Interannual and Decadal Variations of Planetary Wave Activity, Stratospheric Cooling, and Northern Hemisphere Annular Mode

YONGYUN HU AND KAI KIT TUNG

Department of Applied Mathematics, University of Washington, Seattle, Washington

(Manuscript received 20 August 2001, in final form 3 December 2001)

ABSTRACT

Using NCEP–NCAR 51-yr reanalysis data, the interannual and decadal variations of planetary wave activity and its relationship to stratospheric cooling, and the Northern Hemisphere Annular mode (NAM), are studied. It is found that winter stratospheric polar temperature is highly correlated on a year-to-year basis with the Eliassen–Palm (E–P) wave flux from the troposphere, implying a dynamical control of the former by the latter, as often suggested. Greater (lower) wave activity from the troposphere implies larger (smaller) poleward heat flux into the polar region, which leads to warmer (colder) polar temperature. A similar highly correlated antiphase relationship holds for E–P flux divergence and the strength of the polar vortex in the stratosphere. It is tempting to extrapolate these relationships found for interannual timescales to explain the recent stratospheric polar cooling trend in the past few decades as caused by decreased wave activity in the polar region. This speculation is not supported by the data. On timescales of decades the cooling trend is not correlated with the trend in planetary wave activity. In fact, it is found that planetary wave amplitude, E–P flux, and E–P flux convergence all show little statistical evidence of decrease in the past 51 yr, while the stratosphere is experiencing a cooling trend and the NAM index has a positive trend during the past 30 yr. This suggests that the trends in the winter polar temperature and the NAM index can reasonably be attributed to the radiative cooling of the stratosphere, due possibly to increasing greenhouse gases and ozone depletion. It is further shown that the positive trend of the NAM index in the past few decades is not through the inhibition of upward planetary wave propagation from the troposphere to the stratosphere, as previously suggested.

1. Introduction

There has been some evidence indicating that the stratosphere has been steadily cooling during the past few decades (Randel and Wu 1999). Observations show that the Arctic polar vortex has become colder and stronger (Labitzke and Naujokat 2000), that it persists longer (Newman et al. 1997; Waugh et al. 1999), and that there have been fewer major sudden warmings. [There were only two major warmings from 1990 to 2000 (1990/91 and 1998/99; Labitzke and Naujokat 2000).] The strengthening of the polar vortex is consistent with the trend of the Northern Annular Mode (NAM) toward the high-index phase since 1968 (Thompson and Wallace 1998).

It is well established that the zonal-mean temperature in the winter polar region of the stratosphere is determined by the balance of two factors: radiative cooling and dynamical heating (Andrews et al. 1987). The latter is caused by downwelling in the stratospheric polar region of a global-scale wave-driven meridional circulation, and thus the magnitude of this heating depends on planetary wave activity generated in the troposphere. This dynamical heating is responsible for forcing the winter polar temperatures to be above radiative equilibrium during the polar night. Fusco and Salby (1999) and Salby et al. (2000) found that on interannual timescales stratospheric ozone and temperature in the Arctic polar region in winter is regulated by the upward Eliassen–Palm (E–P) flux across the tropopause, and that the two have a strong correlation. A natural question is: Is the long-term stratospheric cooling similarly caused by a long-term decrease of planetary wave activity in the stratosphere?

Based on their climate model simulations of the doubling CO₂ scenario, Rind et al. (1998) and Shindell et al. (1998, 1999), speculated that the planetary wave activity from the troposphere to the stratosphere might have declined due to the increasing greenhouse gases in the atmosphere. They proposed a plausible feedback mechanism responsible for a decrease of planetary wave activity in the stratosphere: Both increasing greenhouse gases in the atmosphere and ozone depletion in the stratosphere would enhance the meridional temperature gradient, with warming in the troposphere and cooling in the stratosphere, in the tropopause region in the sub-
tropics, where the tropopause is higher on the equatorial side than on the poleward side. Such an enhanced north–south temperature gradient can lead to a stronger zonal-mean wind, which presumably impedes upward propagation of planetary waves from the troposphere into the stratosphere. Note that the vertical shear, which is actually the quantity that is supposedly enhanced by the north–south temperature gradient in this mechanism, was not mentioned. In an earlier work, Chen and Robinson (1992) found in their linear model simulations that vertical wind shear near the tropopause is critical for controlling the passage of planetary waves from the troposphere to the stratosphere. They showed that a weaker vertical wind shear tends to enhance upward planetary wave propagation across the tropopause. As we will discuss later, different authors dealt with vertical wind shears at different latitude bands and formed very different conclusions.

Recent studies on the so-called Arctic Oscillation or NAM provide another view on the possible relationship between planetary wave activity and the stratospheric cooling. A recent survey linking NAM, planetary wave activity, greenhouse effects, and climate change is found in Hartmann et al. (2000), who emphasized that NAM may be an internal mode of natural variability of the atmosphere, acting on decadal as well as on interannual timescales. This internal mode is characterized by its deeply vertical coherent structure extending from the surface to the stratosphere, irregular oscillation on broad timescales, and positive trend toward the high-index phase since 1968 (Thompson and Wallace 1998; Thompson et al. 2000; Wallace 2000). Using observational data and general circulation model (GCM) output, Limpasuvan and Hartmann (1999, 2000) found that the NAM oscillation between high and low index phases is a result of the internal coupling between the zonal flow and the planetary waves. They showed that when data were composited according to high and low NAM indices, at the tropopause-level planetary waves are refracted away from the Arctic polar region in the high-index phase, whereas they are more readily focused into the polar waveguide in the low-index phase, where they decelerate the polar jet. The authors attributed the difference to the vertical shear and speculated that the trend toward the high NAM index in the past three decades may be related to greenhouse warming through the mechanism of Shindell et al. The latitude band, which is important for controlling the different vertical propagation properties of planetary waves in the two different phases of NAM, was identified to be the subpolar latitudes of 60°–80°N. This region however is too far north of the latitude band of increasing vertical shear for Shindell et al.’s mechanism to work.

Our purpose in the present paper is twofold. First, we inquire whether planetary wave activity from the troposphere into the stratosphere has changed systematically in the recent few decades as suggested above. Second, we are concerned with the relationship between wave-driven dynamical heating and the long-term stratospheric cooling trend. We focus attention on the trends in January-mean temperature in Arctic polar region and the wave-driven dynamical heating during November–January (NDJ). This allows us to study the dynamical heating separate from the ozone heating effect, which becomes more important in late winter. In addition to carrying out data analysis, we also provide theoretical justification on how wave forcing controls polar temperatures, NAM, and mean angular momentum, and on why these January-mean quantities are determined by cumulative wave forcing during preceding and current months (NDJ).

2. Data

The data used in this study are the reanalysis data from the National Center for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR). Wavenumbers 1 and 2 are Fourier decomposed from 51-yr (Nov 1949–Jan 2000) geopotential heights. To avoid dealing with the boundary problems of the NCEP–NCAR reanalysis model, our analysis does not go above 20 mb. The daily data of 41-yr (1958–98) NAM index at 50 mb, computed from the NCEP–NCAR data, were provided by M. Baldwin. Detailed information of the NAM index can be found in Baldwin and Dunkerton (1999) and Baldwin and Dunkerton (2001). There are three time lines that must be noted in linear trend statistics. The reanalysis data at stratospheric levels before 1958 are probably not reliable due to the lack of sufficient observational data in the upper atmosphere. In many recent works, 1968 is the year that is considered the starting year for stratospheric cooling and for the positive trend in NAM index. Satellite data have been included in generating the reanalysis data since 1979. When put into the perspective of a longer time series, it becomes apparent to us that the starting year (namely 1968) chosen by some of the previous authors played an important role in the linear trends that were found and reported. The statistical test for significance should have to be different if one has the additional degree of freedom to choose the starting year for a trend. Nevertheless, this problem does not impact our results because our goal is to reconcile the self-consistency of trends in wave activity and in mean flow, in whatever time period these trends are found. For example, self-consistency requires that if a significant trend in NAM is found for the period 1968–2000 and if this trend is speculated to be caused by wave forcing, then there should be a trend in the latter for the same period.

3. Trends of planetary wave activity

a. Planetary wave amplitudes

We first examine the long-term variations of planetary wave amplitudes in the stratosphere. Figure 1 illustrates
the variability of wavenumber-1 amplitudes, averaged over NDJ, at 200, 100, 50, and 20 mb along the 60°N lat circle (amplitudes and E–P fluxes averaged over Dec–Feb are similar to that over NDJ). Seasonal averages (NDJ) are more appropriate for the study of interannual and decadal variations. On interannual timescales, wavenumber-1 amplitude varies substantially, and the variations at the four levels are vertically coherent. For periods of 1968–2000 and 1979–2000, wave amplitudes at all four levels do not show significant trends, though wave amplitudes increase at 20 mb and slightly decrease at 200 mb. Additional calculations show that there are no significant trends in wavenumber-1 amplitudes at levels below 200 mb over any period. NDJ-mean wavenumber-2 amplitudes are shown in Fig. 2. The amplitudes at all the four levels exhibit no statistically significant trends in any of the three periods: 1958–2000, 1968–2000, and 1979–2000. An exception seems to be wavenumber-1 amplitudes from 1958 to 2000. Over this period, the linear trends in wave amplitudes at 20 and 50 mb are about 2.18 and 1.45 m yr⁻¹, at significance levels above 97.5% and 99%, respectively. The trends vary with latitudes, with the largest slope near 60°N.

One can conclude that in the NCEP–NCAR data there is no evidence for a decrease in planetary wave amplitudes in the lower stratosphere in the past 51 or 43 yr. The increase in wavenumber-1 amplitudes at some levels from 1958 to 2000 may not be reliable.

b. Eliassen–Palm fluxes

In this subsection, we study the evolution of the E–P flux from the troposphere to the stratosphere in the past few decades since it is the E–P flux and its convergence that measure the overall irreversible wave driving of residual meridional circulation and dynamical heating in the stratosphere.

Following Dunkerton and Baldwin (1991) and Salby et al. (2000), we first define a box from 50° to 90°N latitude and from 100 to 20 mb in height. The size of the box should be consistent with the region over which

![Graph](image-url)

**Fig. 1.** NDJ mean wavenumber-1 amplitudes at four levels at 60°N.

![Graph](image-url)

**Fig. 2.** Same as Fig. 1 except for wavenumber 2.
the zonal mean quantities are to be averaged (see later). For the space covered by this box, E−P flux across 100 mb ($F_{100mb}$) measures the overall wave activity coming from the troposphere, E−P flux across 20 mb ($F_{20mb}$) is the amount going into the upper stratosphere, and E−P flux across the boundary of 50°N ($F_{50N}$) is the meridional flux out of the box. The combination of the three components, that is, $F_{net} = F_{20mb} - F_{100mb} - F_{50N}$ is the “net” E−P flux out of this box, which, by virtue of the divergence theorem, represents the total E−P flux divergence averaged over the box. In each year, E−P flux is averaged over 3 months (NDJ). In calculating E−P flux, the quasigeostrophic version of E−P flux in spherical geometry is used (Edmon et al. 1980; Andrews et al. 1987). Figure 3 shows $F_{100mb}$, $F_{20mb}$, $F_{50N}$, and $F_{net}$ for the 51 yr. Similar to the results by Fusco and Salby (1999) and Salby et al. (2000), the upward E−P flux across 100 mb has large interannual variabilities. Comparison of $F_{100mb}$ with $F_{20mb}$ suggests that about half of the upward E−P flux penetrates the top of the box and goes into the upper stratosphere. Negative net E−P flux means that planetary wave activity always tends to decelerate the westerly mean flow. The most important point to note here is that the plots do not exhibit evidence of either decreasing E−P flux from the troposphere or decreasing E−P flux divergence in the stratosphere in the past 51 yr. This is consistent with the part of the results of Zhou et al. (2001) for early winter.

c. Vertical wind shear in NCEP–NCAR data

The mechanism of Shindell et al (1998) would have led to a declining wave propagation from the troposphere into the stratosphere over the past few decades, which is not the case in the NCEP–NCAR data as reported in the previous section. It turns out that in the NCEP–NCAR data for the past few decades there is indeed a trend toward larger meridional temperature gradients, and hence larger vertical shears, near the tropopause region, but this occurs in the data in the subtropics rather than the midlatitudes in their double CO$_2$ experiment. The subtropical tropopause is not the gateway controlling vertical propagation of planetary waves into the polar waveguide in the stratosphere.

Figure 4a illustrates the vertical wind shears of zonal-mean zonal wind in the upper troposphere (250 mb), averaged over NDJ, as a function of years. This level is the one considered by Rind et al. (1998), Shindell et al. (1998), and Shindell et al. (1999). At this level the Tropics is warm, while middle and high latitudes are cold. Because of the strong meridional temperature gradient between 30° and 40°N, the vertical shear at 35°N is about three times larger than the vertical shears at 45°, 55°, and 65°N. The shear at 35°N has a positive trend from 1958 to 2000, with a slope of 0.004 m s$^{-1}$ km$^{-1}$ yr$^{-1}$ and significance above 99.99%. From 1968 to 2000, the trend is also about 0.004 m s$^{-1}$ km$^{-1}$ yr$^{-1}$, with significance above 95%. From 1979 to 2000, the shear is increasing slightly, but the trend is not significant. This appears to be consistent with the suggestion by Shindell et al. (1998), Limpasuvan and Hartmann (2000), and Hartmann et al. (2000), as far as the effect of global warming on the vertical shear is concerned. Vertical shears at 45°N, 55°N, and 65°N do not show any significant trend. These latitude bands are more important for the propagation of planetary waves from the troposphere into the polar stratospheric waveguide. The lack of long-term trends in these regions is consistent with our finding of no significant trends of wave activity in the stratosphere.

Figure 4b shows the vertical wind shear near 100 mb. Shears at 35° and 45°N show negative values because the Tropics becomes cold, while middle latitudes are relatively warm. This is the case studied by Chen and Robinson (1992), although their imposed shear anomaly covers a broader region (more poleward) than that in
the data. The vertical shear at 35°N has significant trends from 1958 to 2000 and from 1968 to 2000, with slopes of about 0.01 m s\(^{-1}\) km\(^{-1}\) yr\(^{-1}\) and 0.008 m s\(^{-1}\) km\(^{-1}\) yr\(^{-1}\) and significance above 99.5%. From 1979 to 2000, the shear is slightly increasing, but not significant. Vertical shears at 45°, 55°, and 65°N do not show any significant trend.

d. On the refraction index argument

It is apparent that there has been a great deal of confusion over the effect of increasing north–south temperature gradient (and, by the thermal wind relationship, hence vertical shear) on the vertical propagation of planetary waves into the stratosphere. The mean zonal wind and its vertical shear have opposite effects on the index of refraction. Rind et al. (1998) and Shindell et al. (1998) focused on zonal wind magnitudes when it is the vertical shear that is directly related to the expected increase in the north–south temperature gradient. Limpasuvan and Hartmann (2000) and Hartmann et al. (2000) attributed the smaller index of refraction of the high NAM phase to its larger vertical shear, which they thought impedes vertical wave propagation, but the effect is likely due to the larger zonal wind magnitude in the high-index phase. Positive vertical shears actually increase the index of refraction and enhance upward wave propagation.

In the original theory of Charney and Drazin (1961), a uniform zonal-mean wind was used in the calculation of the index of refraction. It led to the conclusion that the vertical propagation of stationary planetary waves are impeded by strong westerly zonal wind and that no stationary planetary wave can propagate in a mean zonal wind stronger than 38 m s\(^{-1}\). However, the largest wave-number-1 amplitudes are often found near the base of the winter stratospheric jet, where the zonal-mean wind exceeds 60 m s\(^{-1}\). We now know that the shears of the zonal wind play an important role in some regions for ducting the waves.

The square of the index of refraction, which characterizes the propensity for stationary planetary wave

---

**Fig. 4.** Vertical shears of zonal-mean zonal wind, averaged over NDJ, at (a) 250 and (b) 100 mb. Each plot shows wind shears at four latitudes: 35°, 45°, 55°, and 65°N. The shears at 250 mb are averaged from the vertical shears between 200 and 250 mb and that between 250 and 300 mb. The shears at 100 mb are averaged from the shears between 70 and 100 mb and 150 and 100 mb.
propagation, is given (see Andrews et al. 1987), in spherical quasigeostrophic form, by

\[ n_z^2(y, z) = \frac{\bar{\Omega}_\phi}{\bar{\Omega}} - \left( \frac{k}{a \cos \phi} \right)^2 - \left( \frac{f}{2NH} \right)^2, \tag{1} \]

where

\[ \bar{\Omega}_\phi = \frac{2\Omega}{a \cos \phi} - \frac{1}{a^2} \left( \frac{\bar{\Omega} \cos \phi}{\bar{\Omega}} \right)_\phi - \frac{f^2}{\rho_0} \left( \frac{\rho_0}{N^2} \right)_{\phi}. \tag{2} \]

is the meridional potential vorticity (PV) gradient. Here, \( k, N, H, f, \rho_0, a, \Omega, \) and \( \phi \) denote the zonal wavenumber, buoyancy frequency, scale height, Coriolis parameter, background air density, earth’s radius, earth’s rotation frequency, and latitude, respectively. It is expected that planetary waves are able to propagate in regions where \( n_z^2 > 0 \) and are refracted from regions where \( n_z^2 < 0 \), and that the larger the \( n_z^2 \) in a region, the easier it is for planetary waves to propagate there (Matsuno 1970).

Expansion of the third term on the right-hand side of (2) yields

\[ -\frac{f^2}{\rho_0} \left( \frac{\rho_0}{N^2} \right)_{\phi} = \left( \frac{f^2}{HN^2} + \frac{f^2}{N^2} \frac{\partial N}{\partial z} \right) n_z - \frac{f^2}{N^2} n_{\phi}. \tag{3} \]

Here, we have applied \( \rho_0(z) = \rho_0 \exp(-z/H) \). The first term on the right-hand side of (3), involving \( \pi_{\phi} \), is the dominant one, except near the jet maximum. The quantity multiplying \( \bar{\Omega}_\phi \) is generally positive above 5 km, a region of interest to us. Substituting this into (2) and combining with (1), one can find that a larger (positive) vertical wind shear \( \pi_z \) should enhance the meridional PV gradient and lead to a larger \( n_z^2 \), thus tending to enhance wave propagation, rather than impede them (Tung and Lindzen 1979). This interpretation is opposite to that by Limpasuvan and Hartmann (2000), due possibly to a sign error.

The work by Chen and Robinson (1992) dealt with negative vertical wind shears near the tropopause. When a negative shear anomaly is superimposed on their standard wind profile, which makes the vertical shear near the tropopause more negative, they found that E–P flux convergence was reduced by as much as 20% in the stratosphere in high latitudes and as much as 40% in the upper tropospheric polar region. As far as the vertical wind shear is concerned, Chen and Robinson’s conclusion, that less negative vertical wind shear enhances wave propagation across the tropopause, is consistent with ours.

From the thermal wind relation, \( \partial \pi_{\phi}/\partial z = -(R/H f) (\partial T/\partial y) \), it then follows that an increase in the north–south temperature gradient \( -\partial T/\partial y \) in the upper troposphere, as noticed by Rind et al. (1998) and Shindell et al. (1998, 1999) in their simulations, should lead to both larger \( \pi_{\phi} \) and \( \pi_z \), and possibly larger \( \pi_{\phi} \). To distinguish the relative importance of these terms in (1) and (2) and to quantify the difference of \( n_z^2 \) between low and high NAM phases, we calculate \( n_z^2 \) for stationary wavenumber 1 using Jan-mean, zonal-mean zonal winds and temperatures in NCEP–NCAR data for two years, 1987 and 1989, chosen to represent the low and high NAM phases, respectively.

Figures 5a and 5b show the Jan-mean zonal-mean winds in the two years. The plots from the two single years are similar to the composite plots in Fig. 3 in Hartmann et al. (2000). The low-phase year (Fig. 5b) has a stronger subtropical tropospheric jet but weaker zonal winds at high latitudes in the troposphere and stratosphere as compared to the high-phase year (Fig. 5a). Plots of refraction indices (Figs. 5c and 5d) in the two single years are also similar to the composite plots [see Fig. 8 in Limpasuvan and Hartmann (2000)]. In both years, \( n_z^2 \) maxima of interest are located around the subtropical tropopause (30°–40°N) when \( \pi_{\phi} \) is taken away. This says that the maxima of \( n_z^2 \) in Fig. 5e and 5f near the polar region are caused by the weak \( \pi_{\phi} \), rather than by \( \pi_z \), which is rather weak in that region. This is different from the interpretation by Limpasuvan and Hartmann (2000), who inferred that the difference of \( n_z^2 \) between low and high NAM phases is a result of the difference of \( \pi_{\phi} \), near the subtropical tropopause.

To distinguish the relative importance of the term involving \( \pi_{\phi} \) from other terms, we plot \( a^2 \bar{\Omega}_\phi \pi_{\phi} \) instead of \( -\partial T/\partial y \) (as shown in the main text), which is more important than the con-
tributions from terms involving $\pi_z$ and meridional derivatives.

In conclusion, while it appears that the observed $\pi_z$ trend is consistent with the speculation of enhanced north–south temperature gradient due to global warming, the effects of $\pi$ and its vertical derivatives on planetary wave propagation into the stratosphere is not as previously proposed. First, such increases of $\pi_z$ and $\pi$ occur only in the subtropics, which is not the main gateway for wave propagation into the stratosphere. Second, if the greenhouse effect is sufficiently strong so that the tropospheric jet is moved from the subtropics to higher latitudes, such as that in the doubling CO$_2$ simulation by Rind et al. (1998), the enhanced vertical wind shears (not $\pi_z$) at midlatitudes would lead to an enhancement of vertical wave propagation rather than suppression. Indeed, Rind et al. (1998) observed an increased wave activity in the low stratosphere in their simulations. However, unlike the case in models, the NCEP–NCAR reanalysis data do not show significant trends in $\pi$, or $\pi$ at middle and high latitudes over the past few decades. Therefore, the refraction index argument is seen to be consistent with our result that planetary wave activity from the troposphere into the stratosphere has not declined in the past few decades.

4. Relation between dynamical heating and stratospheric cooling

a. Theoretical considerations

Before we present more observational results, it is necessary to show theoretically the relationship between planetary wave fluxes and the dynamical heating of the zonal-mean state.

Under the quasigeostrophic approximation applicable to the extratropical region, the zonal-mean temperature equation is (Andrews et al. 1987, p. 129)

$$\frac{\partial \theta}{\partial t} = -\theta_0 w^* + \overline{Q},$$

where $\theta = T(p_o/p)^{\kappa_c}$ is the potential temperature, $\theta_0 = d\theta_o/dz$ is the vertical gradient of the background state of the potential temperature, $\overline{\psi}$ denotes the zonal mean, $w^*$ is the vertical velocity component of the transformed Eulerian mean circulation, and $\overline{Q}$ is the zonal-mean net radiative heating. The first term on the right-hand side of Eq. (4) is the so-called dynamical heating induced by adiabatic compression of descending air in the polar region. It is related to the residual meridional motion through the continuity equation

$$\frac{\partial}{\partial y} (\rho_o w^* \cos \phi) + \frac{\partial}{\partial z} (\rho_o w^*) = 0,$$

with $y = a \sin \phi$.

The nondivergent nature of the meridional circulation (5) implies the existence of a streamfunction $\Psi$ such that

$$\rho_o w^* \cos \phi = -\frac{\partial \Psi}{\partial z}, \quad \rho_o w^* = \frac{\partial \Psi}{\partial y}. \quad (6)$$

Let $\langle \cdots \rangle$ denote aare weighted meridional average from a latitude $\phi$ to the pole; that is,

$$\langle A \rangle = \frac{\int_{-a}^a A \cos \phi \, d\phi}{\int_{-a}^a (1 - \sin \phi) \, d\phi} = \frac{\int_{-a}^a A \cos \phi \, d\phi}{(1 - \sin \phi)} \quad (7)$$

Equation (4) becomes, assuming that $\Psi$ vanishes at the pole,

$$\frac{\partial}{\partial t} \langle \rho_o \theta \rangle = \frac{\theta_0}{a(1 - \sin \phi)} + \langle \rho_o \overline{Q} \rangle. \quad (8)$$

The quasigeostrophic form of the zonal momentum equation is

$$\frac{\partial \vec{w}}{\partial t} - f \vec{w}^* = \frac{1}{\rho_o(z)} \nabla \cdot \vec{F},$$

where $\nabla \cdot \vec{F} = \partial \vec{w}/\partial y (\rho_o w^* \cos \phi) + \partial \vec{w}/\partial z (\rho_o f \delta \theta^*/\partial \phi)$ is the Eliassen–Palm flux divergence, which embodies the irreversible, net effect of wave activity on the mean flow. As pointed out by Newman et al. (2001), over monthly or seasonal timescales $\partial \vec{w}/\partial t$ is about two orders smaller than the other terms, and the dominant balance is therefore between the Coriolis torque and the wave driving

$$-f \rho_o \vec{w}^* = \nabla \cdot \vec{F}, \quad (10)$$

which is the same as

$$\frac{f}{\cos \phi} \frac{\partial \Psi}{\partial z} = \nabla \cdot \vec{F}. \quad (11)$$

Integrating (11) with respect to $z$ yields

$$\Psi = \frac{\rho_o}{\theta_0} \vec{w}^* \theta^* \cos \phi + \delta, \quad (12)$$

where $\delta = -(1/\alpha) (\alpha/\delta \psi) \int_{-a}^a \rho_o w^* \cos \phi \, dz$. Figure 6 shows the 3-month mean of the two quantities on the right-hand side of (12). The quantity, $\delta$, which is determined by the meridional derivative of momentum flux at latitude $\phi$, is about 2 orders of magnitude smaller than the heat flux term in the lower stratosphere. Equation (8) then becomes, to a high degree of accuracy,

$$\frac{\partial}{\partial t} \langle \theta \rangle = \vec{v} \cdot \vec{F} \cdot \cos \phi \right|_{\phi} + \langle \overline{Q} \rangle. \quad (13)$$

We see that at level $z$ the dynamical heating for the mean temperature averaged over an area from latitude $\phi$ to the north pole is given by the poleward heat flux $\vec{v} \cdot \vec{F} \cdot \cos \phi$ at latitude $\phi$. Note that the dynamical heating is neither directly given by $E-P$ flux divergence, $\nabla \cdot \vec{F}$; nor by latitudinally integrated upward $E-P$ flux across the tropopause (Fusco and Salby 1999), $F_{100mb}$; nor by
the momentum flux (Limpasuvan and Hartmann 2000). The derivation can also be given from the Eulerian mean temperature equation with perhaps a little more ease. Newman et al. (2001) derived a similar result from a different starting point. Note that the wave forcing drives the mean temperature tendency, not the mean temperature itself.

Integrating (13) from \( t_1 \) to \( t_2 \) gives

\[
\left( \overline{\theta} \right)_{t_2} - \left( \overline{\theta} \right)_{t_1} = \int_{t_1}^{t_2} F_{\phi,z} \, dt + \int_{t_1}^{t_2} \left( \overline{Q} \right) \, dt, \tag{14}
\]

where

\[
F_{\phi,z} = \frac{\sqrt{\theta^f \cos \phi}}{a(1 - \sin \phi)}, \tag{15}
\]

denotes the poleward heat flux at latitude \( \phi \) and \( z \). Equation (14) states that over any time period \( t_2 - t_1 \), \( \left( \overline{\theta} \right)_{t_2} \) (or more precisely, the difference between it and \( \left( \overline{\theta} \right)_{t_1} \)) is determined by the cumulative wave-driven heating during current and preceding periods, not just over the contemporaneous one. This was qualitatively addressed by Salby et al. (2000) and demonstrated by Newman et al. (2001). We further average \( \left( \overline{\theta} \right)_{t_2} \) and \( \left( \overline{\theta} \right)_{t_1} \) over \( t_2 \) and over \( t_1 \) to yield monthly mean \( \left( \overline{\theta} \right)_{Jan} \) and \( \left( \overline{\theta} \right)_{Nov} \). Then, the corresponding dynamical heating should be the accumulation of monthly running-time mean of \( F_{\phi,z} \) from 1 Nov to 31 Jan. To conform to common practice, we shall use the NDJ mean of \( F_{\phi,z} \), instead of the cumulative running-time mean, there being no significant difference between the two forms of time average.

b. Dynamical heating versus temperature

Figure 7 illustrates Jan-mean polar temperatures at five levels: 100, 70, 50, 30, and 20 mb, area-weighted over 60°–90°N, over the past 51 yr. The temperatures are vertically coherent and exhibit large interannual variations, ranging from about 200 to about 220 K. On decadal timescales, the mean temperatures at all the levels show significant negative trends from 1968 to 2000. At 50 mb, the trend is about \(-0.17 K yr^{-1}\) at significance level above 97.5%. The polar mean temperature at 50 mb reaches a minimum of 201 K in Jan 2000. The net decrease of the polar mean temperature is about 5.6 K in the past 33 yr. This result is consistent with the stratospheric cooling trend for winter–spring of \(-5 K \) (19 yr)\(^{-1}\) reported by Pawson et al. (1998), Randel and Wu (1999), and Thompson et al. (2000) from the Microwave Sounding Unit channel-4 (MSU-4) data. Randel and Wu (1999) in particular found a \(-4 to -8 K\) Arctic cooling during Jan–Apr (peaking in spring) since 1985. The smaller cooling trend in our result is probably because ozone depletion is not yet a large factor for Arctic cooling during polar night (e.g., Jan), unlike the case after final warming. Note that the significance of the trend is sensitive to the starting year chosen because of the large interannual variability. For example, if one chooses the starting year after 1970, one will not find a significant trend.

In the summer season, the stratosphere is dynamically less disturbed, and hence has temperatures close to radiative equilibrium. Therefore, any temperature trend induced by radiative cooling should be more readily detected from summer temperatures. To verify whether a cooling trend exists in the Arctic polar region, we have calculated Jul-mean temperatures, averaged over the same polar region, at the same five levels. The results are plotted in Fig. 8. Compared to the Jan-mean temperatures, the interannual variabilities of the Jul-mean temperatures are much smaller, within 3 K at these levels. The temperatures first increase, reaching a maximum in Jul 1968, then monotonically decrease to year 2000 at 100, 70, and 50 mb. As marked in the figure, the trends at these levels from 1979 to 2000 are about
Fig. 7. Jan-mean temperatures at 100, 70, 50, 30, and 20 mb, area-weighted over 60°–90°N, as a function of years. Legends are marked in the plot.

1668 V OLUME 15JOURNAL OF CLIMATE

FIG. 7. Jan-mean temperatures at 100, 70, 50, 30, and 20 mb, area-weighted over 60°–90°N, as a function of years. Legends are marked in the plot.

20.06, 20.07, and 20.05 K yr⁻¹, respectively, with significance all close to 100%. The trends are close to the results from CLIMAT TEMP data reported by Gaffen et al. (2000). Higher up, at 30 mb the temperature does not show a significant cooling trend. In fact, a warming trend is found at 20 mb, with significance above 90%. Randel and Wu (1999) and Gaffen et al. (2000) also showed similar warming trends at levels above 30 mb. It seems that stratospheric cooling is mainly in the lower stratosphere.

Though the winter trend is not as systematic as the summer trend, the long-term temperature variation in winter is not inconsistent with the summer trend. For example, if one calculates the winter trend starting from 1977, one would obtain a trend with a slope of −0.05 K yr⁻¹ (circle-dash line in Fig. 7), which is close to the summer trend.

To examine the relationship between temperature and dynamical heating due to planetary waves, in Fig. 9 we plot the anomalies of the Jan-mean polar mean temperature, as a function of years, against the normalized anomalies of heat flux $F_{60°N,50mb}$ averaged over NDJ. Consistent with the prediction in (13) or (14), the interannual variability of the temperature is indeed driven by the interannual variability of the cumulative wave-driven heat flux. For the period from 1958 to 2000, the correlation coefficient between the two is about 0.73 (the nearly anticorrelation before 1958 appears to be due to bad data). From 1968 to 2000, the correlation is about 0.90. From 1979 to 2000, the correlation is about 0.84. These are all very high correlation coefficients. The correlation of interannual variations between the polar mean temperature at 50 mb and $F_{60°N,50mb}$ is close to the results in Salby et al. (2000) and Newman et al. (2001).

Though there is a strong correlation of interannual variations between the polar mean temperature and wave-driven heat flux $F_{60°N,50mb}$, the two appear to behave differently on longer timescales. From Fig. 9, one can see that the long-term trends of the two diverge. The trend in the temperature is steep and statistically significant, while the linear regression of $F_{60°N,50mb}$ has a nearly zero slope. Therefore, the stratospheric cooling, $\Delta \theta$, is probably a result of radiative cooling, $\Delta \theta_r$, possibly due to greenhouse gas effects and ozone depletion, rather than a result of a decrease of dynamical heating induced by planetary waves. [That is, the long-term trend is more consistent with the balance: $\Delta \theta = \alpha (\Delta \theta_r - \Delta \theta) = 0$ in (13) rather than a balance between the first and second terms.]

To compare with the result by Salby et al. (2000), in Fig. 10 we plot the normalized anomalies of NDJ-mean $F_{60°N,50mb}$ against the NDJ-mean upward E–P flux from the troposphere, $F_{10mb}$. Except in a few years when the two are out of phase (e.g., 1958, 1974), the interannual variations of $F_{60°N,50mb}$ and $F_{10mb}$ are very consistent. From 1968 to 2000, the correlation coefficient between the two fluxes is about 0.80. This is not surprising because $F_{60°N,50mb}$ is part of the overall planetary-scale wave driving from the troposphere.

As mentioned above, the cooling trend in the Jan-mean polar mean temperature is about three times larger than the summer cooling trend. The large interannual variation of dynamical heating may affect the linear trend obtained for winter. In order to reduce the influ-
ence of the interannual variations of dynamical heating on the cooling trend, we use the method of least squares regression to minimize the sum of squared residual

$$\sum_{i=1}^{34} [T_{50\text{mb}}(t_i) - \alpha F_{60^\circ N,50\text{mb}}(t_i) - b - \gamma t_i]^2,$$  \hspace{1cm} (16)

where $\gamma$ is the trend we expect after minimizing fluctuations of dynamical heating, $\alpha$ is the optimized coefficient for reducing fluctuations from $F_{60^\circ N,50\text{mb}}$, and $t_i$ denotes years. Figure 11 illustrates normalized $T_{50\text{mb}}$, $F_{60^\circ N,50\text{mb}}$, $T_{50\text{mb}} - \alpha F_{60^\circ N,50\text{mb}}$, and $b + \gamma t_i$. One sees that fluctuations of $T_{50\text{mb}} - \alpha F_{60^\circ N,50\text{mb}}$ (solid line) are less than those of $T_{50\text{mb}}$ (dotted line). The trend for normalized $T_{50\text{mb}} - \alpha F_{60^\circ N,50\text{mb}}$ is $\gamma \approx -0.29$. Transforming it to the unnormalized value gives $\gamma_{T_{50\text{mb}}} = -(0.29 \times 4.6)/9.8 \approx -0.14 \text{ K yr}^{-1}$ with significance above 97.5%, where $\sigma_{T_{50\text{mb}}}$ and $\sigma_{\gamma}$ are the standard deviations of $T_{50\text{mb}}$ and $b + \gamma t_i$. The obtained winter cooling slope does not change very much even after the interannual fluctuations of dynamical heating are minimized. What is gained is that visually it is clearer from Fig. 11 that

---

**Fig. 8.** Jul-mean temperatures at 100, 70, 50, 30, and 20 mb, area-weighted over 60°–90°N, as a function of years. Legends are marked in the plot.

**Fig. 9.** Anomalies of Jan-mean temperature at 50 mb, area-weighted over 60°–90°N, vs years, against normalized NDJ mean poleward heat flux, $F_{60^\circ N,50\text{mb}}$ (normalized by its rms). The dash-dot line is the trend in temperature, and the dashed line is the linear regression of $F_{60^\circ N,50\text{mb}}$. For comparison with the temperature anomaly, the normalized E–P flux is multiplied by 5.7.
there is a downward slope in $T_{50mb} - aF_{60^\circ N,50mb}$ (solid line).

5. E–P flux versus NAM index and angular momentum

Now we turn our attention to the relationship between planetary wave activity and the zonal-mean flow. Since the NAM pattern consists mostly of the anomalies of zonal-mean geopotential heights, it can be considered a proxy for the zonal-mean state. The positive trend of the NAM index is thus seen to be consistent with the strengthening polar night jet (Thompson and Wallace 1998; Wallace 2000), which is in turn consistent with polar cooling.

Figure 12 shows the normalized anomalies of the Jan-mean NAM index at 50 mb as a function of years, against $F_{100mb}$. On interannual timescales, $F_{100mb}$ and the NAM index have an antiphase relationship, with a correlation coefficient of about $-0.73$. The anticorrelation becomes stronger in more recent years probably because of better data. For the periods: 1968–98 and 1979–98, the correlation coefficients are $-0.78$ and $-0.88$, respectively. On decadal timescales, however, the two diverge. The NAM index shows a significant positive trend over 1968–98, with slope 0.04 yr$^{-1}$ and confidence close to 95%. This result was first obtained by Thompson et al. (2000). Over the same period, the linear regression of $F_{100mb}$ has a slope that is not statistically different from zero.

Limpasuvan and Hartmann (1999, 2000) and Hartmann et al. (2000) focused on the interaction between the NAM index and eddy-momentum flux (which is equivalent to the horizontal component of the E–P flux here). In Fig. 13 we replot the NAM index together with normalized $F_{50N}$ (the latter was already shown in Fig. 3). The two are generally anticorrelated on interannual timescales, and the anticorrelation is as good as the relationship between the NAM and the vertical E–P flux in Fig. 12. Their long-term trends are even more di-
 verified than those in Fig. 12 for the vertical E–P flux. Incidentally, it should be noted that in the stratosphere, the momentum flux is not an important factor in the momentum budget; see Fig. 3.

In Limpasuvan and Hartmann (1999, 2000) and Hartmann et al. (2000), the E–P flux vectors were composited according to the high-low phases of NAM indices on a winter-to-winter basis. Thus, what was found by them actually represents the interannual relationship between the NAM index and the E–P flux vectors. The anticorrelation between the NAM index and the E–P flux shown above reflects this relationship on interannual timescales. It suggests that the dynamical heating caused by upwelling wave E–P flux can account for most of the year-to-year variations of the NAM index (treating the latter as a zonal-mean quantity). Over decadal timescales, however, the positive trend in the NAM index, which is consistent with decreasing polar mean temperature, is not explainable by the E–P flux trend. We thus suggest that the positive trend in the NAM index is probably caused by radiative effects, a different mechanism from that causing the interannual high NAM index phases.

The interaction between planetary waves and the zonal-mean state in the stratosphere can also be demonstrated in the zonal-mean angular momentum budget. From the transformed Eulerian-mean momentum equation, Tung (1986) and Dunkerton and Baldwin (1991) have deduced that the tendency of zonal-mean angular momentum over a time period is approximately given by the E–P flux divergence averaged over the simultaneous period; that is,

\[
\frac{\partial \langle M \rangle}{\partial t} = F_{\text{net}} - \langle X \rangle,
\]

where \(M = \rho_0 a \cos \phi (\overline{u} + \Omega a \cos \phi)\) is the angular

**Fig. 12.** Jan-mean NAM index at 50 mb vs years, against NDJ mean total upward E–P flux at 100 mb (normalized by its rms). The dash–dot line is the trend in NAM index, and the dashed line is the linear regression of \(F_{100\text{mb}}\).

**Fig. 13.** Jan-mean NAM index at 50 mb vs years, against Jan-mean horizontal E–P flux crossing the box boundary at 50°N (normalized by its rms). The dash–dot line is the trend in NAM index, and the dashed line is the linear regression of the horizontal E–P flux, with a nearly zero slope.
Fig. 14. Angular momentum difference between Jan–mean and Nov–mean within the box vs years, against NDJ mean total upward E–P flux at 100 mb. Both are normalized by their rms. The dash-dot line is the trend in angular momentum difference, and the dashed line is the linear regression of $F_{100\text{mb}}$.

Angular momentum, $\langle M \rangle = \int_0^\beta \int_{z_1}^{z_2} Ma \cos \phi \, d\phi \, dz$ is the mean angular momentum integrated over latitude and height, and $\langle X \rangle$ represents the mean advection of $M$. However, we found that zonal-mean angular momentum does not correlate with E–P flux divergence as well as its correlation with the upward E–P flux, $F_{100\text{mb}}$. It is probably because the mean advection of angular momentum $\langle X \rangle$ is not negligible. Instead, we shall focus on the correlation between angular momentum tendency and $F_{100\text{mb}}$, since the latter measures the overall wave driving coming from the troposphere, as pointed out by Salby et al. (2000).

Figure 14 shows normalized anomalies of angular momentum difference between Jan mean and Nov mean as a function of years, against normalized $F_{100\text{mb}}$ anomalies (angular momentum is averaged over the same box as that for E–P flux in Fig. 3). Similar to the relationship between the NAM index and E–P flux, the angular momentum difference and upward E–P flux are also anticorrelated on interannual timescales. From 1958 to 2000, the correlation coefficient is about $-0.72$. For periods of 1968–2000 and 1979–2000, the correlation coefficients are $-0.82$ and $-0.85$, respectively. Such strong anticorrelations mean that the interannual variability of angular momentum tendency is coupled with planetary wave activity. Again, the two show different trends for longer timescales. The angular momentum difference has a trend over the period of 1968–2000, with a slope of about $2.0 \times 10^8$ kg m$^{-1}$ yr$^{-1}$ and significance close to 95%, in contrast to the nearly zero slope of the linear regression of $F_{100\text{mb}}$.

6. Conclusions

Using 51-yr NCEP–NCAR reanalysis data, we have studied the interannual and long-term variations of planetary wave activity, stratospheric cooling, and NAM. Our results demonstrate that there is no evidence indicating a decrease of planetary wave activity from the troposphere into the stratosphere over decadal timescales. Both E–P flux across the tropopause and planetary wave amplitudes in the lower stratosphere do not show significant changes in the past few decades. This disagrees with the speculation that planetary wave activity in the stratosphere might have been reduced by altered climate conditions in the upper troposphere due to the greenhouse effect.

Our results show that on interannual timescales the variation of the Jan-mean polar mean temperature is strongly driven by poleward heat flux. On timescales of decades, however, the two are not coupled. The temperature has a significant cooling trend in the recent 30 yr, while both the poleward heat flux and the upward E–P flux across 100 mb do not have any significant trend. This suggests that the cooling trend in the polar temperatures is probably a result of radiative cooling, due possibly to the greenhouse effect and/or ozone depletion, and not as a result of declining planetary wave activity.

Similarly, on interannual timescales the variabilities of the NAM index and angular momentum are all strongly anticorrelated with the interannual variability of the upward E–P flux across 100 mb. For longer timescales, the significant, positive trends in the NAM index and angular momentum are not accompanied by a decrease of the upward E–P flux from the troposphere. Therefore, these positive trends can also be attributed to radiative cooling. An evidence supporting this argument is that the NAM trend is most significant in the lower stratosphere (Hartmann et al. 2000), where the cooling trend in the polar mean temperature is largest and most significant.

In the present paper, we have mainly focused on the variations of planetary waves, polar mean temperature, and NAM in early and middle winter. We have not
touched upon the issue of stratospheric cooling in late winter and spring. As briefly mentioned above, stratospheric cooling trends in early versus late winter are very different due to the radiative effect of ozone, which becomes more important during and after the final warming, when the sun returns. It has also been suggested that recent stratospheric cooling in late winter and early spring might be partly attributed to less ozone transport from the Tropics to the Arctic polar region due to a decrease of planetary wave activity (Coy et al. 1997; Shindell et al. 1998). Whether or not this speculation is true remains to be addressed.

Acknowledgments. We are grateful to James Holton, Dennis Hartmann, Walter Robinson, and Murry Salby for their comments on an early draft of this paper. We thank Mark Schoeberl and two anonymous reviewers for their constructive suggestions, Mark Baldwin for providing us with the data of the NAM index, and Katie Coughlin for helpful discussion. NCEP reanalysis data is obtained from the NOAA–CIRES Climate Diagnostics Center, Boulder, Colorado, from their Web site at http://www.cdc.noaa.gov/. This work is supported by the National Science Foundation, Division of Atmospheric Sciences, Climate Dynamics, under Grant ATM 9813770.

REFERENCES


Wu, F., 1999: Cooling of the Arctic and Antarctic polar stratospheres due to ozone depletion. J. Climate, 12, 1467–1479.