Progress in Research of Stratosphere–Troposphere Interactions: Application of Isentropic Potential Vorticity Dynamics and the Effects of the Tibetan Plateau

REN Rongcai¹ (任荣彩), WU Guoxiong² (吴国雄), CAI Ming², SUN Shuyue¹,² (孙舒悦), LIU Xin³ (刘新), and LI Weiping³ (李伟平)

1 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
2 Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida 32306, USA
3 University of Chinese Academy of Sciences, Beijing 100049, China
4 Qomolangma Atmospheric and Environmental Observation and Research Station, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China
5 Beijing Climate Center, China Meteorological Administration, Beijing 100081, China

(Received March 21, 2014; in final form July 18, 2014)

ABSTRACT

This paper reviews recent progress in understanding isentropic potential vorticity (PV) dynamics during interactions between the stratosphere and troposphere, including the spatial and temporal propagation of circulation anomalies associated with the winter polar vortex oscillation and the mechanisms of stratosphere–troposphere coupling in the global mass circulation framework. The origins and mechanisms of interannual variability in the stratospheric circulation are also reviewed. Particular attention is paid to the role of the Tibetan Plateau as a PV source (via its thermal forcing) in the global and East Asian atmospheric circulation. Diagnosis of meridional isentropic PV advection over the Tibetan Plateau and East Asia indicates that the distributions of potential temperature and PV over the east flank of the Tibetan Plateau and East Asia favor a downward and southward isentropic transport of high PV from the stratosphere to the troposphere. This transport manifests the possible influence of the Tibetan Plateau on the dynamic coupling between the stratosphere and troposphere during summer, and may provide a new framework for understanding the climatic effects of the Tibetan Plateau.

Key words: isentropic potential vorticity theory, stratosphere–troposphere interactions, Tibetan Plateau


1. Introduction

Besides the material exchanges across the interface between the stratosphere and the troposphere (or the tropopause), the main stratosphere troposphere interactions lie in the dynamical coupling between the two layers. Typical dynamical interactions in winter hemisphere include the leading stratospheric oscillation mode and the associated changes in the meridional circulation, which are driven primarily by breaking planetary waves in the extratropical stratosphere (Matsumo, 1979). Anomalous circulation signals in the stratosphere propagate downward during this oscillation, thereby affecting both the circulation and the climate in the troposphere (Baldwin and Dunkerton, 2001; Thompson et al., 2002; Cai, 2003; Hu, 2006; Deng et al., 2008; Li et al., 2008; Xiong et al., 2009; Chen et al., 2013; Lu and Ding, 2013b). Because of

Supported by the National Natural Science Foundation of China (41347105 and 41275088), China Meteorological Administration Special Public Welfare Research Fund (GYHY201006001), and National (Key) Basic Research and Development (973) Program of China (2010CB426600 and 2010CB954000).

*Corresponding author: gwwu@ias.ac.cn.

©The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2014
the longer timescale of stratospheric anomalies and their temporal lead (2–3 weeks) relative to that in the troposphere, stratospheric signals used to be regarded as another useful predictor besides ENSO for the climate variability in the troposphere (Baldwin and Dunkerton, 2001). However, this general picture of stratosphere–troposphere interactions is insufficient for the practical application of stratospheric signals in predictions because changes in the tropospheric circulation are also affected by short-timescale variations. The dynamical coupling processes are also much more complex than our current understanding.

As the largest source of planetary waves in the Northern Hemisphere, the Tibetan Plateau (TP) can largely dominate the location and strength of the East Asian westerly jet and the formation of the East Asian trough (Zou et al., 1992a, b). A number of studies by Chinese scientists have shown that the thermodynamic forcing provided by the plateau is a key factor in the initiation and maintenance of the Asian summer monsoon, the formation of climate patterns in Asia, and the global atmospheric circulation (Ye et al., 1957; Wu et al., 2002, 2005; Liu et al., 2007, 2012; Wu et al., 2007, 2012a, b; Duan et al., 2011; Liu et al., 2013). Due to the powerful thermal forcing of TP in summer, the vertical transport and the exchange of mass and constituents by deep convections near the TP are also an important pathway for global stratosphere–troposphere exchange (STE) (Zhou and Luo, 1994; Zhou et al., 1995, 2006; Fu et al., 2006; Park et al., 2007; Mao et al., 2008; Xu et al., 2008; Zhan et al., 2008; Bian et al., 2011, 2013; Chen et al., 2012). The effects of this transport can in turn substantially influence the regional climate and environment (Chen et al., 2006; Hu et al., 2009; Lü et al., 2009). However, the strongest irreversible STE events are related to stratospheric polar vortex oscillation (PVO) processes accompanied by an anomalous meridional circulation (Brewer–Dobson circulation) (Brewer, 1949; Dobson, 1956) or other large-scale dynamical processes such as irreversible mixing and tropopause folding (Yang and Lü, 2003; Chen et al., 2006; Guo et al., 2007; Lü et al., 2008; Liu et al., 2009). The dynamics of stratosphere–troposphere interactions around the TP therefore forms a basis for understanding regional and global STE processes. However, a definitive consensus on the dynamical impact of the TP impact on stratospheric PVO and stratosphere–troposphere interactions is still lacking due to limitations in data availability and research techniques.

According to the isentropic potential vorticity (IPV) theory (Hoskins et al., 1985; Hoskins, 1991), the atmosphere can be divided into three layers from the troposphere to the stratosphere: the “underworld” where isentropic surfaces are entirely within the troposphere and intersect with the ground; the “middle world” where isentropic surfaces intersect with the tropopause but not with the ground; and the “over world” where isentropic surfaces are entirely within the stratosphere.

Potential temperature and potential vorticity are conserved under adiabatic frictionless conditions. Potential vorticity (PV) surfaces can be approximately regarded as material surfaces, and are therefore often used to define the tropopause in studies of stratosphere–troposphere interactions. Moreover, the polar vortex edge that separates high PV cold polar air from low PV warm air outside is characterized by a high IPV gradient (i.e., the baroclinic zone). Large-scale irreversible mixing of cold and warm air across this boundary occurs during stratospheric oscillation events when the polar vortex collapses. These events are typically associated with anomalous meridional circulations, including anomalous tropical “upwelling” and extratropical “downwelling” (Holton et al., 1995). Dynamical coupling of the stratosphere and troposphere in the midlatitude “middle world” (where isentropic surfaces and the tropopause intersect) is characterized by large-scale vortices (e.g., blocking highs and cut-off lows) and tropopause folding events. Furthermore, the “underworld” (where isentropic surfaces intersect the ground) represents the main source and sink of PV for the entire atmosphere because the sum of mass-weighted IPV within a layer surrounded by closed isentropic surfaces does not change even in the presence of diabatic heating or friction (Haynes and McIntyre, 1987, 1990; Hoskins, 1991). Studies of the atmospheric circulation in the TP region have sugges-
ted that one of the most significant impacts of the TP on the atmospheric circulation is its role as a source or sink of IPV (Liu et al., 2001; Wu et al., 2002).

Recently, we employed the IPV theory to study the stratosphere–troposphere interactions associated with PVO processes, and identified the systemic meridional and vertical propagation of circulation anomalies from the stratosphere to the troposphere during the oscillations, and proposed a new stratosphere–troposphere coupling mechanism under the global mass circulation framework (Cai and Ren, 2006, 2007; Ren and Cai, 2006, 2007, 2008). Our recent studies have also shown the lagged effect of El Niño–Southern Oscillation (ENSO) variability in the stratosphere (Ren and Xiang, 2010; Ren, 2012a, b; Ren et al., 2012).

In this paper, we provide a systematic review of the above studies of stratosphere–troposphere interactions and summarize the thermal and dynamic effects of the TP as revealed by applications of IPV theory. We also discuss the climatological significance of the TP with respect to stratosphere–troposphere coupling.

2. Propagation of stratosphere–troposphere circulation anomalies associated with PVO

2.1 A PVO index

Stratospheric PVO is an oscillation in the extratropical stratosphere between zonally symmetric strong westerly winds and zonally asymmetric weak westerly winds. The zonally symmetric case corresponds to an anomalously strong stratospheric polar vortex, while the zonally asymmetric case may correspond to an anomalously weak stratospheric polar vortex, or a vortex that shifts away from the polar region. In this regard, the traditional zonal averaging may not always properly reflect changes in the strength, location, and zonal structure of the polar vortex. By contrast, the distribution of IPV can accurately characterize the pattern, intensity, and location of the polar vortex, as well as the westerly jet along the vortex edge (coincident with the strong IPV gradient) (Baldwin and Holton, 1988; Waugh and Randel, 1999). We therefore constructed a semi-Lagrangian coordinate system with potential temperature ($\theta$) as the vertical coordinate and IPV as the latitudinal coordinate. To do this, we calculated the area enclosed by each selected IPV contour, and assigned each IPV line a latitude value (named IPV latitude, or PVLAT) when the spherical area surrounded by the IPV line equals that surrounded by this latitude circle. Different from conventional Eulerian or spherical coordinates, the PVLAT varies following the temporal variation of IPV. In this $\theta$-PVLAT coordinate, the zonal means are the averaging along the PVLAT (or IPV lines), which are very like the “Lagrangian” averaging and effectively represent the averages on the air mass with similar properties. The contrasts between the air mass within the polar vortex and the air mass outside are therefore more pronounced (Fig. 1), which helps us to effectively capture the key thermodynamic processes associated with PVOs (Ren and Cai, 2006; Cai and Ren, 2007).

EOF analysis of daily IPV anomalies in the $\theta$-PVLAT coordinate system indicates that the first EOF mode explains as much as 69% of the variance in the daily IPV anomalies over the entire Northern Hemisphere, reflecting the intensity oscillation of the stratospheric polar vortex, or the out-of-phase relationship of IPV and temperature anomalies between inside and outside of the polar vortex. The “semi-Lagrangian” coordinate system acts as a natural filter for excluding the effects of advection, so that the time series of the leading EOF mode of PVO is quite smooth with less synoptic-scale disturbances, clearly demonstrating the contrasting features between quiet summers and disturbed winters with on average 1–2 PVO events in every winter season (Cai and Ren, 2007). The maximum lead/lag correlation (−0.91) between the PVO and the NAM (Northern Hemisphere annular mode) is at 20 hPa, when PVO leads the NAM by around 12 days. The maximum correlations tend to propagate downward with time (Cai and Ren, 2007). Since the peak phase that the PVO captures is the most salient features of the changes in stratospheric polar vortex intensity, the downward propagation of the maximum correlations between PVO and the NAM index may therefore reflect the evolution fea-
Fig. 1. Zonal means of (a) PV (PVU; 1 PVU = 10^{-6} m^2 s^{-1} K kg^{-1}), (b) zonal wind (m s^{-1}), and (c) temperature (K) at \(\theta = 650\) K on 2 February 1996. The curves with solid dots are obtained by averaging along PVLAT, while thin curves are obtained by averaging along regular latitudes. [From Ren and Cai, 2006]

2.2 Propagation of circulation anomalies in the stratosphere and troposphere

Kodera et al. (1990) identified downward propagation of stratospheric signals in the extratropics using monthly data. They found that stronger zonal mean westerly winds in the extratropical upper stratosphere in December can sustain westerlies in the lower stratosphere and are associated with stronger westerly winds in the extratropical troposphere during the following February. Later studies noted that the positive and negative phases of the AO (Arctic Oscillation), the NAO (North Atlantic Oscillation), and the leading modes of variability in the extratropical troposphere happen to correspond to the positive and negative phase of the NAM in the stratosphere. The AO and NAO may thus be considered as surface signals of NAM variability. Links between the stratospheric NAM and the surface AO have been attributed to the systematic downward propagation of anomalies in zonal-mean zonal wind (Kodera and Kuroda, 2000; Kodera et al., 2000) and geopotential height (Baldwin and Dunkerton, 1998, 1999). The circulation anomalies in the stratosphere lead the circulation anomalies in the troposphere by approximately 2–3 weeks. Several studies have also showed some signs of poleward propagation of zonal-mean zonal wind anomalies during NAM events (Feldstein and Lee, 1998; Dunkerton, 2000; Kuroda, 2002). Kodera et al. (2000) further showed some signs of simultaneous poleward and downward propagation of zonal wind anomalies by analyzing trajectories on a phase plane defined by the two leading EOF vectors of zonal wind anomalies.

We used the PVO index and the “Lagrangian” zonal means in \(\theta\)-PVLAT coordinates described in Section 2.1 to analyze the spatiotemporal evolution of the circulation anomalies from the stratosphere to troposphere and from the tropics to high latitudes, associated with PVO events. We found that the circulation anomalies indeed exhibit a systematic downward propagation, which not only exists in the extratropics but also in all latitudes from the tropics to the polar region, and for both positive and negative PVO events. Besides, co-existed with the downward propagations are simultaneous poleward propagations of circulation anomalies in the stratosphere, which also prevail in both positive and negative PVO events. Moreover, the simultaneous propagation of temperature anomalies always leads that of other circulation anomalies. The arrival of these anomalous signals in the polar region corresponds to the occurrence of a positive or negative PVO event. The average time required for these anomalies to propagate from the tropics to the polar region is about 50–60 days (half of the PVO event cycle) (Fig. 2; see also Fig. 7 in Cai and Ren, 2007).

More importantly, synchronized with the pole-
ward propagation in the stratosphere (Figs. 2 and 3a) is an equatorward propagation of circulation anomalies in the lower troposphere (Fig. 3b). Despite of the difficulties in capturing the propagation signals in the lower troposphere due to the intersections of the isentropes with the ground, the equatorward propagation of the tropospheric anomalies still can be identified from the composites for different latitude zones (Figs. 3 and 6 in Cai and Ren, 2007).

The success in capturing the systematic propagation features shown above has benefited from the use of the semi-Lagrangian PVLAT coordinate as well as the PVO index we constructed to represent the leading oscillation of the stratospheric polar vortex. However, the \( \theta \)-PVLAT coordinate system also has some drawbacks due to the potential for isentropes to intersect the ground. To further confirm the existence of these propagating signals, and to avoid the difficulties in zonal averaging in the \( \theta \)-PVLAT coordinate due to intersections of isentropes with the ground, we performed lead/lag regressions of the circulation anomalies in pressure coordinates against the same PVO index. The results show that the temperature anomalies at pressure levels located in the stratosphere consistently propagate poleward, while the temperature anomalies at pressure levels located in the troposphere consistently propagate equatorward. We even identified changes in the direction of propagation on pressure levels that intersect the tropopause. Moreover, the downward propagation of temperature anomalies is typically confined in the stratosphere and tends to diminish around the tropopause. As a result, the vertical extent of downward propagation at higher latitudes is typically deeper relative to the vertical extent of downward propagation at lower latitudes, consistent with changes in tropopause height with latitude (Ren and Cai, 2007).

2.3 Out-of-phase temperature anomalies in the stratosphere and troposphere

Consistent with the downward propagation of zonal wind anomalies indicated by earlier studies, Fig. 3 also shows that zonal wind anomalies in the extratropics are consistently in phase between the stratosphere and troposphere during PVOs (contours in 

Fig. 2. A series of vertical-time cross-sections showing lead/lag regressions of isentropic temperature anomalies (shading; K) and zonal wind anomalies (contours; m s\(^{-1}\)) against the PVO index. The abscissa shows the lead-time relative to a PVO event and the ordinate of each panel is the isentropic vertical coordinate (ranging from 400 to 650 K). The anomalies are averaged over the PVLAT band indicated by the text to the right of each panel. The dotted black, solid black, and white contours show positive, zero, and negative values, respectively.
Fig. 3. 15-day running mean composite isentropic anomalies (temperature in K in shadings; zonal wind in m s\(^{-1}\) in contours) as a function of PVLAT (abscissa) and time (ordinate) averaged between (a) 400–650- and (b) 270–290-K isentropic surfaces poleward of 30\(^\circ\)N and 270–300-K isentropic surfaces equatorward of 30\(^\circ\)N. The time axis (ordinate) covers two PVO cycles. Zero contours are omitted for clarity. The solid black line in (b) indicates the position of 30\(^\circ\)N. The contours in (b) are plotted at 0.1, 0.4, 0.7 m s\(^{-1}\) (solid white), and –0.1, –0.4, –0.7 m s\(^{-1}\) (dashed black). The composites shown here are based on the relative intensity of the PVO index (see detailed descriptions in Cai and Ren, 2006, 2007).

Fig. 3). By contrast, in the polar region, the temperature anomalies in the stratosphere tend to be out of phase with those in the troposphere (shadings in Fig. 3). In other words, following the arrival of warm (cold) anomalies into the polar region in the stratosphere (shadings in Fig. 3), which leads to a polar warming event, temperature anomalies in the troposphere are anomalously cold (warm) and begin to propagate equatorward (Fig. 3b). Linear regression of temperature anomalies on pressure levels also confirms this out-of-phase relationship between the stratosphere and troposphere (Ren and Cai, 2007). This implies that temperature anomalies of the same signs do not propagate directly downward into the troposphere. The question is how to understand these out-of-phase temperature anomalies in the context of the downward propagation of zonal wind and geopotential height anomalies from the stratosphere to troposphere.

Further analysis in Cai et al. (2007) by applying the IPV theory (Hoskins et al., 1985) indicated that a positive/negative IPV anomaly in the upper atmosphere will induce a positive/negative vorticity anomaly and a negative/positive temperature anomaly in the troposphere (i.e., a cold low or a warm high). While in a relatively shallower surface layer, the polarity of the IPV anomaly is determined by the stability of the layer rather than the vorticity field, so that temperature anomaly of one sign is accompanied with vorticity anomaly of the opposite sign (a warm low or a cold high). Therefore, the out-of-phase coupling of
IPV anomalies (Fig. 4d) between the stratosphere and troposphere naturally leads to an out-of-phase coupling of the temperature anomalies (Fig. 4a), as well as an in-phase coupling of geopotential height (Fig. 4e) and zonal wind anomalies (Fig. 4c). However, it should be noted the in-phase relationship of geopotential height and zonal wind anomalies does not imply a direct downward propagation from the stratosphere to the troposphere, but rather reflects the different responses of the atmosphere to IPV forcing in the upper and lower layers (as indicated in Fig. 4c by the minimum in geopotential height anomalies near 300 K).

3. The stratospheric PVO and the global isentropic mass circulation

According to the theoretical result by Matsuno (1971) about the Stratospheric Sudden Warming (SSW), the critical altitude at which the planetary wave can reach will successively decrease following the weakening of westerlies due to wave breaking, seemingly related to the downward propagation of stratospheric circulation anomalies in the extratropics. However, the systematic downward propagation during positive PVO events (stronger polar vortices) cannot be explained by using the Matsuno theory. Even if the recovery of circumpolar westerlies after an SSW event is assumed to progress from upper to lower layers and thus may imply a downward propagation (Lu and Ding, 2013a), it is unclear why downward propagation still exists during positive PVO events that are not preceded by SSWs. Further, why does downward propagation consistently exist in all latitudes? What is responsible for the opposite equatorward propagation in the troposphere? Is there any relation of this equatorward propagation with the poleward propagation in the stratosphere? Cai and Ren (2006, 2007) and Ren and Cai (2008) applied the global mass circulation theory (Johnson, 1989) to investigate the stratosphere-troposphere dynamic coupling and proposed a physical explanation for the observed vertical and meridional propagations.

3.1 The global mass circulation in isentropic coordinates

The global mass circulation which was proposed by Johnson (1989) and his collaborators (Gallimore et al., 1981; Townsend et al., 1985) by using the climatological mean isentropic mass stream functions, is a Hadley-like, hemispheric scale, single cell meridional

**Fig. 4.** Vertical-time cross-sections of 15-day running mean composite anomalies of (a) isentropic temperature (K), (b) static stability (s$^{-2}$), (c) zonal wind (m s$^{-1}$), (d) isentropic PV (PVU), and (e) Montgomery potential (m$^2$ s$^{-2}$) averaged between 60° and 90°N. Positive anomalies are shaded. The abscissa is the timeline of a composite PVO event in days. The ordinate is the isentropic vertical coordinate in K. The vertical scale is enlarged slightly between 270 and 300 K as indicated by the heavy horizontal line. [From Cai and Ren, 2007]
circulation. It consists of a warm branch that upwells in the tropics and then goes poleward in the upper layer, and a cold branch that sinks in high latitudes and then back equatorward in the lower layer. They indicated that the global mass circulation is forced by the meridional heating/cooling gradient, as well as the existing wave effect. The strength of the mass circulation is therefore stronger in the winter hemisphere than in the summer hemisphere, and is generally stronger in the Northern Hemisphere than in the Southern Hemisphere. Benefited from the increased resolution of data and the development of research methods, the global mass circulation has recently been further developed (Pauluis et al., 2008; Cai and Shin, 2014), particularly in the realization of diagnosis on daily distribution and variability of the mass circulation. Based on the equation of isentropic mass changes, daily changes of isentropic mass consist of adiabatic mass convergence/divergence along isentropic surfaces and diabatic mass transport across isentropes (Cai and Ren, 2007; Cai and Shin, 2014). In contrast to the conventional representations of the meridional overturning circulation (the residual circulation and the Brewer–Dobson circulations), the mass circulation is able to explicitly resolve the zonal variations of mass transport on a daily basis.

3.2 Intraseasonal variability of the global mass circulation and stratosphere–troposphere coupling associated with PVO

Cai and Ren (2006, 2007) used the spatiotemporal evolution of isentropic mass anomalies to demonstrate the variation of the mass circulation during PVO events. They found that, accompanying the PVO, the isentropic mass anomalies also exhibit a systematic and simultaneous poleward and downward propagation; particularly, the mass anomalies in the troposphere also propagate oppositely equatorward, in agreement with the poleward warm branch and the equatorward cold branch of the mass circulation. The positive/negative tropospheric mass anomalies always begin to propagate equatorward following the arrival of positive/negative stratospheric mass anomalies in the polar region. This manifests the successive strengthening (weakening) of the upper warm branch and the lower cold branch of the mass circulation on seasonal timescales. Namely, the mass circulation strengthens during negative PVO events and weakens during positive PVO events (Figs. 5a and 5b). Moreover, the poleward propagation always seems to appear earlier in the upper stratosphere and later in the lower stratosphere, or the poleward propagating signals in the upper layer always lead that in the lower layer. This then explains the downward propagation of circulation anomalies existing in all latitudes.

More specifically, because the peak phase of negative PVO events (weaker polar vortex) (around day 75 of the abscissa in Fig. 5) just corresponds to the time when warm (positive mass) anomalies arrive in the polar region and just prior to the strengthening of

![Fig. 5. Latitude-time cross-sections of 15-day running mean composite mass anomalies (kg) (a) above 550 K, (b) below the lowest isentropic surface (270 K), and (c) integrated over the total column. Positive anomalies are shaded. The abscissa in (a–c) is the timeline of a composite PVO event in days, and the ordinate in (a) and (b) is PVLAT. [Adapted from Cai and Ren, 2007](Image)]
cold branch, the out-of-phase relationship of tempera-
ture anomalies between the stratosphere and the trop-
osphere is therefore inevitable. Meanwhile, the en-
hanced mass transport into the polar region by the
strengthened warm branch can lead to increases in sur-
face pressure (because it is before the strengthening of
the cold branch or before the cold air mass moves out
of the polar region) (Fig. 5c). At this time, a cold
surface high exists in the surface layer below a warm
high in the upper layer (stratosphere or upper trop-
osphere). Similarly, at the peak phase of a positive
PVO (stronger polar vortex) event, the warm branch
is anomalously weak with negative polar mass anoma-
lies. In this case, a warm surface low exists below
a cold upper-level low. These facts explain why the
geopotential height and the zonal wind anomalies in
winter polar region always exhibit an in-phase rela-
tionship between the upper and the lower layers. On
the other hand, the fact that surface cold high (warm
low) must weaken with height while upper warm high
(cold low) must strengthen with height (determined
by static stability relation), also explains the minimum
values of geopotential height anomalies in the middle
layer when there is an in-phase coupling between the
upper and the lower layers (ex. the minimum of geopo-
tential anomalies around 300 K in Fig. 4e). There-
fore, the downward propagation as well as the oppo-
site meridional propagation of circulation anomalies
from the stratosphere to troposphere, is primarily de-
termined by the coupling and successive changes of
the upper and lower branches of the mass circulation
during the PVO.

Cai and Ren (2006, 2007) further indicated that
the seasonal variability of mass circulation is ac-
companied by advancement of successive steepening
(strengthening) or leveling (weakening) of a series of
isentropic surfaces (or baroclinic zones) from the equa-
tor to the polar regions and from the upper to the
lower layer, and also the meridional exchange of cold
and warm air masses (Fig. 6). Specifically, due to the
earth’s rotation, the westerly flow associated with the
vertically sloped baroclinic zones that prevail from the
troposphere to stratosphere and from the subtropics
to the pole is a physical barrier for a direct merid-

3.3 The global mass circulation and the
Brewer–Dobson circulation

The discovery and confirmation of the Brewer–
Dobson circulation serve to explain the distribution
and transport of water vapor and ozone (Brewer, 1949;
Dobson, 1956). The Brewer–Dobson circulation up-
wells in the tropics, extends poleward, and then sinks
in polar region to complete the vertical and merid-
ional transport of water vapor and ozone. Since it
is inevitable that the Brewer–Dobson circulation also
transports air mass, a back equatorward branch was
also identified then (Brewer, 1949).

Based on the wave–mean flow interaction theory,
Haynes et al. (1991) derived the “residual circulation”
driven by an extratropical “air pump” due to wave forcing in the stratosphere. This theoretically proved existence of the Brewer–Dobson circulation in the stratosphere, and on the other hand, indicated the significance of wave forcing to the Brewer–Dobson circulation. However, the Brewer–Dobson circulation was defined on long-term mean basis, and the “residual circulation” was derived in a zonal-mean framework. In contrast, the mass circulation is based on the original total flow, and it can explicitly define the zonal structure of mass distribution and transport. Therefore, the mass circulation can be defined accurately on a variety of timescales, and by definition, it manifests the effects of both the meridional gradient of diabatic heating and wave activity. Moreover, components of mass transport by diabatic (vertical) and adiabatic (horizontal) circulation can be distinguished.

4. Interannual variability of the stratospheric PVO

The major variability of the stratospheric circulation is dominated by internal variability on seasonal timescales; however, significant interannual variability is also observed. The main factors responsible for interannual variability in the stratospheric circulation include ENSO, volcanic eruptions, and the quasi-biennial oscillation (QBO). Volcanic eruptions are irregular and the impact of the QBO is modulated by ENSO (Wei et al., 2007). ENSO, as one of the major sources for the interannual variability of the climate system, becomes a key impacting factor for the interannual variability of the stratospheric circulation. Early evidence indicated that Northern Hemisphere stratospheric vortex tends to be anomalously weak/strong during warm/cold ENSO winters (van Loon et al., 1982; Hamilton, 1995; Sassi et al., 2004; Taguchi and Hartmann, 2006; Manzini et al., 2006; Camp and Tung, 2007). Some evidence indicates that the strongest effects of ENSO on the stratosphere may appear months after the mature phase of ENSO (Chen et al., 2003; García-Herrera et al., 2006). Our studies, which are based on multiple sea surface temperature (SST) and reanalysis datasets, indicated that the responses of the extratropical stratosphere to ENSO forcing prevail in both the concurrent winter of mature ENSO and the next winter season after mature ENSO, and the polar anomalies in the next winter are much stronger and with a deeper vertical structure than that in the concurrent winter. This lead/lag coupling relationship between ENSO and the stratosphere is especially significant on the lead timescale of ENSO (3–5 yr), implying that the major impacts of ENSO on the extratropical stratosphere lie in its delayed effects (Ren and Xiang, 2010; Ren, 2012b; Ren et al., 2012). Actually, the delayed effects of ENSO on the troposphere have been indicated in earlier studies. For example, strongest temperature anomalies in the tropical troposphere always lag the ENSO peaks by
around one to two seasons (Newell and Weare, 1976; Angell, 1981; Yulaeva and Wallace, 1994; Huang and Huang, 2009). Although the tropical eastern Pacific SST anomaly is weaker during the summer following an ENSO peak, the response of upper tropospheric geopotential height tends to be even stronger than that in the previous summer (Kumar and Hoerling, 2003). The delayed impact of ENSO on the stratosphere is associated with planetary wave activity and interannual anomalies in the global mass circulation that are stimulated by ENSO. In specific, accompanying an ENSO event on approximately 3–5 yr timescale (Fig. 7a), stratospheric mass anomalies in 430–700-K isentropic layer exhibit an interannual timescale poleward and downward propagation, which appears in low latitudes in the initiating stage of the ENSO event (Figs. 7b and 7c), and persists in the midlatitudes from the previous summer before ENSO peak to the next summer after ENSO peak, though minor mass and temperature anomalies do appear in the polar region during the concurrent winter. Till the next winter season after ENSO peak, the most significant and strongest mass and temperature anomalies arrive in the polar region, completing a cycle of strengthening or weakening mass circulation successively from the tropics to the polar region on the interannual timescale.

Nevertheless, when an ENSO event exhibits a much shorter (less than 3 yr) or a much longer (more than 5 yr) timescale, the peak of ENSO may not appear in winter, and the coupling relationship between ENSO and the stratosphere may also be different from that described above. For example, an ENSO event that peaks in autumn will mainly affect the stratosphere in the following winter (Ren, 2012b). Both ENSO and stratospheric variability are phase-locked to the winter season, which defines the seasonality of their coupled relationship (Ren, 2012a). In general, due to the limitation of data length and numbers of observed ENSO cases, the statistical significance and stability of this coupling are not yet fully assured. This lack of certainty is particularly for the coupling processes associated with different types of ENSO. The effect of warm pool ENSO on the extratropical stratosphere may largely be associated with the phase of the QBO (Xie et al., 2012).

5. Influences of the Tibetan Plateau on stratosphere–troposphere coupling

The source/sink of PV in the “underworld” is affected directly by diabatic heating and friction at the surface. The Tibetan Plateau (TP) becomes an important source of sensible heating during spring and summer (Ye et al., 1957; Yu et al., 2011a, b), so that many of the isentropic surfaces near the TP intersect with the ground. The strong surface sensible heating over the TP drives an “air pump” that forces deep convection, leading to strong convective latent heat release over the TP during summer (Wu et al., 1997). Budget analysis of PV fluxes over the TP in July shows that the diabatic heat source over the TP induces a cyclonic circulation near surface. The surface friction effects of the TP represent a source of negative PV in the upper troposphere. Surface friction and diabatic heating together induce a strong anticyclonic circulation (the South Asian high or SAH) in the upper troposphere and lower stratosphere over the summer TP (Liu et al., 2001).

Numerical experiments with and without TP sensible heating show that the anticyclone at 200 hPa resulted from the source of negative PV over the TP, can stimulate a Rossby wave train that extends from the eastern coast of Eurasia to the eastern Pacific and North America. This wave train can pass the influence of the TP to remote areas via energy dispersion and has significant effects on the local circulation (Fig. 8) (Liu et al., 2001; Wu et al., 2002). The 200-hPa pressure level is located in the stratosphere in high latitudes, implying that diabatic heating over the TP in summer may influence not only the tropospheric circulation over Asia, but also the stratospheric circulation in high latitudes.

Figure 9 shows vertical cross-sections of the climatological mean July distributions of potential temperature (thinner lines), IPV (thicker lines), and meridional IPV advection (shadings) along the eastern flank of the TP (90°–115°E) and central Pacific (150°E–135°W), as well as the corresponding differences be-
Fig. 7. (a) Auto-regression of the Niño3 index, and lead/lag regressions of (b) stratospheric mass anomalies in the isentropic layer between 430 and 700 K (shaded; kg m$^{-2}$) and (c) stratospheric temperature anomalies averaged between 30 and 100 hPa (shaded; K) against the winter filtered Niño3 index. The lead time of the index varies from –18 to +18 months. Dashed line indicates the 90% confidence level. [Adapted from Ren et al., 2012]

Fig. 8. Differences in 200-hPa stream function between the control experiment (CON) and an experiment with no surface sensible heating over the summer Tibetan Plateau. Red vectors mark the Rossby wave trains. [From Wu et al., 1997]

tween these two regions. Diabatic heating over the TP raises the tropopause over the southern TP, while the isentropic surfaces are concave downward. The tropopause break is more severe in the TP region than in other regions, including the central Pacific. Isentropic surfaces that typically slope upward from south to north become much steeper and denser near the TP (thinner lines in Fig. 9a). Meanwhile, the powerful negative PV source in the upper troposphere over the TP yields strong negative PV anomalies in the upper troposphere and results in a steepening of the slopes of PV contours over the TP (especially in the 350–150-hPa layer, where PV contours over the TP are almost vertical in pressure coordinates). These approximately vertical PV contours represent the dynamical tropopause that separates the tropospheric low PV air to the south from the stratospheric high PV air to the north. The isentropic surfaces are approximately perpendicular to the PV contours over the TP. Northerly winds prevail over the east flank of TP, indicating the
Fig. 9. Vertical cross-sections of potential temperature (thin lines; K), potential vorticity (thick lines; PVU), and meridional PV advection (shaded; PVU s$^{-1}$) averaged over (a) 90°–115°E and (b) 150°E–135°W. The differences between these two regions are shown in (c). The topography of the plateau over 75°–105°E is shown as grey shading, and the approximate location of the tropopause is shown as green dashed line.

eastern equatorward branch of the SAH. These northerly winds tend to transport midlatitude stratospheric air with high PV into the subtropical Asian troposphere along the steep isentropic surfaces. The meridional gradients of isentropic and PV surfaces are much weaker over the central Pacific, so that meridional PV transport is mainly limited to the stratosphere (Fig. 9b). Comparison of Figs. 9a and 9b shows that the East Asian region is more strongly affected by the advection of high PV stratospheric air during summer than the central Pacific region, primarily because of the existence of the TP. High PV air can stimulate low-level cyclones and deep convection, which may explain the strength of monsoon precipitation over East Asia relative to other monsoon regions.

The analysis detailed above strongly suggests that the TP is not only a main pathway for STE, but also has significant impacts on stratosphere–troposphere dynamic coupling in nearby regions (including East Asia). Further investigation of the stratosphere–troposphere interactions over the TP region may become an important means of understanding how the TP affects the summer climate of East Asia.

As mentioned in Section 1, the TP is the largest source of planetary waves in the Northern Hemisphere during winter. Accordingly, the effects of the TP are fundamental factors in determining the location and strength of the East Asian westerly jet, as well as the formation of the East Asian trough (Zou et al., 1992a, b). However, the role of the TP in coupled stratosphere–troposphere variability (including the PVO) remains poorly understood. It is well known that the TP is a relatively weak heat sink during winter, although it can also serve as a heat source in some areas (Yu et al., 2011a, b), while the springtime increase in surface sensible heating over the TP is more rapid than that over other regions. Nevertheless, there is a general lack of studies focusing on the potential impacts of thermodynamic processes over the TP on stratosphere–troposphere coupling. In particular, further studies of the mechanisms by which the TP influences regional and the global stratosphere–troposphere interactions are needed.

6. Key issues in the studies of stratosphere–troposphere coupling

Studies of the stratosphere and STE have long
been limited by a lack of reliable long-term observations in the stratosphere and ambiguity in key parameters, such as the definition of the tropopause. Theoretical understanding of wave–mean flow interactions explains the mechanism of stratospheric warming events, but cannot provide a reasonable explanation for the meridional and vertical propagation of circulation anomalies during PVO processes. Although we have provided a qualitative physical explanation within the framework of the global mass circulation, no theoretical proof has yet been produced. The lack of understanding regarding stratosphere–troposphere coupling also limits our understanding of interannual variability in the stratosphere itself. In particular, the scientific community has yet to achieve consensus regarding the impacts of ENSO, the QBO, and volcanic eruptions on the stratosphere. The inability of current numerical models to provide realistic simulations of stratospheric processes is another important limiting factor (Charlton et al., 2007; Ren et al., 2009, 2012; Liu Y. Z., 2012; Rao et al., 2014). Improvement of numerical models in turn requires reliable observational data and sound theoretical foundations. The lack of data over the TP is particularly severe, so the improvement of numerical model performance in the vicinity of the TP represents an even bigger challenge.

More and more satellite data have begun to cover the stratospheric layer in recent years, and simulations of the stratosphere in numerical models have greatly improved by increases in model resolution and improved representations of dynamical and chemical processes. These developments will substantially advance our understanding of the stratosphere. Stratospheric research has also been attracting increasing attention in China. A key project recently implemented by the National Science Foundation of China (NSFC) titled “Land-air Coupling Processes over the Tibetan Plateau and Their Global Climate Effects” has adopted “stratosphere–troposphere interactions and their effects” as one of its major scientific topics. This project will provide an invaluable opportunity for studies on relationships between the TP and stratosphere–troposphere interactions. We therefore have reason to expect our understanding of the role of the TP in modulating regional and global stratosphere–troposphere interactions and climate change to develop rapidly in the near future.

REFERENCES

—–, and —–, 2007: Meridional and downward propagation of atmospheric circulation anomalies. Part I:


Mao Wenxuan, Wang Weiguo, Bian Jiahun, et al., 2008: The distribution of cross-tropopause mass flux over the Tibetan Plateau and its surrounding


