Dynamic Coupling of the Stratosphere With the Troposphere and Sudden Stratospheric Warmings

KEVIN E. TRENBERTH — Department of Meteorology, Massachusetts Institute of Technology, Cambridge, Mass.

ABSTRACT — Numerical time integrations of a nine-layer, quasi-geostrophic, highly truncated spectral model of the atmosphere are used to study tropospheric-stratospheric interaction with particular regard to sudden stratospheric warmings. The model is global, extends to 0.05 mb (71 km) with roughly 10-km resolution in the stratosphere, and includes an annual heating cycle. Model integrations simulating the months of December and January were made (1) without nonzonal forcing and (2) with nonzonal heating and orography included, to represent Southern Hemisphere and Northern Hemisphere winters, respectively. The presence of nonzonal heating in the winter hemisphere brought about an increase in circulation intensity and produced a stationary perturbation having a strong westward slope with height extending high into the stratosphere. This feature is similar to the Aleutian system. It was accompanied by considerably warmer temperatures in the polar night stratosphere and a weaker stratospheric westerly jet.

Sudden stratospheric warmings occurred as a result of large increases in the intensity of planetary scale waves in the troposphere, which in turn produced surges of upward propagating energy. The energetics of the warming occurred in two phases. A change from a baroclinically direct to a driven circulation occurred as the stratospheric temperature gradient reversed. This coincided with a change from enhancement to absorption of the vertical energy flux. The mechanism of the warming was similar to that described by Matsuno.

Nonlinear interactions between the progressive long wave and the nonzonal heating were primarily responsible for the tropospheric events that produced the transient upward flux of energy and thus the warmings. A seasonally coupled index cycle in the long waves was also of significance, while interactions with other waves and orographic forcing were of secondary importance in the long-wave energetics of sudden warmings.

1. INTRODUCTION

The discovery of the sudden stratospheric warming aroused much interest, and it has been the subject of many investigations, but many aspects of this phenomenon are still unexplained. Individual warmings have been illustrated and described by Teweles and Finger (1958), Teweles (1958), Craig and Hering (1959), Finger and Teweles (1964), Quiroz (1969), Johnson (1969), Barnett et al. (1971), and others, for the Northern Hemisphere (N.H.); and Wexler (1959), Palmer (1959), Palmer and Taylor (1960), and Phillpot (1969) for the Southern Hemisphere (S.H.).

There are large variations in the time of occurrence and intensity of the sudden warmings in the N.H. They often take place weeks before the sun returns, after which slow radiative cooling again sets in. Several warmings may take place in one winter, but there is usually only one major event that eliminates the strong meridional temperature gradient. The Antarctic stratosphere cools well into winter, but the meridional temperature gradient often relaxes slightly before the vortex breakdown, which occurs in spring with some regularity. Major warmings have not been observed until spring in the S.H., although smaller midwinter warmings have been detected (Julian 1967, Belmont et al. 1968).

Another feature of the warming is the appearance of strong winds preceding the event (Quiroz 1969, Johnson 1969). They produce strong horizontal advection of temperature and result in large upward vertical motions in the region of the high-latitude warming, thereby acting to diminish the magnitude of the warming (Mahlman 1969, Quiroz 1969).
Many authors have associated blocking or low-index situations with sudden warmings. Craig and Hering (1959), Julian (1961), Tewele (1963), Miyakoda (1963), Labitzke (1965), Julian (1965), and Murakami (1965) have commented on various aspects of this—usually on the presence of anticyclonic blocking upstream from the region of the initial stratospheric warming.

This paper attempts to explain many of the above features with particular attention centered on the observed differences between the hemispheres in the sudden stratospheric warmings. Considerably warmer polar night temperatures, weaker westerlies, and the existence of a marked quasi-stationary perturbation are mean characteristics of the N.H. winter stratospheric circulation not found in that of the S.H. We shall also endeavor to explain these differences.

Sudden warmings have been considered to be due to solar radiation, direct or indirect; to a manifestation of baroclinic, barotropic, or inertial instability; or to a combination of these. More recently, forcing from below has been considered a cause. Radiation effects are small at the time of many major midwinter warmings and are generally not sufficient to bring about the observed changes. Investigations into hydrodynamic instability of the polar night jet, which concentrate on the release of the basic flow energy but exclude energy propagation from below, have not been successful. Such instability studies (Fleagle 1957, 1958, Murray 1960, Charney and Stern 1962, Lindzen 1966, McIntyre 1972) have been unable to explain the large-scale changes in such short times as observed, although instability does appear to be present. Matsuno and Hirota (1966) and Hirota (1967) found that barotropic growth of a wave could occur with an elongated basic current, but the source of the growing wave energy was from the unspecified external forces maintaining the basic state.

The lack of success in explaining the phenomenon, plus the observational evidence in the energetics studies of individual sudden warmings (Reed et al. 1963, Miyakoda 1963, Muench 1965, Julian and Labitzke 1965, Perry 1967, Dopplick 1971, and others), has led to the conclusion that the energy propagated up to the region from below is not only not negligible, but forms the primary source of energy for the large-scale warming. Furthermore, an instability theory of the breakdown would not account for the observed greater stability of the more intense S.H. polar night vortex. However, in situ sources of energy may serve to enhance such events.

The energetics studies already cited further show the energy transport to be brought about by the longer waves (i.e., waves 1–3). Miller (1966) shows that the flux takes place chiefly at high latitudes. Miyakoda et al. (1970) found a narrow belt of strong upward energy flux at high latitudes when a sudden warming was occurring.

The sudden warming problem may be considered in three stages:

1. The mechanism of the production of the very long-wave energy that is propagated vertically and ultimately produces the sudden warming.
2. The mechanism of the vertical propagation of the energy.
3. The absorption of this energy and the mechanism of the warming itself.

Descriptions of the sudden stratospheric warming and the differences between the hemispheres provided the basic stimulus for this investigation, with the realization that a full knowledge of the mechanisms involved may eventually allow these events to be predicted. Stages 2 and 3 of the warming have been considered by previous studies.

Eliassen and Palm (1961) showed that the vertical energy flux would be upward in a stationary planetary wave provided the wave sloped westward with height, realizing a poleward transport of heat. They also showed that the energy flux could not cross the singularity at the zero-velocity zonal wind.

Using a quasi-geostrophic baroclinic model, Charney and Drazin (1961) showed that unstable short waves would be external and, thus, could not penetrate far into the atmosphere. They found that forced stationary waves could propagate provided $c < U_1$, where $c$ is the phase speed of the wave, $U$ is the uniform basic zonal wind speed, and $U_1$ is a critical value of $U$. $U_1$ increases with the wavelength of the disturbance. Thus, the longer waves were shown to be more likely to propagate vertically while short waves would be trapped, provided the wind is westerly relative to the phase speed.

In expanding Charney and Drazin’s treatment to consider the effects of the lack of North–South wind shear, Dickinson (1968a) established that planetary waves were absorbed by the mean flow at a critical level rather than reflected. Vertically propagating waves are refracted by the strong westerlies and channeled into ducts in the atmosphere. The ducting allows waves to propagate at high latitudes, but, because of absorption at the easterlies in low latitudes, propagation is unlikely there. Dickinson (1968b) also considered the limitations of the $\beta$ plane used by Charney and Drazin and found similar criteria for propagation, but with greatly differing quantitative requirements. Thus, two or more planetary wave modes were found to be always able to propagate through any observed westerlies in the troposphere or stratosphere.

Dickinson (1969a) found that photochemical relaxation attenuated planetary wave propagation, thus, perhaps, explaining the observed absence of such large fluxes of energy implied by his earlier paper (Dickinson 1988b). Matsuno (1970) considered the observed basic state and found that waves 1 and 2 could propagate into the stratosphere. He was able to duplicate certain features of the observed winter stratospheric flow and found a major sink of the upward flux of energy to be at the line of zero zonal wind in low latitudes.

Charney and Drazin (1961) also found that second-order changes in the zonal flow due to vertically propagating waves should vanish. Thus, eddy fluxes tend to be balanced by the mean meridional circulation, for a steady state, in the absence of critical levels. More recently, Dickinson (1969b) showed that the presence of photochemical relaxation or singular lines would enable a net forcing of the mean flow.
Matsuno (1971) has considered instances when this does not hold so that changes in the zonal flow may ensue. The presence of transient planetary wave upward propagation, or the existence of a critical layer (where \( c = U \)) intercepting the vertical flux, violates the balance. The former produces a vertical gradient of the poleward heat transport accompanying the upward propagating waves, which induces a meridional circulation and weakens the westerlies. If this process continues, a critical layer may be formed that would have similar effects, but concentrated in a shallow layer. Matsuno demonstrated the mechanism with a simple model that assumed a lower boundary to be baroclinic with the stationarity maintained by the increasing and sustained vertical flux of energy remains unexplained.

2. METHOD OF APPROACH

A major attempt to simulate a sudden warming was made by Miyakoda et al. (1970) using the Geophysical Fluid Dynamics Laboratory (GFDL), nine-level, general circulation model. After initializing with real data, they compared the evolution of the model with the observed atmosphere. This simulation was able to reproduce a split of the polar vortex into two vortexes, but these subsequently reunited without producing a warming. In the real atmosphere, however, the vortexes remained separate and a warming ensued. The model failed to predict the intensity of the Aleutian Low in the troposphere, and the vertical structure of this feature was poorly reproduced. As a result, the Aleutian High in the stratosphere was poorly predicted. A blocking-type ridge over the Eastern Pacific was also not predicted.

Byron-Scott (1967) and Clark (1970) demonstrated with their numerical models that sudden warmings would occur in the stratosphere when abrupt changes in the forcing from below took place. Here, we shall attempt to illuminate this point further in a more quantitative fashion by considering realistic mechanisms in the troposphere that might produce such changes.

Indications of the sudden nature of the change are provided in most major warmings, which show evidence of over-compensation in the temperature field and relaxation to lower values afterward. The intensity of the warming and the vortex breakdown may then be expected to be proportional to the magnitude of the sudden increase in the upward flux of energy, and to how long it is maintained.

a. Sudden Warming Hypothesis

We now pose the question, "What are the mechanisms that produce such an increase in the vertical flux of eddy geopotential energy?" A closely related question is, "How and why is energy put into the long waves?"

Charney and Eliassen (1949) and Smagorinsky (1953) demonstrated the importance of orography and nonzonal heating to the long waves when one considers long periods of time. Results of other steady-state models that consider the response to stationary forcing have been summarized by Saltzman (1968), and the most comprehensive results (Sankar-Rao 1965, Sankar-Rao and Saltzman 1969) indicate that nonzonal heating is dominant. Mintz (1968) and Kasahara and Washington (1969, 1971) confirmed the dominant role of the nonzonal heating with time-dependent numerical models of the atmosphere. Holopainen (1970) computed the energetics of the stationary long waves from observed data and found winter disturbances to be baroclinic with the stationarity maintained by nonzonal heating, while the mountain effect was small. However, Saltzman and Fleischer (1960) and Saltzman and Teweles (1964) have also shown the importance of nonlinear interactions with cyclone-scale waves in providing a source of long-wave energy. Nevertheless, to explain stage 1 of the warming, we must consider non-steady-state situations so that transient forcing or transient waves are of major importance. Some answers which we consider as hypotheses are:

H1. The zonal westerlies become stronger with the onset of winter, thereby enhancing the effects of boundary forcing due to orography. Latitudinal shifts in this current may serve to change the magnitude of this forcing and produce an effect at higher levels. Such a short-term shift in the westerlies may be associated with an index cycle.

H2. Possibly a greater effect is that due to the gradual increase in the land-sea heating contrast causing a large perturbation in the tropospheric zonal flow. The phase and amplitude of the heating is a function of the time of year, and short-term fluctuations are very likely present, perhaps associated with the concept of an index cycle. If variations in the degree of activity of small-scale cyclones are present, then (since latent heat release on this scale is in a pattern determined by the large-scale steering flow) the pattern of latent heat release may affect the larger scale directly.

H3. A further source of variation is the concept of an index cycle operating on the scale of the long wave itself. This index cycle is interpreted to represent fluctuations in the baroclinic rate of generation of eddy kinetic energy (Trenberth 1973, henceforth referred to as T).

H4. A similar source of variation in the gain of long-wave energy may be from nonlinear interactions with other scales. Since the cyclone-scale waves are steered by the larger scale flow, the possibility of systematic gains in this way are likely; variations may then be caused through an index cycle operating on the scale of the short waves.

To summarize these hypotheses in terms of an energetics diagram, we have (\( LW = \text{long wave}, SW = \text{short wave} \)):

\[
\begin{align*}
H1 & : KZ \to KE_{lw} \text{ by topographical distortion of zonal flow}, \\
H2 & : GE \to AE \to KE_{lw} \text{ by nonzonal heating}, \\
H3 & : AZ \to AE \to KE_{lw} \text{ by baroclinic processes}, \\
H4 & : AZ \to AE \to KE_{gw} \to KE_{lw} \text{ by baroclinic and barotropic processes plus nonlinear interactions}.
\end{align*}
\]

All these hypotheses include some element of an index cycle, a term that has been used loosely here, but which is...
qualified in T, where the possibility of H3 alone is also considered. The concepts above have been simplified for the purposes of discussion, and the mechanisms were considered separately. Further interesting possibilities may exist if some sort of reinforcement of these effects is established at any time.

b. Reinforcement and Quasi-resonance

The trend for the winter atmospheric structure to form ducts for propagating waves (Dickinson 1968a) allows for the possibility of quasi-resonance to exist in the atmosphere. A form of resonance may occur if the forcing field is structured in a manner to reinforce self-excited baroclinic waves. Nonzonal heating may have this effect, as found by Sankar-Rao and Saltzman (1969).

Thus, further hypotheses for producing a large change in the tropospheric long waves can be formulated. In effect, this is the same as considering our former hypotheses as reinforcing one another. Thus, we consider H2 and H3 or perhaps H1 and H3 together.

Since the phase of the nonzonal heating changes with time, it may serve to reinforce the orographic contribution in a similar way. Thus, we can consider any combination of our stated hypotheses.

Presumably, each of these factors plays a role of varying consequence each year, and, thus, we may perhaps account for the observed year to year changes in the sudden warmings.

Since large-scale perturbations are virtually nonexistent in the mean flow of the S.H., the mechanisms discussed above will be largely absent. The greater symmetry of the Antarctic vortex could be expected to give smaller changes in forcing and, hence, produce effects only at high levels or of small amplitude. These have been observed (Julian 1967, Belmont et al. 1968).

c. Relation With Tropospheric Blocking

We mentioned earlier the association of sudden warmings with anticyclonic blocking upstream in the troposphere. Observations of intense cyclonic activity at the time of a warming (Teweles 1958, Finger and Teweles 1964, Labitzke 1965) and the ensuing blocking appears to be consistent with the hypotheses involving an index cycle. In this way, blocking and the warming may be regarded as part of the same process, not that the blocking is somehow caused by the warming.

d. Experiments

To gain some insight into the mechanisms proposed and how they affect the vertical flux of geopotential energy into the stratosphere, we devised a number of experiments. The method chosen is to incorporate the basic physics into a numerical model and perform a series of controlled experiments with external forcing. The model set up for this purpose, the initialization, and the first experiment (experiment A) have been described in T.

The model is nine-layer, quasi-geostrophic formulated on a sphere, with a highly truncated spectral representation of the variables. The model is global, includes an annual heating cycle, and extends to 0.05 mb (about 71 km) with roughly 10-km resolution in the stratosphere. Tropospheric resolution is roughly equivalent to a three-layer model. Only three waves (2, 4, 6) are explicitly included. A capability exists for including orographic forcing and nonzonal heating in wave 2. The latter is a function of the time of year and reaches a peak three weeks after the solstice (see T).

Experiment A was for the months of December and January with nonzonal forcing excluded, thereby serving to simulate a S.H. winter. The second and third experiments cover part of the same period; experiment B has nonzonal heating, and experiment C has orography, included. The fourth integration, experiment D, again covers December and January and includes both forms of forcing, thereby serving to simulate a N.H. winter.

3. RESULTS

The performance of the model during experiment A has been described in T, and one aspect of this integration of importance with regard to the present investigation was noted. This was the presence of some form of index cycle in the long waves that was coupled to the annual heating cycle. A minor warming in the stratosphere occurred as a result.

a. Experiment B: Nonzonal Heating

The response of the model to nonzonal heating was masked to a large degree by nonlinear effects. In particular, large variations in AE were apparently caused by changes in the baroclinicity of the wave as its slope changed with height. Hence, when the temperature perturbation was reinforced by the heating, increased poleward heat transport resulted, thereby increasing the baroclinic conversion of energy. Such a large response may have been partially caused by the concentration of the heating at essentially a single tropospheric level.

The wave failed to become quasi-stationary, and the result was a transient wave 2 that underwent large fluctuations in baroclinicity with a period of roughly 3 weeks. The results of this experiment are not presented, but discussion relating to it will be given when we consider experiment D.

b. Experiment C: Orography

Experiment C consisted of a 50-day integration so that the month of December was again simulated. The mountain torque served to enhance the surface friction by a factor of roughly 2/3 north of 40° N. The drain on the westerlies was offset somewhat by an increased flux of angular momentum into the region by the large-scale eddies. In the model, the baroclinicity of wave 2 resulted in a transient wave, alternatively reinforced and canceled by the orographic forcing.

There was little change in the energetics from that shown in experiment A (fig. 9 of T), although the intensity of the circulation in the N.H. troposphere decreased.
slightly. The mean exchange for the entire month between KE and KZ due to orography was $9 \times 10^{-7}$ J cm$^{-2}$ s$^{-1}$ from KE to KZ in the N.H. and $10^{-7}$ J cm$^{-2}$ s$^{-1}$ in the same direction in the S.H. Hence, the exchange was insignificant.

c. Experiment D: Nonzonal Heating and Orography

Experiment D was a complete integration, from day 40 to day 124. An adjustment period of 20 days was again included before the months of December and January were simulated. In the remainder of this section, we compare the results of this run with those given in T for experiment A.

Zonally Averaged Fields. In figure 1, we present the fields of zonal wind and temperature and a schematic diagram of the meridional circulation for the dates shown. This figure is the same as that given in the integration with eddies excluded (fig. 4 of T) and for experiment A (fig. 8 of T).

In the N.H., the tropospheric westerly jet is slightly stronger, while the stratospheric jet is about 10 m/s weaker. At high latitudes in the stratosphere, the easterlies are stronger and more extensive. Although the easterly component in low latitudes of the S.H. increased somewhat, westerlies were present at the Equator to 60 km and were slightly greater than in experiment A.

In the temperature field, the most notable difference is in the increase in N.H. polar temperatures at 10 mb by more than 20°K on both January 2 and February 1.

The meridional circulation shows considerably more character, although the overall features are similar. A large Hadley cell still exists in the stratosphere on January 2, and the cells on February 1 tend to be in the opposite sense, as in experiment A. However, the decrease in the strength of the symmetric circulation is apparent in this region. In contrast, the Ferrel cells in the troposphere were weaker.

Certain features of figure 1 are realistic when compared to the observed circulation. Of particular note are the decreased westerlies in the N.H. winter (experiment D) compared to the S.H. winter (experiment A). Similarly, the stronger easterlies at high latitudes in the N.H. winter lower stratosphere are an observed difference. This accompanies the much warmer temperatures in the stratospheric polar night that are observed and simulated by the model. At the pole, the increase in the 10-mb mean temperature from day 40 to 124 in experiment D over experiment A was 18°K.

The tropospheric jet in the N.H. was as much as 5 m/s stronger in late January in experiment D. This is also an observed difference. However, the increased dominance of the direct cells in the troposphere seems in error. Both these features are related to the momentum budget.

Energetics. Figures 2A and 2B show the four-box diagrams (fig. 3 of T) of the energetics for experiment D. This figure is equivalent to figures 9A and 9B of T.

In the N.H. troposphere energetics, we find a number of new features. Most notable is the large increase in the intensity of the circulation in both months. This was
due mainly to increased baroclinic conversions of AE to KE by large-scale quasi-horizontal eddies, but the increased dominance of the direct cells in the symmetric part of the circulation, $\Delta Z \to KZ$, was also a factor. Since a decrease in the intensity occurred in experiment C, it is obvious that the change was brought about by nonzonal heating. The increase in intensity was mainly in wave 2, although a slight increase occurred in wave 4.

There has been a pronounced increase in AE for both months as well as the expected decrease in GE. Such values are more in line with those observed in the N.H. The change was due to the greater role of wave 2 in the circulation.

These features may be seen more clearly in figure 3, which shows the daily variations of the different forms of energy, averaged over the globe and integrated in the
similar energetics in experiment D. The greatest change has been an increased flux from below by means of the vertical propagation of energy. AE and KE increased markedly in both months. Note that an increase in CA is associated with the increased $V\Phi E$ in December while CE is not greatly affected.

In summary, the introduction of orography had little effect on the energetics, in agreement with the comments by Mintz (1968) and Kasahara and Washington (1969, 1971). However, the introduction of nonzonal heating produced a large increase in the intensity of the circulation and improved the simulation of some features of the atmosphere observed in the N.H. The transient wave 2 played an important role. Whether these kinds of effects really exist in the atmosphere for the long transient waves is not entirely clear and will be discussed further at the end of this section. As a result of the changes in the troposphere, there were pronounced effects on the stratosphere and a general increase in the role of eddies in the circulation.

**Angular Momentum Budget.** Unlike experiment C, the mountain torque in experiment D acted to increase the westerly angular momentum of the atmosphere. This signifies that, in the mean flow, there is a high-pressure cell to the east of the mountain, thus demonstrating the dominance of the nonzonal heating in determining the location of the stationary component of the flow. However, since the response in the model is unlike that observed in the region 40°–60°N, it seems that either the phase relation between the orography and nonzonal heating (or the representation of these forms of forcing) is incorrect or the structure of the wave is in error.

Much of the error can be assigned to the fact that only one scale and phase of orography was represented in the model, and the choice of phase was dominated by the Himalayas. Had the northern Asiatic mountains been included, an angular momentum sink would have resulted, as is present in the atmosphere (Newton 1971). The lack of a complete representation of the mountains allows air to flow through regions where mountains should be, and the dominant effect of the Himalayas as a dam (Mintz 1968) is not included. This in turn causes the Ferrel cell in the troposphere to weaken in experiment D, since there was apparently no longer a general circulation requirement that eddies transport momentum to maintain the tropospheric jet. However, the wave structure and behavior of wave 2 are also open to question and will now be examined in more detail.

**Behavior of Waves.** The model failed to produce the quasi-stationary features most prominent in the N.H. winter troposphere, but there was a substantial stationary component.

Figure 4 shows the mean streamline pattern for the period December-January at the six lowest levels of the model. The mean contours were evaluated for each month separately; however, since the patterns contained almost identical features and phases, they were combined. The longitude values shown are relative to our datum chosen at 35°E, so that the mountain peak is at 45° in wave 2.

A curious feature of the tropospheric patterns is the

---

**Figure 3.** Variation of the global forms of the vertical integral of energy in experiment D.

vertical. This figure corresponds to figures 6 and 7 of experiment A. (See T.) The scale for AE has been changed, and, to decrease the clutter, we did not include A2 since it is similar to AE in variability.

In many ways, the curves are similar to those in experiment A except for wave 2. AZ and KZ variations are now quite marked. As in experiment A, the intensity of the circulation increased from December to January, and AZ reached a maximum at about the same time.

While the effect of orography on the energetics of experiment C was small, the mountain torque was important in the angular momentum budget. In experiment D, the presence of mountains was the cause of the other major change in the winter tropospheric energetics and resulted in a barotropic contribution to the growth of eddies. This reversal in CK is approximately offset by the loss in KE to KZ by the orography. Further detail on the reasons for this response are considered later in subsection 3c.

We now briefly consider the lower stratospheric N.H. energetics (fig. 2). In December, the energetics are similar in experiments A and D. In January, however, there is a marked increase in the driving from below through increased vertical propagation of eddy geopotential energy. The conversions CE and CA are roughly twice those of experiment A. Hence, the increased intensity in the tropospheric circulation has increased the driven circulation aloft.

In the upper stratosphere in winter, both months have
appearance of a pronounced wave 4 component. It is due to the very slow movement of wave 4, which had a net movement of only 35° westward from day 60 to day 124. This behavior is unlike that of the other two waves, which were both progressive. A linear baroclinic analysis (Trenberth 1972) revealed that wave 6 would be in the short wavelength region of instability, wave 2 would be in the long wavelength region, and wave 4 near the transition. Both wave 2 and wave 6 moved with mean speeds close to those found in that study. At the transition, a wave is stationary and corresponds to the resonant wavelength.

Figure 4 shows that wave 2 in the N.H. is not of very great amplitude in the troposphere, but becomes increasingly marked above 200 mb. The slope of the stationary component is westward with height throughout. At high latitudes in the N.H., the 800-mb closed low-pressure region is at 157° (i.e., near 12° longitude). This is somewhat east of its observed location in the atmosphere.

Orographically induced disturbances have little or no phase change with height, whereas nonzonal heating may maintain stationary disturbances with either westward or eastward slope with height (Saltzman 1965, Murakami 1967). Mean tropospheric charts for the N.H. in January (e.g., Willett and Sanders 1959) reveal that the Aleutian Low slopes strongly westward with height. The Aleutian anticyclone is a prominent stationary perturbation that extends to high levels in the stratosphere (e.g., Finger et al. 1966) and has been related to tropospheric cyclogenesis by Boville (1960). Since the Aleutian Low seems to be maintained primarily by nonzonal heating (Holopainen 1970), the evidence suggests a direct association with the Aleutian High aloft. This system has a large wave 1 component.

The stationary component in the model in experiment D has a similar structure, but with an even greater slope with height (fig. 4). Both are baroclinic systems maintained by nonzonal heating and propagation of energy to higher levels. Associated with these features are the significantly higher temperatures (table 1) that exist in the winter N.H. stratosphere (experiment D) compared to the S.H. stratosphere (experiment A). These higher temperatures are attributed to the northward heat flux associated with the upward energy flux (Eliassen and Palm 1961).

![Figure 4](https://example.com/figure4)

**Figure 4.**—Mean streamlines for period December-January for experiment D. Longitude values are such that 0° corresponds to 35° E.

<table>
<thead>
<tr>
<th>Mean temperature (°K) at the pole from day 40 to day 124</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure (mb)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Experiment A</td>
</tr>
<tr>
<td>Experiment D</td>
</tr>
</tbody>
</table>

April 1973 / Trenberth / 313
The primarily transient nature of wave 2 may otherwise be interpreted to mean that the transient component of the wave was greater in magnitude than the stationary component. This is the case during the S.H. winter, but the reverse is true in the N.H. winter, although the presence of both components of the long waves in the N.H. is well established. Deland and Johnson (1968) and Bradley and Winn-Neilsen (1968) considered the vertical structure of the transient very long waves and found a strong westward slope with height, so that they are of a baroclinic nature. In particular, the latter study found three vertical pressure modes to be present. One mode, while progressing eastward at 15°/day, was divergent and, for wave 2 in the N.H., sloped westward with height by about 36° longitude between 850 mb and 500 mb. Hirota (1968) found the same speed and a strong westward slope for the transient component of wave 2 in the stratosphere. Phillpot (1969) found wave 2 to have similar movement with a pronounced amplitude at 30 mb in the S.H. This case was notable since the wave was primarily transient, whereas, in the N.H. studies, the waves are quasi-stationary and the transient component has to be extracted. A stratospheric warming later grew from this perturbation. The structure and movement of these waves bears a strong resemblance to the wave 2 performance in the model.

A question arises, however, as to why a quasi-stationary wave was not produced here. Kasahara and Washington (1969, 1971) and Holopainen (1970) indicate that the stationarity is due to nonzonal heating. In our model, we noted the large change in baroclinicity associated with the nonzonal heating. This change in baroclinicity appears to be caused by differential rates of advection with height, which produced changes in the slope of the wave with height. Hence, it seems to be associated with the vertical resolution of the model and the vertical distribution of the nonzonal heating. Other important factors are as follows:

1. The increased westerly component in this model due to the mountain torque.
2. The small change in the tropospheric static stability values as a result of the linear baroclinic analysis. (See Trenberth 1972 or T.) The study showed that this should not affect the wave speed, but it may have altered the response of the long wave to heating. The study also indicates that the presence of westerlies at the Equator would increase the eastward speed of the waves.
3. The structure of the lower stratosphere with stronger winds in lower latitudes than observed.
4. The lack of an Ekman boundary layer in the model.
5. The lack of a variable static stability, which excludes the stabilizing effect of the vertical flux of heat in baroclinic systems. The quasi-geostrophic formulation may have also omitted important influences on the motion.
6. Truncation of the spectral modes, which restricts the number of degrees of freedom in each wave. As a result, the hemispheres are not entirely independent.

Many of the above suggestions and reasons for not obtaining a better simulation of the stationary waves can be easily implemented. Other suggestions would require an expanded model (e.g., items 4 and 5). However, in such a simple model, the improvement of one aspect of the flow may well be detrimental in other respects (e.g., item 2 or a change in heating parameterization to improve item 1 or 3).

4. DYNAMIC COUPLING BETWEEN THE TROPOSPHERE AND STRATOSPHERE

The performance of the model has been evaluated by comparing features of the flow with those of the observed atmosphere. Attention was centered on average properties with a view to bringing out the characteristics of the model. Some discrepancies were considered in T as a result of the analysis of experiment A. Further deficiencies were evident when nonzonal forcing was introduced, most notably in the performance of individual waves. In general, however, the model simulation of the atmosphere has been successful, and differences are chiefly in the degree of response rather than in the response itself. With this evaluation in mind, we now consider the experiments with regard to the sudden warming phenomenon in all its phases.

a. Sudden Stratospheric Warmings in the Model

To define the phenomenon as it occurred in the model integration, we present a time series of the temperature at the North Pole in experiment D as a function of height (fig. 5). This figure corresponds to figure 12 of T for experiment A. The mean temperature from day 40 to day 124 has been extracted and is shown in table 1 along with those of experiment A.
The differences between the experiments are greatest at 10 mb, but significant changes also occurred at 300, 120, and 0.1 mb. These are the same kinds of differences observed between the N.H. (experiment D) and the S.H. (experiment A). In the troposphere, the heating field is acting to produce the lowest polar temperatures after January 12, whereas the minimum was reached during December. This was also noted in experiment A.

In experiment D, large fluctuations with a period of about 18 days occurred in the 10-mb polar temperature, and warmings of over 40°K took place within a 2-week period. The patterns in the two experiments are almost identical until day 60, although a different mean value has been extracted. The warming on day 58 occurred in all four experiments.

To examine the warmings at 10 mb in greater detail, we consider the complete temperature field time series for the N.H. shown in figure 6. At 2 mb, a similar time series is not very revealing since the temperature gradient was poleward at all times. At 10 mb, however, complete reversals in the temperature gradient occurred and were accompanied by increased easterlies at high latitudes. Both of these features are generally regarded as constituting a major warming. The largest warming peaked on day 80.

Coupled with the warming at high latitudes is a cooling trend at low latitudes (fig. 6). Both of these changes were being offset by an induced meridional cell, with upward motion at the pole and sinking at lower latitudes. This meridional flow has been noted in warmings in the atmosphere (Quiroz 1969, Mahlman 1969) and is apparently the source of the easterly acceleration of the zonal flow at high latitudes that accompanies the warming. Clearly, the warming is not caused by subsidence. Figure 7 shows the change in temperature at 10 mb as a function of latitude due to advection of the temperature field by the eddies. The warming is obviously produced by the northward flux of heat by the eddies. Diabatic effects were continually acting to cool the high-latitude region, but figure 7 also shows that most of the cooling has a dynamic source.

At 2 mb, the polar temperature fluctuated more rapidly than at the lower level, but warmings of 10°K over 4 or 5 days were common. Most of these changes were due to the eddies, but the meridional cell induced by the warming at 10 mb was also of significance.

An example of the streamlines and temperature field at 10 mb in the midst of a warming is given in figure 8. The northward heat flux is the source of the warmings, and, in a quasi-geostrophic model, it is directly related to the conversion of AZ to AE (CA). Therefore, in the following sections the energetics are examined in detail to find the antecedent of the flux.
b. Energetics Variability

The variability of the energy components in the atmosphere have not been studied in great detail. Krueger et al. (1965) presented 10-day mean values for the lower part of the troposphere and the model values are not unreasonable in comparison. Dopplick (1971) considered daily values of zonal and eddy forms of energy in the lower stratosphere for 1964. His computations reveal an almost cyclic oscillation in AE throughout the first 3 months of 1964 with a period of roughly 2 weeks and a range of values comparable to those of the model. The model values had a period of about 18 days (fig. 3).

Holloway and Manabe (1971) presented time series for the rate of change of eddy kinetic energy in the N.H. of their global model. This also showed a marked cyclic oscillation with a period of roughly 2 weeks and a range greater than found here. Therefore, the model energy terms show large-scale variability similar to that observed and found in more sophisticated models.

To determine the relation between the variations in the troposphere and the higher levels, we consider the vertical flux of eddy geopotential energy through 200 mb and 20 mb in figure 9. Global, N.H., and S.H. values are shown.

In the S.H., the flux through 200 mb is reasonably constant and, as expected from the theory, is small through 20 mb. Dopplick (1971) presented values of this flux through 100 mb and 10 mb for 1964. The model shows excellent agreement with his summer values once allowance is made for the different levels of evaluation.

However, figure 9 shows the flux through 200 mb in the N.H. to be extremely variable. At 20 mb, the variations are also quite marked and appear to correspond to fluctuations at the lower level with a lag of a few days. This point will be expanded and clarified later.

Dopplick’s calculations for winter indicate that the atmospheric values are slightly larger and have greater range than those in figure 9. Nevertheless, the overall features are similar, and Dopplick’s values also showed the largest surges to be regularly spaced at 2- to 3-week intervals.

c. Energetics of Sudden Warmings

Earlier, we subdivided the problem of sudden warmings into three stages. To evaluate the role of the vertical flux of energy, we consider these in reverse. The remainder of this section will be concerned only with the N.H. (winter), unless otherwise stated.

The manifestations of stage 3 of the sudden warming have been presented, but an explanation of the mechanism was not made. This will be done here. Figure 10 presents a time series of the vertical flux of eddy geopotential energy for experiment D by (A) Wave 2 component and (B) all waves.

Figure 10.—Vertical propagation of eddy geopotential energy ($10^{-5}$ J cm$^{-2}$ s$^{-1}$) for experiment D by (A) Wave 2 component and (B) all waves.
energy as a function of height, obtained for the N.H. in experiment D. Figure 10A shows the wave 2 component and figure 10B shows the total for all waves. The levels at which this quantity was evaluated are shown at right. The 1000-mb value shown is the contribution to KE by the orographic distortion of the flow. Upward propagation is positive. The vertical gradient of this quantity shows the net contribution of this form of energy to a given layer.

We first note the strong resemblance between the patterns in the stratosphere above 20 km; this shows the overwhelming dominance of wave 2 in this region. This was true in all energetics terms so that only the contribution of the wave 2 component was of significance there.

A presentation similar to that in figure 10B is shown in figures 11 and 12 for the conversions CAZ and CEK, respectively. In these two figures, the conversion takes place in the layer centered on the level given at the right. The units here are misleading because of the different scale that 1 mb occupies at high and low levels; however, the presence of a double maximum in the values is revealed in the troposphere and in the 10- to 2-mb region. The diagrams corresponding to figures 10B and 11 are shown for experiment A in figures 13 and 14, for comparison.

In section 3, we noted the increase in CA and \( \nabla \Phi E \) for experiment D in the upper stratosphere, while CE did not change much. Figures 10–14 show the interrelation of these quantities and how they relate to sudden warmings. Eliassen and Palm (1961) showed that stationary waves that propagate energy upwards are accompanied by a poleward heat transport and a westward slope with height. In a poleward temperature gradient, for quasigeostrophic motion, this implies that CA is positive if \( \nabla \Phi E \) is positive. Figures 7 and 10–14 show this also to be true for transient wave propagation.

In experiment D, the 2-mb temperature gradient remained poleward so that CA had the same sign as the northward flux of heat. This was also true at 10 mb until day 54 and from day 62 to 75, but, at other times, the coldest temperatures were in midlatitudes (fig. 6). Then, the hemispheric value of CA for a poleward heat flux will be composed of upgradient heat transports (negative CA) at high latitudes, and downgradient heat transports (positive CA) in low latitudes. A comparison of figure 7 with figure 13 at 10 mb shows the high-latitude region to dominate the hemisphere from days 54 to 62 and 75 to 84.
A sudden warming is thus revealed to consist of two phases:

1. In a poleward temperature gradient with westerly winds, a large surge in upward-propagating energy is accompanied by a northward flux of heat; that is, for $\Phi E$ upwards, $CA$ is positive. The relation with $CE$ is not as distinct, but positive values generally follow positive $CA$ with a slight lag. The induced meridional circulation opposes the high-latitude warming trend and produces an easterly acceleration of the zonal winds in this region. In general, the vertical flow of energy does not modify the zonal flow to the second order (Charney and Drazin 1961), but transient wave propagation of energy is accompanied by a vertical gradient of northward heat flux that produces a net warming at high latitudes (Matsuno 1971). This phase of the warming is the same as the first stage described by Matsuno. If the upward flux is sustained for any period, a reversal in the temperature gradient will result from this process.

2. At this point, $CA$ and $CE$ become negative and the upward flux of energy is not absorbed but instead induces baroclinic growth of the eddy ($KZ \rightarrow AZ \rightarrow AE \rightarrow KE$). This, in turn, results in a divergence of the upward energy flux in the layer. Many examples exist in the model, as may be seen in figure 10 where the upward flux through 10 mb was generally greater than that through 40 mb.

Phase 2 constitutes a major warming; the source of energy for the warming is a convergence of the upward energy flux into the layer ($\Phi E \rightarrow KE \rightarrow AE \rightarrow AZ$). We have described phase 2 as it probably pertains to the atmosphere, but, because of the unrealistic structure of the lower stratosphere, the model did not behave in quite this manner. The lack of sustained upward energy flux also meant that the warming did not extend over the entire hemisphere of the model.

Eliassen and Palm (1961) discussed the nature of the vertical flux of energy as a function of the zonal wind field, and in particular with regard to the change from divergence to convergence of the flux where the vertical shear of the wind reverses above the tropospheric jet. In both phases of the warming described above, energy may still propagate, but it will be either enhanced (phase 1) or damped (phase 2). It will be unable to propagate only when a critical layer is formed. These aspects of stage 2...
of the warming (i.e., the theory of the vertical propagation of energy) seem to be explained by the theory. Both phases have been found in the atmosphere by Dopplick (1971).

The manner in which we have delineated the warming is different from the aspects stressed by Matsuno (1971). His description required the first stage to continue until a "critical layer interaction" took place. The critical layer concept arises from a singularity in linear theory, where $U$ (the zonal wind speed) equals $c$ (the phase speed of the wave). (See, for example, Charney and Drazin 1961.) In our model, the waves were transient and the theory does not apply. Because of the unrealistic structure of the model winter lower stratosphere, easterlies were frequently present at high latitudes in the stratosphere (fig. 1). Therefore, the real part of $c$ did equal $U$ at some part of the atmosphere. However, this did not seem to play a major role in the model.

The division used above not only distinguishes minor from major warmings but also separates the role of the vertical flux of energy into two phases and divides the baroclinically direct energy cycle prior to the warming from the driven energy cycle during the latter stage. However, the formation of a critical layer is undoubtedly important for a complete changeover and downward propagation of the warming.

d. Tropospheric Source of Warming

By relating sudden stratospheric warmings and the vertical flux of energy that causes them to large changes in the level of activity of the tropospheric very long waves (figs. 10-12), we may be able to establish a means of predicting the events. Figure 10 indicates that the lag in time between the tropospheric activity and the resulting modification of the upper atmosphere may be from 3 days to 1 week. The actual response of the upper atmosphere is dependent upon its initial state, but either phase 1 or phase 2 kinds of events should be expected. Once these begin, however, the induced meridional circulation must also be considered in determining the net effect at a given level. The level at which the warming is a maximum, and therefore the level that dominates those above, may be difficult to determine as it is dependent upon many factors. In our model, it was reasonably clearcut because of the finite resolution in the vertical. This aspect will not be considered further here.

The importance of the transient wave in the sudden warming (in particular, the transient component in time of the very long wave), has been discussed in the previous section. The importance of this wave lies primarily in the modification it causes in the zonal flow (Matsuno 1971). In anticipation of this, we established a number of hypotheses (sec. 2) that considered possible sources of rapid changes in the intensity of the circulation in the troposphere. These changes in activity were considered part of the familiar index cycle.

In the model, a marked equatorward movement of the tropospheric jet was associated with a slight decrease in the jet strength and increased conversion of AE to KE (=CE). These patterns are also found in the atmosphere when a high index situation changes to low index (Willett and Sanders 1959). In experiment D, the existence of a marked index cycle was associated with the periodic response of wave 2, as the traveling wave interacted with the stationary forcing. However, the period of 2–3 weeks is less than that generally associated with the index cycle.

As we have seen, all events that affect the long waves may be important, and variations in their degree of importance may account for the variations in the sudden warmings that are observed from year to year. The effect of orography, $H_1$, and the contribution from the interactions between other waves, $H_4$, seem to be of secondary importance and may play this lesser role. The primary source of the long-wave activity in the model was a nonlinear response of the traveling wave to the nonzonal heating. Thus, a combination of $H_2$ and $H_3$ is the cause of stage 1 of the sudden warming. However, as noted in subsection 3c, there is some doubt about the applicability of the events in the model to those in the atmosphere.

Nevertheless, there does seem to be a good possibility that the atmosphere exhibits a response to nonlinear interactions between the stationary and transient wave component. A possible alternative is an interaction between the wave and transient forcing, such as suggested by $H_2$ in section 2.

The model does indicate another possibility, however. The existence of an index cycle on the scale of the long wave, $H_3$, may be an important mechanism in the atmosphere. (See T.) In that study, we noted the tendency for a change in the dominant wave of the circulation in response to the change in the baroclinic growth-rate characteristics of the flow as the seasons progress. An increase in the intensity of the circulation during January was most noticeable in experiment A, and it also occurred in experiment D. In the former case, it was directly associated with a minor warming.

If this is the primary cause of sudden warmings in the atmosphere, it should also be observed in the S.H. Because of the presence of large-scale stationary features in the N.H., however, the initial state of the stratosphere in each hemisphere is quite different, and phase 1 would have to persist much longer in the S.H. to overcome the lower polar temperatures. This may result in only a minor warming, which occurred in experiment A and has been observed in the S.H. Nevertheless, the seasonally increasing land-sea heating contrast is probably also a major factor in the N.H.

Other possible effects mentioned in section 2 were not of great importance in the model but cannot be entirely ruled out. Quasi-resonance could be important for wave 4 and, as such, may serve to enhance wave 2 at times through nonlinear interactions. A major source of wave 1 energy appears to be the nonlinear interactions in the atmosphere (Saltzman and Teweles 1964) and in the GFDL model (Manabe et al. 1970). Hence, while these studies show this to be unimportant in wave 2, it may be of significance in wave 1.

April 1973 / Trenberth / 319
5. CONCLUDING REMARKS

In this paper, we studied the troposphere-stratosphere interaction with particular attention given to sudden stratospheric warmings. The model used is clearly deficient in a number of ways but appears suitable for the problem chosen for investigation here.

Our first experiment was described in T. Further experiments indicated that while mountains are of significance in the angular momentum budget, they play a lesser role in the energy budget. In contrast, the effect of nonzonal heating on the model atmosphere energetics was marked.

We were able to simulate certain features of the stationary flow found in the N.H. when nonzonal heating was included, and the model successfully reproduced the same kinds of differences actually observed between the hemispheric winter circulations.

The model atmosphere failed to produce quasi-stationary features in the N.H. troposphere, and the progressive wave interacted periodically with the stationary forcing. In this way, large changes in tropospheric activity produced large changes in the forcing of the stratosphere through the vertical flux of energy, and events of a sudden warming nature were produced. Therefore, changes in the intensity of a wave (i.e., the transient component in time) were confirmed to be of major importance in producing a modification of the zonal flow. These changes were identified with an atmospheric index cycle in the model N.H. The mechanism of the warming was found to be much the same as that outlined by Matsuno (1971), although different aspects were stressed here.

A detailed knowledge of the kind of events in the troposphere that initiates a large flux of energy into the stratosphere may enable some form of prediction scheme to be set up that will allow large-scale changes in the stratosphere, particularly sudden warmings, to be anticipated. This study indicates that we should look for increases in the intensity of large-scale systems such as the Aleutian Low or Icelandic Low.

We have also made a first attempt to evaluate the role of shorter scale waves, orography, the land-sea heating contrast, seasonal cycles, and combinations of these in producing such changes. Other effects, such as the vertical structure of the atmosphere and the upper boundary condition, particularly with regard to the location of a trapping layer at some high level perhaps changing because of extraterrestrial heating, may be very important in determining the exact location and degree of response.

In establishing the characteristics of the model, we analyzed the deficiencies in detail and suggested likely causes. Some of these defects may be easily removed without major changes in the model.

The results given by this model should be of considerable value in designing further such models and in designing experiments for more sophisticated models. Indeed, this approach seems most useful in bridging the gap between linear theory and the extremely complex primitive-equation models.

ACKNOWLEDGMENTS

This work is part of my doctoral thesis at Massachusetts Institute of Technology. My thanks to Edward N. Lorenz for his advice and guidance. My stay at M.I.T. was made possible by a New Zealand Research Fellowship, for which I am grateful to the New Zealand Meteorological Service. The calculations were performed at the M.I.T. Computation Center and were supported by the Atmospheric Sciences Section, National Science Foundation, under NSF Grants GA-10276 and GA-28203X.

REFERENCES


320 / Vol. 101, No. 4 / Monthly Weather Review


[Received July 5, 1972; revised October 27, 1972]