Strengthening of the Tropopause Inversion Layer during the 2009 Sudden Stratospheric Warming: A MERRA-2 Study

KRZYSZTOF WARGAN AND LAWRENCE COY

(Manuscript received 3 November 2015, in final form 25 January 2016)

ABSTRACT

The behavior of the tropopause inversion layer (TIL) during the 2009 sudden stratospheric warming (SSW) is analyzed using NASA’s Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), and short-term simulations with the MERRA-2 general circulation model. Consistent with previous studies, it is found that static stability in a shallow layer above the polar tropopause sharply increases following the SSW, leading to a strengthening of the high-latitude TIL. Simultaneously, the height of the thermal tropopause decreases by around 1 km. Similar behavior is also detected during other major SSW events between the years 2004 and 2013. Using an ensemble of general circulation model forecasts initialized from MERRA-2, it is demonstrated that the primary cause of the strengthening of the TIL is an increased convergence of the vertical component of the stratospheric residual circulation in response to an SSW-induced acceleration of the mean downward motion between 75° and 90°N. In addition, 6% of the strengthening in 2009 is attributed to an enhanced anticyclonic circulation at the tropopause. A preliminary analysis indicates that during other recent SSW events there was a significant increase in the convergence of the vertical residual wind velocity throughout the middle and lower stratosphere. The static stability increase simulated by the model during the 2009 SSW is 60%–80% of that seen in MERRA-2. The underestimate is traced back to a tendency for the forecasts to underestimate the resolved planetary wave forcing on the stratosphere compared to the reanalysis.

1. Introduction

The tropopause inversion layer (TIL; Birner et al. 2002; Birner 2006; Gettelman et al. 2011) is a distinct feature of the extratropical lower stratosphere. It is characterized by a temperature inversion and a sharp maximum in static stability in a shallow layer (1–2 km) above the thermal tropopause. Connections between the tropopause static stability structure and the occurrence of multiple tropopauses as well as implications for the distribution of trace gases are subjects of ongoing research (Hegglin et al. 2009; Schwartz et al. 2015). The presence of the TIL is believed to impact the upper-tropospheric and stratospheric dynamics by influencing wave propagation. For example, Zhang et al. (2015) demonstrated that the TIL significantly inhibits the upward propagation of inertia–gravity waves, which in turn leads to an intensification of the TIL itself. The mechanisms responsible for the formation and maintenance of the TIL are thought to involve dynamical (Son and Polvani 2007; Erler and Wirth 2011) and radiative processes (Randel et al. 2007). A relationship between the strength of the TIL and the occurrences of double tropopauses in the extratropics was explored by Peevey et al. (2014), who demonstrated the role of the warm conveyor belt in establishing it. In addition, Kunkel et al. (2014) explored a possible contribution to the TIL from gravity waves using baroclinic life cycle simulations. Birner (2010) identified the vertical convergence of the downward branch of the residual circulation as the leading factor in the TIL formation in the middle latitudes during winter. We note that these mechanisms are not necessarily independent; for example, a dynamical change due to wave forcing will also be manifested in the residual circulation.

Sudden stratospheric warming (SSW) events are large-scale disturbances of the wintertime polar vortex, characterized by a rapid increase of polar temperatures...
with a reversal of the 60°–90°N temperature gradient in
the stratosphere (only Northern Hemisphere warmings
are considered here) and a weakening of the mean zonal
flow. A warming event is considered a major SSW if the
60°N mean zonal winds at 10 hPa reverse to easterly.
Major SSW events typically occur once every two or
three boreal winters.

Generally, the high-latitude TIL is found to be weak
or even absent during winter (Tomikawa et al. 2009;
Gettelman et al. 2011). However, using global posi-
tioning system (GPS) radio occultation data, Grise et al.
(2010) found a correlation between static stability above
the polar tropopause and the northern annular mode
(NAM) index: the TIL becomes stronger during SSW
events. Son et al. (2011) found a similar increase in the
tropopause temperature and sharpness (defined as the
cross-tropopause gradient of static stability) during
the major SSW event in 2009. To our knowledge, the
processes involved in this SSW-related strengthening
of the TIL have not yet been studied in detail. The goal
of the present study is to analyze the strength of the TIL
in the context of the dynamics of the polar lower strato-
sphere during the 2009 SSW, using NASA’s Modern-Era
Retrospective Analysis for Research and Applications,
version 2 (MERRA-2; Bosilovich et al. 2015), supple-
mented by an ensemble of general circulation model
(GCM) simulations. Specifically, we attempt to iden-
tify the leading mechanism involved in the SSW-
related increase of lower-stratospheric static stability
observed in MERRA-2 and in the model. The 2009
SSW was chosen for this case study because the in-
crease of the TIL’s strength associated with it was
exceptionally strong compared to other recent SSW
events. In a later part of the paper, we place the
analysis in the broader context of other boreal winters
in the past decade.

Older data assimilation systems failed to represent the
TIL correctly (Birner et al. 2006) owing to excessive ver-
tical smoothing. However, more recent reanalyses do re-
produce the near-tropopause static stability structure seen
in high-resolution independent observations. Gettelman
and Wang (2015) used reanalysis data from ERA-Interim
and the Modern-Era Retrospective Analysis for Re-
search and Applications (MERRA, the predecessor of
MERRA-2) to produce a multiyear climatology of the
TIL. They found a good agreement of the static stability
profiles from reanalyses and those derived from GPS
data. Differences resulted from limited vertical resolu-
tion of the reanalyses, which leads to systematic quan-
titative discrepancies between the TIL’s strength in
these different datasets.

The paper is organized as follows. Section 2 describes
MERRA-2 and provides definitions of the tropopause,
static stability, and TIL’s strength used in this study. The
results of reanalysis diagnostics and model forecast ex-
periments, including the 2009 SSW case study and a
discussion of other recent major SSW events are pre-
sented in section 3. Section 4 is devoted to discussion
and conclusions.

2. Data and methodology

This section describes the data and provides the ter-
minology and conventions used throughout this study.

a. MERRA-2

MERRA-2 is the latest multiyear (1980–present) re-
analysis produced by NASA’s Global Modeling and
Assimilation Office (GMAO) using the Goddard Earth
Observing System, version 5, data assimilation system
(GEOS-5). MERRA-2 constitutes an update of the
original reanalysis, MERRA (Rienecker et al. 2011),
with significant changes made to the GCM and the ob-
serving system. The model updates, described by Molod
et al. (2015), pertain mostly to tropospheric physics but
also include a modification of the gravity wave drag
parameterization scheme, which has significant impacts
on stratospheric dynamics, for example, on the repre-
sentation of the quasi-biennial oscillation.

In the period considered in this study (2004–2015),
MERRA-2 assimilates between 2 and 5 million obser-
vations per 6-h cycle (approximately 3 million in 2009),
mainly radiances from the Atmospheric Infrared Sounder
(AIRS) on the Aqua satellite, the Infrared Atmospheric
Sounding Interferometer (IASI, starting in September
2008), the Cross-Track Infrared Sounder [on the Suomi–
National Polar-Orbiting Partnership (Suomi–NPP) satel-
lite, from April 2012 onward], the Advanced Technology
Microwave Sounder (on Suomi–NPP, starting in No-
vember 2011), and microwave data from the Advanced
Microwave Sounding Unit and a series of High-Resolution
Infrared Radiation Sounder instruments. Of these, the
AIRS and IASI instruments provide hyperspectral ra-
diance observations in 2009. Of particular importance
for the tropopause region are temperature profile obser-
vations from radiosondes, aircraft temperature
measurements, and high-vertical-resolution GPS ra-
dio occultation data also assimilated in the period of
interest. Online bias correction schemes are applied to
the radiance and aircraft observations. The above
list of observations is not exhaustive. Details of the
MERRA-2 observing system, including the complete
specification of input observations, will be provided
by the GMAO in a future publication.

The reanalysis is generated at a 0.635° × 0.5°
longitude–latitude resolution at 72 layers between the
surface and 0.01 hPa. The bottom 32 layers are terrain following, and the rest (pressures 164 hPa and lower) form a constant-pressure grid. Assuming, the surface pressure of 1013 hPa, the layers in the upper troposphere and lower stratosphere are centered around the following approximate geometric heights: 6.61, 7.41, 8.23, 9.34, 10.49, 11.62, 12.76, and 13.9 km, yielding the vertical resolution of ~1 km near the tropopause and in the lower stratosphere. A preliminary climate-focused validation of the reanalysis is given in Bosilovich et al. (2015). In this study, we use the 3-hourly MERRA-2 output fields on native model levels (GMAO 2015). Note that the terms “MERRA-2” and “analysis” are used interchangeably in what follows.

b. Definitions

We adopt the standard definition of the thermal tropopause as the lowest level at which the temperature lapse rate is less than or equal to the threshold value of 2 K km⁻¹, provided the mean lapse rate between that level and every level within 2 km above it does not exceed the threshold (WMO 1957). Whenever suitable, following Birner et al. (2002) and Birner (2006), we employ the tropopause-based (TB) vertical coordinate, in which height is defined relative to the local tropopause.

Regardless of the vertical coordinate used (TB or geometric height above the surface), the MERRA-2 and forecast output temperature and wind profiles are interpolated by cubic splines to a set of constant-height levels 1 km apart unless stated otherwise. This spacing is about the same as the vertical resolution of GEOS-5 in the lower stratosphere. The interpolation is done first, and derived quantities such as static stability and potential temperature are calculated in the new coordinate for each profile separately before any zonal and area-weighted averaging is done. Vertical interpolation is necessary for calculating averaged profiles with respect to the tropopause. The quantity \( N_{\text{MAX}}^2 \) defined in the next paragraph could be computed without it, but we chose to interpolate in this case as well for consistency. We verified that the interpolation step introduces almost no bias (less than 0.5%) and very small standard deviation (2%) to the resulting statistics.

We use the Brunt–Väisälä buoyancy frequency squared as a measure of static stability:

\[
N^2 = \frac{g}{\theta} \partial \zeta \theta = \frac{g}{T} (\Gamma_d + \partial_z T),
\]

where \( \theta \) is potential temperature, \( g = 9.81 \text{ m s}^{-2} \) is Earth’s gravitational acceleration, \( T \) is temperature, the dry adiabatic lapse rate \( \Gamma_d = 0.0098 \text{ K m}^{-1} \), and the subscript \( z \) denotes differentiation with respect to height. Guided by a number of previous studies (e.g., Birner 2010), we find it useful to define the strength of the TIL, \( N_{\text{MAX}}^2 \), as the maximum of \( N^2 \) within a 3-km layer directly above the tropopause. The results are not very sensitive to the choice of the layer thickness, as the maximum is attained fairly close to the tropopause [within 1–2 km (Gettelman and Wang 2015)].

3. Results

a. High-latitude tropopause during major SSW events

While the focus of this study is the strong SSW event in 2009, we will first establish a broader context by briefly examining the behavior of the TIL’s strength during the 10 boreal winters (December–February) between 2006 and 2015 as represented in MERRA-2.

Figure 1a shows the evolution of \( N_{\text{MAX}}^2 \) anomalies averaged north of 75°N during major SSW events in 2004 (Manney et al. 2005), 2006 (Manney et al. 2008), 2009 (Manney et al. 2009; Harada et al. 2010), 2010, and 2013 (Coy and Pawson 2015), compared to the one-sigma envelope calculated from MERRA-2 over all 12 winters. The anomalies are computed by subtracting the 2004–15 mean \( N_{\text{MAX}}^2 \) for each analysis time (eight times per day) separately, effectively “de-seasonalizing” the time series. There were two other major SSWs during that period, in 2007 and 2008, but they were weaker and briefer than the others considered here, and the change in \( N_{\text{MAX}}^2 \) was also weaker in those winters. In the major SSWs winters of 2004, 2009, 2010, and 2013, \( N_{\text{MAX}}^2 \) reached or exceeded the threshold of one standard deviation above the mean for at least 10 days. In those four winters, there were steep increases of \( N_{\text{MAX}}^2 \) initiated within 2 weeks or less before the times of the zonal-mean zonal wind reversal at 10 hPa, 60°N. In 2006, no such increase is evident, but \( N_{\text{MAX}}^2 \) remained high during much of January and in the first half of February. The largest increases are seen in 2009 and 2013. These findings are consistent with the results of Grise et al. (2010), who found a negative correlation between the strength of the TIL and stratospheric dynamics variability represented by the NAM index.

The time series of the mean tropopause height anomalies averaged over the same latitudes are shown in Fig. 1b along with their 2004–15 mean and the one-sigma envelope. Again, the anomalies are obtained by subtracting the 2004–15 mean for each analysis time. In all of the SSW years considered here, the tropopause altitude dropped around the time of the wind reversal, although here too the exact timing is not fixed and, more
generally, the tropopause height varies by as much as \(\sim 2\) km especially in February. Nevertheless, in all five SSW years, the heights decrease to at least 1 km below the average toward the end of the winter. Using GPS radio occultation data, Son et al. (2011) also identified a simultaneous increase of the Arctic tropopause pressure and static stability during an SSW. We will further analyze this behavior of the tropopause height in the following subsection.

b. The 2009 SSW

A very strong vortex split occurred in the winter of 2009 (Manney et al. 2009) with the criteria for a major SSW fulfilled on 24 January. The evolution of high-latitude (75°–90°N) area-averaged temperature and static stability anomalies from MERRA-2 during the 2008/09 winter is shown in Fig. 2 as a function of TB height. The anomalies are calculated separately for each level by subtracting the 2006–15 average MERRA-2 values. As seen in Fig. 2a, the upper stratosphere warms by up to 60 K after 15 January, initiating a rapidly descending pattern of positive anomalies that reach the tropopause before the end of the month. Figure 2b shows increased static stability along the lower edge of the temperature anomalies, as expected from the definition of \(N^2\) [Eq. (1)]. A comparison with Fig. 1a confirms that once the anomalies reached the layer 3 km above the tropopause (on 22 January), there was a sharp rise in the mean \(N^2_{\text{MAX}}\) by \(\sim 1.5 \times 10^{-4} \text{ s}^{-2}\), significantly above its typical winter values expressed as the 2003/04–2014/15 one-sigma envelope around the mean. Starting on 4 February, \(N^2_{\text{MAX}}\) began to decrease. It returned to its typical values (within the envelope) after 15 February.

As seen in Fig. 1, the increase of static stability above the tropopause is associated with a drop in the tropopause
From Eq. (1), we have the following formula for the temperature lapse rate differential:

\[ d(\frac{\partial}{2} T) = \frac{1}{g} (T dN^2 + N^2 dT). \]

Since both, \( dN^2 \) and \( dT \) were positive during the onset of the 2009 SSW, the temperature lapse rate decreased (\( \frac{\partial}{2} T \) increased) at a given altitude so that the 2 K km\(^{-1} \) threshold for the tropopause was attained at a lower altitude and the thermal tropopause moved downward together with the static stability anomaly. The situation is illustrated in Fig. 3, which shows area-averaged 75°–90°N temperature profiles on 17 and 29 January (before and during the SSW). Here, the averaging is performed in geometric height rather than in the TB coordinate in order to show the different tropopause heights on the two days. Note that on 17 January the tropopause is at \( \sim 9.2 \) km and there is no inversion above it. On 29 January the temperature in the lower stratosphere is much higher, there is an inversion, and the lapse rate above 8 km is less than 2 K km\(^{-1} \); the tropopause moves down to 8 km. Furthermore, the two profiles are almost identical in the upper troposphere, implying that the apparent increase in temperature below the tropopause in Fig. 2(a) is a consequence of the fact that the tropopause was located lower on 29 January, after the onset of the SSW, than it was prior to it.

We will now discuss the horizontal structure of near-tropopause static stability. Figure 4 shows typical distributions of \( N^2_{\text{MAX}} \) and relative vorticity at the tropopause. The plots were generated from MERRA-2 fields on 17 and 29 January 2009, before and during the SSW, respectively. Both maps of \( N^2_{\text{MAX}} \) display rich synoptic-scale structures. Before the SSW occurred, the highest values of \( N^2_{\text{MAX}} \) are found in the middle latitudes. This is consistent with previous studies, which showed that the wintertime TIL peaks between 50° and 60°N (e.g., Birner et al. 2006, Fig. 2), and it is much weaker in polar winter, if it exists at all (Gettelman et al. 2011; Tomikawa et al. 2009). After the onset of the SSW, the strength of the TIL north of 75°N is increased. This general pattern is also seen in maps of \( N^2_{\text{MAX}} \) averaged over 2-week periods before and during the SSW (not shown). Evident in Fig. 4 is a very close alignment of the regions of high and low \( N^2_{\text{MAX}} \) with negative (anticyclonic) and positive (cyclonic) relative vorticity, respectively. We find that this relationship between \( N^2_{\text{MAX}} \) and relative vorticity indicates the close connection between the TIL and the static stability anomaly.
and cyclonicity is quite general in that it holds for other months as well, irrespective of the presence of SSWs. This finding provides strong evidence for the dynamical mechanism of the TIL formation seen in model simulations discussed by Wirth (2003), Wirth (2004), Wirth and Szabo (2007), and Erler and Wirth (2011). These model-based studies demonstrated that the vertical convergence of the vertical wind component associated with the formation of upper-level anticyclones induces an increase of $N_{MAX}^2$ directly above it. These results have been corroborated by radiosonde temperature data (Homeyer et al. 2010). Given an enhanced planetary wave activity during SSW events, it is reasonable to ask whether anticyclonic wave breaking at the tropopause contributes to the increased $N_{MAX}^2$ observed in Figs. 1a, 2b, and 4. We will argue that such a contribution exists but plays only a minor role. Figures 5a and 5b show the distribution of $N_{MAX}^2$ and relative vorticity at the tropopause between 75° and 86°N in two 2-week periods before and immediately after the SSW. The $N_{MAX}^2$ distributions (Fig. 5a) are single modal, approximately symmetric with only a small hint of fat tails at the high sides. The mean $N_{MAX}^2$ increases by $1.07 \times 10^{-4} \text{ s}^{-2}$ from $4.83 \times 10^{-4}$ to $5.9 \times 10^{-4} \text{ s}^{-2}$ between the two periods. This is a significant increase compared to the sum of the standard deviations of the two distributions: $(0.4 + 0.56) \times 10^{-4} \text{ s}^{-2} = 0.96 \times 10^{-4} \text{ s}^{-2}$. On the other hand, the shift toward lower relative vorticities (Fig. 5b) is only by $1.01 \times 10^{-5} \text{ s}^{-1}$, small compared to their standard deviations ($5.21 \times 10^{-5}$ and $5.12 \times 10^{-5} \text{ s}^{-1}$): the two distributions largely overlap. Corresponding scatterplots relating the two quantities are shown in Fig. 5c. While a (small) decrease in relative vorticity exists, it is clear from Fig. 5c that the intercept and slope of the linear fit (the two straight lines in the scatterplot) also change during the SSW. A back-of-the-envelope estimate based on the slope of the linear fits and the vorticity decrease suggests that the latter contributes only about

---

**FIG. 3.** Area-weighted 75°–90°N mean temperature profiles from MERRA-2 at 0000 UTC 17 (solid) and 29 (dashed) Jan 2009. The horizontal lines indicate the tropopause heights on the same days.

**FIG. 4.** The 60°–86°N spatial distributions of $N_{MAX}^2$ (colors) on two selected days (left) before (17 Jan 2009) and (right) after (29 Jan 2009) the warming occurred. Overlaid are the contours of relative vorticity ($10^{-5} \text{ s}^{-1}$) at the tropopause. Regions of cyclonic and anticyclonic circulation are shown as blue and red contours, respectively.
FIG. 5. The distributions of (a) $N_{\text{MAX}}$, (b) relative vorticity at the tropopause, and (c) a scatterplot relating the two. All MERRA-2 grid points between 75° and 86°N between 1 and 15 Jan 2009 (blue) and between 28 Jan and 11 Feb 2009 (black) are used. The bin sizes used to compute the distributions are $0.2 \times 10^{-4} \text{s}^{-1}$ for $N_{\text{MAX}}$ and $10^{-6} \text{s}^{-1}$ for vorticity. The linear fit lines are shown as dashed blue and white (1–15 Jan) and black and white (28 Jan–11 Feb). The units of the slopes in (c) are $10^{-3} \text{s}^{-1}$. 
0.061 $\times 10^{-4}$ s$^{-2}$ (~6%) to the mean increase of $N_{\text{MAX}}^2$ so that another process must be responsible for the bulk of the change in static stability. This will be discussed in the following subsection. We note that this result is analogous to the observation made by Wirth (2003), who found that the depth of the TIL is only weakly determined by relative vorticity at the tropopause (their Fig. 12). We also performed an analysis similar to the above but using potential vorticity (PV) instead of relative vorticity. The mean PV at the tropopause drops from 4.1 $\times 10^{-6}$ to 2.82 $\times 10^{-6}$ PV units (PVU; 1 PVU = 10$^{-6}$ K kg$^{-1}$ m$^2$ s$^{-1}$) between the two 2-week periods considered (before and after the SSW). That implies that the height of the dynamical tropopause (defined as the surface of constant PV set to a certain threshold value) does not decrease as much as the lapse-rate tropopause used here. It may also indicate transport of low PV air from lower latitudes into the polar region, although one would need to perform this analysis on isentropic surfaces (rather than the non-conservative tropopause) to confirm and quantify that effect. Here we only note that a detailed PV-based analysis of the TIL formation on synoptic scales was done by Wirth (2003).

c. The role of the residual circulation

In the remainder of this study we will work in the transformed Eulerian-mean (TEM) framework of Andrews et al. (1987), for which the log-pressure altitude above the surface is a more appropriate choice of the vertical coordinate than the tropopause-based height. In particular, it follows that, because of variations in the position of the tropopause, the zonal average $N_{\text{MAX}}^2$ (denoted by $\bar{N}_{\text{MAX}}^2$), calculated by simple averaging of $N^2$ in log pressure, is smaller than that calculated relative to the instantaneous tropopause height and should not be quantitatively compared with average static stability values shown in Figs. 1, 2, 4, and 5. Unlike in the preceding sections, no vertical interpolation is applied. We note that Birner (2010) also used model pressure levels as his vertical coordinate in order to facilitate the TEM framework approach. He too found that the static stability structure above the tropopause is not much affected by this choice.

Using the thermodynamic equation in the TEM framework as his starting point, Birner (2010) derived the following approximate formula for the evolution of $\bar{N}_{\text{MAX}}^2$ [his Eq. (4)]:

$$\delta_t \bar{N}_{\text{MAX}}^2 \approx -\bar{N}_{\text{MAX}}^2 \delta_z \bar{\varpi}^*_{\text{MAX}} + g \delta_z (\bar{\theta}^{-1} Q)_{\text{MAX}},$$  

(2)

where the subscript MAX indicates that a derivative is taken at the point where $\bar{N}^2 = \bar{N}_{\text{MAX}}^2$, $Q$ denotes the diabatic heating rate of the zonal-mean potential temperature and the vertical component of the residual velocity, and $\bar{\varpi}^*$ is given by

$$\bar{\varpi}^* = (a \rho \cos \varphi)^{-1} \delta_z \bar{\Psi}^*,$$

where

$$\bar{\Psi}^* = \bar{\Psi} + a \rho \cos \varphi \frac{\delta \theta}{\delta \theta}.$$

Here, $\Psi$ is the mass streamfunction, $a$ is the radius of Earth (6.371 $\times 10^6$ m), and $\varphi$ and $\rho$ denote latitude and air density, respectively. The primed quantities are deviations from the zonal mean. As before, the overbar denotes zonal averaging. Approximation in Eq. (2) arises from neglecting diabatic eddy terms and contributions due to the meridional component of the residual circulation.

First, we focus on the vertical convergence of $\bar{\varpi}^*$ (denoted by $-\delta_z \bar{\varpi}^*$), which appears in the first term of the right-hand side of Eq. (2). Figure 6 shows the evolution of the area-averaged profiles of $-\delta_z \bar{\varpi}^*$ calculated from MERRA-2 north of 75°N during the 2009 winter. Note that positive (negative) values in Fig. 6 correspond to the convergence (divergence) of $\bar{\varpi}^*$. The band of strong convergence propagating from the middle stratosphere between 20 and 30 January followed by a narrow pattern of negative values indicates a rapid descent of downward acceleration of $\bar{\varpi}^*$. The anomalous $-\delta_z \bar{\varpi}^*$ reaches the layer 3 km above the tropopause on 25 January. The increase of the convergence at 10 km is about 1.2 $\times 10^{-6}$ s$^{-1}$ on that day. As seen in Fig. 2b, this coincides with the initial increase of the TIL’s strength. The event is followed by two weaker bands of increased $-\delta_z \bar{\varpi}^*$, but their impact at the layer above the tropopause is less strong.

Next, we will calculate the terms of Eq. (2). However, we find that it is not appropriate to use the MERRA-2 fields for that purpose. The time derivative on the left-hand side of Eq. (2) implicitly contains a temperature tendency term that, in assimilation, is the sum of model-generated tendencies and tendencies due to the insertion of observations. It is a feature of the data assimilation methodology, in which nudging the model fields by observational data brings the state of the atmosphere closer to the observations but it does so at the expense of introducing a nonphysical component in the tendencies. On the other hand, the right-hand side is constructed only from instantaneous fields. The non-vanishing temperature tendency due to analysis (the analysis update) necessarily prevents the balance from closing. One way to circumvent this problem would be to calculate the left-hand side of Eq. (2) using only temperature tendencies due to dynamics and diabatic
processes. However, these tendencies in MERRA-2 are only available on a reduced set of constant pressure levels (as opposed to the 72 model levels), which provide insufficient vertical resolution in the lower stratosphere. Instead, we chose to use a sequence of 23 GCM forecasts initialized from MERRA-2 fields at 2100 UTC on each day from 14 January to 5 February to compute the terms of Eq. (2). We use the MERRA-2 GCM for this purpose. The diagnostics presented are based on the averages over the first five days of these forecasts. Coy and Pawson (2015) demonstrated that 5-day GEOS-5 forecasts faithfully represent the assimilated meteorological field in the lower stratosphere during a major SSW event. The forecast \( N_{\text{MAX}} \) tendencies are computed by differentiating \( N_{\text{MAX}} \) with respect to time. The results are shown in Fig. 7a. The last term of Eq. (2) (involving the heating rate \( Q \) computed from temperature tendencies due to diabatic processes) turns out to be an order of magnitude smaller than the first term on the right-hand side (the vertical convergence of \( w^* \)), and it is omitted in the plot for clarity. It is clear from Fig. 7a that the \( N_{\text{MAX}} \) tendency from the forecast average (solid blue line) closely follows the \( w^* \) convergence term (dashed blue line):

\[
\frac{\partial}{\partial t} N_{\text{MAX}} \approx -N_{\text{MAX}} \frac{\partial z}{\partial w^*} |_{\text{MAX}}.
\]  

Birner (2010) isolated a similar relationship for much of the extratropical winter stratosphere, but that study did not consider dynamical conditions associated with SSW events. Also plotted in Fig. 7a is the time series of \(-N_{\text{MAX}} \frac{\partial z}{\partial w^*} |_{\text{MAX}}\) from MERRA-2 (solid black line). A 1-day box smoother was applied to the latter in order to filter out diurnal variations. Both \( \frac{\partial}{\partial t} N_{\text{MAX}} \) and \(-N_{\text{MAX}} \frac{\partial z}{\partial w^*} |_{\text{MAX}}\) from the forecasts as well as \(-N_{\text{MAX}} \frac{\partial z}{\partial w^*} |_{\text{MAX}}\) from MERRA-2 are mostly positive between 17 and 28 January and show a pronounced peak around 27 January reaching a maximum of about \(4 \times 10^{-10} \text{s}^{-3}\), marking a rapid increase in \( N_{\text{MAX}} \) seen in Fig. 7b, consistent with the downward-propagating anomalous convergence of \( w^* \) in Fig. 6.

Since each forecast is initialized from MERRA-2, the time series of averaged forecast \( N_{\text{MAX}} \) (blue solid line in Fig. 7b) implicitly includes effects of data insertion. To isolate its evolution due to the GCM’s internal processes alone, we performed the time integration of the left-hand side of Eq. (3) from the forecast average as follows:

\[
N_{\text{MAX}}(t) = N_{\text{MAX}}(t_0) + \int_{t_0}^{t} \frac{\partial}{\partial t} N_{\text{MAX}}(s) \, ds,
\]

where \( t \) is time and \( t_0 \) denotes the time at which the first forecast was initialized, 2100 UTC 14 January in this

![Fig. 6. Vertical convergence of \( w^* \) (10^{-6} \text{s}^{-1}) calculated from MERRA-2 and area averaged within the 75°–90°N latitude band and shown as a function of time (between 20 and 30 Jan) and height. A 24-h box smoother was applied to the data. The vertical coordinate is the mean geometric height above Earth’s surface. The mean tropopause and the level 3 km above it are indicated by white lines.](image-url)
The integrand is the solid blue line in Fig. 7a. The result is shown in Fig. 7b as the solid red line. There is an increase of $0.7 \times 10^{-4}$ s$^{-2}$ associated with the vertical convergence of $w^*$ in the forecasts between 15 and 28 January. But this constitutes only about 60% of the increase seen in MERRA-2 over the same period (black line) and the average forecast ensemble (solid blue line), meaning that the remaining 40% of the increase is due to the analysis tendencies (the insertion of observational data). By examining the individual forecasts (blue dotted lines), it is evident that the model simulates an initial drop in static stability before initiating an increase. Note that all the forecast lines lie below the analysis line in Fig. 7b. This is further confirmed by considering the average of only days 4 and 5 of the forecast tendencies, which yields a larger increase in $N_{\text{MAX}}^2$ (yellow line): if the first three forecast days are omitted in the average, then the simulated increase is over 80% of that seen in the analysis. We conclude that, while the GCM does simulate the static stability increase that is almost entirely due to convergence of $w^*$, it also underestimates $N_{\text{MAX}}^2$ compared with the full analysis. This conclusion is corroborated by the results of additional simulations done for December 2008 during non-SSW winter conditions (not shown). We speculate that this initial forecast decrease in $N_{\text{MAX}}^2$ constitutes a “spinup” period, during which the GCM readjusts the dynamics to its typical internal state—biased with respect to the analysis. It should be noted that this result only reflects a quantitative discrepancy between the model forecasts and the assimilation and does not determine which one is closer to the behavior of static stability in the real atmosphere. In particular, it is conceivable that the assimilation overestimates the increase of $N_{\text{MAX}}^2$.

The results of this subsection demonstrate that the GCM-simulated increase in the high-latitude TIL’s intensity during the 2009 SSW is explained by an enhanced vertical convergence of the vertical component of the
residual circulation. The modeled increase is under-estimated compared to that of MERRA-2.

d. Other recent SSW events

This section briefly discusses the behavior of vertical convergence of $\overline{w^{*}}$ and the convergence term in Eq. (3) during five strong SSW events that occurred in the Northern Hemisphere in 2004, 2006, 2009 (discussed in detail in previous subsections), 2010, and 2013. The choice of these recent events is motivated by the fact that MERRA-2 starts assimilating AIRS radiance data in late 2002 and there are no hyperspectral radiance observations available before then. In addition, the GPS radio occultation data are not assimilated until the summer of 2004. In the future, we plan to expand our analysis over the entire MERRA-2 period (1980–present), but first a careful evaluation is needed of the effects that the changes in the observing system have on shallow dynamical structures near the tropopause in the reanalysis. It should be noted that the SSW events considered here vary in both intensity and character: 2009 and 2013 were vortex split events, while the others were vortex displacement SSWs, with the 2004 event producing a split in the lower stratosphere (Manney et al. 2005). A comprehensive overview of all SSWs between 2004 and 2010, including the weaker and briefer ones in 2007 and 2008, can be found in Kuttippurath and Nikulin (2012). The usual criterion for a major SSW includes the 60°N zonal-mean zonal wind reversal at 10 hPa, and all the events discussed in this study satisfy this condition. However, it is understood that this condition is somewhat arbitrary, and other definitions are possible [see Butler et al. (2015) for a comprehensive review of SSW definitions]. We use the time of the zonal wind reversal to mark a distinguishing moment in the development of an SSW, but we also emphasize that a disturbance of the polar vortex begins prior to and lasts beyond that moment, and its duration exhibits considerable climatological variability. In particular, as seen in Fig. 1, the timing of an effect that a midstratosphere vortex disturbance can have on the tropopause is not tightly constrained to the 10-hPa wind reversal. The 60°N, 10-hPa mean zonal wind reversal dates for the five events discussed here are 5 January 2004, 21 January 2006, 24 January 2009, 9 February 2010, and 6 January 2013.

Figure 8 shows the December–February evolution of the $-\partial_z \overline{w^{*}}$ fields averaged between 75° and 90°N in the five winters, analogous to Fig. 6. We found that, given the relatively long time periods (3 months, compared to 1 month in Fig. 6), the 24-h smoother used in Fig. 8 is insufficient to filter out transient features and produce a clear plot. We applied a 40-h smoother instead. The zero-zonal-wind lines (yellow in Fig. 8) provide a measure of the intensity and depth of each SSW. By this metric, the 2009 event emerges as the strongest, deepest, and longest lasting. Despite considerable variability, in each SSW year there is a prominent structure consisting of downward-propagating alternating bands of positive and negative $\overline{w^{*}}$ convergence concentrated within approximately 2-week periods around or shortly prior to the dates of wind reversal (the latter are indicated by solid white vertical lines in Fig. 8). An exception is the 2010 SSW, where these bands become most prominent about 2 weeks before the reversal date (9 February), with the initial maximum of $-\partial_z \overline{w^{*}}$ occurring around 20 January at 40 km. We note that the 10-hPa zonal-mean zonal wind at 60°N in 2010 dropped down to only a few meters per second as early as on 27 January (this is shown as a dashed vertical line in Fig. 8d) with the zero-zonal-wind line reaching about 36 km in the beginning of February, indicating that the polar vortex was strongly disturbed for nearly 2 weeks before the major SSW criteria were formally met. It may be argued that 27 January, not 9 February, is the more appropriate date to mark the beginning of the 2010 SSW. At least some of the bands of positive $-\partial_z \overline{w^{*}}$ in Fig. 8 extend down into the layer 3 km above the tropopause, as was seen previously in the case of the 2009 SSW (Figs. 6 and 8c) This is not evident in 2006. Note that the latter year was an outlier in that there was not a period of sharp increase in the near-tropopause static stability (Fig. 1a).

A composite of all the $-\partial_z \overline{w^{*}}$ time–height plots from Fig. 8 is shown in Fig. 9a. The central time is chosen to be the wind reversal date for 2004, 2006, 2009, and 2013 (solid vertical lines in Fig. 8) and 27 January for 2010. The time lags between −25 and 25 days are plotted. The composite exhibits a clear structure whereby the maximum convergence occurs at about 38 km prior to the SSW and there are three downward-propagating bands reaching the lower stratosphere at about −12, 0, and 20 days. The maximum $-\partial_z \overline{w^{*}}$ in the layer 3 km above the tropopause is attained close to the central date. Figure 9b plots the convergence term [right-hand side of Eq. (3)] for the individual SSW years (colors) and their average (solid black), along with the one-standard-deviation envelope (gray). While the convergence term has high variability both in time and between different SSW years, the mean exhibits a clear pattern: at time −10 days, the values start rising from less than $0.5 \times 10^{-10}$ s$^{-3}$, attain a maximum of about $1 \times 10^{-10}$ s$^{-3}$ within a few days around lag 0 days, and then begin to fall. We emphasize that the sample of five is not sufficient for a rigorous climatological study, but we note
that, at the time the convergence term reaches the maximum, it is about one standard deviation from zero. According to the approximate relationship given by Eq. (3) and corroborated by the model forecast results of the previous section, positive $\langle \mathcal{W} \rangle$ is associated with an increase in $\mathcal{N}_{\text{MAX}}$. The tendency of the latter, derived from MERRA-2, is plotted in Fig. 9b as the dashed black line. There is a qualitative agreement between it and the convergence term (solid black line). In particular, both attain a maximum at about lag 0 days. As
explained in section 3c, the budget given by Eq. (3) is not expected to close, owing to the analysis tendency being contained in the time derivative of $N_{\text{MAX}}^2$ from the MERRA-2 assimilation. In fact, the tendencies due to data assimilation are of the same order of magnitude as the dynamical tendencies. Therefore, it is not surprising that the agreement between the solid and dashed lines in Fig. 9b is not quantitatively closer.

Finally, we have determined that the increase in the TIL’s strength in the polar region during the 2004–13 SSWs is not explained by a decreased relative vorticity at the tropopause. A detailed analysis is beyond the scope of this study, which is devoted primarily to the strong SSW of 2009. We only report that the contribution connected to relative vorticity at the tropopause, as discussed in section 3b, is approximately 6% in 2009 and less than that during the other SSWs.

4. Discussion and conclusions

We used assimilated data from NASA’s MERRA-2 to investigate the tropopause sharpness during the major SSW of January 2009. Consistent with previous studies, we find that static stability above the high-latitude tropopause (and the TIL’s strength) increased following the event. We confirm that similar increases also occurred during the major SSW events in 2004, 2006, 2010, and 2013 with magnitudes qualitatively matching the intensity of the SSWs. We use a sequence of model forecast to investigate the 2009 increase. We find that the approximate...
The formula for the $N_{2,\text{MAX}}$ tendency [Eq. (3)] is satisfied by the average forecast fields during the SSW event and that the $N_{2,\text{MAX}}$ increase is associated with a descending layer of vertical convergence of $\mathbf{w}$ initiated in the upper stratosphere around 15 January, about 12 days before the peak increase of the TIL’s strength. The enhanced $\mathbf{w}$ convergence results from a downward-propagating acceleration of the residual circulation (and a warm temperature anomaly) during the SSW (Andrews et al. 1987).

The model forecasts simulate about 60%–80% (depending on the method of averaging) of the $N_{2,\text{MAX}}$ increase seen in MERRA-2. More generally, the TIL’s strength is underestimated by the GCM compared to the analysis. Some of this underestimate can be traced back to a tendency for the forecasts to underestimate the wave forcing on the stratosphere with respect to MERRA-2. Figure 10 plots the vertical component of the Eliassen–Palm (EP) flux at 100 hPa during January–February 2009 for both MERRA-2 and the 23 forecasts. Here the vertical EP flux (based only on the meridional heat flux, as the vertical momentum flux is negligibly small for the planetary scale waves of interest) is calculated as follows:

$$
\rho_0 a \cos \phi \left[ f - (a \cos \phi)^{-1} (\pi \cos \phi) \right] \frac{\overline{\theta'}/\overline{z}}
$$

(Andrews et al. 1987, p. 128), where $\rho_0$ is the basic-state density; $a$ and $f$ denote Earth’s radius and the Coriolis parameter, respectively; and $\phi$ is the latitude (radians). The zonal and meridional wind components are denoted by $u$ and $v$, respectively. Derivatives are indicated by a subscript. The vertical component of the EP flux at 100 hPa, averaged over 30°–90°N, provides a good measure of the net planetary wave forcing above that
Along with the total vertical EP flux (Fig. 10a), the contributions from waves 1–3 are also shown (Figs. 10b–d), revealing the large wave-2 EP flux associated with this event in agreement with the analysis of Harada et al. (2010). The 5-day forecasts of the total flux (Fig. 10a, blue diamonds) show good agreement with the analyzed values from 20 to 30 January; however, they tend to be somewhat low around 1 February. By 10 days (cyan diamonds), the forecast trajectories can wander farther from the analysis. The large wave-2 EP flux peak (Fig. 10c) is slightly underestimated by the 5-day and shorter forecasts, with the decaying wave-2 peak on 20–22 January fairly well captured at 5 days but with the 5-day EPV flux forecasts being slightly below the analysis values. Both wave-1 and wave-3 EP flux (Figs. 10b,d) have smaller amplitudes than the wave-2 flux; however, the differences between the plots in Figs. 8a and 8c imply that wavenumbers other than 2 also make a contribution to the total flux.

The 100-hPa vertical EP flux forecast error over the 23 forecasts is shown in more detail in Fig. 11, where the mean bias (blue curve) and standard deviation (black curves) of the forecast minus the analysis EP flux values are calculated as a function of forecast length. The bias toward underestimation of the EP flux begins at ~5 days for the total and waves 1 and 2, and at ~7 days for wave 3. The bias drops until ~8 days and remains fairly constant after that time, as does the standard deviation, suggesting some skill may be present out to 8 days.

The limited resolution of the model may be a factor behind the underestimate of the $N_{\text{MAX}}$ increase in the model. Erler and Wirth (2011) find that the optimal
aspect ratio of the horizontal to vertical resolution for a correct representation of the TIL is \( 300 \text{ m} (1^{-1}) \). In MERRA-2, this ratio is much larger: about \( 1000 \text{ m}/0.5^\circ = 2000 \text{ m} (1^{-1}) \) in the upper troposphere and lower stratosphere. In future versions of GEOS-5 the number of vertical layers will be increased to 137. We plan to revisit this issue then. Son and Polvani (2007) and Wirth (2003) found similar shortcomings in their models compared to observations. The former suggested that a radiative mechanism absent in the model might play a role. One such mechanism, proposed by Randel et al. (2007), involves differential radiative heating by near-tropopause ozone and water vapor. Both those tracers are assimilated in MERRA-2; however, the quality of the lower-stratospheric moisture in the reanalysis requires further evaluation. In addition, the radiative mechanism is thought to be more important in summer (see Randel and Wu 2010; Birner 2010).

Our main conclusions are summarized as follows:

- Consistent with previous studies, MERRA-2 produces an increase in the high-latitude static stability above the tropopause during SSW events. There is a simultaneous decrease in the tropopause height.
- The horizontal distribution of high (low) \( N_{\text{MAX}}^2 \) coincides with regions of anticyclonic (cyclonic) circulation in agreement with previous studies.
- The increase in \( N_{\text{MAX}}^2 \) following the 2009 SSW in the model is almost entirely due to vertical convergence of the vertical component of the residual circulation \( \pi^* \). As estimated in section 3b based on the results in Fig. 5c, about 6% of the increase is attributed to the strengthened anticyclonic circulation at the tropopause.
- The strength of the TIL and its increase during the 2009 SSW are underestimated by 60%–80% (depending on the method of averaging the forecasts) in the model relative to MERRA-2.
- Vertical convergence of \( \pi^* \) also increases during the 2004, 2006, 2010, and 2013 SSW events. With the exception of 2006, anomalies of downward-propagating positive convergence reach the layer 3 km above the tropopause, producing a maximum within a several-day period of the time when the 60\(^\circ\)N zonal wind at 10 hPa weakens to near-zero values.

This study highlights and investigates a connection between stratospheric dynamics and the behavior of the tropopause derived from MERRA-2. The importance of the near-tropopause static stability structure for wave propagation and tracer transport, as well as possible implications for stratosphere–troposphere coupling warrants further research. In particular, a closer analysis of all SSW events in the reanalysis record (including the major event in the Southern Hemisphere in 2002) is desirable. We emphasize the need for a better understanding of the model biases that emerge in the present context.

Acknowledgments. MERRA-2 is an official product of the Global Modeling and Assimilation Office at NASA GSFC, supported by NASA’s Modeling, Analysis, and Prediction (MAP) program. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. We thank Dr. Steven Pawson, Dr. Gloria Manney, and Zachary Lawrence for their comments on the original manuscript and our colleagues at the Global Modeling and Assimilation Office who produced MERRA-2. Finally, we would like to express our gratitude to three anonymous reviewers, whose insightful suggestions helped us improve the manuscript significantly.

REFERENCES


