The January 1992 stratospheric sudden warming: A role for tropical inertial instability?

By SUZANNE M. ROSIER¹ and BRYAN N. LAWRENCE²

¹University of Oxford, UK
²University of Canterbury, New Zealand

(Received 15 June 1998; revised 5 February 1999)

SUMMARY

Twelve simulations of the stratosphere and mesosphere have been performed to investigate the near-major stratospheric warming of January 1992. All were initialized close to the observed warming. The resulting simulations fall into two classes: those that accurately reproduce the timing and anticyclonic merger mechanism of the observed warming and those that fail. The model used was a mechanistic stratosphere-mesosphere model, with a prescribed troposphere; thus differences among the simulations must only be due to differences among the initial conditions, which were constructed using different combinations of observed and climatological data. The main differences between the initial conditions of those simulations that succeed and those that fail are in low latitudes. A detailed analysis of two contrasting simulations reveals that the one which fails does not accurately reproduce the observed conveyor belt feeding tropical low-potential-vorticity (PV) air into the Aleutian high. Several factors appear to have combined to make the difference between the successful and the unsuccessful simulation of this sudden warming. The amount of low PV already present in higher latitudes in the initial condition appears to have an influence, as does the strength of the subtropical PV barrier. Also important is the Rossby wave structure which, by determining the degree of zonal asymmetry of the PV barrier, influences the injection of low PV into higher latitudes. In addition, however, we propose that the inertial stability of the tropical upper stratosphere might have played a significant role. In the unsuccessful simulation the tropical upper stratosphere is more inertially unstable than in the successful simulation; in this region more vigorous small-scale roll structures develop which, in stabilizing the flow, might have inhibited the northward transport of tropical air that appears to be a key part of the development of the observed sudden warming.

KEYWORDS: Inertial instability Observations Simulation Stratosphere Sudden warming

1. INTRODUCTION

The stratospheric sudden warming is a particularly dramatic winter event during which high-latitude stratospheric temperatures increase rapidly and the circulation is strongly disrupted. The first observations of such an event were by radiosonde over Berlin (Scherhag 1952) and since then satellite experiments have provided us with global data, enabling a fuller picture of the phenomenon to be obtained (e.g. Quiroz 1975). It has become obvious that there are many types of stratospheric warmings, but all the ‘major’ warmings† belong to one of two classes: ‘wave-two’ warmings, during which the polar winter vortex splits completely into two separated centres; and ‘wave-one’ warmings, during which the Aleutian anticyclone strengthens, and perhaps moves longitudinally, until it replaces the vortex over the pole. Good examples of these two classes are, respectively, that of February 1979 (e.g. McIntyre and Palmer 1983) and that of March 1993 (Manney et al. 1994).

The dynamical mechanism essential to the sudden warming was established by Matsuno (1971), who reproduced a strong wave-two warming using a numerical model which, in essence, represented tropospheric planetary waves propagating upwards through a realistic stratosphere. Later modelling work (e.g. Smith 1992) has shown that the existing stratospheric state affects the ability of tropospheric planetary waves to initiate stratospheric warmings, and supports the work of O’Neill and Pope (1988), who

* Corresponding author; present address: Department of Meteorology, University of Reading, Earley Gate, Whiteknights, PO Box 243, Reading RG6 6BB, UK; e-mail: s.m.rosier@reading.ac.uk.
† Defined as warmings where, in the zonal mean, a reversal of both the poleward temperature gradient (from negative to positive) and the zonal wind (from westerly to easterly) penetrates down to 10 hPa or below.
clearly demonstrated the highly nonlinear nature of the process, including the absence of a simple connection between developments in the troposphere and in the stratosphere.

This paper focuses on a strong stratospheric sudden warming that occurred in the northern hemisphere in the early part of January 1992. Following two previous minor warming events, a strong wave-one type warming proceeded with the merging of two warm-core anticyclones to form a single, intensifying high that displaced the winter vortex from the pole in a near-major warming (Rosier et al. 1994). Such a sequence of events seems to be a common feature of wave-one warmings, seen in many years of observations (A. O’Neill, personal communication) and in simulations (e.g. Pierce et al. 1993). The initiation of the vortex merger, and hence the sudden warming, requires a supply of air having lower potential vorticity (PV) than the air around it and here we propose that this might be influenced by the tropical stratosphere. Downward motion from aloft also plays a considerable role in the mechanism of this warming, but here we concentrate more on the horizontal motion and the influence of the tropics. In particular, in a series of simulations to be discussed we find that one of the factors influencing the accuracy of the evolution of the northern polar stratosphere could be the degree of inertial instability occurring in the tropical upper stratosphere. The simulations were performed with a model of the stratosphere and mesosphere which has a prescribed troposphere, and differed only in the type of data used to initialize the model fields and the date of initialization.

In what follows, we begin by describing the model and datasets used (section 2) and then summarize some of the relevant literature on stratospheric inertial instability (section 3). Next we present a short review of the actual observations of the January 1992 sudden warming (section 4). We then proceed to interpret this event by use of model simulations; in section 5 we discuss the range of initial conditions and, in particular, one successful and one unsuccessful simulation, with a view to diagnosing the most crucial processes in the development of the warming. This discussion is carried further in section 6 before the paper is concluded in section 7.

2. Numerical issues

(a) The numerical model

The UK Meteorological Office (UKMO) Stratosphere–Mesosphere Model (Fisher 1987) is a global, three-dimensional primitive-equation model of the middle atmosphere, having temperature and horizontal wind as prognostic variables. It has been used extensively, although to date mainly for studies tending to focus on the extratropical stratosphere, e.g. Butchart et al. (1982), Marks (1988), O’Neill and Pope (1988), Fairlie et al. (1990), Farrara et al. (1992). Such studies have shown that the model is capable of reproducing stratospheric behaviour with a substantial degree of accuracy, including both major and minor sudden warmings.

The model avoids having to represent the troposphere by imposing a lower boundary condition near the tropopause. The geopotential height of the 100 hPa surface (~16 km) is specified by a sequence of daily fields which are linearly interpolated in time to provide a continuously varying field.

Radiative heating and cooling are computed using the MIDRAD scheme described by Shine (1987). The effect of breaking gravity waves on the mean flow has been parametrized using a standard Rayleigh-friction scheme. The relaxation time-scales are similar to those employed in most models, decreasing from about 100 days in the lower stratosphere to about two days in the mesosphere.
(b) The initial data

The simulations performed in this study were initialized with various combinations of data, because daily data were not available for the whole model domain. Daily data from both the Improved Stratospheric and Mesospheric Sounder, ISAMS (Taylor et al. 1993) and the UKMO stratospheric assimilation system (Swinbank and O’Neill 1994) were used, along with climatological data from the CIRA* dataset (Barnett and Corney 1985a and b). The CIRA data were used primarily in the mesosphere, where daily data were not generally available.

The initial conditions required for each simulation were the temperature and horizontal winds throughout the model domain. In order to produce dynamical consistency among the different datasets a procedure was carried out which began by re-gridding the temperature data to the model grid. This re-gridding included a merging in the latitudinal and vertical directions between the datasets, which was carried out using a quarter–half–quarter filter (i.e. taking \( \frac{3}{4} + \frac{1}{2} \), \( \frac{1}{2} + \frac{1}{4} \) and \( \frac{1}{4} + \frac{3}{4} \) of the two datasets respectively). When only UKMO data were used in the stratosphere the horizontal merging was not necessary; however, when ISAMS data were used an amount of other data was necessary, since during the period of interest ISAMS was restricted in latitudinal coverage. In this latter case the merging was across latitudes 22.5, 27.5 and 32.5°S, except for one combination (see section 5) where fewer ISAMS data were included; in this case the data were merged across 22.5, 17.5 and 12.5°N. In most cases the stratospheric data needed to be merged vertically with the CIRA data across 52, 54 and 56 km. The resulting temperature fields were found to be reliably smooth across the merging altitudes and latitudes.

Geopotential heights were calculated from this hybrid temperature field, and horizontal balanced winds calculated from the heights using a method based on that of Randel (1987). Between 10°N and 10°S, wind values were simply interpolated from the adjacent latitudes and are therefore less reliable in this region.

The full set of initial conditions constructed in this way is detailed in Table 1 in section 5, where the exact combinations are discussed in more detail. Six different initial data combinations were tested, some of which were used for simulations starting on different days; in total twelve simulations were performed. All were initialized in early January and run through to 17 January 1992, and all were supplied with UKMO assimilated data for the lower boundary (i.e. the geopotential height of the 100 hPa surface).

3. INERTIAL INSTABILITY

In an inertially unstable region of the atmosphere, the flow is unstable to small perturbations in the meridional direction. Theory predicts that this will occur when

\[
f \left( f - \frac{\delta f}{\beta y} - \bar{u}_y \right) < 0,
\]

where \( f \) is the Coriolis parameter \( (f = 2\Omega \sin \phi) \), \( \Omega \) is the earth’s rotation rate of \( 7.27 \times 10^{-5} \text{s}^{-1} \) and \( \phi \) is latitude, \( Ri \) is the Richardson number and \( \bar{u}_y \) is the latitudinal gradient of the zonal mean zonal wind). For large \( Ri \), as is always the case for the situations presented here, the second term in the parenthesis can be ignored, and we see that inertial instability occurs when the vorticity arising from the meridional shear of the zonal flow is greater in magnitude than the planetary vorticity and of opposite sign.

* COSPAR (Committee on Space Research) International Reference Atmosphere.
This condition is mostly likely to occur in equatorial regions of the upper stratosphere and mesosphere, where $f$ is small and $\bar{u}_y$ can be large. Given large $\text{Ri}$, from (1) it is easily shown that such regions where inertial instability is possible have anomalous PV, i.e. negative in the north and positive in the south.

Dunkerton (1981) investigated the effect of zonally symmetric disturbances on an equatorial beta-plane; his theory predicted that the structure arising as a result of inertial instability should consist of vertically stacked small-scale ‘rolls’ of alternating positive and negative perturbations to the horizontal wind, correlated in such a way that their overall effect is to transfer zonal momentum meridionally until the zonal wind profile is stabilized. At the time there was little evidence of such structure in numerical models, a fact that was attributed to insufficient vertical resolution. However, the primitive-equation General Circulation Model (GCM) used by Hunt (1981), having vertical resolution of 2.5 km from the middle stratosphere to 100 km, revealed closely packed layering in low winter latitudes of the January mean zonal mean meridional velocity, $\bar{v}$, fields from the middle stratosphere almost to the mesopause, very similar to that predicted by Dunkerton.

Observational support for the existence of such structures came with the Limb Infrared Monitor of the Stratosphere (LIMS) data examined by Hitchman and Leovy (1986) and Hitchman et al. (1987). Hitchman et al. observed persistent, stationary equatorial disturbances consisting of two or three vertically stacked temperature extrema of alternating sign, to which they attached the name ‘pancake structures’; these were observed in the equatorial lower mesosphere in regions of weak or negative inertial stability. Hitchman and Leovy also observed the signature of small-scale equatorial meridional circulations, similar to the predicted inertial cells, in these regions.

The extension of work to consider non-zonally symmetric inertial instability clarified the role of Rossby waves (e.g. Boyd and Christidis 1982; Dunkerton 1983, 1993; Clark and Haynes 1996). Dunkerton (1993) found inertial instability modes to exist in localized regions of anomalous PV, with the largest growth rate occurring for local stationary instability. He asserted that Rossby waves propagating up from the troposphere propagate into the tropics and pull long tongues of air from one hemisphere to the other, thus organizing local regions of anomalous PV and, hence, inertial instability. Intrusions of negative PV into the winter hemisphere were observed to be much more common and more penetrating than positive PV into the summer hemisphere, behaviour which is probably assisted in the upper stratosphere and mesosphere by the usual summer to winter flow at those altitudes. Thus the role of Rossby waves in organizing regions of inertial instability was found to be two-fold: as well as directly organizing local regions of anomalous PV, more indirectly they force the residual mean circulation, advecting summer easterlies across the equator and increasing the cross-equatorial shear ($\bar{u}_y$) to the point of instability. The modelling results of O’Sullivan and Hitchman (1992) and Sassi et al. (1993) provide good examples of such inertially unstable intrusions of anomalous PV. In addition, they both indicate closely packed layering in equatorial meridional wind fields in the upper stratosphere and mesosphere.

The time averaged effect of the equatorial inertial circulations, namely the redistribution of zonal momentum until the zonal wind profile is stabilized, is expected to lead to a suppression of the usual summer-to-winter flow (Hitchman et al. 1987). In addition, O’Sullivan and Hitchman (1992) pointed out that the mixing effect of the inertial circulations is expected to contribute to maintaining the ‘surf zone’ of weak PV gradients in low winter latitudes which results from ‘Rossby wave breaking’ (McIntyre and Palmer 1983, 1984). It should be noted that whilst the effect of these inertial rolls is thus to confine easterlies to the summer hemisphere, such a process in the real atmosphere
cannot be uniquely attributed to inertial instability since a similar effect should result from the westerly wave driving (and associated residual mean circulation) associated with Kelvin and gravity waves (Hitchman and Leovy 1986).

As will be seen, the model simulations used here to investigate the sudden warming of January 1992 show distinct regions of inertial instability in the tropical upper stratosphere and mesosphere; in this region the inertial stability criterion, \( f (f - \bar{u}_y) \), is negative and layered pancake structures are seen in fields of \( \vec{v} \). In addition, low and negative PV is seen to intrude into the winter hemisphere. It is speculated that the degree of inertial instability present might play a role in the evolution of the sudden warming by altering the penetration of this anomalous PV into the Aleutian anticyclone.

4. The stratospheric warming of January 1992

As mentioned in the introduction, the stratospheric warming of January 1992 follows a sequence of minor warmings which has been described by Rosier et al. (1994), Ruth et al. (1994), O’Neill et al. (1994) and others. The warming proceeds by means of a vortex merger event, where anticyclonic material is brought from lower latitudes into the Aleutian anticyclone. The warming occurs, in part, because the Aleutian anticyclone has become large enough to push the normal winter cyclonic vortex far enough from the pole to allow warm midlatitude air to be brought northward by the stratospheric jet between the cyclone and the anticyclone. Strong descent from above also assists the process.

Figure 1 presents synoptic maps of the vortex merger which at 32 km occurs around 9 January. The first panel is for 3 January and shows the warm core westward of its associated anticyclone at around 180°E and 60°N. The second panel, for 8 January, shows the warm core associated with the original anticyclone, slightly eastward of its position on 3 January, along with a second warm pool at around 80°E which is associated with an anticyclone above. The adiabatic warming accompanying strong descent appears to have been responsible for the appearance of this second warm pool. It is the merger of these anticyclones in the middle and upper stratosphere which results in the strong anticyclone present on 12 January which has pushed the polar vortex further from the pole. This third panel shows that the flow brings warm air northwards from the warm pool west of the anticyclone core. This advection contributes substantially to the polar warming.

The repeated injection of anticyclonic air preceding this period is clearly associated with the advection poleward of air from the tropics as can be seen in the ISAMS \( \mathrm{N}_2\mathrm{O} \) measurements discussed by Ruth et al. (1994). This northward advection is associated with asymmetric circumpolar flow (as depicted, for example, in Fig. 6 of Ruth et al. (1994)). That is, tropical air is being brought north by the jet between the Aleutian high and the polar vortex from the jet entrance region south of 30°N.

It appears, then, that to produce a successful simulation of early January 1992 it is necessary to reproduce the poleward advection of tropical air, which contributes to the anticyclonic vortex merger and the resulting sudden warming.

5. Model simulations

(a) Initial conditions

In investigating the January 1992 sudden warming and the sensitivity of the simulation quality to the initial conditions (a ‘quality’ simulation being one which reproduced the salient features outlined above), in total twelve simulations were carried out. All
Figure 1. Polar stereographic maps of the observed northern hemisphere temperature (solid contours, interval 5 K) and geopotential (dotted contours, minimum contour 28.9 km, interval 250 m) at 32 km on 3, 8 and 12 January 1992. The outermost latitude is 2.5°N; latitude circles at 30 and 60°N are shown. The temperature fields are from ISAMS (see text) version 10, and the geopotential fields were constructed by stacking ISAMS temperature thicknesses onto UKMO assimilated geopotential heights.

were carried out with the same lower-boundary data, and differed only in the type of data used to initialize the model and the start date. Table 1 summarizes the six initial data combinations tested and presents the notation used to refer to them.

The notation cites 'U' (UKMO assimilated data), 'I' (ISAMS) and 'C' (CIRA) in approximate order, from most to least, of the amount of each data type in the initial condition. Data were merged latitudinally and/or vertically as described in section 2. The notation for a particular combination has the day in January on which the simulation started added; for example, simulation UC3 was initialized with UKMO and CIRA data on 3 January. Note that in Table 1 a slight shorthand has been used. Where different data appear for 'Strat' (stratosphere) and 'Mes' (mesosphere), these datasets have been merged across 52, 54, 56 km; thus 'Strat' data include some data from the lower mesosphere. Where different data appear for 'N' (north) and 'S' (south), except for
<table>
<thead>
<tr>
<th>Notation</th>
<th>Initial data combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC</td>
<td>Strat: UKMO, Mes: CIRA</td>
</tr>
<tr>
<td>IC</td>
<td>N: ISAMS, S: CIRA</td>
</tr>
<tr>
<td>IUC</td>
<td>N: ISAMS, S: Strat: UKMO, Mes: CIRA</td>
</tr>
<tr>
<td>CIU</td>
<td>Strat: N: ISAMS, S: UKMO, Mes: CIRA</td>
</tr>
<tr>
<td>UIC</td>
<td>Strat: UKMO, Mes: N: ISAMS, S: CIRA</td>
</tr>
<tr>
<td>CUI</td>
<td>Strat: N: ISAMS (to 22.5°N), S: UKMO, Mes: CIRA</td>
</tr>
</tbody>
</table>

'U' (UKMO assimilated data), 'I' (ISAMS) and 'C' (CIRA) are cited in approximate order of the amount of each data type in the initial condition, from most to least. The notation has the day in January marking the start date of the simulation added.

Combination CUI, merging has been performed across 22.5, 27.5, 32.5°S; thus N data actually include data as far south as 32.5°S, as well as northern hemisphere data.

It was originally anticipated that those simulations which incorporated ISAMS data would be more accurate than those without. This was expected for three reasons: better precision and accuracy of the ISAMS temperature measurements than the data incorporated in the UKMO assimilation; better vertical resolution in the ISAMS data; and daily data in the lower mesosphere, as opposed to climatology. However, as will be seen, simulation UC3 (without ISAMS data in the initial condition) reproduced very accurately the subsequent ISAMS observations of the warming via anticyclone merger, whilst simulation CIU3 (with stratospheric ISAMS data introduced into the initial condition) failed. Since all twelve simulations were supplied with the same lower boundary, differences among the simulations must be due to differences among the initial conditions. As will be seen, these differences were in the tropics and subtropics and probably arise from difficulties in calculating winds from temperatures in the tropical region, rather than in the extra-tropics where the better quality of the ISAMS data was expected to improve the simulation.

Figure 2 shows the zonal mean initial temperature field from combinations UC3, IC5, IUC5, CIU3, UIC5 and CUI5. Simulations UIC5 (which had ISAMS data restricted to the mesosphere) and UC3 (no ISAMS data) succeeded in reproducing the vortex merger; simulations IC5, IUC5, CIU3, and CUI5, which all incorporated stratospheric ISAMS data in the initial condition, failed. The corresponding zonal mean zonal wind distributions are depicted in Fig. 3. (Recall that these winds were linearly interpolated between 10° either side of the equator.)

The most obvious difference in the initial conditions between those that succeed and those that fail is the presence in those that fail of an 'upward bulge' in temperature at the equatorial stratopause which is associated with a strong 'nose' of easterlies pushing northward across the equator and a strong latitudinal gradient of the zonal wind, \( \bar{u}_y \). This bulge is present in the ISAMS data; it has nothing to do with the merging of data.
In addition, all the unsuccessful simulations have a much stronger subtropical westerly jet in the middle and upper stratosphere in the northern hemisphere (centred near 20°N and 40 km); this is associated with stronger, more zonally symmetric latitudinal PV gradients, as will be described in simulation CIU3 (subsection (c)).

A comparison of the initial zonal winds of fields UC3 and CIU3 in particular (panels one and four of Fig. 3) is very informative. These two wind fields led to sharply contrasting simulations, yet differ only in the equatorial easterly jet at the stratopause and the strength of the subtropical westerly jet in the northern hemisphere. Otherwise, they have every major feature in common, including the southern mesospheric easterly jet and the three maxima in the structure of the northern westerlies; all are of very similar magnitude between the two fields.

These differences in the initial conditions hint not only at differences in the inertial stability of the tropics but also at differences in the structures which produce the zonal mean fields. We therefore need to examine the full three-dimensional evolution of the simulations. To concentrate on the important points, we have limited our extended discussion to two simulations which best encapsulate the range of our ensemble. We
begin with a successful simulation and then examine an unsuccessful simulation before discussing possible reasons for the differences.

To examine transport in the simulations, maps of PV on constant potential-temperature surfaces were used. Under frictionless, adiabatic conditions PV is conserved on such isentropic surfaces. For the stratosphere and for the scales of interest here the adiabatic approximation is probably acceptable for a number of days. As will be shown, regions of negative PV in the northern hemisphere are of particular interest since such regions should be inertially unstable; conservation of isentropic PV implies that such air originated in the southern hemisphere. The gross features to be discussed here were confirmed as robust by comparison with PV calculations using different algorithms, and also by further simulations in which the model included a dry convection scheme (by ensuring stability of the vertical gradient of potential temperature it is confirmed that the anomalous PV (negative in the north) arises from the anticyclonic vorticity).
Figure 4. As second and third panels of Fig. 1 (temperature and geopotential at 32 km on 8 and 12 January), but data from the accurate simulation UC3 (see text).

(b) A successful simulation: UC3

Simulation UC3 was initialized on 3 January with UKMO assimilated data in the stratosphere and CIRA climatology in the mesosphere. It succeeds in capturing not only the key features in the evolution of zonal mean flow fields but also the observed synoptic evolution of this sudden warming, all with accurate timing.

The simulation begins during a period of enhancement of wave one in both the temperature and geopotential fields. By 8 January, however, the situation is somewhat different; Fig. 4 shows temperature and geopotential fields at 32 km on 8 and 12 January, for direct comparison with the ISAMS data in the second and third panels of Fig. 1.

Clearly the simulation reproduces the appearance of the second warm core, which first appears the day before at about 60°E and subsequently warms, as observed. Correspondingly, the simulated field of geopotential at 40 km is now also characterized by a three-vortex structure, as observed. By 8 January modelled temperatures in the cold core and the second warm core are only slightly (~5 K) warmer than observed.

The simulated warming continues to evolve very much as observed; the two warm cores and the concomitant anticyclones at 40 km merge one day later, and the resulting temperature and geopotential maxima move swiftly northeastward. The second panel of Fig. 4 shows that the model structures are still very similar to observations on 12 January, although by this time maximum observed temperatures are peaking at 260 K, whereas simulated temperatures are still continuing to increase, peaking at 280 K two days later.

The key events in this simulation are highlighted in Fig. 5, which shows polar stereographic maps of PV on the 1250 K isentrope, near 40 km, on 3, 7, 9 and 13 January. On the initial day of the simulation, 3 January, the strong cyclonic vortex located over the pole but towards the Greenwich Meridian can be easily identified from the region of high PV. The Aleutian high is also clearly visible as the well-defined region of low PV centred just west of the date-line. Also apparent are two other localized regions of low PV in the extratropics, one centred at about 30°E and the other at about 45°W where it shows up weakly in the geopotential at 40 km (not shown) as a slight ridging in the equatorward side of the jet stream.
Over the next few days, this structure as a whole rotates slightly eastward until the Aleutian high is just east of the date-line, by which time it has intensified—the corresponding geopotential heights have risen at 40 km. The intensified Aleutian high strips the vortex of its high-PV air, which is drawn out into long, zonally aligned tongues at about 30°N, stretching from about 270°E westward through almost 180° of longitude. The remains of such tongues can be seen on 7 January (second panel of Fig. 5). By this time, vortex stripping has abated slightly but an associated feature is still visible in the sector from 180 to 270°E. At about 20° latitude (on the equatorward flank of the vortex tongues), latitudinal PV gradients have increased considerably. Strong PV gradients are also now visible from 270°E east to 0°E, tending towards higher latitudes as 0°E is approached. Such strong latitudinal PV gradients are expected to act essentially as barriers to local rapid isentropic mixing across them (e.g. McIntyre and Palmer 1984, Trepte and Hitchman 1992), but they allow rapid transport along the PV contours. Thus the associated circulation introduces low-latitude and even southern hemisphere air (low and negative PV) into higher latitudes by advecting it from about 225°E in
towards the Aleutian high along a large inward spiral, as can be seen in the second panel of Fig. 5. Similar behaviour in trajectory studies is reported by O’Neill et al. (1994). In this manner the second anticyclone, and ultimately the Aleutian high, is maintained by injections of low-PV air from low latitudes and the southern hemisphere. PV calculated from the ISAMS data (using balanced winds from the geopotentials calculated from ISAMS temperatures) shows very similar behaviour, with negative PV penetrating to about 30°N. Very similar behaviour is also observed in the ISAMS N₂O fields (Ruth et al. 1994) and in high-resolution tracer advection calculations using UKMO assimilated winds (H. L. Rogers, personal communication).

Examination of the PV in both hemispheres on all days has confirmed that, consistent with the diabatic circulation, at this altitude and at 1850 K (near the stratopause), any cross-equatorial flow is predominantly from south to north (summer to winter), in the sense that significant amounts of negative PV are present in the northern hemisphere but no positive PV in the southern. Lower down, at 850 K, the hemispheres are much more isolated from each other.

Following the merger on 9 January, as the warming proceeds, the strengthened Aleutian high strips the vortex significantly of its high-PV air by drawing out long tongues from the vortex (again from about 270°E westward through almost 180° of longitude); these eventually become mixed irreversibly with the ambient lower-PV air or get lost through the effects of numerical truncation in the model. The last panel of Fig. 5 shows the situation on 13 January. Although, at this altitude, the PV should not be considered conserved over this whole simulation period, day-to-day consistency in the fields indicates that, over as little as two days, the vortex at this time has shrunk considerably as it is drained of its high PV by the growing anticyclone, which also displaces it from the pole towards the Greenwich meridian.

PV fields at 1850 K, near 48 km, (not shown) reveal very similar behaviour to that lower down, i.e. low and negative PV is drawn up from low latitudes at about 225°E to higher latitudes in a long incursion, and assists in the formation of a second anticyclone that subsequently merges with the Aleutian high and displaces the vortex from the pole. However, at this altitude, this all occurs slightly before it does lower down, consistent with the fact that the warming proceeds from the top down, as is common (e.g. Holton 1980), and also perhaps consistent with stronger northward meridional flow at higher altitude. In addition, vertical velocity fields (not shown) reveal descent along a similar spiral to that of the low PV. Thus, the sudden warming in this simulation appears to be initiated in the upper stratosphere with the drawing up of low-latitude air (possibly even summer hemisphere air). At the altitudes initially involved, the diabatic circulation must be assisting in this process.

(c) A less accurate simulation: CIU3

Figure 6 shows the rather different behaviour produced when ISAMS data are introduced into the initial condition; it depicts temperature and geopotential at 32 km on 8 and 12 January from simulation CIU3. This was initialized on 3 January with ISAMS data in the stratosphere in the northern hemisphere and south to 32.5°S, UKMO assimilated data in the remainder of the southern stratosphere and a CIRA mesosphere. Despite the reasonably accurate reproduction of the main vortex/anticyclone structure, simulation CIU3 clearly fails completely to reproduce the appearance of the second warm core that builds at around 60°E in both the ISAMS and UC3 fields. Correspondingly, no secondary high is apparent in the CIU3 geopotential at 40 km.

A day later, the single warm core in CIU3 does begin to warm; however, this occurs rather more in situ than is observed, i.e. no merging of two separate temperature maxima
(or geopotential highs) is apparent. Over the next three days, the core of maximum temperature moves little; by 12 January (Fig. 6, second panel), the overall structure of the CIU3 temperature field is rather more in agreement with observations, although once again the modelled temperature maximum at this stage is still warming whilst in the observations it is cooling. Such an extension of the warming period also occurs in many of the other simulations; it appears to be a feature of the model.

Figure 7 shows the PV at 1250 K from simulation CIU3, for direct comparison with the UC3 fields in Fig. 5. On the initial day the distribution of PV is of course broadly similar, as it should be since these are two different-data versions of the same day. However there are two major differences between the two distributions. The first is the absence in CIU3 of the distinct extratropical low-PV regions at 45°W and 30°E which are seen in UC3. The second, related, feature is the strong subtropical PV gradient in CIU3 which exists at all longitudes at about 30°N, and showed up in the zonal mean as the strong westerly jet (Fig. 3). In UC3 this gradient is not nearly as strong in the region from 270°E eastwards to 90°E—precisely the area where the low-PV regions are already established in UC3.

It seems that these differences in CIU3's initial conditions have reduced the likelihood of establishing a good low-to-high latitude connection like that in simulation UC3 i.e. the long low-PV incursion which ultimately, via the vortex merger, feeds the Aleutian high and helps the warming to proceed.

Over the next few days in CIU3, enhanced low-latitude westerly circulation does begin to draw material northward and towards the Greenwich meridian, but the detail is quite different. The polar vortex is larger in area (compared with UC3) and the incursion of tropical low-PV air only seems to reach high latitudes about a week later (see the PV distribution for 13 January). Thus it would seem that the location, shape and alignment of the cyclonic vortex (i.e. the Rossby wave structure), especially in the sector 0 to 90°E, are rather less favourable to the injection of this low PV towards higher latitudes (and, ultimately, the Aleutian high). In this case the Aleutian high does not grow as rapidly as in simulation UC3 and vortex stripping is much less marked.
Figure 7. As Fig. 5 (PV on the 1250 K isentrope, near 40 km) but fields from the less accurate simulation CIU3 (see text).

It is interesting to note that a positive feedback could be at work here. It appears that the confinement of low PV to lower latitudes (in CIU3) leads, ultimately, to the non-amplification of the Aleutian high and thus the maintenance of a stronger cyclonic vortex, which in turn continues to confine low PV to lower latitudes. In the successful simulation UC3, the injection of low PV into higher latitudes builds up the anticyclone, which erodes the vortex, allowing more low PV into higher latitudes.

6. DISCUSSION

The comparison of several sets of initial conditions in section 5(a) revealed that the main differences between the initial conditions of successful and unsuccessful simulations were: (i) at the equatorial stratopause, where increased $\bar{u}_y$ in unsuccessful simulations implies greater inertial instability; (ii) in the winter subtropical stratosphere, where a stronger jet in unsuccessful simulations implies a stronger PV barrier to northward transport of low PV. Examination of the synoptic transport in sections 5(b) and (c) indicated a probable influence of the differing strength of the PV barrier on the subsequent evolution of the sudden warming, essentially confirming point (ii) above. In
addition it revealed two other potentially important influences, namely the amount of pre-existing low PV in higher latitudes and the Rossby wave structure being favourable to the injection of low PV into the Aleutian high. It is quite possible that these factors, or any one of them alone, might have made the difference between a successful and unsuccessful simulation of the sudden warming. However, it is also worth investigating point (i) above, namely the part that differing inertial stability of the tropical upper stratosphere might have played in the differing simulations. This is particularly so given that the inertially unstable region is precisely the origin of the air which helps push the conveyor belt of low PV towards the Aleutian high; model vertical velocity fields (not shown) reveal strong descent along the spiralling low-PV incursion, indicating that the low-PV air which ultimately builds the Aleutian high in the mid-stratosphere originates from the tropical upper stratosphere.

Figure 8 shows the inertial instability criterion on 3 January for simulations UC3 and CIU3. As expected, the values are approximately constant with height at constant latitude (the term effectively represents the absolute angular momentum). It can be seen that in the initial conditions of both simulations, there is an unstable region in the low latitudes of the winter hemisphere extending from the upper stratosphere to the top of the model domain. There is no inertially unstable region in the middle and lower stratosphere; since inertial instability affects, and can arise because of, ambient cross-equatorial flow, this is consistent with the decreasing cross-equatorial transport seen with decreasing altitude. The zonal mean inertial instability criterion is violated in a slightly larger area of the upper stratosphere in the unsuccessful simulation than in the successful simulation. This is associated with the large symmetric PV barrier at 30°N in CIU3, which is a direct consequence of the zonally symmetric area of air with southern hemisphere values of PV (which itself arises from the way the initial winds are calculated).

In both simulations structures develop in the tropical and subtropical stratospheric and mesospheric meridional wind (Figs. 9 and 10) which are consistent with their identification as the model representation of the pancake structures discussed earlier. Such pancake structures are the signature of inertial instability. The scale and location of these features are consistent with observations (Hitchman and Leovy 1986) and other modelling studies (Hunt 1981; Sassi et al. 1993) and their vertical wavelength is at two gridlengths, which is what would be expected in a model such as this (O'Sullivan and Hitchman 1992). The intruding tongue of anomalous (negative) PV seen in the PV fields (Figs. 5 and 7) is most likely the zonally asymmetric manifestation of inertial instability; such was also seen by Sassi et al. and O'Sullivan and Hitchman. Since in this tropical upper stratospheric region $\vec{\nu}$ is a very good approximation to the residual mean meridional velocity $\vec{\nu}^*$, we see that the inertial instability has disturbed the usual summer-to-winter pattern of mass transport here.

Comparing Fig. 9 with Fig. 10 we see that these inertial 'rolls' set in earlier in CIU3 than in UC3 and are of greater spatial extent. By reference to Eq. (1) we can see that this implies that in the unsuccessful simulation CIU3 the more disturbed meridional flow is acting more vigorously to re-establish stability via the net southward transport of westerly momentum. The stabilizing action of these rolls might affect the subsequent warming in either or both of two ways: (i) the disturbance to the meridional wind might inhibit the northward motion of southern hemisphere air; (ii) the overall stabilizing effect should be to move the PV barrier southward, perhaps tightening it, and zonally 'symmetrize' the low-PV incursion, thus inhibiting the northward transport of low PV in CIU3, and thereby inhibiting the sudden warming. It might be said that (i) is a more local effect and (ii) a more zonal mean effect; in either case the overall effect of the
stabilization of greater inertial instability would be to reduce the efficacy of the conveyor belt mechanism that feeds low PV into the Aleutian high.

Both simulations are inertially unstable to begin with, however, and this is most probably associated with the intrusion of anomalous PV into the winter hemisphere which in UC3 helps the warming proceed; thus it appears that we cannot overlook the importance of the Rossby wave structure in injecting this low PV into the required location to build up the Aleutian high. However, it remains possible that greater inertial instability in the unsuccessful simulation impedes the subsequent warming, since once the stabilization process is underway this could impede further penetration of low PV into high northern latitudes. In reality the occurrence of this sudden warming is most likely associated with an extremely complex interplay between the Rossby wave structure and the effects of inertial instability.

Figure 11 shows the instability criterion on 14 January for simulations UC3 and CIU3. Comparison with the 3 January field (Fig. 8) confirms that over this pre-warming and warming period the flow has been stabilized, presumably by redistribution of zonal
Figure 9. Latitude–height sections of the zonal mean meridional velocity, $\bar{v}$, at all latitudes on 5, 8 and 11 January, from the accurate simulation UC3 (see text). Contour interval is 2 m s$^{-1}$ and regions of southward (negative) velocity are shaded.
Figure 10. As Fig. 9 but from the less accurate simulation CIU3 (see text).
momentum by the inertial rolls. Such behaviour was also noted by Sassi et al. (1993) in their model study. Correspondingly, meridional wind fields after the warming (not shown) reveal that the vertically stacked pancake pattern is less extensive in both simulations.

The precise details of the differences between these two simulations are clearly extremely difficult to diagnose, which is unsurprising given the difficulty in diagnosing the details of the processes arising from inertial instability (Dunkerton 1981), but the possible role of inertial instability here is clear. In the case where the tropics are more inertially unstable, once the inertial circulations have been established the necessary northward advection of low-PV air required to build up the Aleutian high cannot progress to the same extent (although the Aleutian high can still build in situ by disturbances from below and/or above).

Many of the observations described here could be recast in terms of a discussion of Rossby wave propagation. Arguments of this sort have been made for a relationship between the quasi-biennial oscillation (QBO) and stratospheric warmings (e.g. Dunkerton and Baldwin 1991; Kodera 1991). However, we have chosen to stay with synoptic descriptions, as we believe that further progress in this area will be made by describing
accurately the phenomena that need explanation before trying to understand nonlinear, or even linearized, models based on Rossby waves. It would certainly appear that the importance of inertial instability as a potential inhibitor of the northward flow of low-PV air would also apply in discussions of mechanisms leading to a correlation between the QBO and sudden warmings.

7. Conclusions

This paper has focused on the stratospheric warming that occurred in January 1992. ISAMS observations of temperature and derived geopotential height reveal that the ubiquitous cyclonic vortex and Aleutian high in the middle stratosphere were joined by a second warm-core anticyclone. This then moved rapidly north-eastward in the jet stream and merged one day later with the Aleutian high to form a single, intensifying anticyclone that displaced the vortex from the pole.

Several simulations of these developments (from a total of twelve initialized close to the warming) succeed in reproducing the sudden warming via vortex merger with remarkable accuracy of both developments and timing. By contrast, there are also inaccurate simulations which fail to reproduce the appearance of the second warm-core anticyclone and its merging with the Aleutian high. In these simulations warmings are produced, but with different synoptic behaviour and timing.

A survey of the initial conditions of all the simulations has revealed that the main differences between accurate and inaccurate simulations are in the tropics and subtropics. In the inaccurate simulations 〈u〉 is larger at the equatorial stratopause, indicating greater inertial instability in this region; also the subtropical stratospheric jet is stronger, associated with a stronger, more zonally symmetric PV barrier to northward transport. Detailed examination of two contrasting simulations reveals that in the accurate simulation, air originating in the tropical upper stratosphere descends and is brought northward in a long spiralling incursion and injected right into the Aleutian high; in the inaccurate simulation this northward transport is inhibited.

One reason for this inhibited transport is possibly that the initial conditions of the unsuccessful simulation have caused earlier onset of a larger area of inertial instability in the tropical upper stratosphere. As stabilization proceeds in this region, small-scale inertial ‘rolls’ in the meridional flow disturb the usual summer-to-winter drift and perhaps inhibit the northward transport of low PV onto the conveyor belt that ultimately feeds low PV into the Aleutian high. The stabilization should also result in the PV barrier moving southward, and perhaps tightening, together with a zonal ‘symmetrizing’ of the low-PV incursion that helps to push low PV into the Aleutian high. In either case the effect should be to inhibit the poleward transport of low PV.

In addition, however, there are other factors that appear to have been important in the development of this sudden warming, and these must not be overlooked. In the initial condition of the successful simulation more low PV was already present in higher latitudes, which most probably aided the establishment of a good low-to-high latitude connection and the injection of low PV into the Aleutian high. The unsuccessful simulation also revealed a stronger, more zonally symmetric PV barrier at about 30°N which also doubtless helped to confine the low PV to lower latitudes. Both simulations were inertially unstable to begin with, and the presence of the anomalous PV tongue in the winter hemisphere is most probably a manifestation of this; in the successful simulation the Rossby wave structure seems to have been more favourable for the injection of this low PV into the right place to build the Aleutian high. However, once the stabilization process is underway, it remains possible that the more disturbed state
associated with greater inertial instability in the unsuccessful simulation then impeded further penetration of low PV into high winter latitudes. In practice, of course, it is most likely that the development of this sudden warming involved an extremely complex interplay of all these factors and probably others as well; we believe, however, that inertial instability might have played a non-negligible role.

Our main conclusion, therefore, is that the strength and timing of the stratospheric sudden warming of January 1992, which is likely to have depended on the structure of both tropospheric planetary waves and the winter stratosphere, may well have depended also on the tropical wind structure in the upper stratosphere. There may well have been a role for the absence of large regions of inertial instability; the structures that would arise as the atmosphere returns to stability may contribute to inhibiting the northward transport of low PV which was necessary for the warming to proceed. These results may well also have relevance for other sudden warmings, but clearly many other warmings would have to be studied in detail before making a case for the generality of the results presented here.

ACKNOWLEDGEMENTS

We are indebted to Lesley Gray and the Rutherford Appleton Laboratory for making the production of this paper possible. Both authors acknowledge the UK Natural Environment Research Council (NERC) for financial support, and SMR also acknowledges funding from the UKMO. We are grateful to David Andrews, Richard Kenenna and John Knox for helpful discussions regarding the paper and to Keith Shine for generously allowing the time for its completion.

REFERENCES


Hunt, B. G. 1981 The maintenance of the zonal mean state of the upper atmosphere as represented in a three-dimensional general circulation model extending to 100 km. J. Atmos. Sci., 38, 2172--2186


O'Sullivan, D. J. and Hitchman, M. H. 1992 Inertial instability and Rossby wave breaking in a numerical model. J. Atmos. Sci., 49, 991--1002


