identity used to rewrite the convergence of eddy momentum and heat fluxes on the right side of (1.8) as an eddy flux of potential vorticity

\[ \bar{v}' q' = \frac{1}{\rho_0} \tilde{\nabla} \cdot \tilde{F}. \] (1.10)

Assuming now that the waves are linear, one can from (1.6) derive the following to order \( O(a^2) \),

\[ \frac{\partial}{\partial t} \left( \frac{q^2}{2\beta} \right) + \bar{v}' q' = \frac{\partial}{\partial t} \left( \frac{q^2}{2\beta} \right) + \frac{1}{\rho_0} \tilde{\nabla} \cdot \tilde{F} = 0. \] (1.11)

This expresses the conservation of a wave activity \( A \) (i.e. a measure of the amplitude of the waves); in this case, the (negative of the) pseudomomentum, so-called because of its relationship to the momentum of the zonal mean state (1.8). In general in this asymptotic framework these conservation laws have the form

\[ \frac{\partial A}{\partial t} + \frac{1}{\rho_0} \tilde{\nabla} \cdot \tilde{F} = S + O(a^3). \] (1.12)
The flux $\vec{F}$ is known as the Eliassen-Palm (EP) flux; its quasi-geostrophic form is given implicitly in (1.8). The divergence of this flux connects the behaviour of the waves with that of the mean flow. One therefore thinks of the waves as imparting a force on the mean flow either through transience $\partial A/\partial t$ (the local growth or decay of the waves), or through generation or dissipation processes $S$. The latter is absent in (1.11) as a result of neglecting these processes in (1.6).

If one makes the further assumption that the wavelength and period of the waves are short compared to variations in the background state—i.e. the Wentzel-Kramer-Brillouin (WKB) approximation—one can find a form of the wave activity $A$ and its flux $\vec{F}$ such that $\vec{F} = \mathbf{c}_g^* A$. The flux therefore gives information about the group velocity $\mathbf{c}_g^*$ of the wave-packets. Moreover, the pseudomomentum can be very generally expressed in terms of the energy contained by the waves as $k\hat{E}/\hat{\omega}$. Here the wave energy $\hat{E}$ and frequency $\hat{\omega}$ are measured in the reference frame of the winds such that, in particular, $\hat{\omega}/k = c - U$; Rossby waves, which by (1.7) must have easterly phase speeds with respect to the background wind, will therefore always tend to impart an easterly force when they propagate into a region or dissipate there. This asymmetry is fundamental to the variability of the Arctic vortex.

This framework would provide an almost complete theoretical understanding of the waves and their interactions with the mean flow; unfortunately, some of the key assumptions underlying these developments are severely violated in the Arctic vortex. Typical length and timescales of the waves are of the same order if not longer than those of the background flow. Their amplitudes, moreover, are nearly always large enough that non-linearity is important (McIntyre and Palmer, 1983), and during sudden warmings they have the effect of destroying the westerly flow they depend upon for propagation. Perhaps surprisingly, however, fully non-linear wave activity conservation laws analogous to (1.12) can still be derived by (for instance) exploiting the zonal symmetry of the background flow, which implies (via Noether’s theorem) an associated conserved quantity (McIntyre and Shepherd, 1987; Haynes, 1988). These conservation laws reduce to the linear forms in the small amplitude limit. As a result, one can still speak of the eddy-induced forces acting on the zonal mean as being a result of transience and dissipation in the eddy field and remain on solid theoretical footing. However, although these conservation laws can also be used to derive some constraints on the behaviour of the eddies, they do not provide a simple description of their behaviour. The decomposition (1.5) into a zonal mean and an eddy field is therefore clearly useful for understanding the behaviour of the
former; whether it is sufficient for understanding the eddies themselves, however, is less clear.

For completeness, it is noted that similar ideas can be applied to understand the effect of gravity waves on the mean flow. In this case linear and WKB ideas apply to a greater extent, which has been exploited in parameterizing the effects of unresolved eddies on the mean state. These effects are of particular importance in the mesosphere.

1.2 Variability of the Arctic Polar-night Jet

The Arctic stratospheric vortex is highly variable on intraseasonal and interannual timescales, and its behaviour is the subject of a considerable literature. Due to the strength and intermittency of the interactions between the waves and mean flow, much of this behaviour is strongly non-linear and irregular. A wide range of descriptive indices and terminology has thus been proposed in order to categorize and classify this behaviour. Unfortunately, many of these descriptions provide overlapping classifications, and even amongst those which have been more broadly adopted, definitions and methodologies often vary from study to study. Since one of the primary results of this thesis is to refine the use of one such description, namely, the PJO, the scientific merits of adopting this description must be justified. This is more easily done once some of the fundamental concepts and paradigms that underlie the literature have been introduced. This section gives a brief review of some of the most significant results on the variability of the Arctic vortex before focusing on the PJO itself and the novel contributions of this thesis.

The first recognition that the vortex could undergo spontaneous and rapid warming stemmed from high-altitude radiosonde observations out of Berlin (Scherhag, 1952). It took further observations to realize that these episodes of ‘explosive’ warming were in fact of global scale, and that they did not occur at any fixed date with respect to the season. The first successful theory of sudden warmings was due to Matsuno (1971). In contrast to other studies, which sought to explain the phenomenon in terms of some local instability, the instigating factor of the warming was posited to be an impulsive increase in the planetary-scale Rossby waves generated by the tropospheric flow. This suggested that the stratosphere was in some sense responding passively to tropospheric variability, a notion that is supported as well by the considerably lesser amount of mass within the stratosphere as compared to the troposphere below.

The theory, while posed in the conventional Eulerian Mean framework (rather than
the TEM introduced above), was nonetheless based on the same theories of wave, mean-
flow interaction and of Eliassen adjustment described above. Since the study sought only
to explain the onset of the warming, the comparatively slow radiative processes were not
included. The radiative recovery following the initial warming was included in a highly
simplified framework by Holton and Mass (1976). The Holton-Mass model truncates the
meridional dependence to a single mode, and the zonal dependence to a zonal mean and
one zonal wave number, leaving the vertical dependence of the wave and the mean flow
to be computed numerically. As a simple dynamical model of the variability, it has been
highly influential. One of the insights gained was that, in particular parameter regimes,
the model could produce events closely analogous to sudden warmings with a fixed lower
boundary condition; that is, that the dynamics of the vortex itself are able to control the
amount of wave activity that propagates up from the troposphere.

One of the key parameters of the Holton-Mass model is the height $h_0$ of the deforma-
tions of the geopotential surface at the lower boundary, which act as the source of
the waves. As a function of this lower boundary forcing, Yoden (1987) found two types
of stable steady states for values of $h_0$ below a certain threshold: a cold, strong vortex
in which the winds are close to radiative equilibrium values, and a weak, warm state in
which the waves produce steady poleward fluxes of heat to balance the radiative cooling.
The model exhibits hysteresis, in that the two steady states coexist for certain values
of $h_0$. Above some other threshold (typically lower than the point at which the cold
steady state disappears), the warm steady state becomes unstable, and the model was
found to exhibit periodic sudden warmings. This provided a clear analogy to explain the
differences between the two polar vortices: the Arctic vortex, as a result of the more dom-
ninant surface topography in the Northern Hemisphere, is in the vacillatory regime found
by Holton and Mass (1976), while the colder, more stable Antarctic vortex is in the cold
steady state. Perhaps more subtly, this dichotomous paradigm between the cold, strong
vortex and the weak, disturbed vortex is often applied to understand intraseasonal and
interannual variability in the Arctic vortex itself (e.g. Christiansen, 2009, and references
therein).

This simple model neglects two important aspects of the real system: firstly the full
meridional structure and spherical geometry of the polar vortex, and secondly, the sea-
sonal cycle. That these regimes might nonetheless be a relevant classification of the
types of variability in the real atmosphere was demonstrated by a series of model studies
in which these two restrictions were relaxed (Taguchi et al., 2001; Taguchi and Yoden,
Following the modelling approach of Held and Suarez (1994), these studies solved the same set of fluid equations on a sphere used by many comprehensive general circulation models, but with highly simplified representations of physical processes including, for instance, a Newtonian cooling approximation for radiative processes. Performing a parameter sweep of the surface topography, they found that at intermediate values the model produced the strongest interannual variability in the spring as the vortex breaks down, much like in the observed Antarctic vortex. At larger values, intermittent warming events occur throughout the winter season, and monthly averaged temperatures in the polar vortex take on non-Gaussian and even bimodal distributions, both of which are hinted at in observations of the Arctic as well (Christiansen, 2009).

The advent in the past decade of comprehensive models of the middle atmosphere capable of simulating the stratospheric variability with some fidelity, as well as the development of gridded reanalyses of the observed atmospheric circulation, have led to studies focused more on detailed agreement with observations (Charlton and Polvani, 2007; Charlton et al., 2007; SPARC CCMVal, 2010). Significant challenges still remain in achieving this agreement. Firstly, very long datasets are required to reliably characterize the statistics of Arctic polar vortex variability. This is now less of a problem than it once was for models, since long integrations of the order of several centuries are becoming more common, but the three decade record of reliable satellite observations remains marginal for the reliable characterization of many details. (Extending this time series back to the International Geophysical Year program of the late 1950s does improve the statistics to some extent.) Models still have some difficulty in simulating the correct number of sudden warmings, though this appears to be more an issue of achieving the correct tuning of various physical parameterizations rather than overcoming any intrinsic deficiency in the model dynamics (see, for instance SPARC CCMVal, 2010, Fig. 4.9, as well as Chapter 2 below).

**Annular Modes and Stratosphere-Troposphere Coupling**

A major development in the past ten to fifteen years has been the recognition that not only is the variability to some extent determined by the stratosphere itself but that in some cases the behaviour of the stratosphere can influence the surface below. The most compelling evidence for this in the case of stratospheric sudden warmings was made by Baldwin and Dunkerton (2001), who showed composites of the Northern Annular
Mode (NAM)—essentially an index of the zonal mean geopotential height anomalies at each pressure level—following episodes when the stratospheric vortex was either anomalously weak (as in the case of sudden warmings), or anomalously strong. In both cases, anomalies were seen to propagate down from the stratosphere and produce a persistent anomaly of the same sign in the troposphere for several months.

The annular modes are an empirical description of a dominant extratropical mode(s) of variability; namely, in the troposphere, that of the poleward or equatorward shift of the eddy-driven jet, and in the stratosphere, the strengthening or weakening of the polar-night jet (Kushner, 2010). They are defined in terms of the first EOF of the geopotential heights at every pressure level, though the exact method for computing them varies (Baldwin and Thompson, 2009). They are often used in studies of stratosphere-troposphere interactions, because they provide a single height-dependent index which describes the coupled variability in both layers of the atmosphere. Perhaps the clearest motivation for their use, however, lies in the fact the response of the atmosphere to a wide range of externally imposed forcings is found to project onto the annular modes. This result has some theoretical grounding in the fluctuation-dissipation theorem (FDT) (Leith, 1975), which relates the response to an external forcing to the decorrelation of the natural fluctuations of the atmosphere. The decorrelation timescales of the annular modes are thus of particular interest both as a measure of the potential for seasonal predictability (Baldwin et al., 2003), and as a predictive tool for understanding the dynamical response under climate change in the context of the FDT (Kidston and Gerber, 2010).

The present work does not rely heavily on annular modes, though insights from the techniques applied here are compared to some extent to those gained through the annular modes. One disadvantage of the latter is that the geopotential heights in terms of which they are defined depend both upon a vertical integral of temperature, and upon the surface pressure field, which itself depends upon the behaviour of the entire atmospheric column. The steady state response of the zonal mean atmosphere depends on balancing dynamical heating against radiative relaxation (which depends upon temperatures), and on balancing momentum fluxes against the surface friction. The steady state response of the annular modes is thus quite non-local (note that this is also true of the winds aloft). This point will be returned to in Chapter 6.
1.2.1 Stratospheric sudden warmings

Parameter sweep studies of the sort performed by Taguchi and Yoden (2002a) make it evident that the seasonal cycle of the two polar vortices is intimately related to the strength of the wave-driven variability they experience and that sudden warmings are an integral component of the climatology of the Arctic vortex, even though their occurrence is not tied to the calendar year. Nonetheless, it is useful to consider them as independent dynamical events which occur superimposed on a slowly varying background state. The perspective adopted in this section is that sudden warmings are defined by the dynamics of their onset (i.e. the warming itself), in order to distinguish this phase dynamically from the recovery of that warming, to which the focus will shift in the next subsection. The dynamics of the warming itself likely also has some bearing on the timescale of the recovery, as will be discussed below, and are thus worth reviewing.

Much of the fundamental understanding of these dynamics has not changed from the original theory put forth by Matsuno (1971). The amplification of the waves (which is evident in observations prior to warmings) leads, either through transience or dissipation, to strongly enhanced EP flux convergence within the vortex. In the TEM framework, this drives a strong poleward and downward circulation that warms the vortex adiabatically and decelerates the winds. If the wave driving is strong enough, this can reverse the winds and destroy the potential vorticity gradient required for the waves to propagate (though the latter need not follow from the former).

What specific processes lead to the strong wave driving in the vortex during the warming remains unclear. To some extent this is likely because there is a great deal of variability between different events, so that particular mechanisms may play dominant roles in some events but not in others. The original theory of Matsuno (1971) placed a great deal of emphasis on the development of a zero wind line near the top of the vortex, which subsequently descends from the upper stratosphere throughout the event. The zero wind line is a critical layer for the stationary planetary waves. Critical layers lead to strongly non-linear effects, and in some cases strong dissipation and thus wave driving. Later studies (Dunkerton et al., 1981; McIntyre, 1982) argued for the role of subtropical critical layers (which can act as a reflecting surface for the waves) in focusing the wave activity on to the pole; it is in this context that the notion of the ‘preconditioning’ of the vortex is discussed as a prerequisite for warmings. It must also be mentioned that while this wave, mean-flow description of the dynamics is valid for understanding the zonal
mean response, the spatial extent of the eddies frequently stretches over upwards of 30 degrees of latitude, and thus interpreting the details of the EP fluxes in the meridional plane as indicative of wave propagation may well be misleading.

Similarly, there are a number of proposed mechanisms for the initial amplification of the waves. Matsuno (1971) simply posited an increase in the zonally-asymmetric geopotential height near the tropopause, appealing to an unspecified tropospheric process for generating this change. Numerous studies have sought to find systematic tropospheric precursors to the stratospheric warming. Perhaps the most persistent notion in the literature is the connection to tropospheric blocking events (Tung and Lindzen, 1979a,b; Martius et al., 2009). While there does seem to be some evidence for such precursor events, the fact that modern comprehensive climate models can produce good sudden warming statistics (McLandress and Shepherd, 2009) despite producing too few blocking events (Scaife et al., 2010) suggests these precursors are not a necessary condition.

Another proposed mechanism is that the amplification is due to the presence of a linear mode which comes into resonance with the surface forcing (Tung and Lindzen, 1979b; Plumb, 1981). This has been connected to a barotropic mode (Esler and Scott, 2005) in the case of events during which the vortex splits in two (in contrast to those in which the vortex is forced off of the pole, known as displacements). One intriguing aspect of this mechanism is the potentially non-trivial role implied for the stratospheric circulation in producing the waves, not just in integrating their effects.

### 1.2.2 The Polar-night Jet Oscillation

The notion of the PJO has its roots in a study by Kodera et al. (1990), which noticed that lag-correlations of monthly zonal mean zonal wind anomalies near the northern subtropical upper stratosphere appear to migrate poleward and then downward over the course of several months. Temperature anomalies, in thermal wind balance with the wind anomalies, are also observed. Similar behaviour has also been found in a variety of statistical methods related to Empirical Orthogonal Function (EOF) analyses (Kodera et al., 2000; Kuroda and Kodera, 2001, 2004); in particular, the first two EOFs of either the zonal mean zonal wind (Kodera et al., 2000) or the polar cap averaged temperature profile (Kuroda and Kodera, 2004) are found to capture the slow poleward and downward progression. This progression is illustrated in Fig. 1.1 (Kodera et al., 2000, their Fig. 2), which shows linear combinations of the zonal mean zonal wind EOFs of the
Chapter 1. Introduction

Figure 1.1: Reproduced with permission from Kodera et al. (2000), their Fig. 2. Linear combinations (see text) of the first two EOFs of zonal mean zonal wind from 17 winters of analyzed observations.

form $\cos \phi \times \text{EOF 1} + \sin \phi \times \text{EOF 2}$, computed from 17 winters of analyzed observations. Counter-clockwise rotation (increasing $\phi$) is favoured by the vortex. The perspective that emerges from these correlative analyses is that this mode of variability operates steadily throughout the winter season, with similar behaviour for anomalies of both signs; and since somewhat analogous variability can also be found in the Antarctic vortex, this variability was termed the Polar-night Jet Oscillation by Kuroda and Kodera (2001). That the far more rapid stratospheric sudden warmings tend to occur in a particular phase of the PJO was seen as a sort of phase locking of a fast mode onto a slow mode (Kodera et al., 2000).

The claim made here is that much of the behaviour captured by these statistical analyses is in fact most usefully thought of as a dynamical recovery of the vortex from a sudden warming. By way of example, Fig. 1.2 shows polar-cap averaged temperature anomalies as observed by a satellite-borne instrument (the MLS instrument on the Aura satellite, specifically; see Chapter 2), for the winter of 2008-2009. Following the sudden warming which occurred in January of 2009 (apparent as a rapid warming throughout the stratosphere), a vertically-tripolar structure emerges in the temperatures, with a persistently warm lower stratosphere, a cold upper stratosphere, and a warm mesosphere, the latter corresponding to an abnormally elevated stratopause. While the anomaly in the lower stratosphere persists, the elevated stratopause descends over the course of several months.

Although the warming of 2009 was by many measures extreme, this evolution of the temperature anomalies (and the associated wind anomalies) is remarkably common. By extending the two-dimensional analysis of EOFs just described, Chapter 4 demonstrates that it occurs after roughly half of all sudden warmings (or about once every three years).
In contrast to the linearity implied by the correlative analyses reviewed above, the phase progression of these events is always the same, in that the initial amplification of the polar anomaly is far more rapid than the subsequent downward migration, and coincides with a warming event. While the vortex at times does become anomalously cold and strong, such events do not possess the dynamic similarity shared by extended recovery periods such as the one shown in Fig. 1.2.

This asymmetry is a direct result of the easterly pseudomomentum carried by Rossby waves and the related asymmetry of the Charney-Drazin criterion (1.7). The sudden warming itself is induced by the rapid amplification of planetary waves as reviewed in the previous section; the dynamics of the recovery phase, by contrast, are driven by the much slower radiative processes (and the associated Eliassen adjustment). The extended timescales of the PJO are thus intimately related to the radiative timescales; however, the very fact that the evolution during this recovery phase is driven by radiative processes implies that the wave driving is not active (indeed, as will be seen in Chapter 5, this holds as much for the smaller scale gravity waves as it does for the planetary scale Rossby waves). Some of the basic questions this thesis seeks to address can now be posed. What exactly determines the timescale of the persistent lower stratospheric anomaly, and why is the recovery following other sudden warmings much more rapid? What drives the slow downward migration of the anomalies during the recovery phase? What determines the behaviour of the waves during these events? While these questions have been considered, implicitly or otherwise, by many studies of the Arctic literature, it is argued here that the event-based perspective adopted here that exploits the strong dynamical commonalities
seen between these events allows these questions to be posed and addressed much more precisely than would a bulk, correlative approach.

Given this emphasis on an event-based framework, one might question the adoption of the PJO nomenclature. Clearly the PJO as conceived of here does not correspond to a steady, fixed period oscillation and the associated dynamics are not that of a linear harmonic oscillator. However, there is an oscillatory aspect to the behaviour in the following sense: if one considers a fixed altitude in the mid to upper stratosphere, the recovery of the vortex overshoots its climatological state before ultimately returning to its typical configuration. In the language of dynamical systems, a PJO event corresponds most closely to the periodic orbits of the Holton-Mass model, not to any of its steady states.

It is important to emphasize that the occurrence of extended timescale recoveries from sudden warmings is not a novel observation; they dominate well known composite results such as those of Baldwin and Dunkerton (2001) and Charlton and Polvani (2007). That there may be some benefit to considering separately those events with extended timescales has also been suggested in several contexts (notably Harnik (2009) and Siskind et al. (2010)).

This thesis seeks to clearly identify these events, to demonstrate that modern, comprehensive models are capable of simulating them with high fidelity (and thus represent a valuable tool for their study), and to describe the zonal mean component of their dynamics.

**External influences on the PJO**

As a dynamical recovery from a stratospheric sudden warming, PJO events are instigated by the same pulses of planetary wave activity propagating up from the troposphere below as were discussed in Section 1.2.1. There is, nonetheless, a variety of other external factors that may also influence the polar behaviour. For instance, there is evidence that winds in the tropical and subtropical stratosphere—and thus, external factors which influence these winds—can impact the polar behaviour. Observational evidence suggests that the phases of the 11-year solar cycle and the Quasi-Biennial Oscillation (QBO) can influence the evolution of the polar vortex in the Northern Hemisphere (Holton and Tan, 1980, 1982; Labitzke, 1987).

Indeed, much of the literature describing the PJO has been written in the context of
identifying possible solar influences (e.g. Kodera et al., 1990, 2002). Since the ultraviolet region of the solar spectrum varies much more strongly with the solar cycle than does the visible (in which the bulk of the radiative power is found), the impact of solar variability is more pronounced in the ozone layer where this light is absorbed (particularly as a result of positive feedbacks in the photochemical production of ozone, e.g. Haigh, 1994), and a considerable literature on stratospheric variability associated with the solar cycle exists.

The possible influence of tropical wind anomalies on polar variability has been investigated in the context of simplified circulation models by Gray and co-authors (Gray et al., 2001a, 2003, 2004; Gray, 2003). They applied a variety of perturbations to the equatorial winds, and found strong sensitivity to the equatorial winds in the seasonal evolution of the polar jet. In particular, they found at least as much sensitivity to upper stratospheric wind anomalies (where solar impacts are thought to be more significant) as to anomalies in the lower stratosphere (where the QBO dominates). Downward propagating polar wind anomalies in observational datasets have been noted in composites based on the 11-year solar cycle (Gray et al., 2001b), and in recent comprehensive model studies based on modern observations of solar spectral variability (Ineson et al., 2011).

**Downward influence of the PJO**

The weak vortex events as defined by the NAM, and shown by Baldwin and Dunkerton (2001) to induce an equatorward shift in the tropospheric jets, also include a number of stratospheric sudden warmings and PJO events (the correspondence between these three descriptions will be explored in Chapter 4). A similar tropospheric response to the PJO has also been noted in observations (Kuroda and Kodera, 2004) and in a simplified circulation model (Kohma et al., 2010). Given that these methods of defining events in the Arctic overlap considerably, it is not clear how independent these results are. This question is addressed in Chapter 4.

### 1.3 Outline and contributions

Following this introduction is a brief chapter on the data sources analyzed in this body of work. To a large extent the research has been diagnostic in nature; the reanalysis and satellite observation products are publicly available and have been used ‘as is,’ and with a
few exceptions the significant model experiments were designed and carried out by others. My contributions have been in analysing and describing the circulation patterns captured by these data. A somewhat diverse range of data products have been considered; they and their relative merits in the context of this thesis are described in Chapter 2.

The third chapter presents an analysis of effective radiative damping times (both longwave and shortwave) in the middle atmosphere. It is based on Hitchcock et al. (2010)\textsuperscript{3}, with minor revisions to the text to better situate the discussion in the context of this thesis. The chapter first lays a theoretical framework for approximating radiative heating rates in the middle atmosphere by a linear relaxational term, and for a diagnostic method to estimate the optimal damping coefficients. This method is applied to a series of time-slice runs of a comprehensive chemistry-climate model. This work establishes the perhaps surprisingly strong justification for this approximation, but also highlights several particular cases in which the approximation fails. Most relevantly to the thesis, the analysis provides a more precise estimate of radiative timescales in the Arctic polar-night jet, showing them to be significantly longer than is often assumed. The analysis and theoretical lines of arguments are my own; Prof. Shigeo Yoden assisted with the framing and interpretation of the results and with the text.

Chapter 4 turns to the description, identification, and characterization of the PJO events themselves. For this purpose, the chapter presents a novel tool for the visualization of Arctic polar-night jet variability. These ‘abacus’ plots are then exploited to better understand the relationship between the monthly-timescale variability of the PJO and the more rapid processes described by sudden warmings and annular mode events. This analysis is carried out on a range of datasets which vary in their fidelity to the real atmosphere and length of record. In addition to the identification of the PJO events as a ubiquitous feature of the Arctic circulation, the results provide a test of several hypotheses existing in the literature regarding their long timescales, and present several new hypotheses regarding the processes relevant to these timescales. The analysis and framing of the approach is largely my own, though conversations with Dr. Gloria Manney have been instrumental in forming and developing my thoughts.

Motivated strongly by the statistical characterization of Chapter 4, the next chapter focuses on the description and analysis of the zonal mean dynamics responsible for the characteristic evolution of the circulation anomalies during PJO events. Chapter 5

\textsuperscript{3}\textcopyright 2010 American Meteorological Society, used with permission.
has two major components; the first is the presentation and discussion of a composite analysis of PJO events from a higher-resolution, 100-year time-slice run of CMAM. The second is a detailed analysis of the Eliassen adjustment processes that give rise to the anomalous zonal mean circulation pattern characteristic of these events, with particular attention paid to the radiative origins of the persistent lower stratospheric anomaly, and to the gravity wave dynamics that give rise to the anomalous polar stratopause which descends during PJO events. This also represents mainly my own work, though I am indebted to Drs. Isla Simpson and Michael Sigmond for setting up the CMAM simulations analyzed, and to Dr. John Scinocca in particular for discussion of the relevant gravity wave dynamics.

Chapter 6 discusses a series of simplified general circulation model experiments of the type performed by Taguchi et al. (2001), in which the lower stratospheric radiative damping timescales are modified systematically, while the tropospheric damping is held fixed. The motivation here was two-fold; firstly to understand the role of the lower stratospheric radiative damping timescales in the character of the variability produced by the model, and secondly to test the effects of the effective damping timescales computed in Chapter 3. The model is found to produce clear PJO events following major sudden warmings only when the lower stratospheric damping is weakened from conventionally-used settings to more realistic values. The tropospheric response to the PJO events is also analyzed in these events, revealing a significant role for planetary scale eddies and their associated form drag in the dynamics of the tropospheric jet shift. These runs were performed in collaboration with Prof. Shigeo Yoden and his student Shunsuke Noguchi at Kyoto University, and Prof. Masakazu Taguchi at the Aichi University of Education. The experiments were designed in collaboration. The runs were performed by Prof. Taguchi and S. Noguchi. The analysis is largely my own, though all members of the collaboration assisted with the interpretation and framing.

Three manuscripts, describing the results of Chapters 4, 5, and 6, and for each of which I am the first author, are in preparation.

Finally, results from the thesis as a whole are summarized and conclusions are drawn in Chapter 7. Plans for future work are also presented.
Chapter 2

Data

A variety of model simulations, reanalysis products, and observational datasets are used in this thesis; they are summarized and described here for reference. Note that the term ‘data’ is used here to refer to any representation of the geophysical fields of interest, whether they originate from some measurement of the real atmosphere or from a model integration.

2.1 Model simulations

2.1.1 Canadian Middle Atmosphere Model

The Canadian Middle Atmosphere Model (CMAM) is an extension of the Canadian Centre for Climate Modelling and Analysis (CCCma) tropospheric general circulation model (Scinocca et al., 2008), including a comprehensive set of physical parameterizations and a detailed representation of stratospheric chemistry (de Grandpré et al., 2000). The model has been validated against observations and other Chemistry Climate Models (CCMs) as part of both recent phases of the SPARC CCM Validation (CCMVal) activity (Eyring et al., 2006; SPARC CCMVal, 2010).

A variety of simulations with different configurations have been analyzed. Since the radiative heating and gravity wave parameterizations are of particular relevance to the present work, some of their relevant details are reviewed here. These details, with the exception of some technical details, are common to all the simulations considered here.

The radiation scheme employed by the CMAM is described in detail by Fomichev and Blanchet (1995); Fomichev et al. (1998, 2004), and references therein. The longwave
scheme used in the middle atmosphere includes the 15 µm carbon dioxide band, the 9.6 µm ozone band, and water vapour, and accounts for the breakdown of local thermodynamic equilibrium (LTE) in the mesosphere. The tropospheric scheme includes H$_2$O, CO$_2$, O$_3$, CH$_4$, N$_2$O, CFC-11 and CFC-12, clouds and aerosols. The shortwave radiation scheme includes effects of absorption by O$_3$, H$_2$O, CO$_2$, and O$_2$, and includes chemical heating, sphericity, and non-LTE effects in the near-infrared absorption by CO$_2$.

The gravity wave parameterizations used by CMAM include the orographic parameterization of Scinocca and McFarlane (2000), which extends that of McFarlane (1987) to include low level breaking and improved treatment of the anisotropic momentum flux generated by unresolved surface topography. Deposition of the momentum carried by vertically propagating waves occurs when they encounter critical levels, or when their amplitudes saturate. Saturation of the gravity waves occurs when a non-dimensional measure of the wave amplitudes exceeds a critical value Fr$_{\text{crit}}$. Non-orographically generated gravity waves are parameterized following Scinocca (2003).

CCMVal 1 ‘REF 2’ ensemble

For the statistical characterization of the PJO in Chapter 4, we considered the ensemble of three ‘REF 2’ runs from CCMVal-1 which are forced by projected transient greenhouse gases and ozone depleting substances from 1950 to 2100. The first decade (1950 to 1960) is discarded for spin-up, resulting in a total of 420 years of simulation. We have used the runs from the first phase of CCMVal because its Arctic circulation has been shown by several papers (McLandress and Shepherd, 2009; Hitchcock et al., 2009) to compare very closely to reanalyses, both in its mean state and in its variability. (More generally, CMAM was found to be among the best-performing of the models in terms of agreement with observations, according to the process-oriented grading exercise of Waugh and Eyring (2008).) This same close agreement was unfortunately not obtained by the CMAM integrations submitted for CCMVal-2 (Butchart et al., 2011), though the change in the model configuration responsible for this degradation has not been determined. The simulations were run at a horizontal resolution of T31 ($5.6^\circ \times 5.6^\circ$ linear transform grid) with 72 vertical levels from the surface to the upper mesosphere. The sea surface temperatures and sea ice concentrations in each ensemble member were specified from the output of three runs of a fully-coupled tropospheric model forced by the same emissions scenarios.
Forty-year time-slice runs

Four time-slice runs of 40 model years are considered in Chapter 3; they were suited to the analysis therein partly because of the fixed boundary conditions, and partly because the output included detailed information about the heating rates generated by the radiation scheme which is not present by default in CMAM output. In many respects their configuration is similar to the REF 2 ensemble. These runs included a dynamics-only configuration with prescribed ozone and sea surface temperatures (SSTs), an interactive chemistry configuration with prescribed SSTs, a coupled ocean configuration with prescribed ozone, and finally, a configuration with a coupled ocean and interactive chemistry. All runs had a CO$_2$ mixing ratio fixed at 348 ppm.

Ninety-six year time-slice runs

Finally, a time-slice run of 96 years is analyzed in Chapter 5. This was run in ‘dynamics only’ configuration, with a specified ozone climatology and climatological SSTs, but at a higher horizontal resolution of T63. The value of two parameters in the orographic gravity wave parameterization, a multiplicative factor $G(\nu)$ which scales the overall gravity wave momentum flux produced by the parameterization, and Fr$_{crit}$, described above, were adjusted slightly ($G(\nu)$ from 0.65 to 1.0 and Fr$_{crit}$ from 0.375 to 0.22) to recover sudden warming statistics in good agreement with the observations (Sigmond, pers. comm.; see also Sigmond and Scinocca, 2010).

2.1.2 Mechanistic Circulation Model

The simplified circulation model employed in Chapter 6 is the same as that of Taguchi et al. (2001), to which the reader is referred for further details. The model is run at a horizontal resolution of T21, with 42 vertical levels from the surface through the mesosphere. Radiation is parameterized by Newtonian relaxation towards the same equilibrium temperature field used by Taguchi et al. (2001). The model does not include a formal gravity wave drag parameterization; instead a linear frictional force is applied to the winds above 50 km as a sponge layer for the waves and as a means of closing off the stratospheric jets. Linear relaxation is also applied to the surface winds as a simple boundary-layer friction parameterization. Details of the particular experimental configurations used are given in Chapter 6.
2.2 Reanalyses

2.2.1 ERA40

Forty-five years of data from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA40) are used spanning September 1957 through August 2002 (Uppala et al., 2005). The model underlying the data assimilation system has a horizontal resolution of T159 and 60 vertical levels from the surface to 0.1 hPa, though data is provided only up to 1 hPa. The quality of stratospheric temperatures, particularly prior to 1979, is limited to some degree by inhomogeneities in the assimilated observations as will be apparent. The relatively long record is of particular interest, though, and the data is used as is.

2.2.2 MERRA

To bring the reanalysis record up to present day, data from NASA’s Modern Era Retrospective Analysis for Research and Application (MERRA) spanning January 1979 through April 2011 (Rienecker et al., 2011) are also used in Chapters 4 and 5. The underlying model is of somewhat higher horizontal and vertical resolutions (0.5° latitude by 0.67° longitude, finite-volume, and 72 levels to 0.01 hPa, respectively), as well as extending somewhat higher in the middle atmosphere (with data available to 0.1 hPa) than the ERA40 products. Since the underlying model is used for studies of stratospheric transport, some emphasis was placed on the quality of the middle atmosphere representation.

2.2.3 ERA Interim

Annular mode decorrelation timescales computed from the ERA Interim reanalysis (Dee et al., 2011) are presented in Chapter 4. They were computed by Dr. Isla Simpson. The horizontal and vertical resolution, and the temporal coverage are comparable to that of the MERRA reanalysis.
2.3 Satellite observations

2.3.1 Microwave Limb Sounder

The Microwave Limb Sounder (MLS) instrument on the Aura satellite has provided daily temperature profiles from the lowermost stratosphere through the mesosphere based on thermal microwave emissions from several chemical species (Schwartz et al., 2008). We make use of version 3.3 data (Livesey et al., 2011) from August 2004 through January 2011, and follow the recommended data quality screening procedures. The large vertical domain (from 316 hPa to 0.001 hPa) and relatively high vertical resolution (of order 3 km in the stratosphere, degrading in the mesosphere) provides a validation of the reanalysis data during this period, which is of particular importance in the upper stratosphere and lower mesosphere region where reanalyses are known to exhibit biases associated with the elevated stratopause that occurs during PJO events (Manney et al., 2008).
Chapter 3

Radiative Damping

3.1 Introduction

Radiative transfer plays an important role in damping temperature perturbations in the middle atmosphere. In general, this damping is a non-local process in which heat is transferred to and from remote levels of the atmosphere and the surface, and radiated away to space. This process is also non-linear, mostly as a result of the non-linear dependence of radiated power on temperature. In the mesosphere, molecular collisions occur sufficiently infrequently that local thermodynamic equilibrium no longer holds. A realistic representation of radiative heating in the middle atmosphere therefore requires detailed consideration of these effects.

Nonetheless, in the context of simplified modelling or theoretical studies, it is desirable to reduce the complexity of the radiative heating so that its consequences for other processes can be more easily understood. It is common, therefore, to approximate radiative damping as a local, linear relaxation to a reference temperature state (e.g. Holton and Mass, 1976; Taguchi et al., 2001; Polvani and Kushner, 2002). The time scale of this relaxation has important implications for the large-scale circulation of the middle atmosphere. In general, this damping contributes to the dissipation of wave activity (Andrews and McIntyre, 1978), and thus in part mediates wave-mean flow interactions (e.g.
in the driving of the QBO (Holton and Lindzen, 1972)), in addition to controlling the amplitude and structure of the waves themselves. Moreover, the magnitude of the thermal response to meridional overturning circulations induced by eddy forcings (such as gravity wave drag in the mesosphere) is proportional to the radiative damping time (Garcia and Boville, 1994).

Despite the approximations involved, this ‘Newtonian’ relaxation can in some cases describe longwave heating rates in the middle atmosphere surprisingly well. The purpose of this chapter is to explore in some detail where and to what extent these approximations of linearity and locality hold under realistic conditions in the stratosphere and mesosphere. Moreover, the estimates of radiative timescales found here will be applied directly in Chapters 5 and 6.

Accurate estimates of appropriate radiative damping rates, however, rely on detailed radiative transfer codes. Given a reference profile of temperature and radiatively active trace gases, these are used to compute explicitly at each altitude $z$ the change in heating rate $\delta Q(z)$ as a result of a given temperature perturbation applied to the entire column $\delta T(z')$. For instance, Dickinson (1973) computed damping rates $\alpha = \delta Q(z)/\delta T$ from the 1976 US Standard Atmosphere by perturbing the entire column by 0.1 K. Using a similar approach, several authors have computed damping rates based on observed profiles of temperature and ozone (Kiehl and Solomon, 1986; Gille and Lyjak, 1986; Mlynczak et al., 1999). The computed rates show a strong dependence on latitude and season. This dependence can be partially explained by the temperature dependence of the derivative of the Planck function, though the temperature dependence of the transmission functions are also potentially important.

An important consequence of the non-local nature of the radiative damping is that thermal perturbations with short vertical length scales are damped more rapidly than those with longer length scales. Fels (1982) showed that under WKB-like assumptions that are well satisfied in the middle atmosphere, wave-like disturbances are linearly damped at a rate $\alpha(z, n)$ dependent on the local vertical wavelength $2\pi/n$. They are computed simply by specifying a perturbation $\delta T(z') = \epsilon \cos(n(z - z'))$, with amplitude $\epsilon$. In this framework, the damping rates computed by the studies mentioned above correspond to $n = 0$. This scale-dependent parameterization has been updated to include non-local thermodynamic equilibrium (non-LTE) effects in the mesosphere (Fels, 1984), as well as more recent estimates of the quenching rate of CO$_2$ by atomic oxygen (Zhu, 1993), and the effect of curvature in the background temperature profile (Bresser et al.,
1995). Similar scale-dependency can also be defined for the effective damping due to photochemical feedbacks associated with shortwave radiation as well as transport-related effects (Haigh, 1985).

A number of studies have taken a more empirical approach to specifying the temperature perturbations being damped. In this approach, heating rates are computed off-line from observed (or re-analyzed) temperature and trace gas profiles, and effective damping rates are then defined by linear regression. Ghazi et al. (1985) used satellite retrievals of temperatures and ozone mixing ratios at polar latitudes on selected days with strong planetary waves, finding reasonable agreement with the longwave damping computed by Dickinson (1973). Pawson et al. (1992) extended this study to satellite observations of the whole middle atmosphere for several months’ worth of data. Using 17 years of NCEP daily analyses, Newman and Rosenfield (1997) demonstrated that such a linear regression model can capture a large fraction of the heating rate variance throughout the stratosphere, provided that damping times are allowed to vary with height, season and latitude.

We use a similar approach here to investigate the relationship between the temperature and radiative heating rates in the forty-year time-slice runs of CMAM described in Chapter 2. The use of a comprehensive model with external forcings held fixed allows us to attribute disparities between the statistical fit and the model data to the radiative response, as opposed to measurement uncertainties or secular changes in the true atmosphere. We are also able to obtain a relatively smooth function to describe the damping rates as a result of the improved statistics. Most results presented below are from the interactive chemistry run with fixed SSTs, though the fully coupled run gave very similar results. The effects of interactive chemistry are briefly discussed, though attributing the changes in damping rates between the interactive and non-interactive runs is difficult since the predicted ozone climatology is significantly different from the prescribed climatology (and consequently the temperature climatology is significantly different as well).

We find that the assumptions of locality and linearity do hold to a good approximation for longwave heating rates, even in the mesosphere where non-LTE effects become important. We find two notable exceptions where non-local effects become important, and one region in which non-linear effects are important. Firstly, in the lower tropical stratosphere, the spectrum of perturbations to the vertical temperature profile is sufficiently broad that scale-dependent effects present significant difficulty in defining a single
effective damping rate. Secondly, during the final break up of the Antarctic vortex, strong
temperature changes at the top of the decaying vortex dominate the radiative response
to the weak local temperature changes in the vortex core below. This non-local effect
cannot be described by the scale-dependent framework of Fels (1982). The assumption
of linearity fails only near the edge of the polar vortex where temperature variations are
large. These assumptions also hold for shortwave heating rates where ozone is under
strong photochemical control, though the damping rates are a strong function of the
zenith angle, which somewhat complicates the analysis.

This chapter is organized as follows. We first consider the longwave damping rates,
briefly presenting some radiative transfer theory that is useful for interpreting the re-
sults which we then present. Results for the shortwave damping rates are discussed in
the subsequent section, after which the chapter concludes with a linear estimate of the
radiative-photochemical equilibrium temperature, and a discussion of the implication of
these results for mechanistic modelling studies.

3.2 Longwave relaxation rates

3.2.1 Theory

All of the effective damping rates reported in this study have been computed using linear
regression techniques. To gain some insight into these empirical results, we briefly review
some relevant radiative transfer theory. Following Andrews et al. (1987), the longwave
heating rate due to a given species of absorber from a spectral band \( r \) can be written in
the exchange integral formulation

\[
\begin{align*}
    h_r(z) &= \frac{\pi m_a k_r}{c_P} \left[ \int_0^1 \left( B_r(T(z')) - B_r(T(z)) \right) d\Gamma_r(z, z') \right. \\
    &\quad \left. + \int_{z < z'}^1 \left( B_r(T(z')) - B_r(T(z)) \right) d\Gamma_r(z, z') \right] \tag{3.1}
\end{align*}
\]

where \( m_a \) is the mass mixing ratio of the absorbing species, \( c_P \) is the specific heat at
constant pressure, \( k_r \) is the absorption coefficient, \( B_r \) is the Planck function, and the
subscript \( r \) indicates a band-averaged quantity. The escape function \( \Gamma_r(z, z') \) is a nor-
malized first derivative of the transmission function, and for \( z' > z \) or \( z' < z \) gives the
probability that a photon emitted upwards or downwards (respectively) at $z$ reaches $z'$ prior to being absorbed. $\Gamma_r(z, z')$ provides a convenient coordinate system (for every $z$) in which the exchange of radiation between levels of the atmosphere is emphasized over details of the radiative transfer computation. The terms representing cooling to space and exchange with the surface are included in the integrals by setting $B_r(T(z'))$ to 0 where $\Gamma_r(z, z') \leq \Gamma_r(z, \infty)$ and $z' > z$, and to the flux density emitted by the surface $B_r(T_g)$ where $\Gamma_r(z, z') \leq \Gamma_r(z, 0)$ and $z' < z$.

Most radiation emitted by the 15 $\mu$m carbon dioxide band is exchanged between nearby levels so that the escape functions are strongly peaked (Leovy, 1984). If the temperature profile varies weakly on this radiative length scale, it is valid to expand the integrand of (3.1) as a Taylor series in $(z' - z)$. The Planck function and its derivatives can then be taken out of the integrals, and we can write (3.1) in terms of moments of the effective radiation lengths $\ell_{u,d} = \int_0^1 |z' - z|^m d\Gamma(z, z')$, where the superscripts indicate whether the integral is taken above or below $z$:

$$h_r(z) = \frac{\pi m_{CO_2} k_r}{c_p} \left[ - B_r(T(z)) \Gamma(z, \infty) ight. \\
+ \left. (\ell_1^u - \ell_1^d) \frac{dB_r}{dT} \frac{dT}{dz} + \frac{1}{2} (\ell_2^u + \ell_2^d) \frac{dB_r}{dT} \frac{dT}{dz}^2 + \ldots \right]. \quad (3.2)$$

The derivative of the first term in this expansion is sometimes called the Newtonian cooling coefficient. More generally, the (scale-dependent) change in the local heating rate $\delta h_r(z)$ associated with a temperature perturbation $\delta T = \epsilon \cos(n(z - z'))$ can be computed in this context.

One can posit three relevant length scales: that of the background temperature profile $L_c$, of the temperature perturbation $L'$, and of the radiation $L_r$. We take the ratio of $L_r$ and $L'$ to be $O(1)$ and define a non-dimensional number $L_n = L_r / L_c$. We further define $T_m$, the ratio of the temperature perturbation to the background temperature.
Retaining terms to $O(\ln T m^2)$ and $O(\ln^2 T m)$, the damping rate is

$$\alpha(z, n) = -\frac{\delta h_r(z)}{\delta T(z)} = \frac{\pi m_{CO_2} k_r}{c_p} \left[ \frac{dB_r}{dT} \left( \left( \Gamma(z, \infty) + \frac{1}{2}(\ell^u_2 + \ell^d_2)n^2 \right) \right) \right]$$

$$+ \frac{d^2 B_r}{dT^2} \left[ \frac{dT}{dz} (\ell^d_1 - \ell^u_1) + \left( \frac{1}{2} \Gamma(z, \infty) + \frac{1}{2}(\ell^u_2 + \ell^d_2)n^2 \right) \delta T \right] + \ldots \right].$$  \hspace{1cm} (3.3)

The $n^2$ dependence on the vertical wave number is recovered as a consequence of the radiative diffusion term. The expansion breaks down for large $n$ as the characteristic length scale of the temperature profile becomes shorter than the effective radiation length scale. Sasamori and London (1966) showed that in this limit damping rates vary as $n^{0.5}$, though the dependence saturates in the mesosphere as a result of non-LTE processes (Fels, 1984).

The damping rates depend on the background temperature primarily through the first derivative of the Planck function. They (and their scale dependency) will also tend to increase with height as the radiation length scales increase.

Corrections to the linear damping rate arise at higher order in $\ln$ as a result of the lapse rate and curvature of the background temperature profile. In our scaling the latter enters only at $O(\ln^3 T m)$ and has been dropped, but it may be of equal importance to the lapse rate correction due to cancellations between $\ell^u_1$ and $\ell^d_1$ (Bresser et al., 1995). Non-linearities in this expression are a result of the curvature of the Planck function and arise at $O(\ln T m^2)$.

Terms proportional to (odd powers of) odd derivatives of the temperature will force a change in the vertical phase of the temperature perturbation; these too tend to be small in the middle atmosphere as a result of cancellations between the odd moments of the upward and downward radiation lengths (Fels, 1984).

This expression (3.3) neglects the temperature dependence of the absorption coefficients (and thus the escape function). Other studies have reported that they contribute no more than 20% of the total change in heating rates (Zhu, 1993), though to our knowledge no systematic investigation of their role in determining the damping of large scale perturbations has been carried out. The expression can also be used for the 9.6 $\mu$m band of ozone, though the effective radiation lengths will generally be longer, and in the lower-
most stratosphere exchange with the surface (or cloud tops) dominates. The expression also neglects the temperature dependence of the ozone mixing ratio. Pawson et al. (1992) found some evidence that this may destabilize the lower tropical stratosphere. Though we find strong evidence for the importance of scale dependence in this region, we have not found evidence of any such radiative instability in CMAM.

### 3.2.2 Linear regression estimates

To complete a theory of longwave damping rates in the middle atmosphere, we would need a radiative transfer model to compute the escape functions, as well as a theory for the structure of the temperature profile and typical perturbations. We proceed instead in an empirical fashion, by using temperatures and radiative heating rates generated by a GCM. We define the background state as the zonal and climatological mean (denoted by square brackets and a subscript \(c\), respectively), and the perturbation as the daily mean anomaly from this state (defined from 6-hourly model output), \(X' = X - [X]_c\). Effective damping rates are then computed at every day \(d\), latitude \(\phi\) and log-pressure height \(z\) by linear regression

\[
Q_{\text{LW}}' = -\alpha(d, z, \phi)T' + \epsilon, \quad \alpha(d, z, \phi) = -\frac{[Q'T']_c}{[TT']_c}. \quad (3.4)
\]

Fits are computed over all model years and longitudes, giving approximately \(3 \times 10^3\) (non-independent) samples at each point. The use of daily mean anomalies removes the tides. This is appropriate for mechanistic modelling studies which do not include a diurnal cycle; moreover, as discussed below, the effective damping rates agree well with tidal damping rates computed by McLandress (2002). Examples of daily fits are shown in Fig. 3.1(a-c). Monthly averages are then performed over the fitted parameters to smooth the fields. Note that this definition of the heating rate and temperature anomalies from a climatological and zonal mean differs from that used in other studies using this approach which use anomalies from the zonal mean. This allows us to include the damping of zonal mean perturbations, which would be complicated by the presence of secular changes in observed and reanalyzed fields.

To the extent that linear superposition can be applied, this effective damping rate can be related to the scale-dependent rates by decomposing the anomalies into vertical spectral components \(T'_n\) and \(Q'_n\). The effective damping rate can be seen as an average,
Figure 3.1: Regression of longwave heating rates against temperatures on April 15th at (a) 52° N, 5 hPa, (b) 58° S, 0.4 hPa, and (c) 13° S, 40 hPa. The effective damping rate and correlation coefficient are indicated in the legend. (d) Fraction of variance in longwave heating rates explained by local temperature anomalies. The monthly mean for April is shown.

weighted by the fraction of the total power in each wave number $n$

$$\alpha(d, z, \phi) = - \sum_n [Q_n^* T_n^l]_c [T_n^l T_n^l]_c$$

$$= \sum_n \alpha(d, z, \phi, n) [T_n^l T_n^l]_c,$$

where the asterisk denotes complex conjugation. While conceptually useful, this decomposition would need to be local in nature to be used quantitatively. This argument can also be extended to predict the correlation coefficient of the fit, though we will show that the presence of perturbations at a variety of different scales is not the only reason the fit breaks down.

Figure 3.1(d) shows the correlation coefficient as a function of latitude and height for the month of April. In general, above the tropopause, the statistical model fits quite well. In the middle to upper stratosphere, the local, linear approximation captures nearly all of the variance of the longwave heating rates with correlations greater than 0.9 in the
extratropics (cf. Fig. 3.1(a)), and somewhat lower correlations in the tropics. The quality of the fit drops off somewhat in the mesosphere (cf. Fig. 3.1(b)), with minima in the extratropical lower mesosphere and near the stratopause and mesopause in the tropics. In the extratropical lower stratosphere where the temperature varies comparatively little, the fit still captures greater than 60% of the variance.

However, in the tropical lower stratosphere the fit breaks down significantly. Figure 3.1(c) shows a scatter plot at 13° S, 40 hPa on 15 April. The majority of points lie near the fitted line, but significantly stronger cooling rates than predicted by the linear relation occur as well. Inspection of individual model days shows that these regions of strong cooling are typically associated with synoptic scale temperature anomalies, though the presence of high cloud tops may also play a role. These regions of low correlation shift seasonally with the climatological easterlies in these simulations, suggesting that the breaking of westward travelling waves may play a role in generating these synoptic-scale

Figure 3.2: (a,b) Latitude-height plots of monthly mean $r^2$ for January, defined as in Fig. 3.1(d). (c,d) Month-latitude plots of $r^2$ at 10 hPa. Correlation coefficients are from regressions of the longwave heating against anomalies from the climatological temperature including in (a) and (c) a linear term only, while in (b) and (d), both a linear and a quadratic term are included. (e) Longwave heating rates versus temperatures on 15 July at 52° S, 10 hPa. The parameters of the regression line shown are indicated in the legend. (f) Same as (e) on 15 January at 58° N, 10 hPa.
structures.

To test for the importance of non-linearity in the response, we include a quadratic term in a multiple regression

$$Q_{LW}' = Q_0 - \alpha T' - cT'^2 + \epsilon$$

with residuals $\epsilon$ and fitted parameters $Q_0$, $\alpha$, and $c$. Including a quadratic term in the fit describes an additional 5% of the variance near the vortex edge in the middle to upper stratosphere in both hemispheres (Fig. 3.2(a-d)), as a result of the large range of temperatures found there. A roughly constant value of $(9 \pm 3) \times 10^{-4} \text{ K day}^{-1}$ is found for the coefficient $c$ over this region (Fig. 3.2(e-f)). As a result of the large amount of data, the additional degree of freedom statistically improves the fit nearly everywhere. The differences are, however, physically negligible in other regions and seasons. Over the 60 K range of temperatures shown in Fig. 3.2(e-f), the linear damping rate varies by about 0.05 day$^{-1}$ which is of the same order as the linear rate itself, which indicates the potential dynamical importance of this non-linearity for disturbances to the polar vortex.

### 3.2.3 Effective longwave damping rates

The effective damping rates are compared against the parameterization of Fels (1982) in Fig. 3.3 for log-pressure heights from 20 to 70 km. The parameterized damping rates are computed from the area-weighted, monthly-mean, climatological temperatures in the three latitude ranges shown. Damping rates regressed from the model runs are also shown. The error bars are an estimate of the sampling error from the regression technique. They are estimated by computing effective damping rates for each non-overlapping five-year period in the model runs. The error bars are then twice the standard deviation of the mean effective damping rates over the set of five-year periods. Profiles are shown for the tropics (Fig. 3.3(a)), southern winter mid-latitudes (Fig. 3.3(b)), and Arctic winter (Fig. 3.3(c)); damping times are shown for the latter to emphasize lower stratospheric values. The regressed profiles lie within the range of parameterized rates predicted for large vertical wavelengths. Since the regressed rates are a weighted average of scale-dependent damping rates, the additional vertical structure present in the regressed rates may be a result of changes with height in the spectrum of vertical wavelengths present. Although differences may also arise from the climatological ozone profile, the time-dependent mod-
Figure 3.3: Damping rates as a function of log-pressure height averaged over (a) October, 15° S to 15° N, (b) July, 60° to 30° S, and (c) January 60° to 90° N. The dashed lines are computed using the parameterization of Fels (1982) with climatological temperatures taken from CMAM and vertical wavelengths $\lambda$ as labeled. The symbols indicate regressed damping rates, with error bars indicating an estimate of the sampling uncertainty (see text for details).

Monthly means of the linear damping rates computed from the quadratic fit (3.6) are shown in Fig. 3.4 for January, April, July, and October. The plots are shaded where the regression coefficient $r^2$ drops below 0.5 to indicate where the quality of the fit declines in each month. The broad vertical structure of low damping rates near the tropopause rising to a maximum near the stratopause and falling again in the mesosphere is apparent through most latitudes and seasons. The lowest damping rates are found near the extratropical tropopause, particularly in the winter hemisphere, despite its being warmer than the summer mesosphere. Significant meridional and seasonal variations are apparent. In the stratosphere, damping rates tend to peak in the tropics and fall off towards both poles. There is a stronger seasonal variation towards the poles, with the lowest damping rates at the base of the polar vortices.

In the mesosphere the seasonal cycle is reversed, consistent with the seasonal cycle in mesospheric temperatures. The strong peak in damping rates in the winter mesosphere above Antarctica is partly due to high temperatures, but is more directly a result of the presence of a tertiary ozone maximum (Marsh et al., 2001). This feature of the
ozone climatology in CMAM has not been compared in detail to observations (which are limited in this region). Nonetheless, it is notable that ozone has such a strong impact on the polar mesospheric damping rates. The asymmetry between the two hemispheres is a result of higher temperatures and ozone mixing ratios in the Antarctic winter at mesospheric heights. The increase in damping rates in the tropics towards 0.01 hPa apparent in all seasons has also been noted in the extended CMAM (with model lid near 200 km) (McLandress, 2002), and is associated with a large peak in CO$_2$ cooling rates in the thermosphere.

The features of the seasonal cycle reported here are qualitatively similar to those reported by Kiehl and Solomon (1986) and Mlynczak et al. (1999), though the damping times computed here are not as long. The differences near the stratopause and in the extratropics are roughly 20%-50%, but the extremely long damping times previously reported in the lower tropical stratosphere are absent here. These differences are likely due to the constant temperature perturbation used by both of these studies to compute the damping times, which would result in a larger disparity with our results near the
Chapter 3. Radiative Damping

Figure 3.5: Tropical monthly mean damping rates; dark shading as in Fig. 3.4. Regressions were performed using (a) zonal mean anomalies only, (b) zonal wave numbers 1 to 3, (c) zonal wave numbers 4 and 5.

equator as a result of the shorter vertical length scales associated with tropical waves.

The dependence of tropical damping rates (20° S to 20° N) on zonal wave number is shown in Fig. 3.5. This dependence is computed by filtering the anomalies prior to performing the regression. The seasonal cycle of damping rates at the equator is shown for zonal mean disturbances (Fig. 3.5(a)), for zonal wave numbers 1 to 3 (Fig. 3.5(b)), and for zonal wave numbers 4 and 5 (Fig. 3.5(c)). A small semi-annual cycle is present in the lower mesosphere, consistent with the timing of the semi-annual oscillation in temperatures in these runs. Planetary-scale disturbances are damped significantly more strongly than zonal mean perturbations (the differences are considerably greater than the error bars estimated in Fig. 3.3(a)). The dependence on wavelength increases with height, consistent with Fels (1984). We find that the dependence on zonal wavelength is strongest in the tropics.

Turning to the polar regions where the seasonal cycle is much stronger, we find significant differences between the hemispheres (Fig. 3.6). The seasonal cycle of damping rates averaged poleward of 70° N and 70° S are shown in Figs. 3.6(a) and (b), respectively. Enhanced damping rates are present above 0.1 hPa during the polar night in both hemispheres, though the peak is larger in the Antarctic. Effective damping rates are small in both polar vortices due to the low temperatures, but are smaller in the Antarctic, consistent with the stronger, colder vortex. However, the correlation coefficient for the fit differs drastically between the two hemispheres (Figs. 3.6(c,d)). While the correlation remains high throughout the year in the Arctic (with a minimum in the lowermost stratosphere during the summer months), the local relationship breaks down entirely in the Antarctic lower stratosphere (near 100 hPa). The fit fails within the Antarctic vortex,
and is weakest during the spring as the vortex breaks down from above. In contrast with the Arctic, strong thermal variability in this downward propagating region is a feature of the model’s climatology (Figs. 3.6(e,f)). Within the region of strong variability the linear relationship between temperature and heating rates holds, but below this it fails.

The downward influence of temperature variations at the top of the vortex is further demonstrated in Fig. 3.7. Figure 3.7(a) shows a scatter plot of longwave heating rates against temperatures from the model within the vortex at 50 hPa, 80° S on 26 September. The failure of the assumption of a climatological local damping rate is evident. The clusters of points which extend with both positive and negative slope are associated with individual years, during some of which longwave radiation acts to damp perturbations, but in others acts to amplify perturbations. The zonally asymmetric component of the temperature (Fig. 3.7(b)), longwave (Fig. 3.7(c)) and shortwave (Fig. 3.7(d)) heating rates are also shown for one such year. The strong wave-1 disturbance in temperatures between 25 and 40 km evanesces downward into the lower stratosphere. In the region from 25 to 40 km, the longwave heating is strongly anti-correlated with the temperatures. However, from 15 to 25 km, as a result of the large amplitudes above, the

Figure 3.6: (a,b) Damping rates as a function of month and pressure, averaged over 70°-90° N and 70°-90° S, respectively, with shading as in Fig. 3.4. (c,d) Correlation coefficients for the same latitudes. (e,f) Standard deviation of temperatures at the same latitudes. Northern Hemisphere plots are shifted six months to center the winter months.
longwave heating rates and temperatures are positively correlated. Shortwave heating rates (Fig. 3.7(c)) provide a further mechanism for downward radiative influence. From 35 to 40 km, shortwave heating rates are also anti-correlated with the local temperature, as a result of the photochemical control of ozone abundances. Between 30 and 35 km, however, the opacity effect (Ghazi et al., 1985; Pawson et al., 1992) leads to a positive correlation. Below about 25 km, the heating rate remains positively correlated as a combined result of ozone transport and the opacity effect.

The heating rates induced remotely by the evanescent wave are relatively small (about 0.1 K day$^{-1}$ in this case, which was selected for its clarity). However, they can clearly dominate heating driven by local temperature perturbations in the Antarctic vortex. Moreover, the effective damping rates can clearly vary from year to year in this region and season (Fig. 3.7(a)), presumably as a result of the dynamical variability associated with the timing of the final warming. Though we have not demonstrated what role these
non-local rates play in the breakdown of the vortex in this model, this result provides a caution against using a local, climatological radiative damping rate in mechanistic modelling studies of this region.

### 3.3 Shortwave relaxation rates

A similar regression analysis can be performed on shortwave heating rate anomalies in order to diagnose effective damping rates from photochemical- and transport-related ozone perturbations (i.e. Ghazi et al., 1985). The analysis is somewhat complicated by the strong dependence of the effective damping rate on the diurnal cycle. While the regression can be performed on daily, zonal mean temperatures and heating rates, this significantly reduces the amount of data used to perform the fits and reduces their statistical significance. Instead, we diagnose this dependence directly by performing a linear regression for every day of the year as a function of the local time. Sample scatter plots are shown in Fig. 3.8, in the southern mid-latitude spring near the stratopause at
Figure 3.9: Daily averaged shortwave damping rates and correlation coefficients for December (a-b) and June (c-d). A day-night mask is applied when averaging the correlation coefficients. Contour spacing for the shortwave damping rates are 0.02 day$^{-1}$. Contours for the correlation coefficients are shown at 0.2, 0.4, 0.5, 0.6, and 0.8.

local noon (Fig. 3.8(a)) and in the sub-Arctic mid-stratosphere summer at 06:00 local time (Fig. 3.8(c)). Where ozone is under photochemical control, there is a strong, negative, and linear correlation between temperature and shortwave heating. Where it is controlled by transport, background ozone gradients tend to lead to positive correlations between temperature and shortwave heating (Hartmann, 1981; Nathan et al., 1994). In the latter case, the regression coefficient between temperature and shortwave heating is weaker. In both cases, a clear dependence on zenith angle is apparent (Fig. 3.8(b,d)). Error bars in the fits indicate the 95% confidence interval of the regressed slope.

Figure 3.9 shows daily averages of the shortwave damping rates and correlation coefficients for the months of December (Fig. 3.9(a-b)) and June (Fig. 3.9(c-d)). A day-night mask is applied to the correlation coefficients (but not the regression coefficients) when performing the diurnal average. Photochemical damping is present throughout the upper stratosphere and lower mesosphere, with a broad maximum near 1 hPa of 0.08 day$^{-1}$, falling off more quickly below than above. Somewhat larger values of 0.12 day$^{-1}$ are found above the summer pole. Significant positive correlations between temperatures
and shortwave heating rates are found throughout the summer lower stratosphere, with negative effective damping rates peaking at -0.04 day\(^{-1}\) above the summer pole. This is consistent with the unstable, linear, westward-travelling Rossby modes found there by Nathan et al. (1994). The increased damping towards the polar region is in part a result of the length of the polar day. Similar results are seen for June, though the negative damping rates are stronger (peaking at -0.08 day\(^{-1}\)) and the damping above the pole is weaker. The correlation coefficients above the summer pole are weaker during boreal summer than during austral summer, where they maximize in December. No significant correlations were found elsewhere in the lower stratosphere, or above 0.05 hPa.

The values determined here agree well with those determined using similar approaches with observed temperature and ozone fields. Ghazi et al. (1985) found shortwave damping rates of 0.12±0.06 day\(^{-1}\) at 1 hPa computed from three days of data in the late winter near the vortex edge, falling to 0.02±0.02 day\(^{-1}\) at 2 hPa, consistent with the gradients near 60° found here. Equatorial, solstice photochemical damping rates are also comparable with the scale-dependent rates computed by Haigh (1985), lying between the damping rates computed for uniform (\(2\pi/n = \infty\)) and short-wavelength (\(2\pi/n \approx 7.5\ km\))

![Figure 3.10: Sum of monthly mean longwave and shortwave damping rates (day\(^{-1}\)) where the longwave correlation coefficient \(r_{LW}^2 > 0.4\) for (a) January, (b) April, (c) July, and (d) October. Dark shading as in Fig. 3.4. Note the zero contour in the lower Antarctic stratosphere in (d).](image-url)
perturbations. The meridional structure near the stratopause is also in close agreement. Pawson et al. (1992) found photochemical damping of approximately equal strength, though the region of transport-induced positive correlations was not identified.

The photochemical damping is a significant additional source of damping in the upper stratosphere and lower mesosphere. The total effective radiative damping \((\alpha_{\text{LW}} + \alpha_{\text{SW}})\) is shown in Fig. 3.10. At the peak in longwave damping near the tropical stratopause, the daily, zonal mean photochemical damping contributes an additional 30%; this rises to nearly 50% of the radiative damping in the extratropics, and is roughly equal to the longwave damping above the summer Antarctic. Significant hemispheric asymmetry is apparent at the tropical stratopause near solstice conditions as a result of the stronger photochemical damping in the summer hemisphere. The negative damping in the polar summer lower stratosphere is smaller in magnitude than the positive longwave damping. In the Antarctic spring, however, there is effectively no radiative damping as a result of the additional non-local shortwave effects (Fig. 3.7(d)).

3.4 Radiative-photochemical equilibrium temperature

The quality of the empirical fits throughout much of the modelled middle atmosphere suggests the possibility of their use in computing a ‘radiative-photochemical equilibrium’ temperature \(T^*_r\) towards which the linearized radiative parameterizations of CMAM are effectively relaxing:

\[
T^*_r = [\overline{T} c] + \frac{1}{\alpha} \left( [\overline{Q_{\text{LW}}} c] + [\overline{Q_{\text{SW}}} c] \right), \quad \alpha = \begin{cases} \alpha_{\text{LW}} + \alpha_{\text{SW}} & \text{if } \alpha_{\text{SW}} > 0, \\ \alpha_{\text{LW}} & \text{otherwise.} \end{cases}
\]

Since the negative damping rates are a result of transport, only the photochemical component of the shortwave damping rates are included. This estimate of \(T^*_r\) is shown for 15 January in Fig. 3.11. Also shown for comparison is a computation from a time-marched computation of temperatures determined by radiative and photochemical processes (Fels, 1985). Between 10 hPa and 0.1 hPa and south of about 45° N, where the effects of dynamics on zonal mean temperatures are small, the linear estimate matches the computation of Fels (1985) closely. The extremely cold radiative equilibrium in the polar vortex, however, is not captured. This failure is not solely a result of the curvature of the Planck function (though it is clear from Figs. 3.2(e-f) that linear extrapolation will overestimate
Figure 3.11: (a) Linear estimate of the radiative equilibrium temperature for 15 January. (b) Temperatures determined by a radiative-photochemical model on the same date computed by Fels (1985) by time stepping through the seasonal cycle. Note the non-standard vertical scale of (b), and that the Southern Hemisphere is to the right in both panels.

The linear estimate also over-estimates true radiative equilibrium in the upper summer mesosphere, where non-local effects (from the warm stratopause below) are also important.

3.5 Discussion and conclusions

The assumption that radiative heating rates in the middle atmosphere respond linearly to changes in local temperature is satisfied sufficiently well in most cases to define an effective relaxation rate that varies with height, latitude and season. This effective damping
rate can be usefully diagnosed with simple linear regression models. This has been shown before in the stratosphere by studies combining observed or reanalyzed temperatures and trace gas concentrations with off-line radiative transfer models (Pawson et al., 1992; Newman and Rosenfield, 1997). Here we have explored this relationship more thoroughly in a chemistry-climate model with comprehensive physical, radiative, and chemical parameterizations. The model broadly confirms this result throughout the stratosphere and mesosphere, where the linear model explains more than 90% of the variance in longwave heating rates through much of the extratropical stratosphere, and more than 80% in the mesosphere and tropical upper stratosphere. Where it is well specified, the regression model diagnoses effective longwave damping rates in these forty-year time-slice runs to a sampling error of about 5%. The regressed damping rates agree roughly with the parameterization of Fels (1982), but suggest that changes in the spectrum of vertical length scales in temperature disturbances generate a greater meridional and seasonal variation than would be implied by the dependence on background temperatures alone.

Shortwave heating rates are also well modelled by a simple, diurnally averaged linear damping rate near the stratopause where the correlation coefficient is greater than 0.8. In this region they provide a significant correction to longwave damping of temperature anomalies. Positive correlations with the temperature are also found in the lower summer stratosphere as a result of eddy transport of ozone, though they are less well characterized by a climatologically-specified effective damping rate.

The linear regression model also provides a means of estimating the climatological radiative-photochemical equilibrium temperature in the middle atmosphere. This calculation for the runs analyzed here agrees well with the explicit calculation of Fels (1985) where climatological temperatures do not depart drastically from the equilibrium, but significantly overestimates the equilibrium temperature in the polar night.

Understanding where the regression model fails in the chemistry climate model runs analyzed here gives insight into three ways in which a simple Newtonian cooling parameterization fails to capture the response of radiative heating rates to temperature perturbations.

Firstly, in the lower tropical stratosphere, temperature perturbations with small vertical scales are damped significantly more strongly than more typical perturbations with larger vertical scales (Fig. 3.1(c)). The details of this particular failure are almost certainly sensitive to the presence of a QBO (which is absent in these runs); however, the presence of a broader spectrum of vertical scales in the tropics suggests that the neglect
of the scale dependence of damping rates there is of more general concern.

Secondly, in the lower Antarctic vortex (particularly in late winter and spring), the assumption of locality again fails. Here, however, it fails in a way that cannot be explained by the scale-dependent parameterization. The variance in temperatures above the relatively stable Antarctic vortex is large enough to dominate the local influence on longwave heating rates. Moreover, the behaviour of the longwave radiation depends strongly on the state of the vortex from year to year. The assumption of a inter-annually constant, linear damping rate does not hold.

Finally, the large climatological variance in temperatures near the edge of the polar vortex in both hemispheres leads to significant non-linearity in the longwave heating rates as a result of the curvature of the Planck function. Dynamically, this may be the most important departure from the assumptions implicit in the use of an effective linear relaxation identified in this work, given that the linearized relaxation rate can vary by a factor of two near the vortex edge when a realistic parameterization of longwave heating rates is used.
Chapter 4

Statistical characterization

all the eds and eddies and the sad co-fishermen of drag, the agitated hearts and hearticles of what we cannot say there is at present.

Don McKay – Turbulence

4.1 Introduction

In contrast to the relaxational nature of radiative heating described in the previous chapter, the strongly non-linear interactions between the vortex itself and the planetary-scale Rossby waves propagating into the stratosphere from the troposphere below drive the strong and intermittent variability of the Arctic polar-night jet. These dynamics manifest themselves most spectacularly in the form of stratospheric sudden warmings (SSWs), during which the climatological eastward flow reverses on time scales of less than a week. In contrast to the suddenness of this onset, the vortex can in some cases take as long as several months to recover to its climatological state. The definition, description, and characterization of these recoveries, termed here Polar-night Jet Oscillation (PJO) events, are the central focus of this chapter.

As discussed in Chapter 1, a great variety of indices and events have been defined in order to describe and classify the variability of the Arctic vortex. In order to justify the introduction of yet another event, we propose four criteria to evaluate the definition of an event that should ideally be met by any novel classification:

1. be robust to small changes in their definition,
2. produce events that are similar to each other in some sense beyond the criteria used to define them,

3. afford some novel understanding of the behaviour of the vortex which is not accessible through existing definitions,

4. capture similar events in a variety of datasets, including simplified model integrations.

Some of these criteria bear further comment. The first is a commonly applied standard. The second is particularly important in order for composite averages to have some physical meaning. If non-linearity is important, there is no guarantee that the behaviour of the composite obeys the same dynamics as the individual events. The intent of the third is to emphasize (to the extent possible) the understanding of the vortex, rather than of the index itself. Finally, the fourth criterion is essential for comparisons between datasets; this is particularly important for diagnosing the behaviour of models.

To describe the PJO, this chapter makes use of the first two empirical orthogonal functions (EOFs) of polar cap averaged temperature profiles introduced by Kuroda and Kodera (2004). They are exploited here to visualize the variability of the Arctic vortex on daily timescales sufficiently compactly that several decades can be presented at once. This tool is applied to compare the detailed behaviour of the Arctic vortex in satellite observations (from the Aura MLS instrument), modern reanalyses (ERA40 and MERRA), and a comprehensive chemistry climate model (specifically, the REF 2 ensemble of CMAM simulations).

The goals of this chapter are as follows. Foremost, this visualization tool is used to demonstrate that the monthly timescale variability reviewed in Section 1.2.2 occurs most evidently following a subset of roughly one-half of stratospheric sudden warmings, and that these episodes exhibit dynamics both robustly similar to each other, and distinct from the dynamics following other warmings. In so doing, several hypotheses regarding the origin of the long timescales are tested.

In the context of the PJO, the timescale of the variability is likely to be strongly related to the persistence of the circulation anomaly in the lower stratosphere following the sudden warming. There are several ideas for why polar circulation anomalies should exhibit extended persistence in the lower stratosphere. The simplest is that radiative timescales decrease with increasing altitude and are at their longest in the lower stratosphere (Chapter 3). Since warming events tend to begin in the upper stratosphere and
then descend, this vertical gradient in radiative timescales suggests that the persistence of a given warming should be closely related to the depth in the stratosphere to which it descends. Analyzing a series of ensemble forecasts using a simplified general circulation model, Gerber et al. (2009) confirmed this relationship; moreover they found that warmings that reached the tropopause were also followed by a more robust shift of the tropospheric jets. The depth to which the warming descends has been associated with the persistence of the waves which induce the warming. This was noted in detail by Harnik (2009) in the context of looking for periods during which planetary wave reflection from the stratosphere is strong. In particular, those events which were triggered by a brief pulse of waves disrupt only the upper stratosphere and put the vortex into a configuration favourable for reflection. Those events triggered by an extended pulse can disrupt the lower stratosphere as well, and are found in the following to coincide with PJO events. This result is also consistent with those of Zhou et al. (2002) who studied several cases of warm anomalies, finding that those which descended from the upper stratosphere required an extended period of wave driving to disrupt the lower stratosphere. The descent of the warming will play a central role here as well in identifying PJO events.

The radiative timescales responsible for the extended persistence of the lower stratospheric anomaly, however, are only one part of the dynamics. One must also understand why the eddies (which can act on the mean flow on much shorter timescales) should remain quiescent over such extended periods. This suppression does not necessarily follow from the presence of long radiative timescales—indeed, Charlton-Perez and O’Neill (2010) found that adjusting the radiative damping timescales in the stratosphere in a general circulation model similar to Gerber et al. (2009) did not have a strong impact on the timescale of this persistence, suggesting that the suppression of eddies is a non-trivial effect. There are two (not necessarily exclusive) potential paradigms here. The first is that the generation of planetary waves by the tropospheric flow is naturally intermittent, in which case the stratosphere is in some sense simply responding passively to the circulation below (Plumb and Semeniuk, 2003). The second is that the zonal mean stratosphere itself filters or reflects the waves through the Charney-Drazin criterion or some generalization thereof. This filtering is typically associated with the wind reversal which defines the onset of the sudden warming.

Another hypothesis is that the zonal wave number of the eddies which disturb the vortex is relevant. For instance, several studies have suggested that warmings during which the polar vortex splits have longer timescales than those in which the vortex
is displaced off the pole. Evidence for this has been found in observations (Charlton and Polvani, 2007) and in general circulation models (Yoden et al., 1999). One rather heuristic argument for why this should be is that the displacement of the vortex off the pole might in some sense be more dynamically reversible than the splitting of the vortex. An alternative possibility is suggested by recent work that has identified the relevance of a barotropic wave-mode to vortex splitting events (Esler and Scott, 2005; Liberato et al., 2007; Matthewman and Esler, 2011). This mode would be expected to disrupt the lower stratospheric vortex much more efficiently than the vertically propagating modes responsible for vortex displacements [and indeed vortex splits are found to be much more barotropic in character (Matthewman et al., 2009)]. If the zonal mean lower stratospheric disturbance is indeed of leading importance to the length of the recovery, this could imply that the longer timescales associated with vortex splits are more directly a consequence of the vertical structure of the wave driving, rather than the zonal wave number of the eddies.

As alluded to in Chapter 1, a wide variety of descriptive indices and events have been defined and applied to the variability of the vortex, and the extended timescales associated with the events considered here have been noted in a number of contexts. In addition to the existing literature on the PJO itself reviewed in Section 1.2.2, the events are described implicitly in composites of sudden warmings (Limpasuvan et al., 2004; Charlton and Polvani, 2007) and of weak vortex events (Baldwin and Dunkerton, 2001), which they dominate, particularly at large lags. The warm anomaly at mesospheric height (see Fig. 1.2 or below) results in an unusually high polar stratopause, which descends with the anomaly over the course of the event. These elevated stratopause events have been noted and discussed by a number of recent studies (e.g. Manney et al., 2008). The association of these events with extended timescale sudden warmings noted explicitly by Siskind et al. (2010) is confirmed by the results presented here. The relation of PJO events to the elevated stratopause events will be treated in more detail in the next chapter. Finally, extended timescales in the stratosphere are also implied by the long decorrelation timescales of the NAM (Baldwin et al., 2003; Gerber et al., 2010), though as these are a characterization of the whole time series, their connection with the event-based perspective adopted here is not clear. Particularly in light of the third criterion proposed above, an effort is made below to relate the PJO events to these descriptions. We note that a similar comparison between the two EOFs of polar cap averaged temperatures used to define the PJO has been performed on an extremely long
(15,000 year) integration of a simplified general circulation model (Kohma et al., 2010). The classification of PJO events in the present work differs significantly: emphasis is placed in particular on the descent of the anomalies to the lower stratosphere. The variability of the vortex in the real atmosphere and in the comprehensive model can also be expected to differ in some ways from the simplified model as a result of the absence of parameterized gravity waves in the latter.

This chapter is organized as follows. Section 4.2 reviews the two types of events which will be compared in detail with the PJO: stratospheric sudden warmings, and weak vortex events (as defined by the NAM). The former are further classified into vortex splits and vortex displacements; this classification of events in MERRA is updated through to 2010-2011, and is applied to the CMAM simulations. This classification has not previously been applied to a simulation performed for the CCMVal intercomparison project. The novel visualization tool and the definition used to identify PJO events are also introduced.

Section 4.3, which presents the main results, is divided into three parts. In the first, the robustness of the PJO definition is tested explicitly in the reanalyses and the model, demonstrating in detail the connection between the depth of the warming and the persistence of the event. The second focuses on the relationship of sudden warmings to PJO events, demonstrating the extended suppression of planetary waves in the vortex during PJO events, and evaluating the connection between the zonal wave number of the warming and the timescale of the event. The third focuses on the relationship of weak vortex events to the PJO, demonstrating that the equatorward shift of the tropospheric jets is more robust during PJO events.

Several additional points are discussed in Section 4.4. Firstly, the robustness of cold events analogous to PJO events but of the opposite sign is considered in the context of several commonly used metrics. Secondly, several challenges to the detection of changes in the behaviour of the Arctic vortex associated with the large amplitudes and statistical character of PJO events are also discussed. Finally, the conclusions of this chapter are summarized in Section 4.5.

### 4.2 Methods

For ease of reference, we describe here the calculation of the indices and event definitions used below. Many indices require deseasonalized and detrended anomaly fields. To compute these anomalies for the reanalyses and CMAM, we fit a linear trend at each
grid point and day of the year. We then smooth this background (mean and trend) by fitting the harmonics of the annual cycle and retaining only up to the fourth mode. Anomalies are then computed by subtracting this smoothed background from the field. This approach is similar to that outlined by Gerber et al. (2010); since the impact of ozone depletion on temperatures is considerably weaker in the Northern Hemisphere than in the Southern Hemisphere, we do not expect the use of a linear trend through the entire CMAM simulation period to significantly impact our results.

Our primary focus in this chapter is on the variability of the vortex, not on its sensitivity to climate change. Characterizing this variability requires a large number of years. The frequency of sudden warmings, however, increases over the present century in the CMAM runs considered here (McLandress and Shepherd, 2009). We therefore divide them into a present (1960-2010) and a future (2050-2100) period, giving us 150 years in each period after combining the three ensemble members. Where quantities do not change appreciably over the course of the simulations, we include all 420 years.

Similarly, where quantities are established to not differ significantly between the two reanalyses, or the greater vertical domain of MERRA is not of interest, the datasets are merged using ERA40 data up to 31 December 1978 and MERRA data from 1 January 1979 onwards, providing a single merged reanalysis record from September 1958 to April 2011.

To compute temperature anomalies for the satellite data the methodology is modified slightly. Due to the relatively short record, we omit the three winter seasons during the record with large PJO events—2005-2006, 2008-2009, and 2009-2010. The climatology is then computed from the remaining data, and no trend is removed. While this will bias the climatology, it is preferable to the alternative since the large amplitude and similar timing of these events produces a clear artefact in the climatology which then affects the anomalies in other years. This issue is discussed further below.

### 4.2.1 Sudden warmings

Stratospheric sudden warmings (SSWs) are identified by reversals of the zonal mean zonal wind at 10 hPa, 60° N using the criteria defined by Charlton and Polvani (2007). Note, for the purposes of reproducibility, that there is a minor ambiguity in their prescription of how to remove reversals that occur in quick succession; here zonal wind reversals are considered warmings if they were not preceded by easterly winds at any time within
the previous 20 days (Charlton-Perez and Polvani, 2011). In contrast, McLandress and Shepherd (2009) used an interval of 60 days from the previous wind reversal identified as a warming. The former, when applied to the ERA40 reanalysis, reproduces exactly the central dates in Table 2 of Charlton and Polvani (2007), and it is this criterion that is used in the present study.

Sudden warmings are further divided into vortex splits and displacements through an independent implementation of the classification algorithm outlined in Charlton and Polvani (2007), which is based on identifying vortex edges in terms of absolute vorticity contours on the 10 hPa pressure surface. One minor change to their parameter $n_c$ is made here; 21 vorticity contours are constructed instead of 12. This was found to produce more

Table 4.1: SSW classification in MERRA. D indicates a displacement, and S a split. The ERA40 classification is that of Charlton and Polvani (2007) during the overlap period.

<table>
<thead>
<tr>
<th>Central Date, MERRA</th>
<th>Central Date, ERA40</th>
<th>Type, MERRA</th>
<th>Type, ERA40</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 Feb 1979</td>
<td>22 Feb 1979</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>29 Feb 1980</td>
<td>29 Feb 1980</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>04 Mar 1981</td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>04 Dec 1981</td>
<td>04 Dec 1981</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>24 Feb 1984</td>
<td>24 Feb 1984</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>01 Jan 1985</td>
<td>01 Jan 1985</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>23 Jan 1987</td>
<td>23 Jan 1987</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>08 Dec 1987</td>
<td>07 Dec 1987</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>14 Mar 1988</td>
<td>14 Mar 1988</td>
<td>D</td>
<td>S</td>
</tr>
<tr>
<td>21 Feb 1989</td>
<td>21 Feb 1989</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>15 Dec 1998</td>
<td>15 Dec 1998</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>26 Feb 1999</td>
<td>26 Feb 1999</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>20 Mar 2000</td>
<td>20 Mar 2000</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>12 Feb 2001</td>
<td>11 Feb 2001</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>30 Dec 2001</td>
<td>30 Dec 2001</td>
<td>S</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>17 Feb 2002</td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>18 Jan 2003</td>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>05 Jan 2004</td>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>21 Jan 2006</td>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>24 Feb 2007</td>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>22 Feb 2008</td>
<td></td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>24 Jan 2009</td>
<td></td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>09 Feb 2010</td>
<td></td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>24 Mar 2010</td>
<td></td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.1: SSW occurrence frequency in events per decade as a function of month, and in events per year for all events (rightmost columns), for the reanalyses and CMAM datasets, as indicated. The frequency is divided into vortex splitting events (hatched) and vortex displacement events (unfilled). Estimated 95% confidence intervals are shown for the total frequency (both splits and displacements).

reliable classifications in our implementation as compared to subjective inspection. The algorithm was applied to the model and the MERRA reanalysis; we used the subjective classification of Charlton and Polvani (2007) for ERA40. The classification of SSWs in MERRA is summarized in Table 4.1, and agrees with the ERA40 classifications to about the same degree as did the two reanalyses considered in Charlton and Polvani (2007).

Since the occurrences of SSWs in both reanalyses are in close agreement, we compare their combined statistics to those of the CMAM runs. Figure 4.1 compares the modelled to the observed rate of occurrence as a function of month and for all events. The model simulations are divided into early (1960-2010) and late (2050-2100) periods since the number of SSWs in these simulations increases over the rest of this century (McLandress and Shepherd, 2009).

Confidence intervals at the 95% significance level are estimated by assuming a warming may happen each year with a probability $p$; this implies that the occurrence rate is drawn from a binomial distribution with $n$ equal to the number of years, and $p$ is estimated from the actual occurrence rate. This neglects the possibility of multiple warmings in one year, and the possibility of serial correlations. The latter in particular would potentially increase the size of the error bars; these confidence intervals are therefore best considered as an optimistic lower bound, and emphasize the difficulty of estimating these
statistics even with over five decades of reanalysis data. With these caveats in mind, the deficit of model sudden warmings in January, common to many GCMs (Butchart et al., 2011), is nearly significant.

The occurrence rate is further divided into splits (hatched) and displacements (unfilled). This suggests that although the total frequency of warmings in these CMAM simulations is roughly correct, the model does not produce as many vortex splits as the real atmosphere; applying the same statistical model to each type of warming independently permits confidence intervals to be estimated: the real atmosphere produces 1.9 to 4.3 splits per decade, while the model produces 1.0 to 2.1 per decade over the same period. Similar ratios have been noted in other CMAM simulations with similar overall warming statistics, leading us to suspect the confidence interval is reliable for the model in this configuration. However, it may be that this apparent deficit of vortex splits is in fact just an artefact of the short observational record. Indeed Charlton et al. (2007) noted considerable variability in this ratio, and the confidence intervals estimated above for the model and the observations do in fact overlap, if only just.

### 4.2.2 Annular modes and weak vortex events

The NAM indices are computed following Gerber et al. (2010). At each pressure level and each day, the global mean is removed from the zonal mean geopotential height. This residual is deseasonalized and detrended as discussed above. The first EOF is then computed from area-weighted anomalies north of the equator, after Baldwin and Thompson (2009). The principal component time series is then used as the NAM index. Strong and weak vortex events are defined when the index rises above $1.5\sigma$ or falls below $-2.5\sigma$, respectively. Events that occur within 60 days of a prior event are discarded.

### 4.2.3 Polar-night Jet Oscillation events

The PJO is defined in terms of EOFs of daily-mean polar cap averaged temperatures. We use $70^\circ$-$90^\circ$ N for our polar cap averaged temperatures and include data from all seasons for continuity of the index. The spatial structure of the modes is only weakly sensitive to these details.

The first mode describes a vertical dipole with one maximum near 1 hPa, and the other near 0.01 hPa; the second mode also describes a dipole, one-quarter wavelength out of phase with the first, with a lower maximum near 10 hPa (Fig. 4.2). We adopt
Figure 4.2: (a) EOFs of polar cap averaged 70°-90° N temperatures; EOF 1 is shown in red and EOF 2 is shown in green as computed from the four datasets. (b) Root-mean-squared amplitude of the first two EOFs. The shading indicates 2σ variability on seven-year sub-samples of the CMAM simulation (see text).

The sign convention that a positive anomaly in the principal component (PC) time series corresponds to a positive lower maximum, that is, to a warm upper stratosphere for EOF 1 and a warm middle stratosphere for EOF 2. The two modes together explain roughly 85% of the total daily variance, with a relatively large separation $\Delta\lambda \sim 20\%$ of the fraction of variance explained by each mode (see Table 4.2). The two modes in the model and reanalyses agree very closely. The modes from the satellite data show somewhat larger differences. In particular, the first EOF has a small negative feature in the lowermost stratosphere which is not present in either the model data or the reanalyses. It is, however, present in the EOFs computed by Kuroda and Kodera (2004), who considered 21 winters of satellite and radiosonde-derived analyses. The maxima in the upper stratosphere of EOF 1 and in the lower mesosphere of EOF 2 also show slight differences. To give a

Table 4.2: Percentage of variance explained by the EOFs in Fig. 4.2. The uncertainties are estimated by bootstrapping the CMAM data.

<table>
<thead>
<tr>
<th></th>
<th>MLS</th>
<th>ERA40</th>
<th>MERRA</th>
<th>CMAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOF 1</td>
<td>$56\pm6%$</td>
<td>$50\pm3%$</td>
<td>$50\pm3%$</td>
<td>$54%$</td>
</tr>
<tr>
<td>EOF 2</td>
<td>$32\pm6%$</td>
<td>$27\pm3%$</td>
<td>$35\pm3%$</td>
<td>$31%$</td>
</tr>
</tbody>
</table>
sense of the sampling uncertainty associated with the relatively short satellite record, we compute 95% confidence intervals from the model data by separating it into seven year sub-samples (roughly the length of the MLS record). These intervals are shown by the shading in Fig. 4.2, and suggest that the differences may lie just outside the confidence intervals. If we consider the total root-mean-squared temperature amplitude of the two modes, however, we see that the MLS data agrees well through the stratosphere, with only a small departure in the mesosphere, suggesting that the difference in the lower stratosphere is in the relative phase of the first two EOFs, which, despite the relatively good separation in their eigenvalues, is more susceptible to statistical uncertainty (North et al., 1982). If EOFs are computed from MERRA data from the same time period as the MLS observations the same lower stratospheric structure is recovered (not shown), suggesting that this is indeed a sampling issue. The structure of the EOFs is not sensitive to the inclusion of all years of data in computing the climatology for the satellite data, although the corresponding PC time series do change substantially.

The large fraction of variance explained by these two modes demonstrates that temperature anomalies in the Arctic middle atmosphere are very strongly coupled in the vertical. Moreover, it suggests that we can describe much of the vertical structure of the zonal-mean vortex anomalies with just two degrees of freedom (on top of a climatologically varying background). The PC time series ($t_s^1$ and $t_s^2$) corresponding to these two modes can therefore be used to define a trajectory in a two-dimensional phase space (Kodera et al., 2000; Kuroda and Kodera, 2004), which describes the evolution of the vertical structure of polar-cap averaged temperature perturbations in the Arctic. This approach has also been used to study the descending wind anomalies in the QBO (Wallace et al., 1993), though since the PJO is far less regular, examining trajectories plotted directly in this two dimensional space is more challenging.

To better visualize these phase-space trajectories, we transform them to polar coordinates $r$ and $\theta$, defined by $r^2 = t_s^2 + t_s^2$ and $\tan \theta = t_s^2/t_s^1$. We can then construct plots of the daily evolution of these statistically-defined modes, examples of which are shown in Fig. 4.3. The upper two panels show MLS temperature anomalies during the winters of 2007-2008 and 2008-2009, from November through April. The trajectories of the two winters in the phase space are shown in the lower two panels. These trajectories are then visualized more compactly by the coloured ‘ribbons’ labeled (c) and (d) in the middle two panels (so that the time axes can be easily compared).

The width of the ribbon corresponds to the radial component $r$, or in more physi-
cal terms, to a vertically integrated, root-mean-square measure of the departure of the temperatures from climatological values. The colour corresponds to the phase of the trajectory $\theta$, in other words to the location of the temperature anomaly. The sign convention adopted here is as follows. The positive phase of EOF 1, in which the upper stratosphere is anomalously warm, is considered the positive x-axis and defines $\theta = 0$. It is coloured red. The positive phase of EOF 2, in which the middle stratosphere is anomalously warm, is taken to be the positive y-axis and is coloured green. The negative phase of EOF 1 is coloured blue, and the negative phase of EOF 2 is coloured yellow. Phases intermediate to these four key directions are coloured by interpolating linearly in
RGB space. A legend is provided as an aid to the reader in the lower central panel. For reasons that will shortly become apparent (see Fig. 4.4), we refer to these as ‘abacus’ plots.

The winter of 2007-2008 featured four brief episodes of upper stratospheric warming and lower mesospheric cooling between late January and the end of February (Fig. 4.3(a)). Only the final warming (during which the winds at 10 hPa reverse) reaches the lower stratosphere. These episodes are well described by the first two EOFs, though they can only be made out with some difficulty in the phase-space trajectory. The corresponding abacus plot, however, clearly shows four episodes during which the ribbon broadens. The blue colour in the final episode indicates the lower stratospheric warming. The ability to show both the amplitude and vertical structure of the polar warming at daily resolutions is the main advantage of the abacus plots.

The major sudden warming of 2008-2009 is a particularly strong example of the slow recovery phase on which this study is focused. The warming begins in the upper stratosphere, but proceeds down to the lowermost stratosphere by the end of January. At this point a vertical tripole in the temperature anomalies is apparent, with a warm lower stratosphere, a cold upper stratosphere and lower mesosphere, and a warm upper mesosphere. The lower stratosphere remains anomalously warm for the following months. Meanwhile the anomalies above slowly migrate downwards, though they lose amplitude by mid-April as sunlight returns to the pole and the strength of the radiative damping increases (see Fig. 3.6). This rapid amplification followed by a slow rotation in the EOF phase space is apparent in Fig. 4.3(f) due to its large amplitude, and shows up as a rapid broadening of the ribbon and coincident change in colour from red to green, followed by a slow change of colour from green to blue to yellow as the anomalies descend. By the end of the descent the trajectory has performed nearly one complete rotation. This recovery is evidently one coherent dynamical event, and while at the lowermost levels the anomalies remain single-signed through much of the event, the temperature overshoots (and by thermal wind balance, the zonal wind as well) through much of the stratosphere. The ability of these temperature EOFs to clearly capture the vertical and temporal coherence of these events is one of their main strengths.

One point worth stressing is that during the winter months the amplitude $r$ reaches 4 or $5\sigma$ (and persists for a significant fraction of the season) far more often than would be expected for a normally distributed time series (as in the 2009 warming, which was extreme but not entirely unprecedented in the structure of the temperature anomalies;
This is in part due to the fact that the time series $t_s_1$ and $t_s_2$ are normalized to unit variance over the entire year, though even when the variance is also deseasonalized the distributions remain non-Gaussian (Yoden et al., 2002; Christiansen, 2009). As a result, these large anomalies can easily dominate statistical averages when taken over too few years. This issue is discussed further below.

That these long-time scale PJO events are ubiquitous and evident in abacus plots will be seen shortly; however, it is useful to be able to identify these events with an objective algorithm. A PJO event is defined to occur when the trajectory rotates through a specific phase $\theta_c$, provided the amplitude is greater than a threshold $r_c$. This criterion is illustrated in Fig. 4.3(e,f); an event is identified when the trajectory crosses the bold ray counterclockwise. This will be referred to as the ‘central’ date, though as will become clear it need not occur at the midpoint between the start and end dates defined below. At the central date the vertical profile of temperature anomalies has a local maximum at a particular height; this height corresponds to the value of $\theta_c$. To define the duration of the event, we consider it to begin on the first date prior to the central date when the amplitude exceeds another threshold $r_m$ (where $r_m < r_c$), and to end on the first date following the central date when the amplitude falls below this same threshold $r_m$. This lower threshold is shown in Fig. 4.3(e,f) by the thin inner circle. To reduce the impact of small fluctuations of the trajectory near these threshold points, the principal component time series $t_s_1$ and $t_s_2$ are smoothed by a 5-day low-pass filter prior to computing the amplitude and phase used to define these dates.

Unless otherwise noted, a reference phase of $\theta_c = 2\pi/3$ is used. In all cases, threshold amplitudes of $r_c = 2\sigma$ and $r_m = 1.5\sigma$ are used. Sensitivity to the definition of these parameters is discussed further below. Note that this definition is intended to select events based on a particular height to which the maximum in the vertical profile of the temperature anomalies descends. For the standard reference phase of $\theta_c = 2\pi/3$, this local maximum lies at 60 hPa. The central date corresponds to the first date at which this maximum descends through a particular pressure level. For the standard value of $\theta_c$, this is typically some time following the most rapid period of warming, or equivalently, the period of peak planetary-scale eddy flux convergence in the polar stratosphere.

Finally, to correct for the slight phase differences between the various datasets associated with the EOF analysis, the time series for the two reanalyses are computed by projecting their temperature anomalies onto the CMAM EOFs. One could equally well project the model behaviour onto the EOFs from MLS or the reanalyses; the model
EOFs are chosen for their large vertical domain and statistical robustness. This results in only minor changes to the principal component time series, but since the compositing technique depends on their relative phase, this projection facilitates the statistical comparison between the reanalyses and the model, in accordance with the fourth criterion proposed in the introduction to this chapter. To this analysis we now proceed.

### 4.3 Results

Abacus plots for all years of the three observational/reanalysis datasets and for one century (1980 to 2080) of one of the CMAM ensemble members are shown in Fig. 4.4. The abacus plots are annotated with the PJO events identified as described above (a value of $\theta_c = 2\pi/3$ is used). The vertical black bars indicate the duration of the event. The dates of major sudden warmings are also indicated, classified into displacements (horizontal lines) and splits (upward-opening chevrons), as are weak vortex events (downward-opening red chevrons). Further annotations are described below.

There is a great deal of detail in these plots; however, the central point to this study is the ubiquity of the long, slow evolution of the PJO as is apparent in all datasets in the characteristic red-to-green-to-blue tails. It is clear they are strongly associated with major sudden warmings in that nearly every PJO event is initiated by a sudden warming, though there are a significant number of sudden warmings which do not initiate a PJO event, even in mid-winter (e.g. December 2001 and January 2003). They are similarly associated with weak vortex events. These relationships are analyzed in more detail below. Also indicated on the ERA40 abacus plot are the descending warm anomalies (shown as vertical red lines to the right of the abacus ribbon) and cold anomalies (shown similarly as vertical blue lines) studied by Zhou et al. (2002). The descending warm anomalies correspond to the early phase of PJO events. The cold anomalies that they observed to tend to follow the descending warm anomalies correspond to a later phase of PJO events (with the exception of the events beginning in March 1981 and March 1997).

The qualitative agreement between the reanalyses and the satellite observations during the overlap periods is for the most part very good (compare 1979 to 2002 of ERA40 and MERRA; 2005 to 2011 of MERRA and MLS). The phase shift between the CMAM and MLS EOFs is apparent in the abacus plot in that PJO events in the model and reanalyses begin with a longer red phase (consider the three large PJO events in 2006, 2009 and 2010). In all datasets the initial warming begins in the upper stratosphere and
Figure 4.4: Abacus plots for (a) ERA40, (b) MERRA, (c) MLS and (d) 100 years from one member of the CMAM ensemble. Vortex displacement sudden warmings are indicated by horizontal black lines; vortex splits by upward-opening black chevrons. Weak vortex events are indicated by downward-opening red chevrons. PJO events are indicated by a vertical line to the left of the abacus ribbon. See text for further details.
There are some small quantitative differences between the reanalyses. For instance, the sudden warming in February 2001 is classified as a PJO event in ERA40, but the temperature anomalies it induces are not quite large enough to be classified as such in MERRA. There is an artefact during 1975–76 in the ERA40 dataset which projects onto the 1st EOF; this results in the wide red band that persists through the year. This is likely associated with a known error in the bias-correction of data from the NOAA-4 satellite which affects upper stratospheric temperatures during this period, as reported by Uppala et al. (2005). The reanalysis prior to 1979, apart from this anomaly, looks qualitatively quite similar to the satellite-era data.

The qualitative character of the variability in the model vortex also agrees quite well with the reanalyses. The long red-to-green-to-blue tails characteristic of the descending temperature anomalies during PJO events are similarly ubiquitous and follow many of the major sudden warmings and weak vortex events, though warmings that are not followed by this extended recovery period are also apparent. The relative deficit of vortex splitting events noted in section 4.2.1 is also apparent.

As a coarse characterization of the time series of these two modes, Fig. 4.5 shows histograms of two quantities. The first panel shows histograms of the rotation rate in the reanalyses and model (computed here as the change in phase over five days for all dates with amplitudes greater than $1.5\sigma$), showing the typical counter-clockwise rotation of the
EOFs (or the downward propagation of temperature anomalies). The mean rates agree well between all three datasets (ERA40 $2.0\pm0.6$ deg day$^{-1}$, MERRA $2.3\pm0.9$ deg day$^{-1}$, CMAM to present $2.3\pm0.4$ deg day$^{-1}$). Assuming a wavelength of roughly 50 km, these correspond to a mean downward propagation of roughly 300 m day$^{-1}$, significantly faster than typical polar residual vertical velocities ($\approx 50$ m day$^{-1}$). The rotation rate in the model accelerates with climate change, to $2.9\pm0.4$ deg day$^{-1}$ over the last five decades. Although the relationship between the rotation rate and $\bar{w}^*$ is not straightforward, this change may be related to the acceleration of the Brewer-Dobson circulation.

The second panel shows the distribution of phases during sharp amplifications of the temperature anomalies (defined here as a change in $r > 2.5\sigma$ over 5 days). The reanalyses and the model show a peak in this distribution near $\theta = 0$, indicating that amplifications tend to occur during the positive phase of the first EOF. Physically, this indicates that the largest and most rapid warming occurs in the upper stratosphere. Consistent with the single-signed nature of the pseudomomentum carried by Rossby waves (see Chapter 1) and with prior composite studies (e.g. Limpasuvan et al., 2005), warm events tend to develop more rapidly and to larger amplitudes than cold events.

The event definition itself provides a means of testing the relationship between the depth of the warming and its timescale, since the reference phase $\theta_c$ used to identify events requires that the vertical profile of temperature anomalies at some point during the event peaks at a particular pressure level (see, for instance, Fig. 4.3(g)). Figure 4.6 shows the sensitivity of the event definition to the choice of the reference phase $\theta_c$. Panels (a) and (b) show events from MERRA for two choices, $\theta_c = \pi/8$ and $2\pi/3$, respectively. They are then sorted by the amplitude of the event at the central date.

The central date for the events identified in panel (a) corresponds to an anomalously warm upper stratosphere—the temperature anomaly at this phase peaks at 5 hPa. Most of these events are minor warmings confined to the upper stratosphere (the pulse remains red), and the temperature anomaly peaks and then recedes within a timescale of at most a few weeks. Only a few PJO events are also captured, suggesting that the upper stratospheric warming during PJO events is not as strong as during the minor warmings identified here. This is consistent with the separate peaks in the distribution of warming events found by Kohma et al. (2010) in a mechanistic model. Also shown in these plots is the reflective index (the difference between the zonal mean zonal wind averaged from $53^\circ$ to $74^\circ$ N at 10 hPa and 2 hPa) introduced by Perlwitz and Harnik (2004). Following Harnik (2009), periods when the index falls below $-13.4$ m s$^{-1}$ (2 standard deviations) are
Figure 4.6: Sensitivity of event definition to the reference phase $\theta_c$. (a-b) Events from MERRA with reference phase $\theta_c = \pi/8$ and $2\pi/3$. Sudden warmings are indicated as in Fig. 4.4. The black lines indicate PJO events as classified by the corresponding value of $\theta_c$. Maroon lines to the right of the ribbons indicate periods when the reflective index (see text) is greater than $2\sigma$. (c) The number of events identified per year versus $\theta_c$, for both reanalysis datasets, and the past and future periods of the CMAM simulations. (d) The mean duration in days of events identified versus $\theta_c$, for the same datasets.

indicated on the abacus plots. These are clearly the same events as the reflective events identified by Harnik (2009). For the sake of clarity, note that this condition rarely persists for longer than five to ten days; the maroon lines in Fig. 4.6 indicate this intermittency and are not in fact dashed lines.

The events identified in panel (b), in contrast, have extended timescales. The phase progression (red to green to blue to yellow) corresponding to the downward propagating temperature anomalies is also more apparent. As noted by Harnik (2009), the vortex during these events also typically goes into a strongly reflective configuration for a brief period. In these cases, however, wave activity continues to be absorbed by the mean flow
and the warming continues to descend to the lower stratosphere ($\theta_c = 2\pi/3$ corresponds to a peak warming near 60 hPa). These events are also considerably less common than minor warmings; only 13 events are identified compared to the 22 identified in panel (a). In part this is due to their long timescales; at most two such events occur in a season (e.g. 1998–1999), and this is an unusual occurrence.

Figures 4.6(c) and (d) demonstrate the robustness of these relationships. Panel (c) shows the number of events identified per year as a function of $\theta_c$ for both reanalyses and the present and future period of the CMAM simulations. The frequency of events peaks near the $\pi/8$ case shown in Fig. 4.6(a), and falls off steadily as the compositing level descends through the stratosphere. The number of events also falls off for $\theta_c < \pi/8$. This structure is reproduced by both reanalyses and the model simulations. Panel (d) shows the average duration of the events. The time scales lengthen steadily with the depth of the warming, as noted by Gerber et al. (2009).

There is a weak suggestion in Fig. 4.6(c) that the model predicts an increase in the number of PJO events under the projected climate change. This suggestion is checked in detail by Fig. 4.7. The event frequency is plotted in panel (a) for each decade in the ensemble. To test the significance of any trend, we take as a null hypothesis that the occurrence rate is fixed at $p$ events per year over the 140 years of the simulation. As with
the SSW occurrence rate, we estimate the 95% confidence interval on a 30 year sample drawn from a binomial distribution; this interval is indicated by the dashed lines in panel (a). One of the fourteen decades in the simulation lies slightly below the confidence interval, which is to be expected for the given level of significance. There is, therefore, no statistical evidence of a trend in the number of PJO events occurring in these simulations.

This is perhaps expected from the results of McLandress and Shepherd (2009) given that, like the NAM-based events for which they found no trend and unlike sudden warmings for which they did, the PJO indices are computed after removing a slowly varying background trend.

Similarly there is no evidence for a trend in the mean duration of the PJO events in these simulations (Fig. 4.7(b)). On the basis of these results, we can then make use of the entire simulation period of the ensemble to consider, in Fig. 4.8, the seasonal dependence of PJO statistics. Since the statistical characteristics of these indices also match quite closely between the two reanalyses, we merge the two reanalysis time series. Figure 4.8(a) shows the frequency of events by month for the merged reanalyses and the model. Error bars are estimated in the same fashion as described above for SSW occurrence frequencies. Since events typically span several months, an event is considered to have occurred in a given month if any date between its onset and its conclusion falls within that month.
A significant number of events persist through April. These are not necessarily final warmings in the sense that the warming that initiated them often occurred in February (e.g. the warmings in February 1989, February 2001, or even January 1968). Note that the standard deviation on which the conclusion of the events is based is not seasonally dependent, so the temperature anomalies that persist even after sunlight has returned to the Arctic are not negligible.

Figure 4.8(b) shows the average duration of events that occur in each month. The seasonal cycle of event duration is relatively weak, with early winter (November and December) events persisting for somewhat shorter periods than the rest of the extended winter period. This weak seasonal cycle is to some extent an artefact of how we include events in each month; a similar plot showing the duration of only those events whose central dates lie in a given month shows that those events that are identified in February and March do tend to be somewhat shorter than those identified in January, likely because their amplitude attenuates rapidly once the summer season begins, leaving less time for events which begin later in the season to persist (not shown).

The agreement in all cases between model and reanalyses is well within the estimated sampling error. PJO events in the reanalyses occur at a rate of 3.7±1.1 events per decade,

---

**Figure 4.9:** Decorrelation timescales of the NAM for (a) ERA Interim for the period 1979 to 2010, and (b–d) each of the three members of the CMAM ensemble for the period 1960 to 2010. Contour lines are at intervals of five days.
while those in the model simulations occur at a rate of 4.3±0.4 events per decade. The close agreement between both the duration and the frequency of PJO events in the model and the reanalyses raises the question of how closely these timescales correspond to the decorrelation times of the annular modes. Chemistry climate models are known to exhibit long biases in these decorrelation timescales (Gerber et al., 2010), though the statistical significance of this bias in the Northern Hemisphere has recently been questioned (Simpson et al., 2011).

To check this hypothesis, NAM decorrelation timescales as a function of pressure and season are shown in Fig. 4.9. They are computed following the method of Simpson et al. (2011). The timescales computed from the ERA Interim reanalysis for the period 1979 to 2010 (panel (a)) peak during DJF in the winter lower stratosphere near 30 days. Timescales for each of the three CMAM ensemble members for the period 1960 to 2010 (panels (b–d)) during DJF vary from 15 to 25 days, though there is considerable variability despite the use of 50 years of data. There is a peak in April in two of the runs, which resembles the multi-model ensemble mean shown in Gerber et al. (2010). At any rate the apparent biases in these decorrelation times do not correspond with the close agreement in PJO duration seen in Fig. 4.8, which suggests that these long events are not the sole determinant of the decorrelation timescales.

### 4.3.1 Relationship to sudden warmings

Nearly every PJO event as identified here follows a stratospheric sudden warming. However, the reverse is not true—a significant number of sudden warmings are not followed by PJO events.

Making use of the sudden warming classifications described above, we now divide the sudden warming events in the merged reanalyses and in the model into splits and displacements. Figure 4.10 shows abacus plots for every observed sudden warming (displacements in the top panel, splits in the bottom). Of the 21 observed vortex displacement events, six are followed by a PJO event as identified by the event definition, while a seventh (in February 2001) also appears to have a somewhat extended recovery phase, though not of large enough amplitude to meet the PJO criterion. In contrast, ten of the sixteen observed vortex splits are classified as PJO events, while two or three more (January 1958, February 1989, and possibly February 1979) have extended timescales with weaker amplitudes. Interestingly the February 1979 warming, which has been studied extensively...
as an archetypical vortex split (e.g. Matthewman et al., 2009, and references therein), had a fairly weak impact on the vortex temperatures; this is also consistent with the measure of the temperature change reported by Charlton and Polvani (2007).

The probability of these warmings being followed by a PJO event is summarized in Table 4.3. The uncertainties (computed assuming the simple binomial distributions described above) are large. So while there is a clear suggestion that observed vortex splits are more frequently followed by PJO events than are vortex displacements, the short observational record precludes a definitive conclusion. Indeed, two examples of strong PJO events following clear vortex displacement events have occurred in the past decade (in January 2004 and January 2006). Assuming the observed probabilities, 100
years of observations would be required to distinguish these at the 95% significance level.

Of 252 vortex displacements in the three members of the ensemble of CMAM simulations, 91 are classified as PJO events. (Of the 22 shown in Fig. 4.10 selected from one member, the 10 classified as PJO events are roughly representative of this fraction.) Of the 72 vortex splits, 39 are classified as PJO events (again consistent with 10 of the 19 from the one member shown). As noted above (Fig. 4.1), vortex splits in CMAM occur roughly half as often as in the observations, though this difference is not statistically significant. However, the fraction of splits and displacements followed by PJO events agree with the reanalyses to well within uncertainties. Moreover, the long integrations provide sufficient statistics to differentiate the occurrence of PJO events following splits and displacements. The model confirms the suggestion that PJO events occur somewhat more frequently following splitting events. Both fractions are too noisy to detect any suggestion of a trend over the model integration.

As has been noted by many studies (Limpasuvan et al., 2004; Charlton and Polvani, 2007; Liberato et al., 2007), the flux of planetary wave activity entering the vortex following sudden warmings is reduced. To highlight this point, and to note the difference between warmings with short recovery timescales and those followed by PJO events, Fig. 4.11 shows composites of the (absolute) zonal mean zonal wind, area averaged from 50° to 70° N, and the anomalous vertical component of the EP flux, area averaged from 50° to 90° N. The former is representative of the peak zonally-averaged winds of the vortex, and the latter range was chosen simply to encompass the bulk of the upwards flux of wave activity entering the polar vortex. The results are not strongly sensitive to these details. Regions where the EP flux does not differ significantly from the climatology at the 90% confidence level are masked in white; those which do not differ at the 95% confidence level are lightly faded. The left panels are sudden warming events (both splits and displacements) which are not followed by a PJO event; the two reanalysis products have been divided here due to their different vertical domains and the potential for sponge-

<table>
<thead>
<tr>
<th>Event type</th>
<th>CMAM</th>
<th>Reanalyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(SSW)</td>
<td>0.40 ± 0.06</td>
<td>0.43 ± 0.16</td>
</tr>
<tr>
<td>p(split)</td>
<td>0.56 ± 0.12</td>
<td>0.6 ± 0.3</td>
</tr>
<tr>
<td>p(displacement)</td>
<td>0.36 ± 0.06</td>
<td>0.3 ± 0.2</td>
</tr>
</tbody>
</table>
layer issues. On average, the zero wind line descends just below the 10 hPa level required to meet the sudden warming criterion; moreover, the 10 m s$^{-1}$ contour does not descend much lower than the 100 hPa level, and by 15 days following the central date, winds throughout the stratosphere are greater than 10 m s$^{-1}$. The pulse of EP flux responsible for the deceleration of the winds is apparent; following the wind reversal, however, the fluxes are only reduced significantly for perhaps 10 days.

Composite averages of warmings followed by a PJO event show somewhat more persistent easterly winds in the middle stratosphere (15 days compared to 5 days), but on average the zero-wind line does not differ too drastically from the composite average of the shorter timescale events. The 10 m s$^{-1}$ contour, however, descends to the upper troposphere and these weak westerlies persist for 40 days. The initial pulse of wave activity is again apparent, and as noted by Harnik (2009), is of considerably longer duration than during short time-scale events. In strong contrast to the short time-scale events, the vertical EP fluxes into the polar vortex are strongly suppressed for some 60 days following the initial wind reversal. Note that in the non-PJO composite this reduction remains significant for nearly as long as in the PJO composite, however, the magnitude is much reduced. This significance may also be a consequence of false negatives in the algorithm used to define PJO events; this possibility is discussed further below. While the details of the stratospheric circulation anomalies are discussed at greater length in Chapter 5, we note as well the super-recovery of the upper stratospheric jet during this period, consistent with the strong cold anomalies in the middle to upper stratosphere, as in Fig. 4.3(b). It is likely that the coherent, robust circulation patterns in the stratosphere are a consequence (directly or indirectly) of this strongly suppressed upward EP flux.

An important question, then, is why these fluxes should be suppressed for such a long time following the initial warming. The simple application of the Charney-Drazin criterion clearly does not suffice to explain the suppressed waves given the presence of westerly winds throughout the stratosphere for much of the period of suppression (while it is true that easterlies do persist in certain events for considerably longer than the composite average might suggest, inspection of individual events shows suppression of the fluxes even when no easterly winds are present). Moreover, the filtering of waves by the stratospheric flow does not immediately explain the anomaly in the upward fluxes in CMAM, which extends right to the surface. The suppression of upward fluxes in the troposphere might be explained by reflection (due to the cancellation between the
Figure 4.11: Composites of zonal mean zonal wind from 50°-70° N (contour lines, 10 m s\(^{-1}\) intervals; zero contour is thick) and vertical EP flux anomalies from 50°-90° N (colour shading, in units of kg s\(^{-1}\) day\(^{-1}\)). Regions in which the EP fluxes do not differ significantly from climatology (at the 95% confidence level) are masked in white. Composite events are drawn from (a,b) merged reanalyses, (c,d) MERRA, and (e,f) CMAM. Sudden warming events that are not followed by a PJO event are shown in the left panels (a,c,e); those events that are followed by a PJO event are shown in the right panels (b,d,f).
4.3.2 Relationship to weak vortex events

PJO events are also closely related to the strong and weak vortex events shown by Baldwin and Dunkerton (2001) to have a significant impact on the tropospheric annular modes. Abacus plots of the weak vortex events in the merged reanalysis dataset and in one ensemble member of the model are shown in Fig. 4.12. With a definition threshold of -2.5σ, somewhat less than half (15 of 40) of the weak vortex events in the merged reanal-
ysis dataset are also classified as PJO events. The events in the abacus plots are sorted by the minimum value reached by the annular mode index over the 10 days following the -2.5σ threshold central date; increasing this threshold would therefore remove events from the right hand side of the plot. PJO events clearly correlate with the strength of the 10 hPa NAM anomaly; those events that would still be classified by a -3σ threshold are identified by the bold labels; these larger events also correlate more closely with sudden warmings. The fraction of weak vortex events that are followed by a PJO event is quantified in Table 4.4; roughly 40% following -2.5σ events, rising somewhat to 50% following -3σ events.

Strong vortex events are also indicated in Fig. 4.12 by blue chevrons. During many of the largest PJO events, the super-recovery of the vortex seen in Fig. 4.11 is in fact strong enough to meet the 1.5σ threshold. The recovered vortex at this phase of the PJO events is extremely high in addition to being strong (the cold polar temperatures imply large vertical shears by thermal wind balance), so these events tend to have a much stronger signature in the middle to upper stratospheric NAM indices than they do in the lower stratosphere.

A similar picture emerges from the corresponding model abacus plot. Half of the weak vortex events from one ensemble member are shown in Fig. 4.12(b) as a representative sample. Of the 380 weak vortex events identified in the ensemble, 190 were also classified as PJO events. The fraction of weak vortex events followed by PJO events in the model agrees well with reanalyses for both the -2.5σ and -3σ threshold definitions (Table 4.4). The super-recovery of the vortex also has a somewhat greater tendency to be classified as a strong vortex event in the model than in the observations.

If the coupling between the stratosphere and troposphere does in fact stem from the lower stratospheric annular mode anomalies (as suggested by Baldwin and Dunkerton (2001)), then those events that have more persistent such anomalies ought to have a greater impact on the troposphere. The impact of the PJO on the tropospheric circulation has been noted before (Kuroda and Kodera, 2004; Kohma et al., 2010); the close

<table>
<thead>
<tr>
<th>NAM threshold</th>
<th>CMAM</th>
<th>Reanalyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.5σ</td>
<td>0.41 ± 0.06</td>
<td>0.38 ± 0.15</td>
</tr>
<tr>
<td>-3σ</td>
<td>0.52 ± 0.07</td>
<td>0.5 ± 0.18</td>
</tr>
</tbody>
</table>
correspondence between weak vortex events and PJO events, however, makes it difficult to determine if additional information is gained in the PJO perspective. Simplified model studies have indicated that longer timescale variability in the stratosphere does correlate with stronger stratosphere-troposphere coupling (Gerber and Polvani, 2009).

We therefore divide the weak-vortex events into those events that correspond to PJO events and those that do not. Composites are presented in Fig. 4.13. The reanalysis composites shown in panels (a) and (b) show a clear difference between non-PJO and PJO events. The stratospheric NAM anomaly in the latter is larger, and a $2\sigma$ anomaly persists for nearly 30 days. In contrast, the $2\sigma$ anomaly during non-PJO events on average persists for only 10 days. The negative anomaly in the stratosphere, however, does remain statistically significant for roughly the same length of time in each case. The tropospheric impact of the PJO events in the composite is certainly stronger and more coherent than that of the non-PJO events, but even in the merged dataset the number of events is still too small to resolve the tropospheric impact unambiguously.

Here the benefits of the long CMAM simulations are clear. Corresponding composites
for non-PJO events (Fig. 4.13(c)) and PJO events (Fig. 4.13(d)) confirm the difference in the stratospheric signature of the two sets of events; PJO events (essentially by construction) have a much stronger, deeper and longer anomaly above the tropopause. Their tropospheric impact is also considerably stronger; after roughly a 10 day lag from the onset of the stratospheric anomaly, the tropospheric annular modes follow suit, and persistently negative anomalies remain while the lower stratospheric anomaly persists. As with the reanalysis composite, a weak but statistically significant stratospheric anomaly persists for nearly as long during the non-PJO events, despite the much shorter central feature. While the tropospheric impact is negligible for the first 40 days following the non-PJO events, a weak but significant response arises from days 40 to 80. As discussed further below, this may be a consequence of inadequacies in our composite definition. Regardless, it is clear that the tropospheric impact of PJO events is significantly stronger than that of non-PJO events.

4.4 Discussion

Strong vortex events

While the central focus here has been on the evolution following warm, weak-vortex events, events of the opposite sign are also of interest. Figure 4.14 shows an abacus plot of the merged reanalysis data set highlighting several metrics for cold, strong vortex episodes. Strong vortex events are indicated by downward-opening blue chevrons. Reflective periods (as measured by the index of Perlwitz and Harnik (2004)) are indicated by the vertical maroon lines to the right of the abacus ribbons. Finally, years during which the volume $V_{PSC}$ of the vortex in which temperatures drop below the threshold required for the formation of polar stratospheric clouds is particularly large, are indicated in gray (Rex et al., 2006). This zonally asymmetric quantity is a good predictor of the chemical destruction of Arctic ozone under current halogen loading of the stratosphere. The highlighted years are those identified by Rex et al. (2006) as being the coldest in a five year window; the winter of 2010-2011 was also an extremely cold winter (Manney et al., 2011) and has been highlighted as well. $V_{PSC}$ is most sensitive to temperatures in the lower stratosphere.

Although a significant number of strong vortex events do occur during PJO events (which tend to be amongst the coldest periods in the middle and upper stratosphere as
Figure 4.14: Abacus plot for the merged reanalysis. PJO events are denoted by black vertical lines. Strong vortex events are denoted by a downwards-opening blue chevron. The reflective index of Harnik (2009) is indicated by the maroon vertical lines. Years with particularly high integrated values of $V_{PSC}$ are indicated by the gray shading (see text).

indicated by the large amplitude, blue regions of the ribbon, and perhaps explain the relatively weak impact of sudden warmings on seasonal mean temperatures found by Charlton and Polvani (2007)), other strong vortex events appear to occur under quite different circumstances (consider, for instance, the events in 1987–88, 1988-89 and 1991–92). For instance, nearly all of the coldest years as identified by $V_{PSC}$ exhibit a strong vortex event, though only one (1983–84) also exhibits a PJO. Reflective periods do feature in some of the coldest winters (1995–96, 2000–01), but they play a small role in others (2004–05, 2010–11). Although a thorough study of strong vortex events is beyond the scope of this thesis, this lack of similarity between strong vortex events suggests that posing a definition that satisfies the second criterion proposed in the introduction to this chapter would prove non-trivial.

Statistical challenges

The ubiquity of the large-amplitude, coherent circulation anomalies associated with PJO events has a number of implications. Their large amplitude and intermittent occurrence, in particular, presents particular difficulties when estimating the impact of various external influences on the Arctic stratosphere. To illustrate this, we show differences in
Figure 4.15: Differences in decadal, ensemble-averaged temperatures from the CMAM ensemble for two different pairs of decades. Regions which are not found to differ significantly at the 95% confidence level, as estimated by a Student’s t-test, are masked in white.

decadal, ensemble-averaged temperatures from the model simulations, as a function of pressure and day of the year (Fig. 4.15). These are differences between 30 year climatologies (considering the three ensemble members) of the vortex—a sample size of the same length as the observed satellite record. As shown in Fig. 4.7(a), the number of PJO events that occurred in any of the decades considered is quite consistent with the variability expected from the conservative, binomial distribution that does not assume any interannual correlation. That is, the variability during these periods is entirely consistent with this simple model of the natural variability. In particular, the differences seen between the two panels are of the opposite sign, ruling out transient changes in greenhouse gas or ozone concentrations as systematic causes. Moreover, Fig. 4.7(a) shows that fewer PJO events occurred in the 2090s than in the 2080s, and similarly in the 2030s than the 2040s, consistent with the signs of the climatological differences. Nevertheless, a standard Student’s t-test of the statistical significance of the temperature difference suggests that these differences are highly significant. That the differences are associated with PJO events is clear from the characteristic descending dipole structures (cf. Fig. 4.3(b)). The false positive given by the statistical test is a result of the strongly non-Gaussian distributions of temperatures characteristic of the Arctic vortex (Taguchi and Yoden, 2002b;
Nishizawa and Yoden, 2005). Moreover, the long timescales associated with PJO events mean that these differences can have a great deal of structure, and can survive monthly averaging. The need for extremely long datasets for detecting external influences on the Arctic vortex is evident.

The large amplitude of PJO recoveries also poses some difficulties for composite averages, though the implications are not as severe as for detecting external influences. Because of their long timescales and large amplitudes, PJO events will tend to dominate composites, particularly at large lags. Consider the weak-vortex composite presented in Fig. 4.13(c). Since the criteria used to define PJO events does not catch all long timescale recoveries (for instance, the recovery from the February 2001 sudden warming is evidently PJO-like, despite not being classified as such in the MERRA reanalysis due to its smaller amplitude), some of these extended recoveries will be included in the composite. At short lags, the composite average will be dominated by the much more numerous non-PJO events. After the vortex recovers during these short events, however, the circulation has essentially returned to its climatological state, and the evolution from then on will presumably be uncorrelated with the prior behaviour; that is, it will average out to zero in the composite. The anomalies associated with the long-time scale events will, however, still be coherent, and will therefore dominate the average since the anomalies will all be of the same sign. This is consistent with the weak but apparently significant NAM anomaly that persists in the stratosphere (Fig. 4.13(a,c)) and in the troposphere (Fig. 4.13(c)); the weak suppression in upward EP flux seen in Fig. 4.11(e) may arise for similar reasons.

4.5 Conclusions

We have introduced in this study a new tool for the visualization of Arctic polar vortex variability, based on the vertical structure of polar-cap averaged temperature anomalies. This tool makes clear that roughly one-half of stratospheric sudden warmings are followed by a PJO event—an extended recovery phase characterized by a persistent lower stratospheric warm anomaly, a cold anomaly that forms in the middle to upper stratosphere and descends, and an elevated stratopause. These events are clearly represented in reanalyses and are well simulated in the ensemble of CCM simulations considered here. We have also introduced an objective definition which captures these events that is robust to small changes, provides novel scientific insight (summarized below), and is capable
of capturing similar events in a number of datasets. That the events captured are also
dynamically similar is borne out to some degree by the abacus plots presented above,
however, this will be further established in the next chapter. This definition therefore
meets the criteria laid out at the beginning of this chapter.

The duration of events is strongly correlated with the depth to which the initial
warming descends, suggesting that the long time scales are closely related to the radia-
tive damping time scales in the lowermost stratosphere. The extended timescale events
correspond to those identified by Zhou et al. (2002) and Harnik (2009) suggesting that
an extended period of wave driving is indeed required to disturb the lower stratosphere.
Given the strong vertical gradient in radiative damping timescales present in the po-
lar winter stratosphere, this suggests an important role for the radiative processes in
determining these timescales.

PJO events occur somewhat more frequently following vortex splits than they do
following vortex displacements; however, this is not a necessary condition. Indeed, two
of the large PJO events in the past five years occurred following clear vortex displacement
events.

Just as essential for the long timescale of PJO events, however, is the strong suppres-
sion of upward fluxes of wave activity for the duration of the events. These fluxes are
suppressed for much longer during PJO events than they are following non-PJO sudden
warmings. While this suppression is likely related to the disruption of the lower strato-
spheric vortex, a simple application of the Charney-Drazin criterion is not sufficient to
explain it.

PJO events are also strongly associated with weak vortex events as identified by the
NAM index at 10 hPa. They tend to occur somewhat more frequently following weak
vortex events of larger magnitude. As one would expect, the larger and more persistent
lower stratospheric anomalies associated with PJO events are associated with a stronger
and more persistent tropospheric annular mode response.

Finally, as has been pointed out by numerous other studies (Taguchi and Yoden,
2002b; Nishizawa and Yoden, 2005; McLandress and Shepherd, 2009), the large amplitude
and long timescales of PJO recoveries present significant difficulties for distinguishing
the response of the Arctic polar vortex to an external forcing from its natural multi-
decadal variability. Their characteristic pattern of descending anomalies can dominate
even climatologies computed over multiple decades, and differences due only to variability
can appear significant under standard statistical tests.
Chapter 5

Zonal mean dynamics

Time present and time past
Are both perhaps present in time future,
And time future contained in time past.

T. S. Eliot – Burnt Norton

5.1 Introduction

In the previous chapter, the extended time-scale recoveries from stratospheric sudden warmings, termed PJO events, were classified and characterized in reanalyses and in simulations from a comprehensive middle atmosphere model. The regularity of the evolution during these PJO events suggests that the processes responsible are common to all the events, and are thus presumably predictable; indeed enhanced predictability has been found following certain sudden warmings in forecasting case studies (Mukougawa et al., 2009). Since these events have been shown to influence the tropospheric circulation (Baldwin and Dunkerton, 2001, see also previous chapter), this has practical implications for seasonal predictability. It also implies that a detailed understanding of the zonal mean dynamics of a small number of events has more bearing on the general case than might be expected.

Moreover, the statistical characterization of these events in the previous chapter suggests that much of the dynamics underlying the long stratospheric timescales are essentially zonal in character. For instance, counter to the implications of some studies (Charlton and Polvani, 2007; Yoden et al., 1999), much of the apparently longer recovery timescales following vortex splits are likely a result of there having occurred more PJO
events in the observations following vortex splitting events than have occurred following displacement events. An example of each will be considered explicitly in the present chapter.

The signature of PJO events is particularly apparent in the vertical profile of polar-cap averaged temperature anomalies, as made clear in the previous chapter. They are initiated by a rapid and large amplitude warming in the upper stratosphere which subsequently descends to the lower stratosphere as the major sudden warming unfolds. The warm anomaly in the lower stratosphere persists for the duration of the event, up to nearly three months. The upper stratosphere cools rapidly following the initial warming, resulting in strongly negative temperature anomalies. With this cooling comes an anomalously strong, high polar jet, characterized by strong, positive vertical wind shear through the stratosphere. During the warming, the polar stratopause (defined here as a local maximum in the vertical temperature profile) initially descends as the mid-stratosphere warms. The lowered stratopause disappears when the temperature anomaly reverses sign, while a new maximum re-forms anomalously high, at what is normally thought of as mesospheric heights (Manney et al., 2008). As the lower stratospheric anomaly relaxes towards climatological values, the dipolar temperature anomaly above (cold in the stratosphere, warm in the mesosphere) gradually descends. The purpose of this chapter is to better understand these details of the zonal mean evolution.

The stratospheric signature of these events is well known from studies of traditional sudden warming-type events, since the PJO events tend to dominate sudden warming composites at large time lags. For instance, the super-recovery of the mid-stratospheric temperatures was highlighted by the counter-intuitive result of Charlton and Polvani (2007) that, at 10 hPa, winters with sudden warmings are no warmer than winters without warmings; meanwhile composite winds computed by McLandress and Shepherd (2009) at the same level take several months to recover to climatological values. The cooling following the initial sudden warming is understood to be essentially radiative relaxation as a result of reduced planetary scale wave driving (Limpasuvan et al., 2004).

The elevated stratopause that occurs during these events has attracted recent interest. Unlike at other locations and times of the year, the stratopause in the polar night is a dynamical feature, maintained against strong radiative cooling by gravity wave drag (Hitchman et al., 1989). Mesospheric cooling during sudden warmings is understood to be a result of the wind reversal, which prevents orographically generated waves from propagating past the lower stratosphere (Holton, 1983). The re-formation of the strato-
pause at mesospheric heights following the 2006 sudden warming was noted by Manney et al. (2008), who noted a deficiency in the ability of reanalysis products to properly capture the high stratopause. Such episodes can be captured by models with sufficiently high lids and parameterized gravity wave drag (Ren et al., 2011). Case studies of the elevated stratopause have also been undertaken in the context of a short-term, high-resolution forecasting model (Siskind et al., 2010) and in another comprehensive chemistry-climate model (Limpasuvan et al., 2011).

The purpose of this chapter is two-fold. The first goal is to fully describe the phenomenology of the events from the tropopause up to the mesosphere. While particular aspects of this behaviour are implied by existing studies, the complete description is novel. The second goal is to quantify and attribute the structure of the transient Eliassen adjustment responsible for the anomalies in order to clarify the processes responsible for the anomalous circulation. This is done in the context of a comprehensive model in order to have all of the fully self-consistent dynamical fields, which are not straightforwardly available from the observations, and which allow for more precise attribution.

This chapter is organized as follows. The next section briefly summarizes the datasets considered here. The third section presents composite averages of the dynamical fields relevant to the evolution of PJO events. The analysis then turns towards a case study of a single event, supported by the similarity of its behaviour to the composited events. The residual circulation is considered explicitly in the fourth section. The fifth section presents a series of conceptual models for the temperature anomalies associated with PJO, culminating with a diagnosis of the zonal mean evolution in terms of the transient Eliassen adjustment (Eliassen, 1951) to the imposed eddy-driven torques. The section concludes by considering the full transient Eliassen adjustment of a second case study to demonstrate that the conclusions are not dependent on the choice of event. In the sixth section the gravity wave driven descent of the stratopause is considered explicitly, and the seventh section briefly summarizes the conclusions of the chapter. The zonal mean quasi-geostrophic model used for the analysis of the case study is very closely related to those used by Plumb (1982) and Haynes et al. (1991); the details are therefore left to an appendix.
5.2 Data

The analysis in this chapter is largely based on the 96-year time-slice run of the CMAM. This integration was performed at a horizontal resolution of T63, with 72 vertical levels. No interactive chemistry was used for this run, and SSTs were specified climatologically. Similar analyses have been performed on the 40-year time-slice simulations with fully interactive chemistry and an interactive ocean, and similar conclusions were drawn. As discussed in Chapter 2, the orographic gravity wave drag parameterization has been adjusted slightly in these runs to recover the observed frequency and seasonality of sudden

![Abacus plot of the CMAM time-slice run. Each vertical ‘ribbon’ corresponds to a single year. The width of the ribbon indicates the root-mean-square amplitude of the first two PC time series. PJO events are indicated by the vertical black lines to the left of the ribbons. Sudden warmings are classified as vortex displacements (horizontal lines) or splits (chevrons).](image)
warmings. As a result, these features match observations fairly well (not shown). Zonal mean winds and temperatures from the MERRA reanalysis (Rienecker et al., 2011) are also used for the purpose of comparison.

An abacus plot (see Chapter 4) of the CMAM time-slice run is shown in Fig. 5.1. PJO events and sudden warmings are indicated as described in the caption. The phenomenology of these events in the time-slice run agrees well with observations. The identification of the events used here is the same as in the previous chapter. However, to properly resolve the wind reversal which initiates the warming, the definition of the central date used by the composite is modified slightly. We take the first day following the initial date when the winds reverse at 10 hPa as the central date; if no such wind reversal occurs during the event, then it is discarded (four such events are found in CMAM, and three in MERRA). In this way, we are essentially analysing only those PJO events which are also classified as sudden warmings.

5.3 Composites

Figure 5.2 shows composites of zonal mean temperatures and winds during these PJO events from the MERRA reanalysis (10 events) and the CMAM simulation (38 events). The upper two panels show the absolute zonal mean zonal wind (area averaged from 50° to 70° N) with the zero wind line in bold; the initial date of the sudden warming, when the wind reverses at 10 hPa, is apparent. The plots are shaded where the composite does not differ significantly from climatology. Winds in the lower stratosphere (between 100 hPa and 10 hPa) remain weak for 60 to 80 days following the warming in the CMAM composite. The lower stratospheric winds are quite similar in the MERRA composite, though the smaller number of events results in weaker statistical significance. The polar jet re-forms higher and stronger than its climatology, with winds peaking near 0.5 hPa from 20 to 40 days following the initial warming. The jet maximum descends with time in both composites, although the vertical extent of the winds in CMAM is larger. The vertically compressed structure in MERRA is likely to be an artefact of the model top; similar biases have been reported by Manney et al. (2008).

The lower panels show anomalous temperatures from the two datasets. The composite anomaly during the initial warming is more than 20 K (note that individual events can be several times larger than this), which agrees well with the reanalysis. The descending tripole of anomalous temperatures during the recovery period described above is apparent.
Figure 5.2: Composites of PJO events from (a-b) the MERRA reanalysis and (c-d) CMAM. (a,c) Absolute zonal wind, area-averaged from 50° to 70° N, at intervals of 10 m s\(^{-1}\). Zero contour is in bold, and negative values are indicated by dashed contours. (b,d) Temperature anomaly, area-averaged from 70° to 90° N, at intervals of 10 K. In both cases, gray shading indicates where the composite does not differ significantly from climatology at the 95% (light) and 99% (dark) confidence levels.

from lag days 10 to 70. The upper warm anomaly is captured only towards lag day 40 in the MERRA composite, when it has descended into the model domain.

The zonal mean evolution during these events is quantitatively well captured by this CMAM simulation. Some confidence, therefore, can be placed in the relevance of the CMAM composite dynamics to the real atmosphere. We begin with the zonal-mean accelerations induced by the eddy flux convergences, and work in the TEM framework. Figure 5.3(a) shows composites of the zonal wind acceleration, averaged from 50° to 70° N. The deceleration in the stratosphere just prior to the central date is clearly induced by the strong, resolved EP flux convergence (Fig 5.3(b)), which is nearly entirely due to planetary-scale eddies (Fig. 5.3(c)). These EP flux convergences are presented as accelerations for the purpose of comparison with the wind tendencies. As is well known (Matsuno, 1971), the component of the eddy forcing that does not directly decelerate the winds drives a strong residual meridional circulation. Following the wind
reversal, the planetary-scale eddies produce an anomalous EP flux divergence. While there are often brief periods of absolute divergences likely produced by the effects of wave transience immediately following the breakdown of the vortex (see, for instance, below), much of this anomalous divergence is in fact a result of the nearly complete suppression of planetary scale eddy fluxes noted in the previous chapter.

The parameterized gravity waves also respond strongly to the lower stratopheric flow during PJO events. Figure 5.4 shows the anomalous zonal accelerations produced by non-orographic (a) and orographic (b) gravity wave drag. The net upward momentum flux from orographic waves is shown in panel (c). The climatological westward drag due to both types of parameterized waves apparent prior to the central date vanishes in the mesosphere coincident with the initial lower stratospheric wind reversal. The westward winds throughout nearly the entire stratosphere at this point permit eastward-phase speed waves to propagate upwards; as a result the net non-orographic drag at this point becomes eastward. The orographic waves are filtered by the lower stratospheric winds (see the shading in Fig. 5.4(c), which indicates composite zonal wind speeds less than 4 m s$^{-1}$), so their net mesospheric drag vanishes.

As the lower stratospheric winds recover, the filtering effect switches off, and the westward drag in the mesosphere increases. Over the next 60 days the peak drag descends through the mesosphere. This occurs for both types of parameterized waves, though the

Figure 5.3: Composites of PJO events simulated by CMAM. (a) Zonal wind tendencies, (b) total acceleration due to resolved EP flux convergence, and (c) acceleration due to planetary-scale (zonal wave numbers 1 to 3) EP flux convergence. Contours for all quantities are at intervals of 2 m s$^{-1}$ day$^{-1}$. Shading indicates absolute values greater than 3 m s$^{-1}$ day$^{-1}$. Quantities area-averaged from 50° to 70° N.
**Figure 5.4:** Same as Fig. 5.3 but for zonal accelerations due to parameterized (a) non-orographic and (b) orographic gravity wave drag, and (c) the parameterized vertical flux of momentum from orographic gravity waves. Contour intervals for (a) and (b) are 2 m s$^{-1}$ day$^{-1}$. Contours for (c) are logarithmically spaced as labelled. Shading in (c) indicates zonal mean winds less than 4 m s$^{-1}$. Quantities are area-averaged from 50° to 70° N.

Non-orographic waves deposit their momentum higher than do the orographic waves. This descent is perhaps most obvious in the downward tilting contours of the orographic wave flux. The reasons for this descent will be discussed below.

The anomalous vertical residual velocities (averaged over the polar cap) induced by the associated torques are shown in Fig. 5.5(a). The strong anomalous downward flow

**Figure 5.5:** Same as Fig. 5.3. (a) Anomalous vertical residual velocity, with contours at 0, ±0.5, ±1, ±2, ±4, and ±6 mm s$^{-1}$. (b) Anomalous Brunt-Väisälä frequency squared, contour intervals of 2.5×10$^{-5}$ s$^{-2}$. Gray shading indicates absolute anomalies greater than 5×10$^{-5}$ s$^{-2}$. (c) Adiabatic heating in K day$^{-1}$; contours at the same numerical values as (a). All quantities are area-averaged from 70° to 90° N.
in the stratosphere during the warming is apparent, as is the coincident upward flow in the mesosphere. These are consistent with the expected planetary scale drag in the stratosphere and the gravity wave drag in the mesosphere. During the recovery phase, the stratospheric vertical velocities are anomalously weak (as will be clarified below, they remain weakly negative in the net). Above this, the anomalous circulation is downward, following the descent of the enhanced gravity wave drag.

The change in static stability during these events is shown in Fig. 5.5(b). The anomalies are rarely more than 15% of the climatological values; they are therefore quantitatively important to the circulation, as will be discussed below, but are not qualitatively so. Nonetheless, we note the strongly enhanced stability just above the tropopause in the composite. This enhanced tropopause inversion layer (TIL) forms in the polar night (when the climatological TIL is weak) and is clearly associated with the extended recoveries from sudden warmings. The mechanisms by which this transient TIL forms are unlikely to be the same as those responsible for the extratropical TIL outside the polar night (Birner, 2006).

The anomalous adiabatic heating is shown in Fig. 5.5(c), and is well predicted by the anomalous vertical velocities. The initial warming is induced adiabatically as expected, and the descending stratopause is clearly maintained by this dynamical heating.

The anomalous temperature tendencies and diabatic heating are shown in Fig. 5.6(a) and (b). As shown in Chapter 3, the diabatic heating is very strongly correlated with the anomalous temperatures. The heating rates predicted by a Newtonian cooling term
are shown in Fig. 5.6(c). The magnitude is somewhat over-predicted near 5 hPa, which is consistent with the weakly non-linear effects from the curvature of the Planck function (see Chapter 3). The effective radiative timescales here have been computed by regressing the polar-cap averaged zonal mean heating rate anomalies against the corresponding temperature anomalies, following Chapter 3, using only dates during the events. The timescales increase towards the lower stratosphere, and are shown explicitly further below.

The picture that emerges from these composites, then, is as follows. The initial stratospheric warming is induced by the enhanced planetary wave drag, while the coincident mesospheric cooling is a result of the strongly reduced residual circulation. This reduction is a combined effect of the anomalous gravity wave drag and of transient effects from the resolved drag; the relative role of each in CMAM will be clarified below. The subsequent upper stratospheric cooling is a result of the suppressed dynamical driving and strong radiative cooling. Although the anomalous adiabatic heating responsible for the initial warming ceases throughout the stratosphere nearly simultaneously in the composite, the temperature anomaly persists for much longer in the lower stratosphere. The apparent descent of the cold anomaly is associated, therefore, with the vertical gradient in radiative timescales. The stratopause re-forms high in the mesosphere as a result of enhanced and anomalously high gravity wave drag, and the subsequent descent follows that of the region of enhanced drag.

5.4 Residual circulation

To understand these events more quantitatively, we focus henceforth on a particular case study from the CMAM simulation. The approach taken here will be to try to reconstruct the dynamics of this case study through several models of increasing complexity. The details of this particular event are quite similar to those of the composite just discussed, so much of the interpretation carries over to the general case. A similar analysis of other events confirms this claim; where particularities arise they are noted. There are two advantages to analyzing a single event. Firstly, there is a significant seasonal cycle through the extended winter period when these events occur, so by focusing on individual events we can better understand the dynamics in an absolute frame of reference. Secondly, unlike the composite average, this individual event is a solution to the equations of motion integrated by CMAM. Although non-linearities in the zonal mean dynamics are
weak in most senses (an exception will be discussed explicitly below), the structure of the eddy-driven forcings is quite strongly and non-linearly dependent on the zonal mean state. Focusing on an individual event clarifies this dependence. It must be stressed that similar analyses of other events yield the same general conclusions (though there are some particularities which are convenient for the purposes of presentation; these are pointed out below). A second example, in this case for a vortex splitting event, is presented once the full, transient Eliassen response has been developed below.

The particular event discussed here is the PJO event that begins in November of model year 41 (and continues through to March of year 42). It follows a vortex displacement (see Fig. 5.1). Several aspects of its zonal mean evolution are shown in Fig. 5.7. The polar cap temperature anomalies (Fig. 5.7(a)) show the characteristic descending tripolar anomaly following the peak warming in November, with an extended lower stratospheric anomaly, a strong upper stratospheric cold anomaly, and an elevated stratopause. The descent of the anomalous stratopause is interrupted in January when the lower stratospheric winds

Figure 5.7: PJO event from model years 41–42. (a) Temperature anomaly (70° to 90° N), contour intervals of 5 K, zero contour in bold. (b) Zonal wind (50° to 70° N), contour intervals of 10 m s$^{-1}$, zero contour in bold. (c) Acceleration due to EP flux convergence of zonal wave numbers 1 and 2 (50° to 70° N), contour lines at -30 m s$^{-1}$ day$^{-1}$ and 0 m s$^{-1}$ day$^{-1}$. (d) Residual vertical circulation (70° to 90° N), contour lines at -1 mm s$^{-1}$ and 0 mm s$^{-1}$. Note that the shaded contours in (c) and (d) are spaced logarithmically.
accelerate somewhat; this feature is particular to this event, though similar behaviour is not unusual in other events simulated by CMAM. Similarly, the strong precursor to the warming in early November is present in some events but by no means in all of them. The absolute zonal mean zonal winds from 50° to 70° N (Fig. 5.7(b)) reverse in the middle stratosphere in mid-November, after which the polar jet re-forms higher and stronger than climatology, as in the composite (Fig. 5.2(c)). The lower stratospheric winds remain weaker than 10 m s\(^{-1}\) through mid-December, and weaker than 20 m s\(^{-1}\) through mid-January. The acceleration due to planetary scale EP flux convergence (Fig. 5.7(c); compare with Fig. 5.3(c)) shows two initial pulses of convergence responsible for the early-November precursor and the mid-November warming.

The absolute residual vertical velocity (Fig. 5.7(d); compare with Fig. 5.5(a)) shows the large circulation which drives the initial warming. As implied by the composite, the circulation in the stratosphere following the warming is much weaker than climatology; in the net, however, it remains weakly downward nearly all the time. The strong circulation responsible for the descending stratopause is also apparent.

With this zonal perspective in mind, we assume the dynamics of the recovery are essentially that of an Eliassen adjustment (Eliassen, 1951) by which the atmosphere maintains thermal wind balance under imposed zonal mean torques \(F\) and diabatic heating \(Q\). The adjustment is achieved by a meridional circulation which, under quasi-geostrophic scaling, redistributes the effects of the forcings instantaneously. In the TEM framework, the effects of the eddies are thought of solely as a torque and can to some extent be considered as an imposed forcing. The diabatic heating (primarily radiative in the middle atmosphere) is fundamentally relaxational in character (Chapter 3). The response to localized forcing of either type is to drive a residual circulation which locally opposes the effect of the forcing, and forms two cells that close in both directions. The instantaneous influence of these cells tends to be weighted more heavily upwards (Plumb, 1982).

However, if the zonal mean reaches a steady state in response to an imposed eddy-driven torque, the circulation must close entirely downward (Haynes et al., 1991). Following a switch-on forcing in an atmosphere with a constant rate of radiative damping, the time scale on which steady state is expected to be obtained is given by \(\Delta z H/\alpha H^2_R\), where \(\Delta z\) is the vertical distance from the height of the forcing, and \(H_R\) is a Rossby height that depends on the horizontal scale of the forcing. In the real atmosphere (and in the simulations considered here), there is substantial vertical structure in the radiative damping rates. However, one expects from this argument that the residual circulation
during these events will still be influenced significantly by transients.

We diagnose the transient residual circulation using the TEM formulation of zonal-mean quasi-geostrophy on the sphere, following Plumb (1982) and Haynes et al. (1991). This yields a linear equation for the vertical residual velocity of the form

$$\mathcal{L}w_{QG}^* = \mathcal{L}_F F + \mathcal{L}_Q Q,$$

(5.1)

where $\mathcal{L}$, $\mathcal{L}_F$, and $\mathcal{L}_Q$ are linear differential operators defined in the appendix. One particular difficulty in inverting this equation to find $w_{QG}^*$ is specifying a lower boundary condition in pressure coordinates. The condition used here is that of Haynes and Shepherd (1989), which permits changes in the surface pressure. In the TEM framework, an additional surface heat-flux term must be included as a forcing at the lower boundary. These conditions are still unsatisfactory as they do not account for the bottom topography; however we have found that the model is, nonetheless, sufficiently accurate as a diagnostic tool for the present purposes. More discussion of this issue, and a brief outline of the model and the numerical methods used to solve it are given in the appendix for the purposes of reproducibility.

The steady-state downward control vertical residual velocity under the same quasi-geostrophic scaling is given by (Haynes et al., 1991)

$$\bar{w}_{DC}^* = \frac{1}{\rho_0 a \cos \phi} \int_0^\infty \frac{\partial}{\partial \phi} \left( \rho_0 \cos \phi \frac{F}{f} \right) dz,$$

(5.2)

where the symbols are defined in the appendix to this Chapter. In this case the circulation is considered to be driven entirely by the torques; the same circulation could be derived from (5.1) by also imposing the diabatic heating required to maintain steady state.

These two expressions are essentially linear, and can therefore be used to straightforwardly decompose drivers of the residual circulation. The eddy forcings and diabatic heating simulated by the model are used as input in each case.

The results of these decompositions are shown in Fig. 5.8 for the absolute residual circulation over the polar cap, averaged over two separate phases of the event. The decomposition of the circulation is shown for the quasi-geostrophic case in the central panels, and the downward control case in the rightmost panels. The static stability profile, required by (5.1), is taken to be the polar-cap $N^2$, averaged over the corresponding time period. The first phase (top panels; indicated by the solid vertical lines in Fig. 5.7) is
Figure 5.8: Decompositions of the polar-cap averaged vertical residual circulation from (a) 20 to 30 November and (b) 20 to 30 December, model year 41, during the event shown in Fig. 5.7. The leftmost panel in each case shows the total circulation simulated by CMAM (solid), and as estimated by the quasi-geostrophic decomposition (dashed) and by the downward control decomposition (dotted). In the lower left panel the inset shows the lower stratospheric circulation in greater detail (the vertical axis of the inset matches that of the panel). The central panel shows the quasi-geostrophic circulation induced by resolved planetary-scale wave drag, resolved waves of all other wave numbers, parameterized gravity wave drag, the surface heat-flux, and the diabatic heating. The rightmost panel shows the downward control contribution from the four categories of wave drag.

during the initial peak warming period, when the resolved wave driving is at its strongest and most transient, and thus where one would expect the largest discrepancy between the transient and equilibrium circulations. Indeed the downward control circulation overpredicts the magnitude of the circulation everywhere in the stratosphere and mesosphere, particularly so near 1 hPa. The agreement between the quasi-geostrophic circulation and the true circulation is substantially better, although the circulation is still slightly
overpredicted in the mesosphere. In general the downward control estimates of the circulation are larger in magnitude than the transient estimates, because of the effects of the inferred steady-state diabatic heating. By contrast, the non-planetary wave torques play little to no role throughout the model domain in this period, with the exception of the synoptic-scale driven circulation in the troposphere. Also shown, to be explicit, is the circulation induced by the surface heat flux. Despite being imposed at the lower boundary, this drives a significant circulation even in the lowermost stratosphere.

The second phase considered (bottom panels; indicated by the dashed vertical lines in Fig. 5.7) is during the recovery phase of the event, when the planetary wave driving is strongly suppressed in the stratosphere, and the stratopause is in the midst of its descent. This is reflected in the circulation, with strong downward velocities above 1 hPa, and very weak downward velocities from 100 hPa to 10 hPa. Here the downward control estimate is considerably more accurate than during the first period, though careful inspection of the lowermost stratosphere (inset in the leftmost panel) shows that the magnitude of the circulation is underestimated. Both decompositions show that waves of all scales considered here play a role in the stratopause circulation, with the orographic waves contributing most substantially just above 1 hPa. The diabatic heating also plays a significant role throughout the depth of the polar cap. In the lowermost stratosphere, the weak upwards circulation driven by planetary waves is also compensated by the surface heat-flux contribution.

To summarize, the residual circulation shows significant contributions from the transient response to the imposed eddy-driven torques throughout the event. Waves of all scales appear to play some role in the re-formed stratopause following the initial warming, with the orographic wave drag apparently playing the dominant role towards the base of the mesosphere. The net lower stratospheric residual circulation following the initial warming is extremely weak but still downward, with significant contributions from the diabatic heating. While the role of the transient response to the eddies is clearly indicated by this decomposition, the details of the approach to steady state remain to be clarified, as does the relative role of the adiabatic and diabatic heating in determining the temperatures.
Figure 5.9: Vertical profile of (a) radiative damping timescales and (b) Brunt-Väisälä frequency squared. Shown in the left panel are the Holton-Mass profile (red solid line), the regressed damping timescales from the events in CMAM (black line), and the analytical approximation to the latter (red dashed line). Shown in the right panel are the profiles of polar-cap averaged (70° to 90° N) static stability taken from the CMAM event. The solid profile corresponds to the period indicated in Fig. 5.7 by the solid lines, while the dashed profile corresponds to the period indicated by the dashed lines.

5.5 Temperature anomalies

We turn now to the description of the temperature anomalies themselves. The focus initially will be on the stratospheric component of the circulation. As demonstrated in the previous section, following the initial warming, the net adiabatically driven warming in the stratosphere is weak. The first approximation, considered in the next subsection, will be to assume that the temperatures are driven entirely by the relaxational diabatic heating. This is refined somewhat in the second subsection by including the residual circulation induced by the diabatic heating itself. Finally, in the third subsection the full, transient response to the imposed torques will be presented, and the attention broadened to the full middle atmosphere response.

As input to these calculations, vertical profiles of radiative damping rates and the Brunt-Väisälä frequency are required. The profile of effective damping timescales computed from the PJO events simulated by CMAM is shown in black in Fig. 5.9(a). The profile used by Holton and Mass (1976) (which has been used by a number of subsequent studies) is given by the solid red line, and an analytical approximation to the CMAM
rates is given by the dashed red line. The profile of damping rates is given by

\[ \alpha_{CMAM} = \alpha_T + \frac{1}{2} (\alpha_M - \alpha_T) \left[ 1 + \tanh \left( \frac{z - 45 \text{ km}}{14 \text{ km}} \right) \right] \]

\[- \alpha_{LS} \exp \left( -\frac{(z - 14 \text{ km})^2}{2(3 \text{ km})^2} \right) \text{ day}^{-1}, \]  

(5.3)

with \( \alpha_T = 0.04, \alpha_M = 0.22, \) and \( \alpha_{LS} = 0.028. \) Also shown for reference in Fig. 5.9(b) are two profiles of \( N^2, \) computed from the CMAM simulation during the two periods highlighted in Fig. 5.7.

### 5.5.1 Radiative relaxation

In the absence of dynamical heating, longwave radiation will cool the stratosphere towards radiative equilibrium. The net heating rate \( Q \) is then given by the climatological cooling rate \( Q_c, \) in addition to the linear relaxational component due to the temperature anomaly \(-\alpha T'.\) This is much more naturally expressed in an absolute framework (i.e. \( \partial_t T = Q \)); however, since the present focus is on the behaviour of the anomalies, it will be more consistent with later arguments to work in an anomaly framework. In transforming to this framework, there arises an additional term due to the climatological temperature tendency (in other words, the net temperature should stay constant in the absence of heating of any kind). In equations,

\[ \frac{\partial T}{\partial t} = Q \]

\[ \frac{\partial}{\partial t} (T' + T_c) = Q_c - \alpha T' \]

\[ \frac{\partial T'}{\partial t} = Q_c - \frac{\partial T_c}{\partial t} - \alpha T'. \]  

(5.4)

Note that this is equivalent to the fixed dynamical heating approximation in which the absolute dynamical heating is set to zero.

Two particularities of this case make it especially suitable for this simple analysis. Firstly, the initial warming is quite barotropic in nature—peak temperature anomalies occur throughout the stratosphere almost simultaneously. As was pointed out by Zhou et al. (2002), often additional pulses of wave activity are responsible for bringing the warm anomaly down from the middle to the lower stratosphere. The barotropic character is convenient for the present analysis, as we take the anomalies on a date near this peak as
the initial conditions for integrating the above equation. Secondly, this particular event occurred very early in the season, so that much of the recovery phase occurs during midwinter. The background seasonal cycle is therefore quite weak, and the corresponding term is essentially negligible (though it is included for completeness). It plays a more significant role in events that occur later in the season.

The results of three integrations are shown in Fig. 5.10. The temperature anomalies in each case are initialized at the date indicated by the vertical line, but they differ in the choice of radiative damping rates applied. The first panel uses a constant rate of $1/10$ day$^{-1}$. Despite the constant profile, the lower stratospheric anomaly persists for longer than the upper stratospheric anomaly due to the weaker climatological cooling. The minimum in climatological cooling near 200 hPa is apparent. The second panel uses the profile of Holton and Mass (1976). The weaker damping in the lower stratosphere extends the lower stratospheric anomaly, while the stronger damping in the upper stratosphere
reduces the cooling rate and thus the magnitude of the cooling anomaly. Finally, using the profile fit to the effective rates computed from these runs, the persistence of the lower stratospheric anomaly is extended by the timescales which approach 75 days.

Two points bear emphasizing here. Firstly, there is no vertical transfer of information in this model, not even numerical diffusion. The downward migration of the stratospheric cold anomaly, which is captured here, can therefore be explained to a large extent as a ‘phase descent’ of a character similar to that discussed by Plumb and Semeniuk (2003). In this case the apparent descent is in part due to the vertical gradient in damping rates, and in part due to the gradient in background cooling rates which act in the absence of dynamical heating. Secondly, the lower stratospheric anomaly simulated by CMAM itself persists for substantially longer than this simple model predicts (compare the $T' = 0$ K contour in Fig. 5.10(c) and Fig. 5.7(a)), despite the extended radiative timescales.

### 5.5.2 Radiative relaxation with Eliassen adjustment

The under-prediction of the persistence of the lower stratospheric anomaly from radiative considerations alone is consistent with the weak but downward circulation found in Fig. 5.8(b). We therefore extend the simple radiative relaxation used in the previous section by adding in the effects of the Eliassen adjustment induced by the diabatic heating rates themselves,

$$\frac{\partial T'}{\partial t} = Q_c - \frac{\partial T_c}{\partial t} - \alpha T' - S\bar{w}Q_G,$$

where $S = (H/R)N^2$ is related to the static stability by the density scale height $H$ and the dry gas constant $R$.

Here the circulation is predicted using (5.1) from the absolute diabatic heating rates predicted by Newtonian cooling. For simplicity a constant profile of static stability is used ($N^2 = 4 \times 10^{-4}$ s$^{-2}$). The analytical fit to the CMAM damping rates is retained, and the result is shown in Fig. 5.11(a). The effect is to extend the lifetime of the lower stratospheric anomaly; essentially some of the diabatic cooling is going into accelerating the winds rather than into lowering the temperatures.

Figs. 5.11(b) and (c) compare these models in greater detail to the event itself, showing temperature time series at 1 hPa and 200 hPa, respectively. At 1 hPa the temperature anomaly associated with the two pulses of planetary waves (thick black line) has nearly decayed by the beginning of the integration. The temperatures drop rapidly to nearly 30 K below climatological values (thin black line), before returning to climatological
values by the beginning of January. The initial drop in temperature is reasonably well described by all three integrations: the radiative relaxation with a 10 day timescale (dashed red line) and with the CMAM profile (solid red line), and by the radiative relaxation with Eliassen adjustment (blue line). All three integrations cool until they approach their respective radiative equilibrium temperatures; the subsequent warming in the real event requires further dynamical heating. At 200 hPa, where the assumption that there is no torque-induced heating is a better approximation, the difference between the three simple integrations is more informative. Here the warm anomaly in the true event persists significantly longer even than the straight radiative relaxation with the
CMAM profile, better matching the integration with Eliassen adjustment. Note that the timescale of the anomaly is greater than would be expected from radiative relaxation alone, if one assumes that the temperatures are relaxing back to radiative equilibrium. If one worked in the anomaly framework and did not apply the climatological heating rates, one would find that the anomaly decays substantially faster than the radiative timescale might imply; this argument, however, implicitly retains the climatological dynamical heating which is absent in the aftermath of the warming.

Finally, it should be noted that although the barotropic nature of this particular event is not particularly universal, the conclusion that the downward migration of the cold anomaly is ultimately a result of the gradient in both the radiative damping rates and the climatological cooling rates holds more generally.

### 5.5.3 Transient adjustment

The full transient response is now considered. Here we work explicitly with anomalies from the climatology, integrating

\[
\frac{\partial T'}{\partial t} = -\alpha T' - S \overline{w_{QG}}',
\]

where the anomalous circulation \( \overline{w_{QG}}' \) is given by solving (5.1) for the imposed anomalous torques \( F' \) and the anomalous diabatic heating predicted by the Newtonian cooling term. This is equivalent to the model solved analytically by Haynes et al. (1991), but with torques generated by the free running model, and a realistic profile of radiative damping rates. We no longer specify the climatological radiative cooling since it will balance the climatological torques. The conceptual framework here is that, since the radiative heating is linear and relaxational, one can think of the residual circulations induced by the heating as ultimately being associated with the imposed torque that drove the radiative imbalance in the first place. We work in an anomaly framework to retain the connection to the composite picture. For the most accurate results, the time dependent polar-cap averaged static stability profile (two periods of which are shown in Fig. 5.9(b)) is used. This is found to give better results than, for instance, the globally averaged profile.

The results of this integration are shown in Fig. 5.12. The total response, imposing the resolved wave driving at all wave numbers and the parameterized gravity wave drag of both types, is shown in panel (a). This can be compared with Fig. 5.7(a). The details
Figure 5.12: Polar cap temperature anomalies predicted by the quasi-geostrophic transient adjustment of the circulation to (a) all torques, resolved and parameterized, (b) planetary-scale eddy-driving, (c) parameterized orographic gravity wave drag, and (d) parameterized non-orographic gravity wave drag. Contours as in Fig. 5.7(a).

are reasonably well captured; the two initial pulses of planetary-scale waves in November warm the stratosphere, following which the warm anomaly in the lower stratosphere persists nearly to the end of January. The amplitude of the cold anomaly in the upper stratosphere following the warming is somewhat under predicted, but the details of the descending stratopause, including the arrest in its descent at the beginning of January, are well captured. The response can then be decomposed into the response to each of the dominant forcings.

The initial warming, as expected, is nearly entirely explained by the planetary scale EP flux convergence (Fig. 5.12(b)). The mesospheric response to the initial pulse is to cool; that this is caused by the transient upward-closing cell of the residual circulation can be inferred by noting that the anomalous EP flux divergence is everywhere negative over the pole (not shown); the cooling predicted by the transient response is not, therefore, a result of reduced drag. The secondary pulse, however, generates considerable mesospheric warming as well, contributing to the initial formation of the high stratopause. This is a relatively common feature of the events simulated by CMAM in this run; whether this is true of the real atmosphere is not clear from reanalyses, but has been noted by Siskind.
et al. (2010). The lower stratospheric response is in fact more persistent, while the cold anomaly in the upper stratosphere is considerably weaker than the full response.

The reduced orographic gravity wave drag in the upper stratosphere and lower mesosphere following the initial two pulses of planetary waves (and resulting deceleration of the polar jet) results in significant radiative cooling (Fig. 5.12(c)—that this should be thought of as a radiative response is clear from the net orographic drag shown in Fig. 5.4(b). This plays a significant role both in the mesospheric cold anomaly during the initial warming, and in the descending cold anomaly during the PJO event itself. By mid-December the response becomes anomalously warm near the top of the model domain. This warm anomaly descends through December and persists through January, contributing somewhat less than half of the total anomaly seen in the full response.

The remainder of the anomalous mesospheric cooling during the initial stratospheric warming, and the anomalous mesospheric warming during the PJO event, is explained by the non-orographic parameterized gravity waves (Fig. 5.12(d)). Unlike the orographic wave drag, some of the cold anomaly is in fact a result of a net eastward torque during the initial warming as the westward waves are filtered by the wind reversal in the stratosphere.

Some sense of the degree of non-linearity that arises from the change in static stability induced by these temperature changes is given by considering the sensitivity of this response to the use of the time-dependent profile of $N^2$ taken from the free-running simulation. This is shown in Fig. 5.13(a). In general, where the time-dependent stability is larger than the constant profile (during the initial warming, for instance, and near the enhanced tropopause inversion layer), the temperature response assuming the latter is too cold, while where it is lower (in the middle to upper stratosphere during the descent of the cold anomaly) the temperature response is too warm. This ‘direct’ effect (which results from the enhanced adiabatic heating associated with greater stability) dominates the indirect effect of weakening the circulation. Indeed, the error from neglecting this non-linearity reaches 10 K near the lowermost stratosphere, not because the instantaneous change in adiabatic heating is at any point particularly large, but because the radiative damping is so weak.

Similarly, the effect of the extended radiative timescale in the lowermost stratosphere can be estimated by computing the response assuming the Holton-Mass profile which does not include this feature (Fig. 5.13(b)). Most notably, the response of the lowermost stratosphere is too cold in this case, which provides additional confirmation that the effective radiative timescales in this region are significantly longer than assumed by the
Figure 5.13: Difference in the total predicted temperature response as a result of using (a) the climatological polar cap profile of $N^2$ and (b) the Holton-Mass profile of radiative damping timescales instead of the profiles used in Fig. 5.12. Contour intervals of 3 K, zero contour indicated in bold.

Holton-Mass profile. Similarly, the mid-stratospheric cold anomaly is too warm.

To demonstrate that the basic dynamics just described apply to other events, Fig. 5.14 shows the results from a similar calculation for the PJO event that occurs in the winter of model years 93–94 (see Fig. 5.1). The sudden warming that initiates this event was a vortex splitting event, as opposed to the displacement that initiated the previously considered event. The temperature anomalies during the event are shown in panel (a), and the anomalies predicted by the transient quasi-geostrophic adjustment are shown in panels (c-d). As in the vortex displacement event, the transient adjustment captures the full model behaviour with some fidelity. The initial warming is produced as expected by planetary scale wave drag (Fig. 5.14(c)), as is the initial mesospheric warming, though in this event the drag at mesospheric heights persists for longer than in the displacement event. Nonetheless, the majority of the upper warm anomaly (and its descent) is produced by anomalous gravity wave drag (Fig. 5.14). Nearly all of the stratospheric cooling in this case is in fact due to reduced gravity wave drag. As with the displacement event (and emphasized in Fig. 5.13(a)), imposing the time-dependent profile of polar cap static stability is necessary to achieve the close agreement seen between Fig. 5.14(a) and (b),
confirming the quantitative importance of non-linearity to the temperature evolution (not shown).

5.6 Stratopause descent

We consider finally the descent of the stratopause. It is clear from Fig. 5.12 that the anomalous temperatures at the top of the stratospheric jet are maintained by torques of a variety of origins; initially the planetary-scale waves provide some heating, followed by the parameterized non-orographic gravity waves, and finally the parameterized orographic gravity waves dominate as the stratopause descends to more climatological heights. The strong radiative cooling at these heights implies that the temperature anomalies must be actively maintained by the dynamical heating. We focus here on the orographic wave drag, partly because it descends most clearly along with the stratopause, and partly because the anomaly picture emphasized in Fig. 5.12 belies the greater role that the orographic waves play in the climatological polar winter stratopause.

Following the initial warming, the radiative cooling of the stratosphere leads to a high
Figure 5.15: Vertical profiles of (a) zonal wind, (b) upward flux of westward momentum carried by parameterized orographic gravity wave and (c) acceleration due to orographic gravity wave drag. All quantities are averaged from 50° to 70° N. The solid, dashed, and dotted lines are averages over consecutive 10 day periods following 22 December, model year 41.

There is also an indication that the source of the waves is increasing throughout the event, since the fluxes throughout the depth of the atmosphere are increasing. This may be a result of strengthening surface winds at these latitudes.

The descent of the region of wave drag, though similar in some ways to the descending shear zones of the QBO, differs in several key respects. Unlike the tropics, the acceleration of the winds is due to the radiatively-generated strong thermal gradient, not to the
dissipation of waves of eastward phase speed. Moreover, the waves break as a result of saturation processes, not as a result of critical lines (or the strongly reduced vertical group velocities encountered near them). The region of wave drag is thus above the jet maximum, not below. The parameterized waves dissipate, rather than accelerate, the winds, confirming that the high wind speeds are radiatively rather than dynamically produced. In these events this is permitted by the strong filtering of the waves by the stratospheric wind reversal during the warming itself (Holton, 1983). The role of the lower stratospheric winds in this filtering bears re-emphasizing; in some events simulated by CMAM, the recovery of the lower stratospheric winds is interrupted by a further pulse of planetary waves. If this decelerates the winds sufficiently, the parameterized waves are filtered strongly, and the descent of the stratopause above is arrested. The lower stratospheric anomaly therefore has a great deal of influence on the circulation above as well as on the tropospheric response below.

5.7 Conclusions

The focus of the present chapter has been to clarify the dynamics that lead to the characteristic pattern of the zonal mean circulation that occurs during PJO events, with a particular emphasis on the structure of the temperature anomalies. These features are sufficiently robust that many of the details found in the composite analysis can be inferred from the analysis of a single case study. One implication of this robustness is that the circulation of the stratosphere and mesosphere (at least in the zonal mean) should be quite predictable during these events.

The zonal mean dynamics of the initial sudden warming are well known and are corroborated by the present analysis; the rapid warming of the stratosphere is a result of strong EP flux convergence associated with the first two zonal wave numbers. If this warming reaches the lowermost stratosphere, it can persist for several months. This persistence was shown here to be a result not only of the weak radiative cooling in this region, but also of the dynamical heating induced by the Eliassen adjustment to the resulting diabatic cooling. The apparent descent of the cold anomaly through the stratosphere can be explained by the vertical gradient in both the radiative timescales and in the climatological rates of cooling.

The zonal mean temperature anomalies were found to be well captured by considering the full quasi-geostrophic transient Eliassen adjustment problem on the sphere, provided
realistic profiles of radiative damping rates and (time-dependent) static stability are imposed. In particular, suppressed parameterized gravity wave drag following the warming was found to play a significant role in permitting the radiative cooling responsible for the cold anomaly in the stratosphere.

The cooling in the mesosphere during the initial warming phase is induced in the present run by a combination of the radiative cooling permitted in the absence of parameterized gravity wave drag and by a transient circulation induced by the resolved wave drag in the stratosphere below. This results in the disappearance of the stratopause at its climatological height. The stratopause subsequently re-forms well above this height, initially as a result of planetary waves which reach the mesosphere, followed by non-orographic and subsequently orographic parameterized gravity waves. This sequence of events matches that described by other studies of elevated stratopause events (Siskind et al., 2010), though the transient calculations performed here allow us to clearly attribute the temperature anomalies to the different sources of wave driving. The descent of the stratopause is somewhat analogous to the descent of shear zones in the QBO in that the region of wave drag descends with the peak of the jet; there are, however, significant differences in details of the processes responsible. There appears to be some correspondence between the apparent rate of descent of the cold anomaly in the stratosphere and the warm anomaly in the mesosphere. It is clear that the lower stratospheric circulation plays a role in determining the gravity wave flux responsible for the mesospheric anomaly, but whether the roughly equal rates of descent are anything more than a coincidence is unclear.

5.A Zonal-mean quasi-geostrophy on the sphere

The diagnostic equation for the quasi-geostrophic residual circulation on the sphere is derived from the following set

\[
\frac{\partial u}{\partial t} - f v^* = F
\]

\[
\frac{\partial T}{\partial t} + S w^* = Q
\]

\[
\frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} (\cos \phi v^*) + \frac{1}{\rho_0} \frac{\partial}{\partial z} (\rho_0 w^*) = 0
\]

\[
f \frac{\partial u}{\partial z} = -\frac{R}{a H} \frac{\partial T}{\partial \phi}.
\]

Here \(\rho_0\) is the reference density profile, \(f = 2\Omega \sin \phi\) is the Coriolis parameter, \(\Omega\) is the angular velocity of the Earth, and \(a\) is the radius of the Earth.
Solved for $\mathbf{w}^*$ these yield the differential operators of (5.1),

\[
\mathcal{L}(\cdot) = \frac{\partial}{\partial z} \left( \frac{1}{\rho_0} \frac{\partial (\rho_0 \cdot \mathbf{w}^*)}{\partial z} \right) + \frac{N^2}{(2\Omega a)^2} \frac{\partial}{\partial \mu} \left( \frac{1 - \mu^2}{\mu^2} \frac{\partial (\cdot)}{\partial \mu} \right),
\]

\[
\mathcal{L}_F(\cdot) = \frac{1}{2\Omega a} \frac{\partial}{\partial \mu} \left( \frac{\sqrt{1 - \mu^2}}{\mu} \frac{\partial (\cdot)}{\partial z} \right),
\]

\[
\mathcal{L}_Q(\cdot) = \frac{1}{(2\Omega a)^2} \frac{\partial}{\partial \mu} \left( \frac{1 - \mu^2}{\mu^2} \frac{\partial (\cdot)}{\partial \mu} \right),
\]

where $\mu = \sin \phi$.

The lower boundary condition used is that of Haynes and Shepherd (1989) modified for the TEM. The assumption is that geometric velocity normal to the bottom surface vanishes ($D\Phi/Dt = 0$ at $z = 0$). This yields the following condition for the residual circulation in log-pressure coordinates at the lower boundary

\[
\frac{1}{\rho_0} \frac{\partial \rho_0 \mathbf{w}^*}{\partial z} - \frac{g}{4\Omega^2 a^2} \frac{\partial}{\partial \mu} \left( \frac{1 - \mu^2}{\mu^2} \frac{\partial \mathbf{w}^*}{\partial \mu} \right) = \frac{g}{4\Omega^2 a^3} \frac{\partial}{\partial \mu} \left( \sqrt{1 - \mu^2} \frac{\mathbf{v}' S}{S} \right).
\]

Here $g$ is the acceleration due to gravity. Since the lower boundary is not in fact a pressure surface in CMAM, we use meridional heat fluxes interpolated on to the 1000 hPa surface.

The quality of the agreement between the residual circulation in CMAM and that predicted by this quasi-geostrophic approximation is significantly better in this T63 time-slice run than in other, lower resolution runs. The reason for this is unclear.

The method of solving this equation is as follows. The equation is separated following Plumb (1982), yielding the zonal mean Laplace equation in the meridional direction. (This restricts $N^2$ and $\alpha$ to be independent of latitude.) This defines two Hermitian operators of interest, $\mathcal{M}^\dagger \mathcal{M}$ and $\mathcal{M} \mathcal{M}^\dagger$, where $\mathcal{M}(\cdot) = \sqrt{(1 - \mu^2)/\mu^2} \partial_{\mu}(\cdot)$, each of which is associated with a family of eigenfunctions ($\mu B_n$ and $\Theta_n$ in the notation of Plumb (1982)). The meridional dependence of all quantities can then be expanded in terms of these two orthogonal sets. This expansion is carried out numerically and therefore requires that the computed eigenfunctions be orthonormal under the numerical integration operations; while this is straightforward to obtain for each set independently, finding two sets that are both simultaneously orthogonal proved more difficult than was expected. Several initial attempts to compute these functions based on finite-difference approximations yielded sets of functions unusable for the required expansions. Computing these functions based
on expansions in terms of Legendre polynomials yields a set that is adequately orthogonal through roughly the first half of the eigenfunctions. Since these describe the larger length scales expected to satisfy quasi-geostrophic scaling, truncating the expansions is equivalent to filtering out some of the unbalanced component of the circulation. For all the calculations carried out here, the sums are truncated at mode 14.

The resulting vertical equation that remains is then discretized on a grid with 0.5 km resolution. The differential operators become tridiagonal matrices which can be applied and inverted explicitly for every order $n$.

The temperature equation is integrated forward in time using a third-order Adams-Bashforth method. For the full adjustment problem, (5.1) is inverted at every time step.
Chapter 6

Mechanistic circulation modelling

6.1 Introduction

As discussed in Chapter 1, the lower polar stratosphere has been identified as a key region for the two-way coupling between the stratosphere and the troposphere. Circulation anomalies in the stratospheric polar vortices in both hemispheres have been shown to influence the extratropical tropospheric jets, whether the anomalies are caused by, for instance, ozone depletion in the Southern Hemisphere (Thompson and Solomon, 2002; Son et al., 2008) or sudden warmings in the Northern Hemisphere (Baldwin and Dunkerton, 2001).

As demonstrated in Chapter 4, extratropical coupling in the boreal winter is at its strongest and most persistent following weak vortex events that are followed by a PJO event. Moreover, the persistence of the stratospheric anomaly is closely related to the long radiative timescales present in the lowermost stratosphere (Chapter 5).

The present chapter, then, is in part motivated by an attempt to better understand the internal dynamics of these PJO events and their role in stratosphere-troposphere coupling in the context of a simplified, ‘mechanistic’ circulation model, of the type motivated by Held and Suarez (1994). Such models have been an essential tool in characterizing the range of possible behaviour of the fully coupled stratosphere-troposphere system (Taguchi et al., 2001; Yoden et al., 2002), since their simple parameterizations result in relatively fast execution times, permitting key parameters to be swept through plausible ranges. The possible influence of the time-mean strength of the stratospheric vortex on the position of the tropospheric jet was highlighted dramatically by Polvani and Kushner (2002). While the sensitivity was subsequently found to be unrealistically large due to the un-
realistically long decorrelation timescales in their tropospheric jet, Gerber and Polvani (2009) nonetheless confirmed the poleward shift of the tropospheric jet in response to the radiative strengthening of the stratospheric vortex.

Radiative heating is commonly parameterized in this class of models as a simple Newtonian relaxation to a prescribed radiative equilibrium temperature (Held and Suarez, 1994; Yoden et al., 2002). The role of lower stratospheric radiative damping is, therefore, a natural candidate for a parameter sweep experiment. Indeed, a variety of vertical profiles of damping timescales have been employed, and several sensitivity studies have been carried out. Two profiles in particular are commonly employed. A vertically constant timescale was employed by Polvani and Kushner (2002), Gerber and Polvani (2009) and Scott and Polvani (2006), with values ranging from 5 to 40 days. The vertical profile of timescales adopted by Holton and Mass (1976) falls from 25 days in the lower stratosphere to 5 days in the upper stratosphere, and was employed by Taguchi et al. (2001) and Scott and Polvani (2006). The latter found that power in the spectrum of variability produced in a perpetual January run generally shifted towards lower frequencies as the radiative timescales lengthened, though the behaviour was a complicated function of the height of the imposed surface topography. In a study closely related to the present work, the sensitivity of tropospheric annular mode decorrelation timescales to four separate profiles of stratospheric damping times was investigated (Charlton-Perez and O’Neill, 2010, hereinafter CO), revealing relatively little sensitivity of the tropospheric dynamics to the stratospheric radiative timescales.

Chapter 3 demonstrated that this linear approximation can (with some notable exceptions) realistically describe middle atmospheric heating rates in a comprehensive model, and that conversely, realistic effective damping timescales can be diagnosed from such a model with some precision. The radiative timescales found in Chapter 3 (and which were demonstrated in Chapter 5 to be relevant for the persistence of PJO events simulated in the comprehensive model) can reach 70-80 days in the lower Arctic stratosphere (see Fig. 6.1(b) below), which is considerably longer than the values typically used in mechanistic circulation studies. This suggests that it is worth reexamining the effect of radiative timescales on stratospheric vortex variability, and provides an additional motivation for the present chapter.

In contrast to CO, the present parameter sweep produced a broad spectrum of sudden warming event durations, with the longest and most realistic profiles inducing variability quite analogous to the observed PJO. This clearly demonstrates the relevance of the
radiative timescales to the variability. Moreover, this spectrum of variability provides a useful context for investigating the tropospheric impact of sudden warmings. While the warmings themselves are induced by the coupling between the stratosphere and troposphere, in this context they can be thought of as a ‘forcing’ external to the tropospheric jets themselves, whose dynamics are driven primarily by synoptic scale eddies confined to the troposphere.

The dynamics of the tropospheric jets in response to various external forcings is a rich and active area of study (see Kushner, 2010, for a recent review), and a detailed consideration of the dynamics of the tropospheric response is beyond the scope of the present study. The intent here is simply to document the magnitude and duration of the tropospheric response to the spectrum of warmings, and to point out the perhaps under-appreciated role that the rearrangement of planetary scale momentum fluxes and form drag generated by the topography play in this response.

Details of the model configuration, including the vertical profiles of radiative damping timescales, are provided in the second section. A complication arising from the present methodology is that when the radiative damping timescales are changed, not only is the variability changed, but also the time mean circulation. An important task in the present analysis is therefore to distinguish between these two effects. To this end, the time mean response of the circulation is described in the third section. The stratospheric response is found to be dominated by the direct radiative effect, while the tropospheric jets shift equatorward as a result of a synoptic-scale eddy response to the warming of the polar lower stratosphere. A characterization of the stratospheric variability and the associated tropospheric response follows in the fourth section. Conclusions and a summary are presented in the final section.

6.2 Model setup

The model employed here is the same as that of Taguchi et al. (2001), to which the reader is referred for further details. The dynamical core is run at a horizontal resolution of T21, with 42 vertical levels from the surface through the mesosphere. Radiation is parameterized by Newtonian relaxation towards the same equilibrium temperature field appropriate for a persistent-January configuration used by Taguchi et al. (2001). The control runs make use of the same vertical profile of radiative damping rates, adopted from that used by Holton and Mass (1976). We perturb the vertical profile according to
Figure 6.1: (a) Vertical profiles of radiative damping timescales (in days). The control profile is shown in black. (b) Profile of middle-atmosphere radiative timescales in the Arctic winter estimated from CMAM. The two lines show the 95% confidence interval.

The lower stratospheric damping rates are controlled by $\alpha_{LS}$; simulations are performed at the following values: 0.1, 0.15, 0.2, 0.25, 0.3, 0.5, 0.7 and 0.9; these profiles are shown in Fig. 6.1. For comparison, also shown is an estimate of the damping rates based on CMAM simulations (see Chapter 3 for details). The value of $\alpha_{LS} = 0.5$ corresponds to the profile used by Taguchi et al. (2001); this simulation is therefore considered the control run. The layer in the lower stratosphere with long damping timescales is somewhat deeper in the idealized profiles than in CMAM, but it is clear that the lower stratospheric damping timescales considered here are well justified by the diagnosed values. For comparison, the profiles considered by CO had a maximum lower stratospheric timescale of 40 days. The emphasis in the present analysis will be on the weakened damping runs. The parameters determining the upper stratospheric damping and the tropospheric damping are held at values used by Taguchi et al. (2001); $\alpha_{US} = 2.5$ and $\alpha_T = 0.5$, respectively. This results in tropospheric damping timescales of approximately 25 days. Note that this is a stronger damping than the 40 day timescale specified in Held and Suarez (1994).

Two sets of experiments have been performed, with sinusoidal surface topography of zonal wave numbers 1 and 2 specified in the Northern Hemisphere. The meridional profile in all cases is quartic in $\sin \phi$ with a maximum of 1000 m at 45°N; the analytical form is the same as that of Taguchi et al. (2001). Surface friction is imposed as a linear damping

\[
\alpha = \begin{cases} 
   \left( \alpha_{LS} + \frac{1}{2} (\alpha_{US} - \alpha_{LS}) \left[ 1 + \tanh\left( \frac{z-35 \text{ km}}{7 \text{ km}} \right) \right] \right) \times 10^{-6} \text{ s}^{-1} & \text{if } z > 10 \text{ km}, \\
   \alpha_T \times 10^{-6} \text{ s}^{-1} & \text{otherwise.}
\end{cases}
\]
on the lowest model level with a rate of 0.5 day$^{-1}$. Rayleigh drag is also imposed above 50 km as a sponge layer. The model simulations are run for 10,200 days with the first 200 days discarded for spin-up. Two additional simulations at the weakest damping rate $\alpha_{LS} = 0.1$ are run for 100,200 days to improve composite statistics as discussed further below. Except where noted, the 10,200 day runs are used for consistency.

Although some concerns have been raised regarding simulations performed at this relatively low horizontal resolution (Gerber et al., 2008), the use of topography in the present runs results in tropospheric NAM timescales of the order of 20 days, comparable to the real atmosphere. In contrast, the Southern Annular Modes (SAM) in the absence of topography are found to be of the order of 40 days. These timescales are further discussed below.

### 6.3 Time-mean response

#### 6.3.1 Stratospheric changes

The time-mean changes induced by weakening the radiative damping are summarized for the wave-one topography runs in Fig. 6.2 and for the wave-two topography runs in Fig. 6.3. The coarse details are quite similar for the two sets of experiments. The direct impact (and the dominant one, in the present case) of weakening the radiative damping in the lower stratosphere is to warm the poles, which are subject to adiabatic heating, and to cool the tropics, which are subject to adiabatic cooling. Panel (a) in each figure shows the control run climatological temperatures (shaded contours), and the change induced by weakening the damping rates (contour lines). The polar regions in the lower stratosphere warm, and the tropics cool, as expected. In balance with this weakened equator-to-pole temperature gradient, the polar-night jet weakens (Panel (b) in each figure). In addition to this direct response, however, the planetary-scale eddies also adjust, providing a negative dynamical feedback. The residual (TEM) circulation weakens (Panel (c)) as a result of the weakened planetary-scale wave driving (Panel (d)). While the waves will also be directly affected by the weakened damping rates, were this the dominant effect they would be expected to propagate deeper into the stratosphere. This feedback induces deeper changes in the stratosphere, warming the tropical upper stratosphere where the upwelling decreases, and cooling the high-latitude upper stratosphere where the downwelling decreases. The latter is present in both hemispheres but is much stronger
Figure 6.2: (a) Climatological zonal mean temperatures from the wave-one control simulation (contour lines) and the difference in the weakened damping run (filled contours), in degrees K. Contour lines are at an interval of 10 K. (b) As (a), but for zonal mean zonal winds. Contour lines are at intervals of 10 m s\(^{-1}\). (c) As (a), but for the TEM stream function. Contour shading and lines are logarithmically spaced between 0.1 and 1000 kg m\(^{-1}\) s\(^{-1}\). (d) Difference in EP fluxes (quivers) and their convergence (filled contours) between the weakened damping run and the control run.

in the Northern (winter) Hemisphere. The Southern Hemisphere thus provides some indication of the magnitude of the direct response.

The wave-two climatological jet is considerably stronger than the wave-one jet, and the changes in the winds induced by the weakened damping rates are more barotropic. This can be understood as a consequence of the greater tendency of the shorter wave-length waves to be refracted equatorward. The climatological eddy fluxes in the wave-two control run are weaker than those in the wave-one control run, and do not reach as high altitudes or latitudes in the stratosphere. While the overall fluxes are weaker in the weakened damping run, more flux reaches high latitudes leading to greater flux convergence,
which shifts the downwelling branch of the meridional circulation poleward (Fig. 6.3).

The tropospheric jets in both series of experiments also shift equatorward in both hemispheres. This is consistent with the response to transient stratospheric warming in the real atmosphere (Baldwin and Dunkerton, 2001), and with the response found to time-averaged diabatic forcings in other studies (Polvani and Kushner, 2002; Gerber and Polvani, 2009). This shift is considered in greater detail below; here we note simply that the shift in the Southern Hemisphere jet is stronger than that in the Northern Hemisphere in the wave-one simulations, despite the weaker temperature changes, and that the shift in the jets in the wave-two simulations is stronger than that in the wave-one simulations.

To consider these changes somewhat more quantitatively, Fig. 6.4 shows several summary statistics for these runs as a function of $\alpha_{LS}$. Confidence intervals at the 95% significance level are estimated by a modified Students’ t-test, in which the degrees of freedom are modified to take serial correlations into account. Details can be found in the appendix.
Under quasi-geostrophic scaling assumptions, the temperature anomaly in steady state is determined by the balance between radiative cooling and adiabatic heating. The direct effect of weakening the damping rates can then be computed assuming the adiabatic heating remains unchanged from the control run,

\[
\langle T - T_{rad} \rangle^t = \frac{\langle S w^* \rangle^t}{\alpha} \approx \frac{\langle S w^* \rangle^t_c}{\alpha}.
\]

Here the stratification is given by \( S = H N^2 / R \), where \( N^2 \) is the Brunt-Väisälä frequency, \( H \) is the scale height, and \( R \) is the dry gas constant. Time averages of a quantity are indicated by \( \langle \cdot \rangle^t \), and the subscript \( c \) indicates the quantity is computed from the control run.

Figure 6.4(a) shows the polar cap averaged (60° to 90° N) temperatures at 100 hPa for the wave-one simulations. The full response is indicated by the squares. The solid line shows the temperature expected if the adiabatic heating is held fixed at the control run value for this region. The polar cap temperatures for the weakest damping runs do not increase as much as expected by this direct effect as a result of the weakened overturning circulation. To give a sense of whether it is the change in circulation or the change in stratification that is driving the temperature change, the crosses show the temperatures expected if only the stratification is held fixed at the control run value \( \langle S \rangle^t_c \). At the weakest (and strongest) damping rates, most of the dynamical feedback arises from the weakened residual circulation. There is a narrow parameter regime between \( \alpha_{LS} = 0.25 \) and 0.3 for which the enhanced vertical temperature gradients induced by the warming lead to a weak positive dynamical feedback.

Similar quantities are plotted in Fig. 6.4(b) for the wave-two topography simulations. Although the overall meridional circulation decreases with weakened damping, the polar warming here in fact exceeds the direct effect as a result of the circulation moving poleward into the vortex. As with the wave-one simulations, much of the dynamical feedback is attributable to changes in the circulation, though the enhanced stratification plays a non-negligible role.

Figure 6.4(c) shows the maximum of the TEM stream function at 70 hPa for both sets of simulations. For damping rates near the value of the control run, the wave-one simulations have a stronger Brewer-Dobson circulation than do the wave-two simulations. At both extremes the opposite is true. In both cases the overturning circulation decreases with weakened damping.
Figure 6.4: (a,b) Polar cap (60° to 90° N) averaged temperatures for (a) the wave-one topography simulations and (b) the wave-two topography simulations. Full model responses are shown by the squares. The crosses show the expected response if the stratification is held fixed. The solid line shows the expected response if the adiabatic heating is held fixed. (c) Maximum in the TEM overturning stream function at 70 hPa, and (d) vertical EP fluxes from 50° to 90° N at 100 hPa, for both sets of simulations. Error bars indicate 95% confidence intervals for all quantities.

The planetary-scale EP flux at the base of the stratosphere is shown in Fig. 6.4(d). Perhaps surprisingly, the wave-two fluxes are stronger than the wave-one fluxes for all values of the damping. In both cases the fluxes decrease with weakened damping for much of the parameter regime considered, though this does not hold for the wave-one run with the strongest damping.

In summary, while the dynamical feedbacks play a significant quantitative role in determining the climatological response in the stratosphere, the dominant change to the time-averaged zonal-mean state can be understood through the direct radiative effect of the weakened lower stratospheric damping rates, which require a larger polar cap temperature anomaly to balance the dynamically driven adiabatic warming, thereby resulting in a warmed lower stratosphere.
6.3.2 Tropospheric changes

This time-mean, lower stratospheric warming is very similar to the time-mean stratospheric warming considered by Haigh et al. (2005) and subsequent related studies, although in this case the warming is induced in the context of a fully simulated stratosphere with ‘realistic’ variability. As expected, this warming induces an equatorward shift in the tropospheric jets which can be seen in Figs. 6.2 and 6.3. This shift involves both an equatorward shift in the surface winds, and an enhanced vertical shear, in balance with an increased equator to pole temperature gradient on the equatorward side of the jet, and the reverse on the poleward side of the jet.

The shift in zonal mean torques required to maintain the shift in surface winds against friction is usefully diagnosed by vertically integrating the angular momentum budget. In the terrain-following sigma coordinates used by the model, the relevant terms are (e.g. Laprise and Girard, 1990)

\[
\frac{\partial}{\partial t} \left( \frac{p_s \mathcal{L}}{a \cos \phi} \right) + \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} \left( \cos \phi \left[ p_s v \mathcal{L} \right] \right) + \frac{\phi_s}{a} \frac{\partial}{\partial \lambda} p_s + a \cos \phi \left[ p_s \mathcal{F} \right] = 0. \tag{6.2}
\]

Here \( \mathcal{L} = a \cos \phi (u + \Omega \cos \phi) \) is the angular momentum, \( \mathcal{F} = -k u_s \) is the surface friction (the sponge layer torques are neglected), \( u \) and \( v \) are the zonal and meridional winds, respectively, \( p \) is the pressure, the subscript \( s \) indicates a surface quantity, and the rest of the symbols are standard. Square brackets indicate the vertical integral \( [\cdot] = \int_1^0 (\cdot) \, d\sigma \), and overlines a zonal mean. Rewriting these in terms of the surface wind,

\[
U = L_t + M + D, \tag{6.3}
\]

where

\[
U = \frac{p_s u_s}{p_{s0}}, \quad L_t = \frac{\partial}{\partial t} \left( \frac{p_s \mathcal{L}}{a \cos \phi k \Delta \sigma p_{s0}} \right), \quad M = \frac{1}{a^2 \cos^2 \phi} \frac{\partial}{\partial \phi} \left( \frac{\cos \phi p_s v \mathcal{L}}{k \Delta \sigma p_{s0}} \right), \quad D = \frac{\phi_s}{a \cos \phi k \Delta \sigma p_{s0}}.
\]

Here \( \Delta \sigma \) is the thickness of the surface layer in the model and \( p_{s0} \) is a reference surface pressure, taken here to be 1000 hPa. The transient term is then neglected in the time mean.
CHAPTER 6. MECHANISTIC CIRCULATION MODELLING

6. Mechanistic circulation modelling

Figure 6.5: Vertically integrated, zonal mean angular momentum budget, shown in units of the surface wind (see text) for (a) the wave-one and (b) wave-two control runs and the change seen in (c) the wave-one and (d) wave-two weakened damping run. In all cases, the thick blue line shows the zonal mean surface wind, while the thin blue line is the surface wind implied by the balance. The budget is subdivided in two ways: into the total momentum flux convergence (black dashes) and the form drag (black dots), and into the synoptic-scale momentum flux convergence (red dashes) and the sum of the form drag and the planetary-scale flux convergence (red dots).

The time mean balance for the two control runs is shown in Fig. 6.5(a,b). The surface wind produced by the model is shown by the thick blue line, while the wind predicted by (6.3) is shown by the thin blue line. The form drag $D$ is a dominant component of the balance in the Northern Hemisphere. The drag is considerably stronger in the wave-two topography control run.

The net effect of the topography can be seen as the sum of the momentum flux convergence due to the orographically generated waves as well as the form drag. This separation is straightforward in the present runs as a result of the sinusoidal topography used. Assuming that planetary waves generated by wave-wave interactions between the baroclinic eddies can be neglected (supported by the absence of any such momentum flux convergence in the Southern Hemisphere balance), the momentum flux convergence...
can be decomposed into that generated by the two gravest zonal wave numbers $M_{1,2}$ and the rest, which will be dominated by the synoptic scales $M_s = M - M_{1,2}$. The net forcing by the topography is then given by $M_p = D + M_{1,2}$. These two terms are shown in red in Fig. 6.5. The synoptic scale momentum fluxes in each hemisphere in the wave-one control are quite similar; the planetary waves contribute a significant fraction of the total momentum flux convergence. This additional momentum flux convergence is, however, more than compensated for by the form drag. The wave-two topography control run exhibits considerably strengthened planetary-scale momentum fluxes, as also implied by Fig. 6.4(d). The synoptic scale fluxes are also stronger, and the sum of the two are balanced by a much enhanced form drag. The changes in synoptic scale fluxes between Figs. 6.5(a) and (b) are consistent with the change expected from the decreased upper-tropospheric temperatures in the wave-two control run (with respect to the wave-one control). Note, however, that the shift in the surface winds between these runs does not project onto the annular mode (the centre of the jet weakens and both of its flanks accelerate, not shown).

The time-averaged change induced by the weakened damping rates is shown in Fig. 6.5(c,d). In both cases and both hemispheres the response of the surface wind is dipolar, corresponding to an annular mode type equatorward shift. The synoptic scale fluxes weaken, resulting in more convergence on the equatorward flank of the jet and less on the poleward flank. This change in synoptic scale eddies is consistent with that found by other studies (e.g. Simpson et al., 2009). The synoptic scale response is nearly the same in both hemispheres in the wave-one runs, while the response (particularly on the poleward side of the jet) is considerably stronger in the Northern Hemisphere in the wave-two runs. The net effect of the topography is to counteract the shift in the synoptic scale momentum fluxes: although the planetary scale fluxes reinforce the changes induced by the synoptic scale eddies, the form drag more than compensates. This negative feedback is also stronger in the wave-two runs than it is in the wave-one runs, such that the change in the surface winds in both sets of simulations are approximately the same. The negative feedback is implied by the tendency for the tropospheric jet to be co-located with the topography, and for the annular mode decorrelation timescales to be reduced in the presence of topography (Gerber and Polvani, 2009).

Note that although the net effect of the topography is to counteract the surface wind response, the reduced planetary-scale heat fluxes must be balanced by enhanced diabatic heating or cooling; in the present setup where the diabatic heating is given by
Newtonian cooling, this implies a change in temperature, and a corresponding change in the shear. As can be seen in Figs. 6.2 and 6.3, these in fact reinforce the surface wind changes. Moreover, since the tropospheric damping rates are the same in all experiments, the larger change in fluxes in the wave-two simulations implies the larger temperature anomalies found therein.

While the stratospheric reduction in planetary wave fluxes might best be understood through Charney-Drazin type filtering as a result of the weakened zonal winds, it is less clear why the planetary scale fluxes should also weaken in the troposphere. The likely candidates are either enhanced reflection, or reduced generation. This question is considered further below.

6.4 Variability

6.4.1 Abacus plots

To analyze the stratospheric variability in these runs, two types of indices are used: that of the PJO and of the NAM. The PJO is indexed here by the first two EOFs of the polar-cap (70° to 90° N) averaged temperature anomalies. In observations (Kuroda and Kodera, 2004) and in comprehensive chemistry-climate models (Chapter 4), these modes describe deep vertical dipolar temperature anomalies. In the sign convention adopted in this thesis, the first mode describes a warming of the upper stratosphere and a coincident cooling of the mesosphere, while the second describes a warming of the lower stratosphere and a cooling in the upper stratosphere. These two modes collectively describe roughly 90% of the total polar-cap averaged (deseasonalized) temperature variance. The structure of these EOFs in a run equivalent to the wave-one topography control run has been shown in Kohma et al. (2010), and agrees well with the calculations shown here. The vertical structure of these EOFs in the control runs and weakened damping runs is shown in Fig. 6.6. Emphasis is placed on the structure up to the upper stratosphere as the details in the mesosphere will be affected by the use of Rayleigh friction as a crude gravity wave drag parameterization above 1 hPa.

Figure 6.6(a) shows the first and second modes in red and green, respectively, for the control run (solid lines) and the weakened damping run (dashed lines). As in the observations, the first mode describes upper stratospheric anomalies, while the second mode captures lower stratospheric anomalies. In the weakened damping run, the am-
The first two EOFs of polar-cap averaged temperature anomalies for the control run (solid lines) and a weakened-damping run (dashed lines) for the (a) wave-one and (b) wave-two simulations. The first EOF is shown in red, and the second in green. (c) Fraction of variance explained by the first (red) and second (green) EOF.

The amplitude of the first mode has decreased somewhat, and the amplitude of the second has increased and shifted downwards. Figure 6.6(b) shows the same EOFs for the wave-two topography. The EOFs of the control run have much weaker amplitudes in the lower stratosphere. The EOFs of the weakened damping run, however, agree well with those of the weakened damping wave-one run.

The fraction of the variance explained by the two modes as a function of the lower stratospheric damping rate is summarized in Fig. 6.6(c). The fraction explained by the first mode is shown in red for all runs, while that explained by the second is shown in green. For both sets of simulations, the fraction explained by the second EOF increases as the damping rate decreases, again consistent with a more variable lower polar stratosphere, with the second mode describing a somewhat higher fraction in the wave-one simulations than in the wave-two simulations. The total fraction of variance explained by the two modes is roughly constant across all runs.

The principal component (PC) time series $t_{s1}$ and $t_{s2}$ of these two modes are not dy-
namically independent. Trajectories in the two-dimensional phase space defined by these two modes typically rotate counter-clockwise, corresponding to downward propagation of temperature anomalies. Following Chapter 4, the trajectory can be transformed to polar coordinates $r$ and $\theta$, defined by $r^2 = ts_1^2 + ts_2^2$ and $\tan \theta = ts_2/ts_1$. The ‘abacus’ plots are then generated as before. Rather than use the PC time series computed directly from each run, the anomalies are projected on to the EOFs of the wave-one control run in order to more directly compare the behaviour in different runs, which is the primary purpose of this section. Specifically, this choice makes evident the weaker variability in the wave-two control run below.

Abacus plots for the wave-one control run and weakened damping run are shown in 6.7(a) and (b), respectively, for days 200 to 10,200. Each vertical ribbon corresponds to 400 days of model time, which runs up and to the right in these plots. Also shown (chevrons) are the dates of weak vortex events as computed from the NAM index at 10 hPa. A threshold of $-2.5\sigma$ is used here. Note that this results in a relatively constant number of events being identified in each run, despite the fact that runs with stronger damping generate substantially weaker events. This is explored further below.

The character of the variability in these two runs is immediately apparent from the abacus plots. In the control run, the vortex is much more frequently disturbed by minor warmings (apparent in red and green pulses) with timescales of about twenty days. While there are some more stable periods (following weak vortex events near days 6,700, 9,100, and 9,400, for instance), these are the exception, not the rule. In contrast, the weakened damping run shows a very regular, long timescale response to the weak vortex events, evidenced by the slowly narrowing ribbons following the chevrons which slowly shift from red to green to blue. These events are strongly reminiscent of the PJO events identified in Chapter 4. While there are some periods characterized by shorter timescale events (days 6,200 to 6,900, for instance), again the dominant behaviour is of the long timescale recoveries.

The relationship of these modes of variability to the dynamical driving of the vortex by planetary waves is also shown on these plots. The horizontal black lines indicate local maxima in the total EP flux convergence between 10 hPa and 1 hPa, from 50° to 90° N. The time series has been smoothed by a 15 day low-pass filter, and only maxima corresponding to a deceleration of more than 10 m s$^{-1}$ day$^{-1}$ are shown. These pulses drive a warming of the upper stratosphere, as can be seen by the tendency of the ribbon to widen and become more red following the pulses. They occur much more frequently
Figure 6.7: Abacus plots (see text) for wave-one (a) control run and (b) weakened damping run. Weak vortex events (as defined by the annular mode index at 10 hPa) are indicated by chevrons. Pulses of EP flux convergence are shown by the horizontal lines.

in the control run, and are clearly suppressed following the weak vortex events in the weakened damping run.

Similar abacus plots are shown for the wave-two topography runs in Fig. 6.8. In this case variability of any kind is clearly suppressed in the wave-two control run. The weak vortex events are strongly confined to the upper stratosphere (almost no positive anomalies of the second mode are apparent), and the induced anomalies do not persist for longer than about a month at most. The lack of sudden warmings in runs with wave-two topography at this horizontal resolution has been reported by Taguchi et al. (2001) and Gerber and Polvani (2009). In the weakened damping run, however, the zonal mean structure of the weak vortex events closely resembles that of the events in the wave-one weakened damping run.

Since the wave driving in the upper stratosphere is generally weaker in the wave-
two runs than in the wave-one runs, the threshold is less frequently reached and the equivalent suppression following the weak vortex events in the weakened damping run is not apparent in these plots.

### 6.4.2 Stratospheric composites

To demonstrate this suppression, Fig. 6.9 shows composites of the zonal mean zonal wind at 60° N (contour lines) and of the vertical EP fluxes from 50° to 90° N (filled contours) following weak vortex events in these four runs. The filled contours are faded where they do not differ significantly from climatology at the 95% confidence interval. The zero wind line is in bold. The effects of the weakened damping rates on the warming events are two-fold. Firstly, reduced damping allows the deceleration induced by the same eddy fluxes to descend further in the stratosphere, since the relaxational diabatic heating rates are weaker. Secondly, once the circulation is disrupted in the lower stratosphere, the anomaly can persist for longer assuming some further dynamical activity does not
Figure 6.9: Composites of weak vortex events in the (a) wave-one and (b) wave-two control runs, and in the (c) wave-one and (d) wave-two weakened damping runs. Contour lines indicate the absolute zonal mean zonal wind at 60°N at intervals of 10 m s\(^{-1}\). The zero wind line is in bold. Filled contours indicate the absolute vertical EP flux from 50° to 90°N in kg s\(^{-2}\). The contours are spaced logarithmically. For the filled contours, periods that are not statistically significant at the 95% confidence level are partially masked in white.

The vertical fluxes shown are the total fluxes, to better compare the four runs. The enhanced fluxes which trigger the event are apparent in all four runs and are somewhat weaker in the weakened damping runs and are in fact strongest in the wave-two control, consistent with the time averages seen in Fig. 6.4(c). Most importantly, the strong suppression following the events in the two weakened damping runs is clear, as the upward flux is reduced by nearly an order of magnitude in the lower stratosphere. Notably, the suppression extends all the way down to the surface, so it is not a matter of the perturb it again. Both of these effects can be seen in the wind contours in Fig. 6.9. Inspection of the 20 m s\(^{-1}\) contour shows that it descends only to 10 hPa in the wave-one control run, but it descends nearly to 100 hPa in the weakened damping run, and extends out nearly 60 days following the central date. Similar behaviour is seen in the wave-two weakened damping run, though the events in the control run are so weak that the 20 m s\(^{-1}\) contour is hard to distinguish.
waves simply being refracted equatorward. The suppression is clearly coincident with the weakened lower stratospheric winds, though the absence of a zero wind line for much of this period precludes the direct application of the Charney-Drazin criterion as an explanation. Calculations with a linear, steady-state stationary wave model (Harnik and Lindzen, 2001) forced by the same topography do suggest that the weak winds and enhanced stability in the lower stratosphere can explain roughly 50% of the reduction in upward EP flux, with the majority of this reduction arising from the reduced potential vorticity gradient (not shown). The tropospheric fluxes are not suppressed in the linear model. A more complete understanding of this suppression, though clearly desirable, is beyond the scope of the present study.

6.4.3 Tropospheric response

The wind and temperature anomalies in the lowermost stratosphere also induce an equatorward shift in the tropospheric jets, as in the time-averaged response. Composites of the tropospheric zonal wind anomalies at 300 hPa are shown (filled contours) in Fig. 6.10

![Figure 6.10](image_url)

Figure 6.10: Same as Fig. 6.9 but for the difference in zonal mean temperatures between 300 hPa and 100 hPa (contour lines; intervals of 2 K), and zonal mean zonal wind anomalies at 300 hPa (shaded contours).
for the same four runs. The difference between the temperature anomalies at 100 hPa and 300 hPa is also shown (contour lines) as a proxy for the anomalous upper tropospheric static stability. As in Fig. 6.9, the filled contours are faded where the zonal winds do not differ significantly from the climatology. The equatorward shift in the jet coincides closely with the enhanced upper tropospheric stability at high latitudes. It is much more persistent in the weakened damping runs than it is in either of the control runs. The magnitude of the shift, however, depends more strongly on the wave number of the topography than it does on the duration of the event. Moreover, there is a significant shift even in the wave-two control run events, which do induce a brief period of enhanced stability in the upper troposphere despite the very weak stratospheric signature. As expected from the stronger equatorward refraction of wave-two planetary waves, this enhanced stability is at a lower latitude than in the other runs. Note that while this strong correlation between the upper tropospheric stability and the shift in the tropospheric jet is consistent with, for instance, Simpson et al. (2009), changes in the lower stratospheric winds may also be playing an important role.

A very strong tropospheric precursor to the warmings is also seen in all four runs, in which the jets shift poleward during the period of enhanced upward EP fluxes. This has been noted in some studies (Cohen and Jones, 2011), though the effect seems especially strong in these runs. The correspondence between the tropospheric jet location and the enhanced wave fluxes does raise the possibility that the generation of planetary waves may be affected by the latitude of the jet. This would provide a mechanism for a tropospheric feedback on the stratospheric timescales: the equatorward shift of the jet induced by the upper tropospheric anomaly would in turn reduce wave generation, permitting the anomaly to persist. This feedback is clearly not the primary determinant of the stratospheric timescales given the similar magnitudes of the tropospheric shift in the control runs, but it may serve to extend them.

The terms responsible for the redistribution of angular momentum associated with the tropospheric jet shift are considered in Fig. 6.11 in more detail. Here composites of the terms in the budget described by (6.3) are shown for the two weakened damping runs, in which the tropospheric response is most persistent. Here, in order to improve the statistics, we have considered the 100,200 day runs. The shorter 10,200 day runs show a similar response, but the uncertainties are considerably larger. The terms are time averaged over the period during which the jet has shifted in the composite; lags 30 to 90 following the central date for the wave-one experiment, and lags 5 to 100 for
the wave-two experiment. Estimates of the 95% confidence interval for the surface wind response and for the two flux convergence terms are shown. The confidence interval for the surface wind response estimated by the budget is omitted for clarity; note however that despite the relatively large uncertainties, in both cases the response is well predicted by the assumption of steady state. This is consistent with the steadiness of the response seen in the composites; the transient term $L_t$ can be neglected. The jet shift is driven by a change in the synoptic scale eddies, consistent with the dynamics of the annular mode. The net effect of the form drag and planetary-scale momentum fluxes is to damp this response. The structure of the transient response shown in Fig. 6.11 is very similar to that of the time mean response shown in Fig. 6.5(c,d), which gives some confidence to this interpretation despite the large uncertainties.

**Annular mode timescales**

The vertical profile of NAM timescales is shown for a subset of the wave-one and wave-two experiments in Fig. 6.12. A 95% confidence interval is included on the weakest damping run in each case for reference; see the appendix for details of how these timescales and confidence interval are estimated. Somewhat unexpectedly, very little sensitivity to the radiative-damping timescales is seen in the stratospheric annular mode timescales in the wave-one topography runs, despite the clear difference in the character of the variability seen in Fig. 6.7 and in composites shown below. Closer inspection of the principal component time series at stratospheric levels suggests that these decorrelation timescales
correspond to different physical processes as the damping timescale is varied. For the weakest damping runs, this timescale does correspond roughly to the timescale of the large PJO-type events apparent in Fig. 6.7(b). For stronger damping runs, however, the decorrelation timescale corresponds to the typical time between large geopotential anomalies, despite the relatively rapid decorrelation of the large anomalies themselves (not shown). In contrast, the wave-two set of experiments exhibit the expected annular mode timescale profiles, which correspond closely to the radiative damping timescales in the stratosphere. The timescales saturate for the stronger damping runs, which may be consistent with the lack of sensitivity seen by CO. This difference in behaviour between the wave-one and wave-two topography runs is perhaps explained by the nearly complete repulsion of wave activity by the vortex for stronger radiative damping cases in the latter case. Note that while the timescales themselves are quite sensitive to the method by which they are computed, their sensitivity to the radiative damping is not.

The tropospheric NAM timescales in all cases are relatively realistic and are comparable to those reported by Gerber et al. (2008). They are perhaps somewhat extended in the wave-two topography weaker damping runs, but the differences are relatively small, with at least one of the weaker damping runs exhibiting similar annular mode timescales to the strongest damping run. They do not show any sensitivity to the jet location (see below).
6.4.4 Summary of transient response

The responses described in detail above are summarized in Fig. 6.13. Panel (a) shows the maximum pressure to which the 5 m s\(^{-1}\) contour descends during the weak vortex events as a function of the lower stratospheric damping timescales. The warmings descend more deeply into the lower stratosphere as the damping is weakened. The wave-one events tend to descend more deeply than do the wave-two events for all the runs except the weakest damping run, consistent with the greater tendency for wave-two planetary waves to be refracted equatorward. As the radiative timescales lengthen and the depth to which the warming descends increases, so too does the duration of the enhanced upper tropospheric static stability (Fig. 6.13(b)). This metric is, however, relatively insensitive to the weakening of the radiative damping. Coincident with this upper tropospheric enhanced stability, the tropospheric jets shift equatorward (Fig. 6.13(c)). As noted above, however, the magnitude of the shift (as opposed to its duration) is quite insensitive to the radiative damping timescales. If anything the magnitude is more strongly determined in the present runs by the wave number of the imposed topography. In contrast, the time

![Figure 6.13](image_url)

**Figure 6.13:** (a-c) Composites of the weak vortex events in all runs. (a) Maximum pressure to which the 5 m s\(^{-1}\) contour descends. (b) Duration of the enhanced upper tropospheric/lower stratospheric stability, as indicated by the difference in polar-cap averaged (70\(^\circ\) to 90\(^\circ\) N) temperature anomalies at 300 hPa and 100 hPa being greater than 2 K. (c) Shift (in degrees) of the latitude of the wind maximum at 300 hPa. (d) Time averaged position of the jet maximum at 300 hPa.
mean jet shift varies strongly with the damping timescale, shifting by upwards of two degrees poleward from the control run to the weakest damping run (Fig. 6.13(d)). It is not entirely clear why this sensitivity should be absent in the transient case, though it may simply be too weak to clearly resolve with the present statistics. One possible explanation is competing effects: while the stability changes induced by the warmings are stronger in the weaker damping runs, they are generally centred above the pole, and are thus further away from the jet (which has moved equatorward). While the amplitude of the forcing may be increasing, its projection onto the annular models may be decreasing, resulting in a response whose magnitude remains roughly constant as the jet latitude shifts.

6.5 Conclusions

In the present chapter, a parameter sweep experiment was performed by weakening the lower stratospheric damping rates while holding fixed the tropospheric configuration. The central result is that, in contrast to the results of CO, the character of the sudden warmings changes drastically as a function of the radiative damping timescales. Weakening the radiative damping results in warmings that disrupt the vortex lower in the stratosphere, and that persist for longer. Although the modified damping rates do affect the climatology of both the troposphere and the stratosphere, the changes in the stratosphere are dominated by the direct radiative response and the time mean change in the wave driving of the stratosphere is in comparison relatively weak. It is therefore unlikely that the time-mean circulation changes are primarily responsible for the changes in the variability of the wave driving. The change in character of the variability is ultimately stratospheric in origin, suggesting that the stratosphere does play an important role in mediating the waves responsible for its variability. Even if the statistics of the stratospheric waves are responding to some change in their generation by the tropospheric flow (as would be implied by the slaved-stratospheric variability paradigm of Plumb and Semeniuk (2003)) this change must still ultimately be influenced by some aspect of the stratospheric circulation, since only the stratospheric damping has been modified.

The results of this chapter suggest, therefore, that the structure of the circulation anomalies produced by the initial warming influences the subsequent evolution of the circulation in two ways. Firstly, during the extended recovery period following the deepest warmings, further dynamical forcing is strongly suppressed. The long stratospheric...
recovery is therefore predominantly radiative in character, as was found in Chapter 5. All else being equal, stronger damping would result in a more rapid return to climatological conditions. The extended recovery is, however, as much a result of the absence of eddies. The most natural hypothesis is that this absence is due to the weakened westerlies in the lower stratosphere. However, a steady, linear wave model can only explain a fraction of the suppression, primarily as a result of the weakened potential vorticity gradient, not of the winds. If the weakened potential vorticity gradient is in fact responsible, this may suggest that the reduced planetary-scale fluxes in the troposphere are due to enhanced reflection; note however that the vortex during the recovery is in a strongly absorbent configuration according to the index of Perlwitz and Harnik (2004) and Harnik (2009), with strong positive vertical shear throughout the polar stratosphere. An alternative possibility suggested by the present experiments is that the equatorward shift of the tropospheric jet induced by the warming in turn reduces the wave generation.

The second effect of the lower stratospheric warming is, as just mentioned, to induce an equatorward shift in the tropospheric jets. The duration of the shift is strongly correlated with the duration of the lower stratospheric warming above. The shift in surface winds appears to be in balance with the form drag and momentum flux convergence (consistent with rapid adjustment period seen in Figs. 6.10(c) and (d)), and is driven by a shift in the synoptic scale momentum flux convergences. The form drag and planetary scale momentum flux convergences act in the net to significantly counter the effects of the synoptic scale eddies in the vertically integrated angular momentum budget (though the planetary scale heat fluxes also act as a positive feedback on the shift in the upper tropospheric jets by amplifying the vertical wind shears). The negative feedback on the surface winds from planetary-scale topography is consistent with the well-known reduction in annular mode timescales associated with its imposition in simple models. That the topography acts to damp the response indicates that the direct downward control influence of the stratospheric wave driving is not contributing directly to the jet shift in these runs.

The spectrum of warmings generated by this set of experiments in principle provides an opportunity to test the applicability of fluctuation-dissipation type ideas to the tropospheric response. However, the magnitude (as opposed to the duration) of the composited jet shift was found to be relatively insensitive to the stratospheric damping rates. It is plausible that a detailed consideration of the ‘external’ forcing induced by the warmings and of the autocorrelative structure of the annular modes would explain this lack
of sensitivity as a result of competing effects. Such considerations are beyond the scope of this work. The emphasis on a composite analysis in the present work, however, does provide a caution against the simple interpretation of decorrelation timescales computed from annular mode autocorrelations: in the stratosphere, the correspondence between this timescale and the recovery timescale for sudden warmings is not robust (this was also noted for observations and comprehensive models in Chapter 4), while in the troposphere, the extended responses to the sudden warmings seen in the event composites do not appear to affect the decorrelation timescales. In this sense the present study agrees with the results of CO, which also found that the tropospheric time scales were unaffected, but it stands in contrast to sensitivity of the timescales to sudden warmings in a comprehensive model diagnosed by Simpson et al. (2011). Given the potential importance of the decorrelation timescales to the response of the tropospheric circulation to climate change (Kidston and Gerber, 2010), a better understanding of the relationship would seem desirable.

6.A Uncertainty estimation

6.A.1 Time-averaged quantities

Serial correlation is taken into account in the confidence interval estimation by reducing the degrees of freedom from the sample size $n$, following Zwiers and von Storch (1995). The number of effective degrees of freedom $n_e$ is computed from the autocorrelation function $\rho(t)$ of the given quantity using (their Eq. 4),

$$n_e = \frac{n}{1 + 2 \sum_{t=1}^{n-1} (1 - \frac{t}{n}) \rho(t)}.$$  \hfill (6.4)

The sum in the denominator was found to converge adequately if terms were retained up to the lag where the autocorrelation function first falls below $e^{-2}$.

6.A.2 Annular mode timescales

Annular mode timescales $\tau(p)$ were computed by performing a least-squares fit of the computed autocorrelation functions at each pressure $p$ to the following form,

$$\rho(t) = e^{-t/\tau} \cos(\alpha t) + \epsilon,$$  \hfill (6.5)
using the first 50 days of lag. Including the cosine modulation was found to significantly improve the fit over a simple exponential.

A sense of the sampling error in these timescales was obtained by computing the autocorrelation function for each non-overlapping 1000 day period of a run independently. The uncertainty in the sample mean was taken as a confidence interval for the timescale.
Chapter 7

Conclusion

To Generalize is to be an Idiot.

William Blake

7.1 Summary

The focus of this body of work has been the identification and description of Polar-night Jet Oscillation (PJO) events, an intraseasonal mode of variability present in the Arctic polar-night jet. The statistical (Chapter 4) and dynamical (Chapters 3 and 5) behaviour of this mode has been investigated in satellite observations, reanalyses, a comprehensive chemistry-climate model, and a simplified general circulation model (Chapter 6). The main results are summarized and synthesized in the next subsection, followed by a discussion of their implications for the motivating themes mentioned at the beginning of Chapter 1. The thesis concludes with a discussion of potential future directions.

7.1.1 Statistics of the PJO

The extended time-scale recoveries from sudden warmings which have here been termed PJO events were defined and identified in detail in Chapter 4. This analysis exploits a novel diagnostic tool, based on polar-cap temperature anomalies, introduced to visualize the daily variability of the Arctic stratospheric polar vortex over multiple decades. This visualization clearly illustrates the ubiquity of PJO events. These are robustly characterized by an anomalously warm lowermost stratosphere which persists for several months, as an anomalously cold middle stratosphere and an elevated stratopause form above
which both descend over this same period. This mode of variability was characterized in four datasets: temperature observations from the MLS experiment on the Aura satellite, the ERA40 and MERRA reanalyses, and an ensemble of three transient simulations from CMAM. In particular, the comprehensive chemistry-climate model was found to reproduce many features of the observed and reanalyzed datasets in detail. They can also be produced in a mechanistic model provided realistic radiative damping rates are specified in the lower stratosphere (Chapters 3 and 6).

The timescale for the recovery of the polar vortex following sudden warmings was found to correlate strongly with the depth to which the warming initially descends (Chapters 4 and 6). Indeed, this observation is the basis of the criteria used to identify PJO events outlined at the end of Section 4.2.3. They occur following roughly half of all major sudden warmings, and are associated with the extended-timescale suppression of wave-activity fluxes entering the polar vortex. They occur more frequently following vortex splits than they do vortex displacements, but the former is not a necessary condition for their occurrence; two of the most recently observed PJO events followed displacements, and experiments with the mechanistic model are capable of producing clear PJO events with both wave-one and wave-two topography.

They are also closely related to weak vortex events as identified by the annular modes, and were found to be more likely to occur the larger the amplitude of the NAM anomaly initiating the weak vortex event. Those weak vortex events that are followed by a PJO recovery are also followed in the reanalyses and comprehensive model by a more persistent equatorward shift of the tropospheric jets. This was is found to be the case in the mechanistic model, though the magnitude of the tropospheric response does not correlate with the persistence of the stratospheric anomaly.

Finally, as a result of their large amplitudes and extended timescales, natural variability in the frequency and timing of PJO events can easily produce anomalies in the climatologies of the vortex computed from records of length equal to that of the satellite record that appear significant under standard statistical tests.

7.1.2 Dynamics of the PJO

Eliassen adjustment

The zonal mean dynamics of PJO events are driven by the interaction between the torques imposed by both planetary-scale Rossby waves and small-scale gravity waves, and rad-
iative heating. That the latter can be well approximated by a local, linear relaxation to a reference temperature state was established in Chapter 3. Indeed, a linear regression model can capture more than 80% of the variance in longwave heating rates throughout most of the stratosphere and mesosphere, provided that the damping rate is allowed to vary with height, latitude and season. Additional damping due to shortwave radiation can be similarly diagnosed; while they are in general not as well represented by a linear relaxation, they contribute significantly to the net effective damping in the upper stratosphere as a result of feedbacks between temperatures and photochemical processes.

In particular, the damping rates in the Arctic stratosphere relevant to the PJO events were found to be well described by this technique, with the quantitative exception that the linearity of the relaxation breaks down near the edges of the polar vortex as a result of the large amplitude of the temperature fluctuations and the curvature of the Planck function. Notably, they are significantly longer than those commonly used in simplified modelling studies of the coupled stratosphere-troposphere system.

The character of the torques induced by the waves during PJO events were clarified by the composite analysis presented in Chapter 5. The initial warming in the middle stratosphere is caused (as is well known) by planetary-scale wave drag. The high stratopause that re-forms following the sudden warming is also initially formed by planetary-scale wave drag at mesospheric heights, while its descent follows that of the parameterized non-orographic gravity wave drag and subsequently the parameterized orographic gravity wave drag. This sequence of events is also robustly found in individual events.

The evolution of the zonal mean temperatures in the polar cap are found to be well captured by transient quasi-geostrophic Eliassen adjustment to the imposed eddy-driven torques, assuming the radiative damping timescales estimated in Chapter 3. In particular, an estimate of the transient response using a quasi-geostrophic model gave significantly more accurate estimates of the residual circulation during these events than the downward control estimates.

The persistence of the lower stratospheric anomaly was found to be radiative in nature as expected, but the analysis clarifies that its longevity is a result of three factors: the long radiative timescale, the weak climatological cooling rates, and the Eliassen adjustment to the radiative heating itself. The apparent descent of the cold anomaly was found to be a result of the vertical gradient in the radiative cooling; this was sufficient to explain the apparent descent without requiring any vertical coupling. The relevance of the lower stratospheric radiative damping to the persistence of the PJO is further evidenced by
the results of Chapter 6, which demonstrated the strong sensitivity of the stratospheric variability to these timescales.

The full thermal response to the torque induced by each category of waves (most relevantly the planetary-scale Rossby waves, parameterized orographic gravity waves, and parameterized non-orographic gravity waves) was decomposed numerically, following the analytical calculation of Haynes et al. (1991). That this calculation is capable of reproducing the full model response provides a strong confirmation that the radiative timescales diagnosed in Chapter 3 are accurate and physically relevant.

**Filtering of wave drag**

These results clearly show the relevance of the radiative processes to the persistence of PJO events. However, just as important is the implied suppression of wave drag following the initial warming. Planetary-scale waves are found to be strongly suppressed during PJO events in the reanalyses and the comprehensive model. The mechanistic model experiments in Chapter 6, in which the character of the wave driving changes significantly in response to a purely stratospheric perturbation suggest strongly that the suppression is caused by the stratospheric circulation. This is not adequately explained by the Charney-Drazin criterion, since the lower stratospheric winds are westerly (albeit weakly so) through most of the recovery phase.

Moreover, the decomposition of Chapter 5 indicates an important role for gravity waves during the recovery as well. In particular, reduced gravity wave drag was found to contribute significantly to the cold anomaly in the mid stratosphere. Moreover, the descent of the stratopause was found to be a result of strong parameterized gravity wave drag above the high polar-night jet maximum, which arises following the return of westerly winds in the lower stratosphere. The strong westerly shear in the vortex during this period permits the waves to propagate up through the jet without saturation since their amplitudes are inversely proportional to the wind speed. Above the jet maximum the winds no longer increase with height and the wave amplitudes quickly saturate. This erodes the jet from above, leading to the descent of the stratopause.

**Stratosphere-troposphere interactions**

The tropospheric response to PJO events found here in the reanalyses, the comprehensive model, and the mechanistic model is consistent with previous studies. Specifically, the
Northern Hemisphere tropospheric jet shifts equatorward during the recovery phase of the events. The duration of the shift is closely correlated with lower stratospheric anomaly in all datasets; in particular it follows the period of enhanced upper tropospheric/lower stratospheric stability. The magnitude of the shift saturates quickly during the events, and is sensitive in the mechanistic model to be sensitive only to the wavelength of the topography, not to the duration of the event. The shift in the surface winds (responsible in part for the equatorward shift of the jets) is driven by changes in the synoptic scale eddy momentum flux convergence, while it is damped significantly by the net effects of the topographic form drag and the planetary scale momentum flux convergences. The component of the jet shift associated with vertical wind shear, on the other hand, is driven roughly equally by the modified temperature gradients induced by synoptic- and planetary-scale meridional heat fluxes.

Some evidence was found to suggest that the tropospheric jet shift could also play a role in reducing the generation of planetary waves, further extending the stratospheric dynamical timescales.

The Arctic lower stratosphere

The Arctic lower stratosphere thus emerges as a critical region for the behaviour of the PJO; the radiative timescales there correlate strongly with the persistence of the PJO, which is consistent with the control it exerts over the behaviour of both the resolved and parameterized waves. This region also appears to play a critical role in influencing the synoptic scale eddies in the troposphere below.

7.2 Discussion

The understanding of PJO events that emerges from this work is of some relevance to a number of broader scientific questions. The possible implications of these events to three broad areas of current interest are discussed below. Firstly, understanding the PJO is essential for the detection of changes in the climatology of the Arctic polar stratosphere induced by some external influence and the correct attribution of those changes to their true cause. Secondly, PJO events represent a prospective source of improvements for seasonal forecasting of surface weather and climate in the Northern Hemisphere extratropics. Finally, the PJO may well be relevant to the transmission of
the effects of a variety of stratospheric forcings to the surface. Each of these topics is discussed in turn.

7.2.1 Detecting and attributing changes in the Arctic vortex

The polar Arctic stratosphere is expected to respond to changes in greenhouse gases and ozone depleting substances projected over the next century. In the longer term, there is good agreement amongst simulations of the twenty-first century by chemistry climate models that tropical upwelling in the stratosphere associated with the Brewer-Dobson circulation will strengthen (SPARC CCMVal, 2010). Where the return circulation will occur, however, is less clear; moreover, the details of the stratospheric response can influence the near-surface response to climate change (Sigmond and Scinocca, 2010). In the shorter term, there are concerns that current levels of carbon dioxide are sufficient to cool the polar vortex in dynamically undisturbed winters and lead to unprecedented chemical ozone loss (Rex et al., 2006; Manney et al., 2011). Relevant external influences are not limited to slowly changing, secular factors; as discussed in greater detail in the next subsection, the behaviour of the vortex is expected to be sensitive to natural external influences such as solar variability or volcanic eruptions, or tropical modes of variability such as the QBO or the El Niño-Southern Oscillation (ENSO). Detecting that the behaviour of the vortex is indeed systematically changing, or attributing any such changes to any one of these external influences hinges upon a solid characterization of the undisturbed climatology and variability (notionally, the climatology and variability exhibited in the presence of fixed boundary conditions). The PJO events discussed here play a significant role in this undisturbed variability, as indicated by the mechanistic model simulations in Chapter 6.

That this is a particularly challenging task, requiring long time series in order to characterize the variability has been pointed out before in the literature. Extremely long integrations with simplified general circulation models show clear non-Gaussianity and bimodality in distributions of monthly temperature averages (Taguchi and Yoden, 2002b; Nishizawa and Yoden, 2005), both of which pose difficulties for standard tests of the statistical significance of changes or trends. The near lack of any observed sudden warmings through the 1990s, followed by a relative excess in the 2000s (Charlton and Polvani, 2007), also suggests the presence of significant power in the variability at decadal timescales.
The PJO events studied here comprise a significant fraction of the variability of the Arctic vortex, both in the sense that they occur roughly every three winters, and in the root-mean-squared sense. Their behaviour differs quite fundamentally from a first order autoregressive, red-noise process of the sort that is often assumed in time series modelling, and any statistical method applied in order to reliably detect changes in the behaviour of the vortex should be sensitive to this behaviour.

Considering changes in polar cap temperatures or winds as a function of the season in order to detect differences between sample sizes of the order of the observed satellite record (roughly three decades at present), can easily be misleading, as is clearly established by Fig. 4.15. This same conclusion would hold whether considering trends (Nishizawa and Yoden, 2005), or years divided by, for example, the phase of the QBO. The difficulty associated with the non-Gaussianity of the distributions has been pointed out before, what is novel here is that the anomalies also have strong serial correlations over timescales of several months, which leads to the coherent structures seen in the anomalies in Fig. 4.15. These coherent structures can easily (and misleadingly) reinforce the interpretation of the spurious statistical results.

As an alternative, for instance, a simple binomial model was proposed for PJO occurrence frequency in Chapter 4; without invoking any serial correlation, this model is capable of describing the full decadal variability seen in the extended CMAM ensemble quite accurately. This may suggest that looking at the frequency of the occurrence of such events is a more straightforward and reliable means of detecting changes, though further work would be required to establish the statistical power of such a technique.

7.2.2 Seasonal forecasting

From the point of view of the paradigms of Arctic variability outlined in the introduction, one novel aspect suggested by the results of this work is that the most stable state (in the sense of being relatively insensitive to small perturbations) of the vortex is in fact, to draw from the language of dynamical systems, a stable orbit and not a fixed point. The vortex during PJO events is neither in a strong and cold state, nor is it disturbed and weak: rather, in the phase space of the Holton-Mass model, PJO events would correspond to the vacillatory solutions. PJO events do not persist indefinitely, but the two or three month pattern of circulation anomalies outlined in detail in Chapter 5 is regular enough that one almost does not need a model to predict the behaviour of the stratosphere once
the PJO event has begun.

While the processes driving the circulation are predominantly radiative in the stratosphere, the reason for this enhanced predictability must be the suppressed planetary waves. It is clear simply from the nature of sudden warmings themselves that the waves can amplify rapidly, and that the details of their amplification are strongly sensitive to small perturbations. This renders predictability difficult; in contrast, the radiative processes active during these events are relaxational and thus fairly insensitive to perturbations. A properly initialized forecasting model should be able to accurately predict the stratospheric behaviour.

That this predictability might extend to the surface is also supported by the strength (and persistence) of the tropospheric response seen both in Chapters 4 and 6, though predictability of the surface climate at these timescales must be probabilistic in nature. The coupling between the stratosphere and the troposphere during stratospheric sudden warmings, and the long dynamical decorrelation timescales of the stratospheric annular modes, have been argued to suggest that it may be possible to exploit the stratospheric circulation for the purposes of improving seasonal forecasting (Baldwin et al., 2003). Improvement in the forecasting of the surface climate variability has been noted upon the inclusion of stratospheric information in idealized hindcasting experiments (Douville, 2009) and in case studies of particular warming events (Kuroda, 2008; Mukougawa et al., 2009). In particular, there seem to be strong implications for the forecasting of anomalous winter conditions over Europe and North America (Bell et al., 2009). Although this skill is expected following sudden warmings, clear indications of a systematic improvement in more comprehensive studies remain elusive (Yoden, pers. comm.; Sigmond, pers. comm.). Narrowing the emphasis to forecasts begun on the central date of PJO events (as defined in Chapter 4) may provide a clearer signal.

### 7.2.3 Surface influence of stratospheric forcings

The coupling of the tropospheric circulation to the stratosphere is also of interest as a possible pathway for transmitting the influence of a variety of stratospheric forcings to the surface climate. Perhaps the most compelling example of this is the atmospheric response to changes in solar radiation associated with the 11 year solar cycle, which has motivated many previous studies of the PJO. As outlined in Section 1.2.2, the solar cycle is expected to influence the tropical upper stratosphere. Although the upper stratosphere was shown
in Chapter 5 not to be relevant to the lower stratospheric behaviour during the recovery phase of the PJO, these tropical anomalies may well influence the timing and structure of the planetary-waves which induce the initial warming. Once the lower stratosphere is disturbed, the results of Chapter 6 confirm the robustness and causal direction of the subsequent tropospheric response. The pathway for true downward influence from the upper stratosphere is thus most likely to be found in the modulation of the initial upward pulse of the planetary waves.

The implied shift of the tropospheric jets was found in response to solar variability a multiple linear regression analysis of reanalysis data by Haigh et al. (2005) (though the mechanism suggested therein was somewhat different), and the whole sequence of events has been shown in recent model simulations by Ineson et al. (2011).

Polar responses to the phase of the QBO and to stratospheric aerosols may similarly be transmitted downward. Moreover, studies have also argued that the effects of ENSO on the Northern Hemisphere extratropics may be modulated in part by the state of the stratosphere (Ineson and Scaife, 2009; Bell et al., 2009), though it is less clear that this modulation acts through large amplitude events like sudden warmings or the PJO (Butler and Polvani, 2011).

A better understanding of the PJO may then permit both improved detection of this influence, and a better understanding of the mechanisms by which the influence is transmitted to the surface. The non-Gaussianity and temporal correlations discussed in the previous subsection will also have implications for detecting these indirect influences on the tropospheric circulation. Detection of the relevance of a forcing acting through this pathway would first require clear evidence that the occurrence statistics of the stratospheric events have in fact responded to the initial perturbation.

7.3 Future work

As discussed at several points in this work, the suppression of planetary-scale vertical EP fluxes during PJO events is necessary for their extended timescales. Moreover, it is reasonable to assume that this period of suppression is a source of hope for efforts to find improved forecasting skill at seasonal timescales from the stratosphere. The most pressing theoretical question raised by this work would thus be: why are the waves suppressed? An argument was made in Chapter 6 that this suppression could not fully be explained by linear theory; and that perhaps the tropospheric response was acting
as a negative feedback on the waves, further reducing their generation. A bottom-up approach to this problem would be to attempt to apply more sophisticated wave models to try and reproduce the EP fluxes produced by the fully non-linear dynamical model.

Alternatively, spectrally-selective nudging experiments of the type used by Simpson et al. (2011) might be useful for constraining particular aspects of the fully non-linear flow in order to isolate the key determinants of the waves. As a first step, simulations in which the stratospheric zonal-mean state is nudged towards that of a previously simulated PJO event would help to clarify whether it is in fact the zonal mean state of the stratosphere that is responsible for the reduction in wave activity.

Following a more pragmatic path, the hypothesis that these events represent periods of highly enhanced predictability should be tested explicitly. The question will be whether forecasts spun off from dates following the initial warming period of a PJO event do in general show improved seasonal predictability. There is an opportunity to test this in the context of realistic historical forecasts; Drs. Michael Sigmond and John Scinocca are involved in performing a set of such experiments to test whether any such improvement can be found over the satellite record (1979 to present). There is interest in performing a similar set of forecasts started from dates derived from the abacus plots presented in Chapter 4. These would hopefully supply just such a direct test.

Finally, while the question of radiative damping timescales is perhaps a relatively minor one, it would be interesting to continue pursuing it. In particular I am interested in quantifying the damping rates using observed temperatures and a more realistic radiative transfer code. The upper troposphere/lower stratosphere is an extremely important region radiatively, chemically and dynamically, and while it may be difficult to quantify the sensitivity of heating rates there to perturbations in temperatures or in radiatively active gases, the fact that the timescales are so long implies that the responses to external forcings can be large. It would therefore be useful to improve the constraints on these sensitivities.


