How Well do we Understand the Dynamics of Stratospheric Warmings?

By Michael E. McIntyre

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, U.K.
(Manuscript received 19 October 1981)

Abstract

Ever since Matsuno's pioneering numerical simulations of the stratospheric sudden warming there has been little reason to doubt that this spectacular natural phenomenon is essentially dynamical in origin. But theoretical modelling, and the use of satellite observations, are only just reaching the stage where there seem to be prospects of understanding stratospheric warmings in some detail and forecasting them reasonably well. An informal discussion of recent progress is given, and suggestions are made for future work, including a way of avoiding spurious resonances in mechanistic numerical models in which tropospheric motions are prescribed *a priori*.

1. Introduction

The stratospheric sudden warming is a large-scale experiment which nature kindly performs for us from time to time. It is one of the crucial tests of our understanding of the dynamics of the middle atmosphere, and indeed of atmospheric dynamics in general. The dynamics, in turn, is one of the necessary ingredients in attempts to understand the stratospheric circulation in general, and to predict the effects of pollution on the ozone layer in particular. Recent progress in the dynamical theory, along with the more uniform global coverage facilitated by infrared satellite observations, have led to better ways of analysing the observations and to more illuminating comparisons with computer simulations, and our understanding of stratospheric warmings is now advancing significantly. There are signs moreover that this understanding may be leading to fresh insights into other, at first sight unrelated, phenomena, for instance the nonlinear behaviour of mid-latitude, tropospheric depressions (Hoskins, 1982).

In this article I do not propose to give a comprehensive review of the literature on stratospheric warmings (for which the reader may consult Quiroz *et al.*, 1975, McInturff, 1978 and Schoeberl, 1978), but rather to concentrate on some recent developments with which I have been in touch. Indeed the subject is moving so fast that an informal discussion is probably the most appropriate thing at present. Some of the work whose implications I shall discuss (most of it not my own) was presented at the IAMAP symposium on the general circulation held at the University of Reading in August 1981. I am indebted to a number of colleagues for permitting me to draw upon the results of their work in press or in progress.

Hardly anyone who has followed observational and theoretical work on major stratospheric warmings can be in much doubt today as to the essentially dynamical nature of the phenomenon. Its enormous depth scale, covering several scale heights, and its "suddenness" compared to estimates of diabatic time scales at least in the lower stratosphere, makes it pretty obvious that dynamically-induced air-parcel descent is required to account for the observed temperature rises. The self-consistency of this view has been well checked by the results of many numerical simulations, including the first such study by Matsuno (1971). These mechanistic, or hypothesis-testing simulations have consistently reproduced at least the final, "sudden" stage of the process in a qualitatively convincing way (see the reviews by Quiroz *et al.*, 1975 and Holton, 1975). And in the simulations, at least, there is no doubt whatever that the large temperature rises are induced adiabatically, by descent of air parcels in the polar cap (*e.g.* Hsu,

1980; Dunkerton *et al.*, 1981); the fact that this is compatible with ascending Eulerian-mean motion is well known by now, Mahlman (1969) having apparently been the first to point it out in the present context. Some of the effects of experimenting with diabatic time scales in the models can be seen from studies such as those of Holton (1976) and Schoeberl and Strobel (1980a). Finally, the weight of observational and theoretical evidence leaves little room for doubt that large-amplitude "planetary waves", in the sense of large, planetary-scale disturbances to the zonal wind in the stratosphere, especially those involving zonal harmonic wavenumbers 1 and 2, are an essential ingredient in the process and not merely an accompaniment to it.

I shall take all the foregoing for granted, then, until proven otherwise, and suggest that the fundamental questions of interest today begin with those in the following list. Of course several of them are questions which we should ask about the behaviour of theoretical and numerical models as well as of the real atmosphere.

1. How and why do planetary-wave amplitudes become anomalously large?
2. When they do become large, for what stratospheric conditions is a major warming likely to occur (and why are major warmings relatively uncommon)?
3. To what extent can we use linear planetary-wave theory for the wave structure? And in particular,
4. to what extent can we think of the principal zonal wavenumbers, 1 and 2, as acting independently of each other? In other words, how much can we explain without invoking nonlinear interactions between different zonal wavenumbers?
5. Are wave-reflection and resonance phenomena important or not (*e.g.* to question 1)?
6. Are "critical lines" important or not (*e.g.* to question 2, or to question 5)?
7. Are shear instabilities involved at any stage, and are they relevant to question 1?
8. What quantities should be monitored in order to be able to forecast warmings?
9. To what extent, and in what sense, does the troposphere behave independently of the stratosphere (for the purposes of question 1 for instance), and how should we represent tropospheric-stratospheric coupling in mechanistic models?

In recent years we have been coming a good deal closer to answering some of these questions. A major reason is the impact of data from satellite-borne infrared radiometers. In particular, the global coverage from satellites has permitted more reliable estimates of the stratospheric circulation to be made on a daily basis, enabling a close dialogue between theory and observation to take place for the first time. The observations are not only giving a much better idea of how, for instance, Eulerian-mean zonal wind profiles change from day to day, but even some idea (Butchart *et al.*, 1982; Chapman and Miles, 1981; Kanzawa, 1980; Kanzawa and Hirota, 1981; O'Neill and Youngblut, 1982; Palmer, 1981a, b; Simmons, 1982b) of the harder-to-estimate quantities which theory tells us must be central to the dynamics. These include isentropic potential-vorticity gradients and associated planetary-wave refractive indices, and the convergence of the Eliassen-Palm wave flux. Of course we have been lucky, scientifically speaking, in that the essential phenomena really do seem to have the deep vertical scales already remarked on, so that they can be seen rather well by the satellite radiometers. Were this not so, the picture now emerging could hardly have the degree of dynamical self-consistency which it seems to have.

There has been at least one other piece of scientific good luck. Nature decided to present the satellites and the FGGE observers with a specially significant sequence of events in January and February 1979. As I shall now argue, that sequence of events, culminating in the major, wave-2-dominated warming of February 1979, contains some particularly important clues about the dynamics.

2. Why wave 2, and why January-February 1979?

It might be asked why we should be specially interested in wave 2, when many of the warmings observed during the past decade seem to have been more or less dominated by wave 1. I think that it is precisely the comparative rarity of wave-2-dominated warmings that makes them unusually interesting. In some ways they are the severest test of our understanding—particularly as regards question 2 on my list. It seems very likely that, in order to get a major warming, stationary planetary waves must not only attain large amplitudes, and have phase tilts of the type usually associated with propagation from

below, but must also be unusually well focused into the high-latitude polar cap, say latitudes $\geq 60°$. The small mass and moment of inertia of that region gives the waves by far their best chance of causing dramatic effects. Now observation and theory both suggest that although stationary waves 1 and 2 can often propagate quite happily up from the troposphere into the wintertime stratosphere, they also have a general tendency to propagate equatorwards, *away* from the polar cap, much as one might expect from a consideration of the spherical geometry of the earth. For instance if one were to start a wave propagating horizontally along a latitude circle, at high latitudes, it would tend to go off at a tangent, along something like a great circle path (Hoskins and Karoly, 1981). The resulting tendency for the waves to avoid the high-altitude polar cap, and propagate into the much larger areas available elsewhere, could be called "defocusing" for want of a better term. It is probably one reason why major warmings do not happen more often. Defocusing tends to be more pronounced for wave 2 than for wave 1, as originally found by Matsuno and confirmed by subsequent studies, most recently the ray-tracing calculations of Karoly and Hoskins (1982) reported elsewhere in this issue.

In order to quantify such things as focusing and defocusing, it is important in practice to use a conservable measure of wave propagation. The use for instance of eddy fluxes of geopotential, as a measure of wave propagation, tends to obscure the issue. They are strongly affected by the local strength of the mean westerlies in a way that has nothing to do with focusing (Eliassen and Palm, 1961; Bretherton and Garrett, 1968). A similar thing happens with measures of wave amplitude such as eddy geopotential height, which in a tight polar-night jet tends to be roughly proportional to the local jet speed (Simmons, 1974, Eq. 10), representing the speed with which air parcels travel through a given streamline or latitudinal displacement pattern (cf. Edmon *et al.*, 1980, §2d). So inspection of the magnitudes of such quantities in a latitude-height cross-section may tell us little more than where the jet is. They are also very deceptive as indicators of such things as vertical propagation times. The eddy fluxes of geopotential do, to be sure, tell us the direction of the planetary-wave group velocity in a meridional plane, to the extent that the notion of group velocity is valid, a question which I shall touch on in section 3. But their magnitudes are potentially misleading as a guide to whether or not the waves are converging onto a given height and latitude.

It is fortunate, therefore, that there is another measure of wave flux which is just as good for indicating the direction of the group velocity (when that notion applies), but which is both easier to compute from observations and is also a true *conservable* measure of the flux of wave activity across an arbitrary zonal-wind profile. That is, it does not converge unless either the waves are building up transiently at the place in question, or there are some dissipative or other departures from conservative motion. This is the so-called Eliassen-Palm wave flux F. It has a number of other useful properties, to some of which I shall refer later. Its horizontal component is proportional to minus the Eulerian northward eddy momentum flux, and its vertical component proportional to the northward eddy heat flux. For a recent review the reader may consult section 2 of the paper by Edmon *et al.* (1980). Examples of the use of the EP flux to describe stratospheric planetary waves have been given by Butchart *et al.* (1982), Dunkerton *et al.* (1981), Kanzawa and Hirota (1981), O'Neill and Youngblut, 1982, Palmer (1981a, b), and Sato (1980). I have been using the term "Eliassen-Palm cross-section" to refer to latitude-height cross-sections showing both F and its divergence.

The pair of EP cross-sections shown in Figs. 1a and 1b gives an excellent illustration of the defocusing of wave 2 under "typical" or "climatological" conditions. Fig. 1a is taken from Dunkerton *et al.* (1981; *q.v.* for further details) and Fig. 1b from unpublished work by C.-P. F. Hsu (personal communication). They were obtained from a pair of model warming simulations using Hsu's (1980) modification of the semi-spectral model developed by Holton (1976), in which wave 2 was forced at an artificial lower boundary in much the same way as in Matsuno (1971). (The extent to which this is a valid procedure will emerge later, when we address question 9 on the list.) Figs. 1a and 1b represent an early stage in the simulations, before the mean state has changed very much. Each simulation was started with exactly the same initial state, a zonally-symmetric state close to the kind of climatological zonal mean which has often been used in such modelling studies, with a broad polar-night jet merging smoothly into the still broader mesospheric westerlies (Hsu, 1980, Fig. 1a). The convergence of F shown by the negative

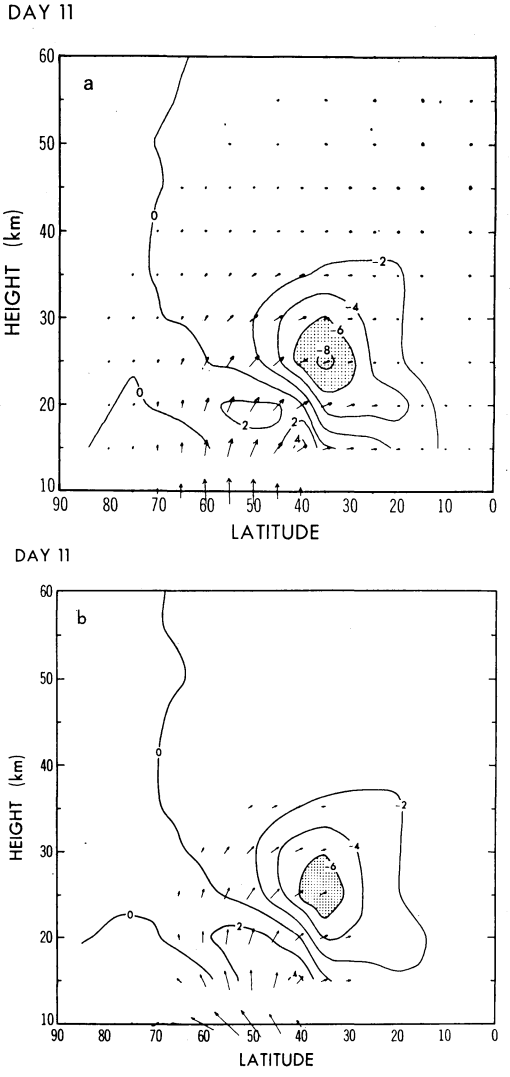

Fig. 1 Eliassen-Palm cross-sections for a pair of model simulations in which planetary waves of zonal wavenumber 2 are generated by applying two different lower boundary conditions. The waves propagate on a basic zonal wind profile typical of what has usually been taken as a representative climatological mean. The arrows represent the Eliassen-Palm wave flux and the contours its divergence, plotted in accordance with the conventions described in Dunkerton *et al.* (1981), from which case (a) is taken. Case (b) is from unpublished work by C.-P. F. Hsu, with kind permission. The arrow scales are such that the arrow patterns look nondivergent if and only if the flux is nondivergent.

contour values is attributable mainly to transience, simply showing where the waves are arriving—a process which is being slowed down by the proximity of a subtropical zero-wind line. Similar cross-sections for earlier times in the simulations show the waves emerging from the boundary and then turning equatorwards. (They also show clearly how the waves speed up as they get into slightly stronger westerlies, and *vice versa,* just as suggested by the ray propagation times shown in Fig. 7 of Karoly and Hoskins, 1982, page 119 in this issue.) The phrase "*turning* equatorwards" refers of course to the way things appear in the meridional cross-section.

The difference between the two simulations is due solely to different lower boundary conditions. The purpose of changing the boundary condition was to try to persuade the waves to focus into the polar cap simply by forcing F to point poleward at the boundary. This was done by imposing a boundary forcing with a southeast-northwest phase tilt, an experiment suggested by O'Neill and Taylor (1979). It is striking how easily the defocusing effect frustrates this attempt. The same phenomenon appears to be implied by the theoretical and observational results shown in Figs. 5b and 6b of Matsuno (1970) and in Fig. 1 of Hirota and Sato (1969). Above 20 km, the waves hardly seem to notice the difference, and turn towards the equator regardless. Indeed the subsequent evolution of the two simulations from the stage shown in Fig. 1 was astonishingly similar, even as regards the timing of the various stages of mean-flow evolution described by Dunkerton *et al.* (*op. cit.*). Essentially similar results have been independently obtained by Butchart *et al.* (1982), using a different numerical model.

Results like these add to the growing body of evidence suggesting that, no matter what the troposphere is doing, conditions in the stratosphere have to be prepared in some special way before a major warming can take place (Butchart *et al.,* 1982; Dunkerton *et al.* 1981; Kanzawa, 1980; Labitzke, 1981; Palmer 1981b; Quiroz et al., 1975, §2b), especially a warming dominated by wave 2. Something is needed which can overcome the defocusing effect and guide planetary waves upwards into the polar cap. For wave 2 we may expect the requirements to be more stringent than for wave 1. Whatever these requirements are, it is clear that they were well satisfied just before the major warming around 20 February 1979. Fig. 3b of Palmer (1981a) shows strong focusing of wave 2 from below on 19 February—the direction of F was tilted well in towards the pole, a state of affairs quite

the reverse of that shown in our Fig. 1a.

This is reason enough for paying special attention to the case of February 1979. But there are further reasons. Without this case, we would have significantly less evidence concerning what the requirements for focusing might actually be. Most observed warmings appear to involve not only planetary waves of very large amplitude, but also more than one wave component simultaneously. It is possible that the mere presence of one large-amplitude wave might help to focus another. If this kind of nonlinear, wave-wave interaction in the stratosphere were an essential ingredient in the warming process, a thorough investigation of it by numerical experimentation would be a daunting task indeed. The strength of such interactions is likely to be sensitive to all kinds of variables, including details of the basic state chosen, and the number of possibilities to be explored before full understanding could be claimed would be enormous. It is here that February 1979 has provided a clue of the first importance in the scientific detective story. As I shall explain, it is a specially clear example in which the focusing of wave 2 seems unlikely, in fact, to have depended crucially on the presence of other wave components. This is a very direct piece of evidence bearing on question 4 in my list.

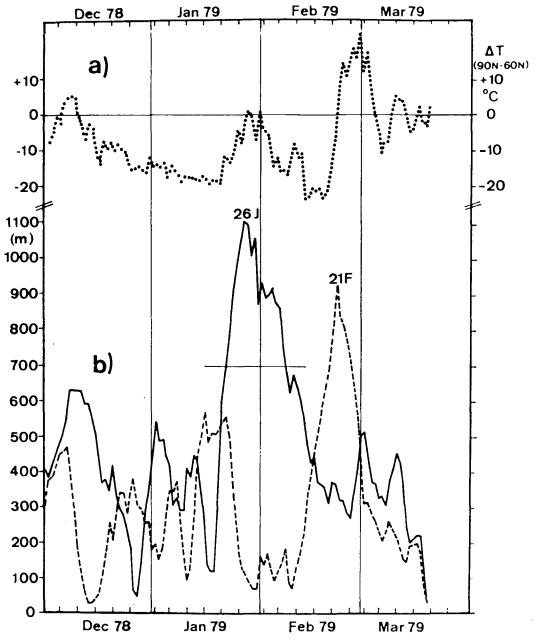

Fig. 2 (a) Difference at 30 mb, around 24 km altitude, between the temperature at the north pole and the zonally averaged temperature at 60°N; (b) amplitudes in metres of zonal harmonic geopotential height waves 1 and 2 (broken line) at 60°N and 30 mb. After Labitzke (1981).

This evidence is strengthened and its value enhanced by the fact that, in many other respects, the events of January and February 1979 seem to follow a pattern not untypical of other major warmings, especially those of 1967-68 and 1970-71. Fig. 2, adapted from Labitzke (1981), shows for instance the observed time variation of the wave amplitudes, and of the latitudinal temperature contrast in the polar cap, at 30 mb during January and February 1979. The upper curve (a) shows the zonal-mean temperature difference between the north pole and 60°N, and the lower curves (b) the geopotential height amplitudes of waves 1 and 2 at 60°N. These curves may be compared with an essentially similar series for the years 1964-81 reproduced in the article by Labitzke (1982, Fig. 1, page 127 in this issue). See also the review by Schoeberl (1978). In the present case, the time evolution of the wave amplitudes follows more or less what Schoeberl calls the "type A" pattern, in which a small wave-2 pulse (occurring in mid-January in this case) is followed by a large peak in wave 1, and finally by another wave-2 pulse of variable strength at about the time of the main temperature rise. Also quite variable are the exact timing of the main temperature rise and the final wave-2 pulse. The present case would appear to be an extreme case mainly in one respect, which is the unusually long delay of about four weeks between the final wave-2 pulse and its wave-1 predecessor. The delay is about double what has been seen in other cases. Tropospheric observations for the same period (*e.g.* Fig. 4 of Quiroz, 1979 or Fig. 3 of Labitzke, 1981) suggest, as Dunkerton *et al.* remark, that "the warming occurred late simply because no wave-2 forcing was available earlier" from the troposphere. Whatever the reason for the delay, one consequence of it was that the final wave-2 pulse, and the warming itself, took place with only a comparatively modest amount of wave 1 present. We can scarcely avoid the impression that to get a major warming of this general type it may not be essential to have waves 1 and 2 both present simultaneously.

What then was the role of the huge wave-1 peak in late January, which gave rise only to a "minor" warming? The foregoing suggests that its essential effect could only have been to *pre-*

condition the basic state so that it was able subsequently to focus wave 2. The idea that effects of this kind might be important has recently been stressed by several authors (Butchart *et al.*, 1982; Dunkerton *et al.*, 1981; Kanzawa, 1980; Labitzke, 1981; Palmer, 1981b). It is certainly an observational fact, evident from even a cursory inspection of stratospheric synoptic maps for January-February 1979, that the basic state was very different after the big wave-1 event from what it was before. Afterwards the polar-night jet was much narrower. The dynamical implications of this warrant further discussion, and I shall take them up in the next two sections. I am not saying, of course, that wave-wave interactions were of no significance at all. It would be naive to think so, and we already know from model experiments that they can matter at least for getting the details right (*e.g.* Butchart *et al.*, 1982; Hsu, 1981; Lordi *et al.*, 1980). However, it now seems very likely that there are real cases in which wave-wave interaction is not the most dynamically fundamental effect, and it seems sensible to try to develop a really good understanding of such cases first.

I have been keeping up my sleeve yet another noteworthy recent development, which once again concerns the stratospheric events of January-February 1979, and which incidentally removes any lingering doubts, if such remain, as to the essentially dynamical character of those events. The main January and February events were successfully forecast, in considerable detail, by a numerical model being developed at the European Centre for Medium Range Weather Forecasts (ECMWF) (Simmons, 1982b). The January event appeared to be quite well simulated starting from 16 January up to the main wave-1 peak in Fig. 2 ten days later, and the February event from 13 February at least up to the splitting of the vortex around 20 February. This is the first truly successful forecast of a real sudden warming which I have heard of. The model was an experimental version of the high-resolution forecasting model at present being used operationally at the ECMWF. It has a hybrid (sigma, p) vertical coordinate system with the top level at $p=10$ mb, instead of the 25 mb ($\sigma=0.025$) top level of the operational model. The operational, sigma-coordinate model did, in fact, capture some aspects of the events in the stratosphere as well, including the splitting of the vortex at 50 mb (Bengtsson *et al.*, 1982). The existence of simulations like these implies, for one thing, new opportunities for the extraction of the dynamically relevant diagnostics which put such a strain on observational data. A realistic, high-resolution simulation can provide "data" which is more dynamically consistent than raw observations.

Daily maps of Ertel's potential vorticity in isentropic surfaces would be one important example. The availability of such maps would open the way to all kinds of refinements in our understanding of the dynamics—things which are more or less well hidden by present-day diagnostics, EP cross-sections included! For instance I suspect that isentropic potential-vorticity maps might make it obvious why there was a second phase of mean zonal deceleration in the polar cap around 26 February, well after the splitting of the polar vortex (Palmer 1981a). A rough estimate of advection times in each half of the split vortex seems consistent with a simple explanation in terms of the advection of potential-vorticity "debris" around the two cutoff lows. This idea might also help to explain the failure of a lower-resolution simulation (Butchart *et al.*, 1982, to be referred to in the next section) to reproduce this further episode of mean zonal deceleration, since following the "debris" in detail after the vortex splits would immediately place a considerable strain on numerical resolution. Such a regime of motion marks the point at which the eddies have largely ceased to be wave-like, dynamically speaking; one may say that they have "saturated", or "broken". It is quite like what happens to ocean waves on a beach. An even closer analogy is the breaking of tides and internal gravity waves in the mesosphere, since in that case there is no question of having two immiscible fluids like air and water, and so the basic gradient (static stability) to which the waves owe their existence is mixed irreversibly. Similarly, irreversible mixing of the isentropic potential-vorticity gradient may be regarded as the distinguishing feature of a breaking planetary or Rossby wave. As soon as such wave-breaking occurs, the detailed potential-vorticity distribution will become very complicated, and "wave-wave" interactions (between very many zonal harmonics) will be prominent in any detailed description. Indeed thinking too literally in terms of "waves", in the dynamical sense, may not then be very profitable.

3. A test of the focusing hypothesis

The main hypothesis suggested by the observations, as discussed so far, is that the unusual

focusing of wave 2, just before the major warming of February 1979, was simply due to an unusual configuration of the basic zonal state. This hypothesis has been directly tested in a beautifully-conceived series of numerical experiments carried out at the U.K. Meteorological Office by Butchart *et al.* (1982). The results provide very strong support for the essential correctness of the hypothesis, and in the process resolve the long-standing question as to why a different overall behaviour was found in Matsuno's wave-2 simulation and its successors. The model used was a more elaborate one than Matsuno's, being a finite-difference, primitive-equation model with somewhat more zonal resolution than would correspond to Matsuno's semi-spectral truncation to one zonal wavenumber. The model did not attempt to represent the troposphere. It was forced in much the same way as Matsuno's, by prescribing the geopotential artificially at 100 mb.

The essential point which these numerical experiments establish is simply that the results are, indeed, sensitive to the basic state adopted.* They clearly vindicate "the need for special care in the choice of initial conditions for model simulations" suggested by Quiroz *et al.* (1975, §2b). If a climatological mean state is used as in Matsuno's simulation and its successors, then the model exhibits just the familiar behaviour first found by Matsuno for wave 2, which begins in the defocused way illustrated by Fig. 1 above and hence cannot produce a warming without first undergoing a long period of mean-flow evolution, typically twenty days or more. The way in which enough focusing is eventually achieved to give rise to a warming in this kind of wave-2 model experiment has been elucidated by Dunkerton *et al.* (1981). The focusing depends upon the partially-reflecting properties of the nonlinear "critical layer" associated with a zero-wind line which moves northwards from the subtropics and reflects the waves back into higher latitudes, a bit like an artificial side wall.

If, on the other hand, the actual mean state on 16 February 1979 is used, with its much narrower polar-night jet, then there is immediate focusing without any preliminary period of mean-flow evolution. Moreover, if care is taken to use a forcing having a zonal phase speed of the order of that observed at 100 mb, which was eastward and significantly different from zero, then even with pure wave-2 forcing the focusing persists long enough to produce a strong warming. This takes only about ten days, despite the fact that the initial state is zonally symmetric. It should be remembered of course that the real stratospheric circulation on 16 February was far from being zonally symmetric, as is obvious from Fig. 2. When the actual initial conditions and the actual 100 mb forcing were used, so far as could be determined from the observations, then the model came closer still to imitating the warming that actually occurred. A warming looking quite like the real one was achieved in less than *five* days.

Butchart *et al.* argue persuasively that some important aspects of the model's behaviour could have been anticipated by inspection of meridional cross-sections of the refractive index squared, R^2, appropriate to a linear, wave-2 disturbance propagating steadily at the prescribed phase speed on the Eulerian zonal-mean state (Charney and Drazin, 1961; Matsuno, 1970). R^2 is the basic quantity entering into the "ray theory" of planetary-wave refraction in a meridional cross-section. It contains a term proportional to the latitudinal isentropic gradient of potential vorticity, divided by the velocity of the mean zonal wind relative to the wave. Rays tend to bend towards regions of large, positive R^2 (Palmer, 1981b; Karoly and Hoskins, 1982, §2d, page 112 in this issue); and Butchart *et al.* found that the EP wave flux in the model tended to behave in a similar way. In particular, the spatial distribution of R^2 seemed to account satisfactorily for the focused and defocused initial EP flux patterns found for the actual and climatological mean states. Results carrying similar implications have been obtained by O'Neill and Youngblut (1982), from an observational study of the January 1977 warming which included some ray-tracing calculations based on suitably smoothed observational estimates of R^2 at different times. It is quite remarkable how well the refractive-index and ray-tracing concepts, and by implication the concept of group velocity, seem to succeed in predicting important aspects of the behaviour of the wave fluxes. Meridional and vertical wavelengths, and temporal rates of change of the mean state, are all too large for the relevance of those concepts to be self-evident—to say nothing of the presence of additional complications such as interference between stationary and travelling wave components, which can cause transient fluctuations in the direction and magnitude of the wave fluxes (*e.g.* Boyd, 1976; Madden, 1975; Palmer, 1981a,

* *Note in proof:* See also Bridger and Stevens (1982).

appendix; Schoeberl and Strobel, 1980a).

It seems certain, then, that computations of R^2 and associated ray paths are going to be an important aid to understanding the results of future numerical experiments on stratospheric planetary waves. A cross-section of R^2, at least, would be virtually indispensable, before one could tell *a priori* whether there was any possibility of a given, narrow polar-night-jet profile focusing wave 2. In this situation the refractive index is very sensitive to the precise shape of the jet velocity profile. For any given profile there is always a "tunnelling" region of negative R^2, or imaginary R, near the pole, which rays by definition cannot enter. Rather, they bend away from it, accounting for some of the defocusing effect already discussed. The negative values are due to a term in R^2 proportional to minus the zonal wavenumber squared, which for the moment I shall call somewhat loosely the "defocusing term" even though it is not in fact the only term which can cause defocusing. Its magnitude quadruples when we go from wave 1 to wave 2, and as the pole is approached it always dominates the term involving the potential-vorticity gradient, as a result of geometrical factors multiplying the two terms.* In order to have positive R^2 somewhere in the polar cap, which is a necessary condition for rays to be able to enter that region at all, the competition between the two terms has to go the other way—the potential-vorticity-gradient term must dominate the defocusing term—somewhere in the polar cap. Because of the geometrical factors this has its best chance of happening if the largest potential-vorticity gradients are concentrated towards the *south* side of the jet maximum, the side furthest from the pole, as suggested by the heavy curves in Fig. 5a below. (Potential vorticity Q is shown on the left and zonal velocity u on the right; the thin curves suggest climatologically typical profiles.) It appears from Butchart *et al*'s results that the mean state on 16 February had essentially this configuration in the lower stratosphere, and that R^2 for wave 2 was not only positive on the south flank of the jet, but had a positive maximum there, both for waves with zero phase speed and for waves with the observed phase speed. Because of the local maximum this is a configuration capable of causing focusing, but as we shall see shortly this is not the whole story! Generally speaking, the competition between the two terms under discussion means that in the polar cap, especially for wave 2, even the sign of R^2 can be sensitive to the precise shape of the zonal velocity profile, which has to be differentiated twice to get the gradient of Q. A recent series of numerical simulations of linear planetary-wave behaviour (Lin, 1982) confirms the expected sensitivity by showing how it takes only small changes in the shape of the jet to change the refractive-index configuration completely and give a drastically different pattern of wave propagation.

Cross-sections of R^2 can hardly be expected to tell us everything, on the other hand. And there is one specially important limitation on their validity which has not yet been discussed. Broadly speaking, the notions of refractive index, ray theory and group velocity are likely to work best in strong westerlies, but worst near "critical lines" where mean zonal wind velocity equals zonal phase velocity. Their validity, even for linear Rossby waves, fails utterly *at* a critical line (Andrews and McIntyre, 1976 Appendix B; Grimshaw, 1980). Moreover, if we use synoptic maps to estimate the nonlinear advective terms for typical planetary-wave amplitudes we find that linear wave theory itself fails utterly as well. In the real atmosphere, a critical line will always be surrounded by a nonlinear "critical layer" in which the waves saturate, or break, in the sense referred to at the end of the last section, and within which R^2 is completely irrelevant.

As originally suggested by the idealized models of Benney and Bergeron (1969), Davis (1969), and others, such a nonlinear critical layer can act as a reflector once the waves have broken

* See e.g. Matsuno (1970, eq. 11). It should be noted that some authors prefer to work with a quantity corresponding to R^2 with the defocusing term omitted, so as to be able to plot cross-sections which apply to more than one zonal wavenumber. The defocusing term must be included, however, if tunnelling regions are to correspond to regions of negative R^2. In Fig. 5c of Karoly and Hoskins (1982, page 117 in this issue), the tunnelling regions for zonal wavenumber n are those with contour values less than n. Note also that Karoly and Hoskins' definitions incorporate the geometrical effects into a transformation to Mercator coordinates. Palmer (1981b) adopts a different coordinate transformation which leads to another definition of refractive index, corresponding to Matsuno's definition divided by the sine of latitude squared. This transformation assumes that static stability is independent of latitude, but is then very convenient since it makes the propagation appear uniformly isotropic in height and latitude.

(quite irrespective of how R^2 behaves in its immediate neighbourhood). The general circumstances under which this nonlinear Rossby-wave reflection can occur have been clarified recently, and they will be discussed in section 5. They appear to include the circumstances of present interest, to a large extent, both for the real stratosphere and for mechanistic models of it. Zonally truncated mechanistic models cannot properly represent the wave-breaking process itself, but surprisingly (and very fortunately) they do manage to imitate the nonlinear reflection in a crude but qualitatively not unreasonable way, as was first shown by Geisler and Dickinson (1974) and further explained by Dunkerton *et al.* (1981, appendix B).

Butchart *et al.* note the possibility that reflection from a critical layer situated to the south of the polar-night jet might have played a role in maintaining the focusing which occurred in their most realistic numerical experiments, just as it did in the late stages of the simulation discussed by Dunkerton *et al.* The most realistic experiments, in which the focusing persisted long enough to induce a strong warming, were just those in which the phase speed of the wave forcing was realistic, as well as the initial mean state. With a phase speed of the right order (Butchart *et al.*, Fig. 1b) one finds, for instance from the mean zonal velocity cross-section in Butchart *et al*'s Fig. 2a, that a critical line did exist in middle latitudes on 16 February, extending from below 30 mb to above 5 mb. Its shape and position were, in fact, remarkably similar to the shape and position of the mid-latitude zero-wind line found a few days before the major warming in Dunkerton *et al*'s model simulation (*op. cit.*, Fig. 2c), in which the waves were nearly stationary. The subsequent evolutions were also quite similar. If we set these facts alongside the theoretical evidence presented by Dunkerton *et al.*, it seems not only possible, but indeed practically certain, that reflection from the associated critical layer must have been taking place in all the experiments with realistic phase speeds.

Not only must reflection from the critical layer have been taking place, but I see no escape from the further conclusion that the reflection must have been the primary reason, if not the *only* reason, for the persistence of focusing which characterized all the experiments with realistic phase speeds. The persistence cannot be explained simply by persistence of the initial local maximum in R^2. That maximum disappeared as the polar-night jet decelerated, at least in those cases for which cross-sections of R^2 were displayed. These included two experiments with realistic initial mean states one of which had zero and the other a realistic phase speed. In the former case there was no critical layer to stop the defocusing effect from reasserting itself as soon as the local maximum in R^2 disappeared, and defocusing was exactly what then happened. The disappearance of the maximum in both cases is hardly surprising when one recalls its sensitivity to the jet configuration.

It would be tempting at this point, and not unreasonable on the evidence so far, to conclude that the foregoing statements apply also to the real warming of February 1979. The critical line was certainly present*, and it is likely that the associated wave-breaking region would have tended to act as a reflector. What is less easy to be sure about is the behaviour of the refractive index, not only because of data problems but also because in the real atmosphere, as opposed to a truncated model, there is less reason to suppose that the Eulerian zonal mean is a good basis for estimating R^2. I shall return to the latter problem in my concluding remarks.

It is worth adding one more remark about critical lines. If a critical line happens to lie within a region where large-scale, isentropic gradients of potential vorticity are anomalously weak in the first place, then the region will act as a reflector even before wave breaking takes place. All critical-layer theories agree on this (*e.g.* Tung, 1979, Eq. 46 with $\hat{\beta}_c=0$). The reason is that the absorption predicted by the usual linear, transient, critical-layer theory (Dickinson, 1970; Warn and Warn, 1976) depends on the development, through advection in the early stages preceding wave breaking, of a certain pattern of eddy potential vorticity in the critical-layer region (Stewartson, 1978; Warn and Warn, 1978, Fig. 2b). That pattern could not develop if there were no large-scale isentropic potential-vorticity gradient across the critical-layer region to start with. Of course when the waves do break, the consequent mixing of potential vorticity tends to ensure that the large-scale potential-vorticity gradients *become* weak even if they were not weak in the first place; this in essence is the nonlinear critical-layer reflection

* Contrary to a tacit assumption by Dunkerton *et al.*, that wave 2 was stationary in the real atmosphere on the relevant days in February 1979.

mechanism. In the present instance, however, there is good reason to suppose that in mid-February the large-scale, isentropic potential-vorticity gradients were already weak, on average, in middle latitudes, as suggested schematically by the heavy curve in Fig. 5a below. It is here that the wave-1 precursor comes into the story.

I should not risk leaving the reader with the impression that Butchart *et al.* claimed to have explained every last detail of the February 1979 warming solely in terms of interaction between the mean state and wave 2. On the contrary, they found that even the modest amount of wave 1 present in the actual initial conditions and 100 mb forcing seemed to be significant for approximating the observed behaviour in a very small region within a radius of ten degrees' latitude or so from the north pole. By way of comparison, the main region of zonal-mean easterlies resulting from the February 1979 warming covered a much larger area, out to a radius of about thirty degrees, *i.e.* to about 60°N. Butchart *et al.* point out that truncation errors due to finite differencing near the pole may have been significant in their simulations, and in view of the fact that the otherwise very similar final stages of Dunkerton *et al*'s pure-wave-2 simulation did not corroborate this detail, it is perhaps a cause for some concern. But it can also be remarked that sensitivity to wave 1 very near the pole is what one might expect in any case from a synoptic viewpoint. The somewhat artificial process of taking zonal means within ten degrees of the pole can obviously give quite variable results with even the slightest departure from wave-2 symmetry.

4. The wave-1 precursor

What then of events prior to 16 February? In particular, how did the mean state take up a "non-climatological" configuration with a narrow polar-night jet favourable to wave-2 focusing? The synoptic maps, for example those for geopotential height at 10 mb between say 23 January and 10 February, show beyond reasonable doubt that the wave-1 event which dominated that period must have caused a great deal of quasi-horizontal mixing of potential vorticity in middle latitudes. Throughout that period, the Aleutian high was well developed as a cutoff high spanning a range of latitudes reaching from the subtropics to about 80°N, as illustrated for 27 January by Fig. 3, taken from Pick (1979). The closed-streamline circulation around this huge system must inevitably have been advecting potential vorticity straight across the planetary gradient, twisting up the isopleths of potential vorticity in each isentropic surface like spaghetti on a fork. Much the same thing would have been happening to the isopleths of ozone and other quasi-conservative tracers. A model calculation giving some idea of the kinematics of the

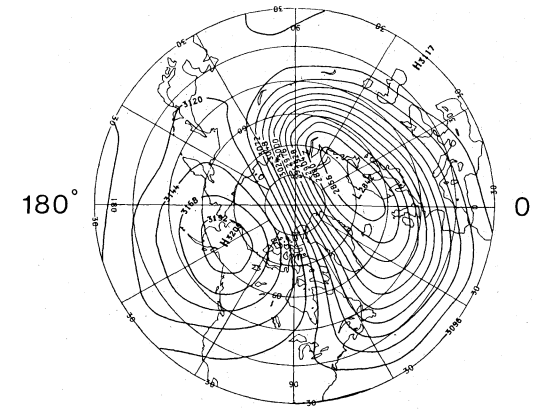

Fig. 3 Breaking planetary wave at 00Z on 27 January 1979, as shown by the height of the 10 mb constant-pressure surface. Contour interval is 24 dekametres; H=32.09 km, L=28.42 km. From Pick (1979). Latitudes north of 30°N are shown.

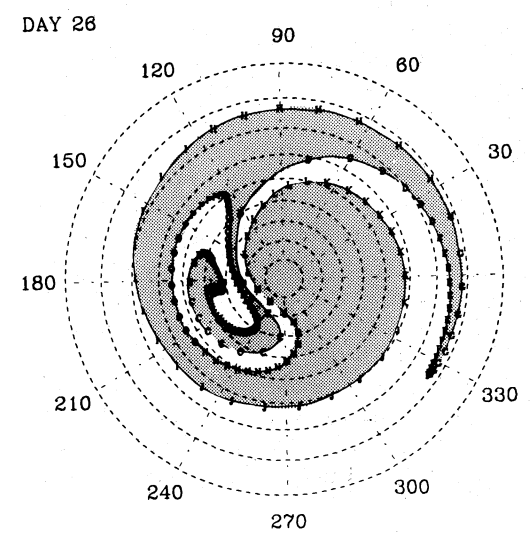

Fig. 4 Shape of a material line, originally coincident with the 30°N latitude circle at an altitude of about 31 km, and then advected by a wind field of a qualitatively similar pattern to that implied geostrophically by Fig. 3. The wind field was generated in a mechanistic model simulation. From Hsu (1981). Whole hemisphere is shown.

process is shown in Fig. 4, taken from Hsu (1981), in which the light areas represent low-potential-vorticity air from the tropics and *vice versa*. The boundary between the two air masses is a material boundary which was initially coincident with the 30°N latitude circle, before a pure wave-1 disturbance was switched on. This of course is another example of what I have been calling a breaking, or saturating, planetary wave. To that extent it is essentially the same as what goes on in a nonlinear critical layer, but now on such a grand latitudinal scale that the term "layer" begins to be something of a misnomer.

Note incidentally that in terms of streamfunction, approximately equal to geopotential height divided by Coriolis parameter, the Aleutian high would have been centered further south in Fig. 3, but that Fig. 3 stops at 30°N whereas Fig. 4 shows the whole hemisphere. Of course the correspondence is intended to be qualitative at most.

Exactly what an actual potential-vorticity map would look like for the real stratosphere, as the wave-1 peak died down during the first half of February, would be extremely hard to guess without the help of a very accurate numerical simulation. By that time, the horizontal mixing would have produced a complicated, sheared-out pattern of small-scale potential-vorticity debris—the phenomenon underlying the so-called "potential enstrophy cascade" (*e.g.* Rhines, 1979)—with some bits and pieces still being pulled round the weakening Aleutian high, and others in the outer part of the displaced polar vortex, which incidentally undergoes considerable fluctuations in shape and position during the period of large wave-1 amplitude. The very important dynamics involved in the whole process seems not always to have been fully appreciated by synopticians—one sometimes hears about the "mere" strengthening of the Aleutian high. Perhaps this has been because of the near-impossibility of drawing isentropic maps of potential vorticity from even the best data analyses and thus seeing directly what is going on. The experimental, high-resolution ECMWF forecasts for late January 1979 would seem to be our best hope at present of being able to say anything quantitative about this, and I am hoping that Dr. Simmons will be able to do something about it before too long!

However, we can guess the qualitative effect on the zonal-mean state easily enough, as already hinted, using the quasi-conservative property of potential vorticity. The ideas date back to the old arguments of G. I. Taylor, C.-G. Rossby and others concerning the mixing of absolute vorticity in barotropic flows, and of course are intimately related, in various ways, to the ideas developed for instance in Dickinson (1969), Davies (1981), Geisler (1974), Green (1970), Holton and Dunkerton (1978), Rhines and Holland (1979), and Rhines and Young (1982). From the synoptic evidence we can expect that most of the mixing was centered on middle latitudes in this case, and therefore that as far as the net effect on the larger scales are concerned, *i.e.* ignoring the small-scale "debris", the isentropic potential-vorticity gradients would have tended on average to be smeared out in middle latitudes, as suggested schematically by the heavy curve in Fig. 5a to which we have already referred. A smearing-out of large-scale mean gradients in middle latitudes implies a sharpening of gradients at the edge of what is left of the polar-night jet, giving rise to a tighter and narrower jet as suggested (again schematically) by the heavy curve in Fig. 5b. As we have already seen, the basic-state

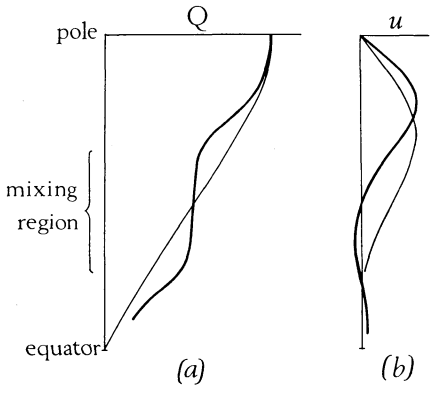

Fig. 5 (a) Schematic latitudinal distributions of Ertel's potential vorticity Q on an isentropic surface in the stratosphere, before and after a large-amplitude planetary wave breaking event centred on middle latitudes. Thin curves are "before", and heavy curves "after". (b) Corresponding polar-night jet profiles u. A broad jet (thin curve) is converted into a narrow jet (heavy curve) with a region of small $\partial Q/\partial y$ to the south of it where the potential vorticity has been most strongly mixed. No attempt is made to suggest the small-scale structure of Q due to "debris" from the wave-breaking event. The profiles may be thought of as representing Eulerian zonal means *after* the wave has largely decayed; see remarks near the end of section 4 about reversible and irreversible changes in the Eulerian-mean state.

configuration suggested by Fig. 5 is precisely the sort of configuration needed to focus an upward-propagating wave-2 pulse, if and when wave 2 decides to amplify subsequently in the troposphere.

Of course an ensuing major warming—the end result of the subsequent pulse, if any, being focused and then breaking *in* the polar cap—can itself be looked at from the viewpoint of potential-vorticity mixing (Davies, 1981). From this viewpoint the difference between a minor warming and a major warming is simply that middle latitudes are mixed in the first case, but the polar cap is mixed in the second.

Presumably the configuration sketched in Fig. 5 would be able to focus wave 1 just as well if not better than wave 2; and a subsequent wave-1 pulse in February 1979 could presumably have led to a major warming just as well as the wave-2 pulse that actually occurred. That is more or less what seems to have happened in the final warming of 1973-74, if the diagram corresponding to Fig. 2 is any guide (Labitzke, 1982, Fig. 1, page 127 in this issue). For that matter, if large wave-1 amplitudes had simply persisted a little longer in the present, January-February 1979 case, there could have been a more or less continuous evolution into a wave-1 dominated major warming. Whether the evolution is continuous or "pulsed" does not seem to be a specially fundamental distinction. There are suggestions in Labitzke's Fig. 1 that wave-1-dominated examples with roughly continuous evolution occurred in February 1980, and in 1969-70. Schoeberl (1978) calls the 1969-70 case a "type B" warming, defining this category by the criterion that wave 1 "maintains a large amplitude for a long period." O'Neill (1980) describes in some detail another example of a strong, wave-1-dominated warming which evolved in a more or less continuous way, taken not from observations of the real atmosphere but from a 13-level general circulation model. See also the nonlinear mechanistic model simulation reported by Hsu (1981).

These remarks have obvious implications for question 8 on my list. The forecasting of a major warming seems certain to depend equally crucially on two separate things. One is an accurate estimate of the initial potential-vorticity gradients in the polar-night jet, together with relevant phase speeds and critical-line positions. The other is an accurate estimate of just how long the anomalous forcing from the troposphere will persist. It is hardly surprising that no simple rule of thumb for predicting major warmings has been found.

A complementary view of the process suggested by Fig. 5 is given by wave, mean-flow interaction theory. The zonal deceleration in middle latitudes suggested by Fig. 5b is precisely what that theory would predict for a not-too-well-focused planetary-wave pulse which saturates somewhere in middle latitudes instead of in the polar cap. Wave 1 can probably do this under a wider range of conditions than wave 2, which may explain why "an intense development of height wave 1" is usually necessary, according to Labitzke (1978), before a major warming can occur.

In an Eulerian-mean description of the effect of a wave pulse like the big wave-1 peak in Fig. 2, an appropriate measure of "where" the waves break or saturate is the convergence of the EP flux integrated over the time of the whole wave event. The time integration gets rid of the purely temporary, reversible changes which may complicate the Eulerian-mean picture from moment to moment as wave amplitudes fluctuate. This ties in with our previous view of Fig. 5 because the convergence of the EP flux is approximately proportional to the isentropic flux of potential vorticity, as is well known (*e.g.* Green, 1970, Eq. 11; Edmon *et al.*, 1980, Eq. 3.5); and the time integration over the wave event picks out the net contribution representing irreversible, downgradient mixing of potential vorticity (cf. Rhines and Holland, 1979; Edmon *et al., op. cit.*, p. 2610; Hoskins, 1982).

The way in which the time integrated EP flux convergence enters the wave, mean-flow interaction theory can be seen directly from eqs. (4.1a, d) of Dunkerton *et al.* (1981) for the rates of change $\partial\bar{u}/\partial t$ and $\partial\bar{\theta}/\partial t$ of the Eulerian-mean zonal velocity $\bar{u}$ and potential temperature $\bar{\theta}$. Those equations are the prognostic members of the set of transformed Eulerian-mean equations presented by Andrews and McIntyre (1976, 1978a; see also Boyd, 1976, eq. below 3.9). If changes in static stability are neglected, the transformed equations for $\partial\bar{u}/\partial t$ and $\partial\bar{\theta}/\partial t$ may be integrated over any given time interval to give equations of the same mathematical form but involving only the net changes $\Delta\bar{u}$ and $\Delta\bar{\theta}$ in $\bar{u}$ and $\bar{\theta}$, instead of the instantaneous rates of change $\partial\bar{u}/\partial t$ and $\partial\bar{\theta}/\partial t$. The time-integrated EP flux convergence now appears in place of the instantaneous convergence.

Wave-mean theory highlights two interesting points about the net mean-flow change $\Delta\bar{u}$ suggested by Fig. 5b. First, the theory predicts not only mid-latitude deceleration where the strongest EP wave flux convergence occurs, but also a general tendency for acceleration to occur north of that location, if there is comparatively little EP convergence there. This point was noted by Palmer (1981b). Second, there is a strong tendency for the main deceleration region to be narrower latitudinally, and deeper vertically, than the region of actual EP wave flux convergence. This phenomenon is nicely illustrated, from observational data for the real atmosphere, by Figs. 5a and 5b of O'Neill and Youngblut (1982). It is one reason why the amount of focusing or defocusing is so important for questions 2 and 8 on my list. As far as the net effect on the $\bar{u}$ profile is concerned, it really does matter at what latitudes the waves saturate. For a given state of the troposphere, the precise degree of wave focusing in the stratosphere could easily make all the difference between getting a major warming such as that of February 1979, and a minor warming such as that of January 1979.

Examples which nicely illustrate these two points have been given by Dunkerton *et al.* (1981) and Hsu (1981), and Fig. 6 recalls one of them. It is taken from a later stage in the same mechanistic simulation that produced Fig. 1a, but still well before the final major warming. The heavy curve marked D_F represents the divergence of the EP flux, in the quasi-geostrophic approximation, rescaled so as to represent the effective Eulerian-mean zonal force per unit mass due to the waves (with dimensions of acceleration). This makes D_F *equal* to the northward flux of quasi-geostrophic potential vorticity (within the approximations usually associated with quasi-geostrophic theory). D_F incorporates the principal Eulerian eddy heat as well as momentum fluxes, so that quasi-geostrophically there is no other wave-induced forcing of the mean state, as described by the transformed Eulerian-mean equations. That is why the response, as measured by the actual mean zonal acceleration $\partial\bar{u}/\partial t$, follows D_F much more closely than either the eddy momentum flux convergence or the Eulerian-mean Coriolis acceleration, shown by the thin curves in the figure. It can be seen, however, that the region of deceleration is narrower in latitudinal extent than D_F itself (and it is, in fact, very much deeper vertically, although the figure does not show this). It can also be seen that positive, westerly acceleration is indeed occurring in high latitudes. This phenomenon was remarked on by Holton (1976) in connection with a similar model simulation, and it can also be seen in Fig. 7b of Hsu (1981), which relates to a different, wave-1-dominated simulation.

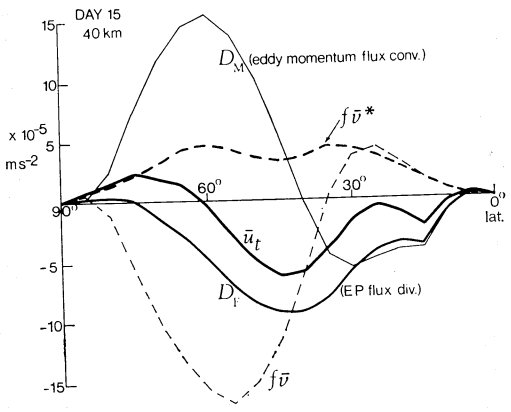

Fig. 6 Terms in equations governing the mean zonal acceleration $\bar{u}_t$, in a situation where the EP wave flux is converging mainly in middle and low latitudes. D_F is the EP flux divergence normalized so as to represent a zonally-directed force per unit mass, and $f\bar{v}^*$ is the Coriolis force due to the residual meridional circulation, which appears in the transformed Eulerian-mean equations. The light curves show the principal terms in the conventional mean momentum equation for comparison. From Dunkerton *et al.* (1981).

Both the high-latitude acceleration, and the narrowing and deepening of the region of deceleration, are immediate consequences of the general way in which a balanced zonal flow responds to a given zonal force D_F per unit mass concentrated at a given height and latitude, a classical problem studied by Eliassen (1951) and discussed in this context by Dunkerton *et al.* (*op. cit.*, §4), by Palmer (1981b, §5), and by O'Neill and Youngblut (1982). In Fig. 7, also taken from Dunkerton *et al.*, the shaded region shows the height and latitude where D_F is greatest in our example. Eliassen's theory tells us that the response will include a meridional circulation whose Coriolis force redistributes the effect of the force D_F, and whose vertical advection re-orients the isentropic surfaces, in such a way as to preserve thermal-wind balance. In the transformed Eulerian-mean formalism, the relevant meridional circulation is what Andrews and McIntyre called the "residual circulation" ($\bar{v}^*$,

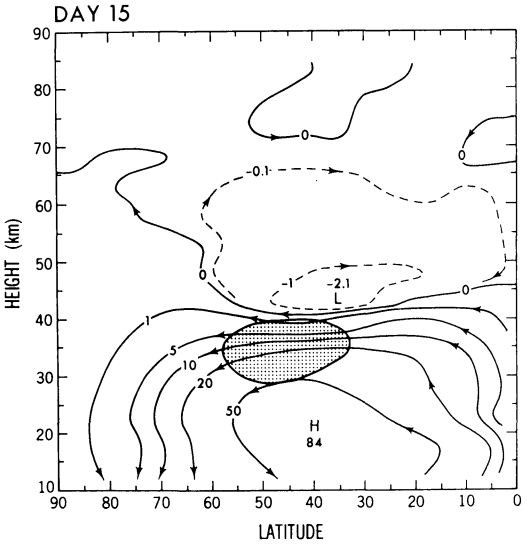

Fig. 7 Region of maximum D_F (shown shaded), for the same situation as in Fig. 6, and mass streamfunction $\bar{\chi}^*$ for the residual Eulerian-mean circulation ($\bar{v}^*$, $\bar{w}^*$). The shaded region represents values of $-D_F$ exceeding 10^{-4}m s^{-2}. $\bar{\chi}^*$ is defined such that $\partial\bar{\chi}^*/\partial\phi = a\rho_0\bar{w}^*\cos\phi$ and $\partial\bar{\chi}^*/\partial z = -\rho_0\bar{v}^*\cos\phi$, where ϕ is latitude, a the radius of the earth, $\rho_0(z)$ a standard density, and z is 7 km times $-\ln$ (pressure/1000 mb). Contour values are to be multiplied by 1.46 kg m^{-1} s^{-1}. From Dunkerton *et al.* (1981).

$\bar{w}^*$). The residual circulation induced by D_F in the present example is also shown in Fig. 7. It extends over many scale heights and has the simple, two-cell structure predicted by Eliassen's theory. The lower cell is acting to tilt the isentropic surfaces anticlockwise in the picture, and the upper cell clockwise. The northward flow through the region where D_F is concentrated extends into the polar cap, where the associated Coriolis acceleration ($f\bar{v}^*$ in Fig. 6) causes the positive, high-latitude zonal acceleration already noted. The Coriolis forces in the two branches of the return flow effectively extend the region of *deceleration* high up into the mesosphere and, apparently, well down into the troposphere, so far as we can tell from a mechanistic model with an artificial lower boundary condition. I should emphasise that none of these general features of the response depend in any way whatever upon linear wave theory. Eliassen's theory applied to the transformed Eulerian-mean equations shows that the response will have the character just described whenever the EP wave flux converges, for whatever reason, onto a sufficiently well localized region in middle latitudes in the manner suggested by the shaded region in Fig. 7.

This picture seems to apply quite well to the 1979 wave-1 episode, provided that we are careful to interpret it in the time-integrated sense. For quantitative purposes the effects of diabatic cooling would have to be added. The relevant period of time appears to be mid-January to mid-February. Fig. 4c of Labitzke (1981) gives a latitude-time section of the mean zonal wind $\bar{u}$ at 10 mb, and Fig. 5 of the same paper gives three meridional cross-sections of $\bar{u}$ during that period. Those figures confirm that $\bar{u}$ did undergo a net change $\Delta\bar{u}$, over the whole time interval, very like that suggested by our Fig. 5b, involving deceleration in middle latitudes and acceleration in high latitudes. But they show also that it happened in at least two distinct stages. (Fig. 3 in the same paper suggests that even that may be a considerable oversimplification, when transient events at higher altitudes are taken into account.) At 10 mb, Labitzke's Fig. 4c shows a net Eulerian-mean deceleration over the first few days, from about 18 to 25 January, in a fairly broad region spanning middle and high latitudes. Much of this Eulerian-mean deceleration is attributable to wave transience associated with the rapid growth of wave-1 amplitude during that period. Viewed synoptically, part of the effect of wave-1 "transience" on the Eulerian mean is just the kinematical effect of displacing the main, cyclonic polar vortex out of line with the latitude circles around which the mean is taken, as is illustrated by our Fig. 3.

Most of the high-latitude acceleration appears to have taken place, somewhat erratically, during a second stage which occupied the period from about 26 January to mid-February. Weak deceleration continued, on average, in middle latitudes during most of that period. From Labitzke's figures and the daily synoptic maps it can be anticipated that the EP wave flux would have been diverging from high latitudes and converging (weakly) into the main region of wave-breaking to the south, when time-averaged over this second stage. Some preliminary EP cross-sections constructed by Dr. Palmer appear to confirm this, in addition showing considerable day-to-day fluctuations as expected.

The EP divergence in high latitudes, time-averaged over the second stage from 26 January to mid-February, was probably due mainly to transience in the opposite sense from before, associated with a net reversible decay of the local wave amplitude as planetary-wave activity propa-

gated (on average) off to the south. I am using the term "reversible" in the sense implied by the discussion given by Andrews and McIntyre (1978b, §5.3 and appendix C). Viewed synoptically, this reversible decay, and the associated Eulerian-mean zonal acceleration, are associated with the displacement of the surviving part of the main cyclonic vortex back towards the pole and back into line with latitude circles. By "surviving" I mean the inner part of the main vortex, which was centered over Scandinavia at the stage shown in Fig. 3, and which largely escaped the effects of potential-vorticity mixing (cf. Hsu, 1980, Fig. 7c), so that it was a material entity which carried its isentropic potential-vorticity distribution back with it. (From a Lagrangian viewpoint this entity *is* the narrow polar-night jet, whether or not it is displaced away from the pole.) Reversibility, in the sense under discussion, is the physical reason why Palmer (1981b) was able to explain an episode of high-latitude, Eulerian-mean acceleration towards the end of the second stage using the theory of linear, conservative waves.

In terms of the Eulerian-mean description, then, the high-latitude EP flux divergence would have cancelled much of the earlier convergence due to the initial growth of the wave, leaving a region of maximum time-integrated convergence in middle latitudes resembling that discussed in connection with our Figs. 6 and 7. In Eulerian-mean language one can say that the net change in the mean state over the whole wave-1 episode was of the kind illustrated by Figs. 5, 6 and 7 because the transience due to the arrival of the wave-1 disturbance proved to be more or less reversible in high latitudes, but more or less irreversible in middle latitudes in virtue of the fact that it was in middle latitudes that the waves saturated and ultimately dissipated. Thus one can summarize the essential information about the wave-1 precursor event, despite the complexities of the actual day-to-day evolution of the Eulerian-mean state which showed a mixture of reversible and irreversible changes, by the simple statement that a convergence of the *time-integrated* stratospheric EP flux took place in middle latitudes, where the waves ultimately dissipated.

One can summarize the same sequence of events in a more Lagrangian, synoptically-oriented language by saying that as the wave-1 disturbance grew to large amplitude in late January the main polar vortex was displaced southward off the pole, as illustrated by Fig. 3, carrying a central core of potential-vorticity contours with it. Potential vorticity contours outside some such central core would have been eroded away by the wind field associated with the cutoff Aleutian high, as suggested by Fig. 4. The result would have been weak potential-vorticity gradients outside, and sharp gradients at the edge of, the surviving core region which went back towards the pole when the wave-1 disturbance decayed in mid-February—again, just the configuration needed to focus the subsequent wave-2 pulse, sketched in Fig. 5. There is of course no fundamental reason why the process should have stopped there; had large wave-1 amplitudes persisted a little longer, the core could have been eroded away entirely, in one continuous operation. We would then have had an example of a "type B" major warming.

A question now arises as to whether the buildup to the precursory wave-breaking episode we have been discussing was facilitated in its early stages by the proximity of a subtropical critical line. That is another question to which I shall return, but I can hardly resist the temptation to point out immediately that three papers elsewhere in this issue (Holton and Tan, 1982; Labitzke, 1982; Wallace and Chang, 1982) put forward observational evidence carrying the suggestion that major warmings may be somewhat more likely to occur in years when the tropical quasi-biennial oscillation is in its easterly phase. One fact which may prove significant, and which can be seen from Fig. 3 of Labitzke (*op. cit.*, page 130 in this issue), is that the two winters in the period 1953-1980 in which the quasi-biennial easterlies occupied the *deepest* layer (January 1963 and January 1977) both produced warmings of exceptional strength. In January 1979 the layer of tropical easterlies was not so deep, but it nevertheless extended from 30 mb up to higher than 7 mb (Labitzke, *loc. cit.*; Coy, 1979b).

5. Resonance?

So far I have concentrated on those questions on my list concerning which the greatest progress has been made recently. One question I have ignored altogether is question 1, namely why wave amplitudes should become large in the first place. That question is avoided—or so it would seem at first sight—by mechanistic model simulations of the kind I have referred to, in which wave amplitude is simply prescribed at the

bottom of the stratosphere. (I have also, so far, been ignoring question 9, by tacitly following the conventional wisdom that the troposphere does act more or less independently of the stratosphere, which of course is a prerequisite for such an artificial lower boundary condition to make any kind of sense at all.)

Observational evidence has long suggested that a good answer to question 1 for the real atmosphere would entail a good answer to the age-old question of what causes tropospheric "blocking" and related anomalies. A theoretician is tempted at once to suggest resonant growth. According to this idea, planetary waves forced by topography or stationary thermal forcing can grow anomalously if the basic state evolves into a configuration such that a free mode of the whole atmosphere exists and is nearly stationary. This possibility has been discussed by a number of authors, most extensively by Tung and Lindzen (1979a, b); see also Clark (1974) and Simmons (1974).

Recently Plumb (1981a, b) has pointed out that for realistic wave amplitudes the strongest growth is to be expected not when the linear condition for resonance is satisfied, but rather when the mean state is initially to one side of resonance, such that if a stationary, topographically-forced wave starts to grow the change in the mean state induced by the wave growth takes the mean state further *towards* resonance. Such "self-tuning" can then lead to further wave growth. It is precisely this positive feedback process that gives rise to the inviscid topographic instability noted earlier by Charney and DeVore (1979, §2) and further studied by Paegle (1979). For general initial conditions, with a free, travelling planetary wave present, Plumb's scenario implies the slowing down and growth of the travelling wave as the mean state approaches resonance. Interference with the stationary, forced wave contributes to the growth of the total disturbance in its early stages and hence to the evolution of the mean state. The model scenario is strikingly reminiscent of the behaviour of travelling waves which is often observed to precede major and minor warmings in the real stratosphere. For instance, such behaviour was observed to precede the large wave-1 event of January 1979 (Madden and Labitzke, 1981; Quiroz, 1979; for several other examples see Quiroz, 1975). This slowing down of the travelling wave is itself a clear indication of an approach to resonance.

Plumb presents a combination of detailed analytical theory and numerical experiments which demonstrate convincingly that nonlinear resonance in this sense plays an important role at least in certain idealized model warming simulations, of the general sort first studied by Geisler (1974). Geisler's model is like Matsuno's but ignores questions of latitudinal propagation and focusing, by restricting the flow to a beta-plane channel bounded by latitudinal, perfectly-reflecting walls, and truncating spectrally in the meridional as well as the zonal direction (see also Holton and Mass, 1976). Clearly there is plenty of scope for a resonant cavity to form in such a model. It can happen for instance if vertically propagating planetary waves are reflected back down from high altitudes in the model, in the manner discussed by Plumb, and by Tung and Lindzen in their second paper (1979b).

In spherical geometry, the defocusing effect illustrated in Fig. 1 suggests that reflection from the tropics or subtropics is likely to be more important than reflection from above, in helping to form a hemispheric or smaller cavity capable of exhibiting resonant behaviour. Wave activity could be sent back along the ray paths suggested by the directions of the EP fluxes in Fig. 1 if there were a suitably oriented reflecting surface in the subtropics. (During resonant growth the *net* wave flux would still be represented in that case by arrows directed as in Fig. 1.) As already mentioned, one of the ways in which such a reflecting surface could arise is through the presence of a nonlinear critical layer, as was proposed in this context by Tung (1979). The conditions under which reflection would occur have been much debated, but it can be shown very generally, following an argument presented by Killworth and McIntyre (1982), that

(a) if a quasi-geostrophic Rossby-wave critical layer is entraining no new potential-vorticity contours (by growing in width, for instance, or by translating sideways), and
(b) if the overall time scale is short enough for the critical-layer region to be considered free of external, zonal-mean forcing (including that due to mean diabatic effects, and to mean viscous stresses, if any, at the edges of the critical-layer region),

then the critical layer cannot sustain wave absorption for much longer than the time for wave breaking to occur. After that it must reflect, at least in a time-averaged sense. (The

sustained absorption observed in general circulation statistics applying to seasonal and longer time scales can be traced to violation of condition (b).) The time for wave breaking to occur depends on the wave amplitude, and can be roughly estimated as the time for an air parcel near the center of the anticyclonic system of closed streamlines in the critical layer, when viewed in a frame of reference moving zonally with the wave, to travel about halfway round the center. Killworth and McIntyre's argument is a rigorous version of that already sketched in connection with Figs. 4 and 5, the essential point being simply that conservation of potential vorticity restricts the net amount by which the mean profile of potential vorticity can change, in a given latitude band, quite irrespective of the details of the wave breaking and any associated potential-enstrophy cascade. This imposes a bound on the time integrated potential-vorticity flux, and therefore on the time integrated EP flux convergence.* The bound is evidently zero if the initial potential-vorticity gradient is zero, which verifies the fact, mentioned earlier, that a critical layer will reflect immediately if *previous* wave-breaking events, or other causes, have already annihilated large scale potential-vorticity gradients in the critical-layer region.

If a reflecting surface exists in the tropics or subtropics, then the higher its latitude in the winter hemisphere, the smaller the resulting cavity, and the greater the potential for a rapid response. The whereabouts of a critical layer depends of course on the phase speed of the waves giving rise to it. But, other things being equal, we would expect poleward critical-layer positions to be more liable to occur when tropical winds are easterly. That is why I am inclined to believe the hint from observations, referred to at the end of the last section, that stratospheric warmings may be facilitated by the presence of a deep layer of quasi-biennial easterlies.

I am not trying to suggest, of course, that nonlinear critical layers are always perfect reflectors, even when condition (b) is well satisfied. We should not forget condition (a), especially in a strongly time-dependent situation. As soon as wave amplitudes grow, as happened in late January 1979 for instance, the critical layer will widen, and eat into the ambient potential-vorticity gradient on each isentropic surface. Killworth and McIntyre's argument suggests that the critical layer will act as a partial absorber as long as it continues to entrain new potential-vorticity contours. Such entrainment would occur, as noted in the statement of condition (a), either if the wave were growing or if its phase speed were changing appropriately relative to the mean flow, making the critical layer translate sideways. Observations suggest (*e.g.* Madden and Labitzke, 1981, Fig. 3) that *both* effects were occurring in late January 1979. So if resonant growth was involved it may well have been slowed down to some extent by the resulting wave absorption. Dunkerton *et al.* (1981, appendix B) argued that the translation effect reduced, but did not eliminate, the reflectivity of the nonlinear critical layer in the mechanistic model simulation studied there. The importance of such effects, which like those studied by Plumb represent an intrinsically nonlinear aspect of the resonance problem, has yet to be assessed quantitatively.

It is of interest to look for resonant behaviour of the kind under discussion in mechanistic model simulations permitting latitudinal wave propagation, such as the one studied by Dunkerton *et al.*, an early stage of which provided Fig. 1a. For models with an artificial lower boundary, as in this case, the precise conditions for resonant behaviour would be unlikely to agree with those for the real atmosphere, but the phenomenon would be fundamentally the same. C.-P. Hsu and I believe we have found an example of such behaviour in this same model simulation. A weak resonance effect seems to be the correct explanation of a spontaneous growth in the EP wave flux from the bottom boundary which Dunkerton *et al.* had noted but had been unable to explain. This growth, by a factor 2 or so (compare Fig. 1a with Dunkerton *et al.*'s Fig. 1c), is quite important in helping to induce the simulated major warming. It appears to be attributable to reflection from the nonlinear critical layer in the model, as it advances from the subtropics towards a position in middle latitudes from which some of the reflected waves constructively interfere with those forced by the boundary. Evidently this is one of the possible mechanisms whereby real atmospheric cavities,

* Moreover the bound still holds even if the potential vorticity is subject to eddy diffusion within the critical layer, provided only that condition (b) is not violated. Note on the other hand that no such bound applies to absorption by *gravity-wave* critical layers, since there is no constraint analogous to that imposed here by conservation of potential vorticity.

as well as model ones, might tune themselves toward resonance.

One check on this interpretation comes from the time evolution of the phase tilt in the lower stratosphere shown in Dunkerton *et al.*'s Fig. 3b. Qualitatively speaking it compares well with the phase behaviour of simple analytical solutions describing wave growth in a resonant cavity with the same artificial lower boundary condition, for instance the solutions presented by Simmons (1974, Figs. 10 ff.). The amplitude behaviour is qualitatively similar also (Dunkerton *et al.*, Fig. 3a; Simmons, *loc. cit.*). From a comparison with the patterns of amplitude behaviour shown in Fig. 18 of Matsuno (1971), in Fig. 4a of Schoeberl and Strobel (1980a), and in Figs. 8 and 16 of Koermer *et al.* (1982), one gains the impression that a similar resonant enhancement may have occurred in those simulations as well. Moreover, there is at least one published example of what looks like an occurrence of essentially the same signature in the real stratosphere. This is apparent in the wave-2 amplitude shown in Fig. 2a of Hirota and Sato (1969), which refers to the lower stratosphere in January 1963. It may be significant that the mean zonal wind shown in the same figure is suggestive of a northward-moving critical layer just when the amplitude has the appearance of growing resonantly.

We now come to a subtle and intriguing question. What, if any, is the connection between the foregoing ideas and tropospheric phenomena such as "blocking"? There seem to be two quite separate versions of the resonance theory in the literature, of which the first is the one we have been discussing:

Version 1: The troposphere and stratosphere act as one big cavity, which on occasion becomes tuned—or rather suitably *detuned,* as we should say in the light of Plumb's work—in such a way as to exhibit resonant behaviour and lead to large planetary-wave amplification.

Version 2: The troposphere acts as a cavity by itself, and decides what it will do largely independently of the stratosphere, to a first approximation. The stratosphere merely responds as to a given forcing from below.

Tung and Lindzen's second paper (1979b), for instance, is concerned with version 1, and their first paper (1979a) relates more directly to version 2. In support of the notion of an independent troposphere involved in version 2 there is the long-familiar fact that many aspects of the large-scale flow in the middle and upper troposphere seem to be captured by simple barotropic models —hence the time-honoured notion of "equivalent-barotropic" flow in the troposphere. The ray-tracing results of Karoly and Hoskins (1982, page 117 in this issue) provide further support by suggesting that the leakage of stationary planetary waves from the troposphere into the high stratosphere is quite modest: many ray paths originating in the troposphere stay within the troposphere or lower stratosphere, because the troposphere tends to have a higher refractive index squared than the lower stratosphere in middle latitudes (see also Fig. 3 of Matsuno, 1970, and Fig. 8b of O'Neill and Youngblut, 1982). Comparatively few rays go high into the stratosphere, even for zonal wavenumbers 1 and 2. No rays for zonal wavenumbers 3, 4, etc. can cross the lower stratosphere at all in middle and high latitudes, which for those wavenumbers and the conditions assumed is a "tunnelling" region where the corresponding refractive index squared is negative (*i.e.* where the quantity plotted in Fig. 5c of Karoly and Hoskins, 1982, page 117 in this issue, is less than 3, 4, etc.). Further theoretical evidence for the approximate independence of the troposphere can be found in recent papers by Held (1982), Hoskins and Karoly (1981), and Simmons (1982a).

I should say at once that I am not trying to suggest that there is any real conflict between these two versions of the resonance theory. As will be explained more fully in the next section, I think they simply represent idealizations of two different aspects of the problem. Version 1 may well help explain at least some of the large, ultra-long wave events seen most prominently in the upper stratosphere, but should probably not be advanced as an explanation of "blocking". Whether version 2 can explain "blocking" is a question which we have not yet touched on. A closer look at that question seems worthwhile, since it will lead to a better appreciation of the nature of the connection between blocking and stratospheric warmings, and in the process to what I believe is a new suggestion for overcoming the problems introduced by artificial lower boundaries (and hence incorrect resonances) in mechanistic models of stratospheric warmings.

Version 2 of the resonance theory involves two separate questions. One is whether the troposphere acts independently of the stratosphere, and the other is whether an independent, equivalent-barotropic troposphere behaves like a resonant cavity. Now the idea that the whole

troposphere, or one hemisphere of it, might behave as a resonant cavity seems inconsistent with tropospheric observations. For instance the idea does not seem to fit in very well with the observed fact that blocking highs in, say, the Pacific are not accompanied very often by those in the Atlantic or elsewhere round the globe. To be sure, the wave-2 peak in late February 1979 shown in Fig. 2 was associated with simultaneous Atlantic and Pacific blocking highs, but this seems to be unusual. Synoptically one has the impression that the simultaneity in such examples is fortuitous, and that certain locations are dynamically special in some way, independent of what is going on in most other parts of what one might have envisaged as the tropospheric cavity.

That impression is reinforced when one looks at the "teleconnection" patterns whose existence was foreseen by pioneers like G. T. Walker and J. Bjerknes, and which have been emerging more and more clearly in recent years through the systematic use of objective statistical techniques on long time series of tropospheric data. Some recent references are the papers by Wallace and Gutzler (1981) and the thesis by Dole (1981). It now seems clear that these patterns are an important key to understanding the variability of the tropospheric stationary-wave patterns on time scales from several days to a few weeks. As is illustrated by Fig. 2, these are the time scales of interest for the dynamics of stratospheric warmings. One of the most remarkable things about the teleconnection patterns is that, provided the time-mean state and the anomalies about it are defined appropriately, using suitable low-pass filtering of the time series, the spatial anomaly patterns in a given geographical location, particularly the large-amplitude anomalies found in the north Atlantic and Pacific, look much the same for anomalies of either sign.

The independence of sign carries a suggestion that some kind of linear theory ought to be relevant. Indications that this is indeed the case have been emerging from recent work on Rossby-wave propagation on a sphere (*e.g.* Hoskins and Karoly, 1981, & refs.). A number of such calculations have suggested that direct, one-way propagation of trains of stationary, equivalent-barotropic planetary waves from sources in the tropics may account for a good many aspects of the observed teleconnection patterns. The calculations use equations linearized about a zonally symmetric state. When plausible estimates of dissipation are used, resonance plays no significant part in these calculations. There is, however, a still more recent twist to the story. It comes from a remarkable series of model experiments by Simmons (1982a), which extends Hoskins and Karoly's work and appears to take us significantly closer to explaining some aspects of the teleconnection patterns, particularly the large amplitudes of anomalies in the north Atlantic and Pacific. I am told that G. Branstator of NCAR (Boulder, Colorado) has done similar experiments independently, with similar results. The experiments of particular interest to us used a simple barotropic spectral model of the troposphere, and took as basic state not a zonally-symmetric state but, rather, an observational estimate of the climatological January 300 mb height pattern. In the model this pattern was held steady by the application of a suitable distribution of vorticity forcing, conceived as representing whatever topographical, thermal, and time-averaged transient eddy effects are needed to maintain the climatological pattern in the real atmosphere. A localized anomaly in the forcing was then introduced by switching on an additional vorticity source. This was tried at various places in the tropics. Strikingly large responses were obtained in the north Atlantic and Pacific regions, for certain (respectively different) forcing locations. There was no special tendency for the Atlantic and Pacific responses to occur together. The responses were larger by up to an order of magnitude, in streamfunction or geopotential height, than those obtained in experiments with the same forcing anomaly using zonally-symmetric basic states.

These results again reinforce the view that resonance is not likely to be involved on a hemispheric or global scale. The large response builds up well before there is time for significant information to get into other parts of the hemisphere and reflect back. The general behaviour and in particular the time dependence found in the experiments do seem to suggest, on the other hand, that the north Atlantic and Pacific diffluent-jet regions may be acting as comparatively *localized* resonant cavities, involving wave reflection in the zonal as well as the latitudinal direction. The reflection might be only "partial" reflection, not describable by ray-tracing calculations. The cavities would no doubt be quite leaky in any case. Even a very leaky cavity could produce resonant enhancement by a factor of 2 or 3 without much trouble, which is already

significant for present purposes. This, then, suggests yet another version of the resonance theory, which as far as I know has not been proposed before:

Version 3: In the northern hemisphere winter the parts of the troposphere over the north Atlantic and Pacific act as separate resonant cavities, which can be excited independently of each other and of the stratosphere, to a first approximation. The stratosphere still responds as to a given forcing from below, and the stratospheric response (for a given state of the stratosphere) will tend to be strongest in wave 1 when the Atlantic and Pacific anomalies happen to have opposite signs, and strongest in wave 2 when they happen to have the same sign.

If this version is a good approximation to the truth then there should be some tendency for the "strong wave 1" and "strong wave 2" conditions in the stratosphere to be mutually exclusive, especially at times when the magnitudes of the Atlantic and Pacific anomalies are at their largest. Time series of stratospheric wave amplitudes should tend to show minima in stratospheric wave 2 at about the same time as large maxima in stratospheric wave 1, and vice versa. Such behaviour is indeed observed, and has often been remarked upon. It is a noticeable feature of Schoeberl's "type A" pattern, of which Fig. 2b is an example. Other examples, at 30 mb as in Fig. 2b, can be seen in Fig. 1 of Labitzke (1982, p. 127 in this issue) and at 100 mb in Fig. 1 of Koermer *et al.* (1982). Of course the wave amplitudes, especially in the high stratosphere, must also be affected by the variability in the responsiveness of the stratosphere itself, which is to be expected for reasons already discussed.*

Version 3 of the resonance theory is not the only idealization which might be capable of explaining the facts under discussion, although my present feeling is that it is the most likely one. An alternative possibility is that large tropospheric responses in the Atlantic and Pacific regions might arise simply from horizontal focusing, and as such might be explicable by ray-tracing calculations in which rays bunch together, or cross each other, forming a "caustic", as they go through the region in question, without reflecting back and forth locally. It would be interesting to carry out the appropriate ray-tracing calculations for the barotropic time-mean state in Simmons' model to see whether or not they can account for the model behaviour in some such way.

In either case, the theory as developed so far is still consistent with the original idea that the anomalies in the north Atlantic and Pacific regions have the nature of a one-way response, with or without resonant enhancement, to something that goes on in specific (and respectively different) locations in the tropics. When one asks what that something might be, tropical sea-surface temperature anomalies come naturally to mind. However, while tropical sea-surface temperature anomalies may well account directly for some of the mid-latitude anomalies, especially the longer-lived ones (*e.g.* Horel and Wallace, 1981), recent studies based on general circulation models have demonstrated that this is unlikely to be the whole story. It appears that realistic teleconnection patterns can be produced by general circulation models even when sea-surface temperature is held *constant* (Lau, 1981; M. L. Blackmon, personal communication). Simmons' work seems to suggest a likely explanation for this, too. In his barotropic model experiments the large extratropical response to a given tropical anomaly is sensitive to small variations in the basic state adopted for linearization. The fact that a large response was found at all is itself an indication of such sensitivity. Simmons notes other evidence for this from his experiments, and confirms the sensitivity directly by varying the basic state. If version 3 of the resonance theory proves to be correct, the effects of this sensitivity can probably be thought of in terms of variability in the leakiness of the Atlantic and Pacific tropospheric cavities.

6. Mechanistic models with realistic lower boundary conditions: a suggestion concerning question 9

Let us now summarize the picture that seems to be emerging. The old idea that the troposphere acts almost independently of the stratosphere, especially the upper stratosphere, to a first approximation, continues to be borne out by theoretical developments. But if resonance in the troposphere is involved it seems more likely

* I suspect that this latter consideration accounts for much of the high-altitude variability of waves 1 and 2 found in the two simulations by Koermer *et al.* (1982), using a topographically forced model in which waves 1 and 2 were forced simultaneously and steadily. Indeed the amplitude behaviour found there is reminiscent of the resonant model behaviour discussed earlier in this section.

to be in the sense of version 3 of the resonance theory than in the sense of version 2. The question remains, in any case, as to whether and how such ideas fit in with version 1, the version envisaged in Tung and Lindzen's second paper (1979b) and developed to a further stage of sophistication by Plumb (1981b). As we have seen, that version also has strong claims on our attention. It seems to entail that the troposphere and stratosphere are *not* independent, with the implication that mechanistic models of the stratosphere with artificial lower boundary conditions may seriously misrepresent reality in this respect.

The view which now seems likely to prove correct is that versions 1 and 3 refer to distinct, co-existing "modes" of wave motion in the real atmosphere, although since nonlinear fluid dynamics is involved the term "mode" is not to be taken in too precise a sense. Version 1 applies to the ultra-long zonal harmonics, waves 1 and 2. These do feel the ground, and extend up through the troposphere and far into the upper stratosphere in winter. Observational and theoretical studies of ultra-long travelling waves suggest that they can organize themselves at least part of the time into free normal modes, implying the existence of deep cavities on a hemispheric scale spanning the troposphere and the stratosphere (*e.g.* Madden, 1979; Schoeberl and Clark, 1980; Salby, 1981). The strongest nonlinearity in the dynamics of these waves occurs not in the troposphere but in the high stratosphere, manifesting itself in the saturation or wave-breaking phenomenon already discussed. As we have seen, wave breaking may itself contribute to the formation, or re-shaping and re-tuning, of an effective cavity. For this and other reasons one would expect the damping and structure of ultra-long-wave free modes to be quite variable (and not necessarily classifiable according to the separable vertical and horizontal structures of tidal theory); but the bottom ends of the gravest modes, so far as they have been detected observationally, tend to have an "external", node-free vertical structure in the lowest few scale heights (*e.g.* Madden and Labitzke, 1981, p. 1252). (This particular free-mode structure, incidentally, is not allowed in mechanistic models in which geopotential height is specified at an artificial lower boundary.)

Version 3, by contrast, refers to tropospheric "long waves" having somewhat smaller horizontal scales, on the whole, and an equivalent-barotropic vertical structure which is comparatively well confined to the troposphere (*e.g.* Fig. 21 of Blackmon *et al.*, 1979). For these tropospheric "long waves" the dynamics is linear or nonlinear according to one's viewpoint. Viewed as anomalies about a "climatological" time mean they seem to behave to some extent linearly, if we take at face value the observational and theoretical evidence already discussed. But what might appear as a linear response in the time-mean formulation would certainly require consideration of a complicated set of nonlinear wave-wave interactions to describe it in terms of *zonal* means and deviations. The translation from the one type of description to the other should contain important clues to understanding the relation between versions 1 and 3, and hence to finding good answers to the last part of my question 9, on how the stratosphere should be coupled to the troposphere in mechanistic models.

The idea that the phenomena envisaged in version 3 are nearly independent of those envisaged in version 1 suggests an analogy with Lighthill's theory of aerodynamic sound generation (*e.g.* Crighton, 1981). In its simplest form, that theory considers the generation of low-frequency sound waves (which we are now going to regard as analogous to the generation of the ultra-long stratospheric planetary waves involved in version 1) by an isolated patch of nonlinear, usually turbulent, fluid motion at low Mach number (which we shall regard as analogous to the large-amplitude tropospheric motion involved in version 3). Lighthill's theory exploits the fact that the turbulence is only a weak radiator of sound, and the reaction of the sound back onto the turbulence correspondingly weak. Thus to good approximation the problem can be solved in two separate stages. The turbulent motion can be computed, or observed, by methods which ignore the presence of the sound field and do not attempt to separate "sound" from other motions—for instance the equations of incompressible flow can be used. This motion is then regarded as known to good approximation, and substituted into the nonlinear terms appearing in the equations for *compressible* flow, giving known source terms from which the acoustic response is subsequently computed. The method works because under the assumed conditions, and with a judicious choice of the mathematical form of the source terms, the acoustic far field is insensitive to errors in representing the turbulent motion. (There is no reason, on the other hand, to suppose that the same would be

true of the near field within the turbulent region itself.)

The analogy, then, suggests a similar exploitation of the idea that large-amplitude tropospheric motions can nonlinearly excite the ultra-long planetary waves envisaged in version 1 while themselves remaining unaffected to some approximation. The simplest realization of the idea could start with the observed height field at say 250 mb. Instead of imposing that field at an artificial lower boundary, it could be multiplied by an empirical, equivalent-barotropic vertical structure $f(p)$ suggested by observations (*e.g.* the vertical structure for the Atlantic and Pacific areas shown in Fig. 21 of Blackmon *et al.*, 1979 —note that this falls off sharply in the lower stratosphere), and the result substituted into the nonlinear terms in the dynamical equations of a suitable model. The model would represent the whole atmosphere including the troposphere and the high stratosphere, as in Schoeberl and Strobel (1981b) and Koermer *et al.* (1982). By "nonlinear terms" I mean those contributions to the advection terms which would be neglected if one were linearizing the equations about some reasonable zonally-symmetric state. The contributions to these nonlinear terms from the self-interaction of the prescribed tropospheric motion would be written, so to speak, on the right of the equations, and from then on, as far as the model is concerned, treated as a *known forcing*.

The main interest would then lie in the projection of this forcing onto zonal harmonics 1 and 2. The key point about forcing the model in this way is that it does not change the tuning of any ultra-long-wave resonant cavities which might exist in the model, as envisaged in version 1, but does allow those cavities to be excited by nonlinear coupling to the tropospheric blocking or other anomalies envisaged in version 3. Following Lighthill, one neglects the cross-terms involving products of the given tropospheric motion with the model's response to it. By analogy with Lighthill's theory, one would not expect the response of the model to this forcing to give an improved approximation to the detailed flow in the troposphere itself, which is analogous to the turbulent near field in the acoustic problems. But it might be reasonable to hope that the ultra-long-wave response, including the nonlinear response in the high stratosphere, would be well represented, like the acoustic far field.

The simplest realization described above would use geostrophic winds for the prescribed tropospheric motion, and because of the assumed equivalent-barotropic structure there would be no nonlinear forcing from the temperature-advection terms. More elaborate realizations are clearly possible, using observed tropospheric winds and temperatures at a number of levels, up to 50 mb or so, and it will be important to find out how much difference this makes. It will also be of great interest to see how much it matters whether or not the tropospheric fields are *low-pass-filtered* to suppress fast-evolving, synoptic-scale motions such as mid-latitude depressions. Since the latter tend to be organized into "storm tracks" on a global scale the associated nonlinear effects might turn out to contribute directly and significantly to the forcing of the ultra-long waves.

7. Shear instability?

The idea that shear instabilities, either barotropic or baroclinic, might play a major role in sudden warmings has been out of favour for many years now, and for quite good reasons on the whole, although suggestions that they might be important have been made from time to time since Charney and Stern's well-known paper of 1962 (*e.g.* Kuo, 1979; J. Frederiksen, personal communication). As far as I know, no-one has ever produced a remotely realistic-looking stratospheric warming simulation of which a major cause was shear instability. In his study of a warming occurring in a general circulation model, O'Neill (1980) tested whether necessary conditions for shear instability had been satisfied, and concluded that such instability was unlikely to have been important. Of course potential-vorticity patterns of the sort suggested by Fig. 4 do undoubtedly tend to be barotropically unstable on small scales. A specific example has been analyzed in detail by Killworth and myself (1982). However, that kind of instability is probably more important for expediting the small-scale mixing of potential vorticity than for large-scale developments. As such it could be said to play a supporting role in the drama but not a leading one.*

Nevertheless, the transient details of what I have been calling the "mixing" of potential

* In particular, the instability makes no difference to the conditions (a) and (b) mentioned in section 5, under which sustained wave absorption by critical layers is impossible—contrary to an earlier speculation of mine reported at the IAMAP meeting in Canberra in 1979.

vorticity depend on the precise history of planetary-wave behaviour as well as on the pre-existing distribution of potential vorticity. While we remain comparatively ignorant over what the potential-vorticity distributions on isentropic surfaces look like during real warmings it is difficult to rule out the possibility that breaking planetary waves might sometimes produce sufficiently large regions of negative potential-vorticity gradient at the right moment for significant large-scale instabilities to ensue. In addition, I have not mentioned diabatic effects, which could in principle directly generate such gradients, albeit rather slowly. In the analysis by Kanzawa (1980) for the period prior to the major warming of January 1973, there are some indications of negative gradients even in the Eulerian zonal mean.

Very recently, O'Neill and Youngblut (1982) made the interesting suggestion, supported by computations of large-scale potential-vorticity gradients from observations, that barotropic instability could have expedited the spectacular warming of January 1977. During that exceptional event, easterly winds appeared in the tropospheric as well as the stratospheric polar cap. The potentially unstable region appeared to be centered on the upper troposphere. It remains to be shown, however, that instability theory would predict growth rates, phase speeds, EP flux patterns, etc., consistent with the observed behaviour. For a pure barotropic instability one would expect horizontal EP fluxes from regions of negative into regions of positive potential vorticity gradient (like Fig. 2 of Edmon *et al.*, 1980, turned on its side). However, the observed directions of the EP fluxes do not seem to bear an obvious resemblance to any such pattern.

It should be cautioned that the zonal mean states predicted by numerical simulations of warmings using *mechanistic models* which are spectrally truncated in the zonal direction could be quite misleading as a basis for estimating the potential for large-scale instability. Such truncated models have a strong tendency to predict negative Eulerian-mean potential-vorticity gradients across the entire region of wave breaking, as was first shown by Geisler and Dickinson (1974). The models exaggerate these gradients because the truncation prevents them from correctly representing the smaller-scale aspects of the mixing process, with the result that the model's imitation of the mixing process takes place on larger scales than is realistic. So the negativeness of large-scale potential-vorticity gradients, and with it the potential for large-scale shear instabilities, would almost certainly be overestimated by any such truncated model.

8. Concluding remarks

Perhaps the most remarkable thread running through the recent developments I have touched upon concerns question 3 on my list. It seems that linear planetary-wave theory is more powerful than one might have thought—provided we treat it fairly by linearizing about a basic state which fits the physical situation as closely as possible. The promising results from simple, barotropic models simulating aspects of tropospheric "blocking" and related anomalies, by linearizing about a climatological time-mean state rather than a zonally symmetric one, are a case in point (section 5). The sensitivity of stratospheric planetary-wave focusing or defocusing to the precise configuration of potential-vorticity gradients in the polar-night jet is another (section 3). This has an obvious bearing on questions 2 and 8 concerning the forecasting of warmings, and in particular it helps answer the old question why major warmings are relatively uncommon.

These discoveries raise the difficult question of what should be meant by a basic state "which fits the physical situation as closely as possible", when as often happens the basic state is observed only in the presence of large-amplitude disturbances. There is no reason to suppose that an Eulerian mean (time mean or zonal mean as the case may be) is a particularly good answer, even though the work of Butchart *et al.* shows that in some circumstances it may be considerably better than no answer. Eulerian averaging will generally smear out the potential-vorticity gradients to which wavelike disturbances are sensitive. The problem is especially acute when wave amplitudes are large, and it suggests that cross-sections of the Eulerian-mean state would not be a very useful aid to forecasting. It should be noted that Butchart *et al.* took advantage of the unusual gap between the wave-1 and wave-2 events in January-February 1979 in order to minimize this particular problem. Their initial conditions were based on the observed conditions for 16 February, near the crossover point in Fig. 2b at which conditions were least far from zonally symmetric.

Of course the question of how the basic state should be estimated may seem beside the point for the purposes of detailed, high-resolution numerical forecasting, to which a division into

"basic state" plus "disturbance" is irrelevant. But a good practical definition might lead to a way of producing two-dimensional observational pictures of potential-vorticity gradients and refractive indices which would be of real help to a forecaster with no access to the full machinery of a high-resolution numerical forecasting facility. And together with the suggestion in section 6 it might lead to the development of a more powerful generation of inexpensive, mechanistic, "wave-mean" models with which to study the scientifically important questions of sensitivity and so on, the answers to which must eventually have an impact on any forecasting procedure.

One simple idea worth trying would be as follows. In the situation suggested by the heavy curves in Fig. 5, the contours of Ertel's potential vorticity in each isentropic surface are tightly spaced in the main part of the polar-night vortex, and will tend (after judicious smoothing to remove small-scale "debris") to have the same overall shape as the vortex itself. The same will be true of the contours of quasi-geostrophic potential vorticity in isobaric surfaces (Charney and Stern, 1962). The idea is to invoke the approximations of quasi-geostrophic theory and to associate with the distorted polar vortex a zonally symmetric basic state which is constructed simply by deforming these quasi-geostrophic potential-vorticity contours back into zonal circles, in each isobaric surface, preserving the area enclosed by each contour (a procedure consistent with the standard approximations of quasi-geostrophic theory). One could then construct a corresponding zonal velocity profile $\bar{u}(y, z)$ by solving a Poisson-like equation in two dimensions, which is a well-conditioned, and by modern standards computationally inexpensive, process. The resulting zonally symmetric polar-night-jet structure would have potential-vorticity gradients closely comparable to the actual large-scale gradients in the distorted polar vortex, and would almost certainly be more relevant to questions of focusing and so on than the Eulerian zonal mean. The precise result would depend on the boundary conditions adopted when solving for $\bar{u}(y, z)$, but the result should be much the same for any reasonable boundary conditions imposed not too close to the polar-night jet, say at the ground, in the mesosphere, and in the tropics. It would be most interesting to see whether a procedure like this could help take simulations like those of Butchart *et al.* (1982) another step closer to reality.

It hardly needs saying that there are many important questions which I have failed even to touch on. Some are discussed in the survey by Murgatroyd and O'Neill (1980). One is the effect of stratospheric warming, and the complementary cooling in the tropics and the summer hemisphere (the dynamical reasons for which were spelt out by Dunkerton *et al.,* 1981, §6) upon the infrared radiation budget of the troposphere (Ramanathan, 1977). Another is the comparison between the northern and the southern hemisphere—a further experiment set up by nature which we should try to understand. It is to be hoped that future modelling studies of planetary-wave events on different basic states will include examples representative of the southern hemisphere, so far as conditions there can be estimated.

There is one aspect of question 6 on the possible role of critical lines which I have not mentioned so far, but which should perhaps be put on record for completeness. It is now being recognized more and more clearly that a descending mesospheric critical line is *not* an important part of the dynamics of sudden warmings (*e.g.* Davies, 1981; Dunkerton *et al.,* 1981; Geisler, 1974; Holton, 1976; Houghton, 1978; Plumb, 1981b).* The essential reason is that the fall-off of mean density with height is more than sufficient to cause rapid saturation and wave breaking on a large scale, once waves with realistic amplitudes can be persuaded to propagate vertically or, even more effectively, to focus strongly into the high-altitude polar cap. Under such circumstances the waves will inevitably break quite irrespective of whether a critical line is present.

It should perhaps be recalled, in this connection, that the only reason why critical lines appear to play a special role in linear theories which assume steady, conservative waves is that such an assumption forces all the EP wave flux to converge onto the critical line. That is, the assumption forces all the wave transience and dissipation to go on inside the critical-line singularity. It is not the critical line *per se,* but rather the transience and dissipation, that is the fundamental feature to be abstracted from such theories when looking for a correspondence with more realistic models and with the real atmosphere. The point was appreciated by Matsuno and Nakamura (1979, §6), and our Fig. 7 provides a good illustration of it. The horizontal cross-flow is not taking place along

* *Note in proof:* Also Grose and Haggard (1981).

a horizontal critical layer, as might be thought at first sight from a superficial recollection of Matsuno and Nakamura's model together with approximate theories relating residual and Lagrangian-mean circulations (*e.g.* Dunkerton, 1978). A horizontal critical line is present, but it is many scale heights further up, at about 80 km, and is quite irrelevant to what is going on. Even if the waves had propagated straight upwards (in fact, they were still somewhat defocused), they would have saturated, or dissipated in other ways, well below 80 km.

Planetary waves which are strongly defocused and propagating equatorwards as in Fig. 1, on the other hand, are by no means so certain to saturate in the absence of critical lines. They are not getting into less dense altitudes so quickly, especially in the case of wave 2, and they are spreading into far larger geographical areas. That is why it is in the subtropical or mid-latitude stratosphere, rather than anywhere in the mesosphere, that critical lines—or, more to the point, any regions of sufficiently weak zonal wind relative to the waves—are likely to be significant for determining the favoured sites for wave breaking. And it is the tendency of such regions to become reflectors after wave breaking, as opposed to absorbers during it, under the circumstances discussed in section 5, which appears particularly significant for understanding how warmings work. Reflection from such regions can play two distinct roles. First, it provides a robust mechanism for counteracting the defocusing effect, as discussed in section 3. The mechanism does not cease to operate when basic velocity profiles change slightly. Second, if the wave-breaking region has a suitable shape and position the reflection from it may cause resonant enhancement of wave amplitudes in the stratosphere as a whole (as anticipated by Tung and Lindzen, and discussed in section 5 under the heading "version 1 of the resonance theory"). This may help bring about the large wave amplitudes which could lead to a precursory minor warming or to a major warming. Observational information about wave phase speeds, as compared to zonal-wind strengths in the mid-latitude and tropical stratosphere during the buildup towards a large-amplitude stratospheric wave event, would therefore be of great interest. Better still, synoptic maps of the wind field viewed in a frame of reference rotating with the angular phase speed of each prominent wave component, especially if accompanied by estimates of the motion of air parcels in anticyclonic regions of closed streamlines (recall Fig. 4), would give an immediate idea of where, and how quickly, the waves are breaking.

So where have we got to with the resonance theory, question 5? (Apart from saying that the answer to question 5 is almost certainly yes.) The theory seems to be evolving with bewildering rapidity, as new observational and theoretical evidence becomes available. There is now no real doubt that nonlinear resonance is significant in at least some mechanistic model simulations of stratospheric warmings (Plumb, 1981b), and in a definitely unrealistic way if artificial lower boundaries are used. It seems very likely that resonant effects will be found in most such models, whether or not they allow for latitudinal propagation, and in section 5 I mentioned some evidence for this already in hand. As far as the real stratosphere is concerned, viewed on the time scales of sudden warmings, the weight of evidence, theoretical as well as observational, now points clearly to the existence of deep resonant cavities on a hemispheric scale, in the sense of what I called version 1 of the resonance theory. The damping and structure of those cavities is likely to be highly variable, but other things being equal they should tend to respond fastest and thus be at their most effective (less damped by diabatic effects, for instance) whenever the tropical quasi-biennial easterlies span an exceptionally deep layer, as they did in January 1963 and January 1977. Since the response tends to be largest at high stratospheric altitudes but comparatively modest in the troposphere, such hemispheric cavities probably make little direct contribution to the most prominent equivalent-barotropic planetary-wave anomalies in the troposphere, the anomalies associated with phenomena such as "blocking". As discussed in sections 5 and 6, there are some indications that the latter phenomena may involve an entirely different set of resonant cavities largely confined to the troposphere, which are more local than hemispheric, and which couple *non-linearly* to the "stratospheric" cavity. This idea certainly merits further investigation. We can be sure that more will soon be known about the tropospheric problem, at least, if only because of the special opportunities afforded by the FGGE data for that remarkable year 1979, along with the recognised need to study the interaction between the tropics and middle latitudes in order to make progress in medium-range weather forecasting.

Acknowledgements

I am grateful to many colleagues who told me about their recent findings and in many cases generously allowed me to share their implications with a wider audience in advance of full publication, or who have provided helpful comments on the manuscript of this article. In particular I should like to thank M. L. Blackmon, N. Butchart, S. A. Clough, R. Hide, I. Hirota, J. R. Holton, B. J. Hoskins, C.-P. F. Hsu, H. Kanzawa, D. J. Karoly, A. Kasahara, J. P. Koermer, N.-C. Lau, K. Labitzke, T. Matsuno, R. J. Murgatroyd, A. O'Neill, T. N. Palmer, R. A. Plumb, R. S. Quiroz, M. R. Schoeberl, A. J. Simmons, H.-C. Tan, P. J. Trevelyan, J. M. Wallace, and C. E. Youngblut. I should like to thank C.-P. F. Hsu also for kindly supplying Figs. 1b and 3b, K. Labitzke for Fig. 2, and D. R. Pick, S. A. Clough, and the Director-General of the U.K. Meteorological Office for Fig. 3.

References

Andrews, D. G. and McIntyre, M. E., 1976: Planetary waves in horizontal and vertical shear: the generalized Eliassen-Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, **33**, 2031-2048.

———, 1978a: Generalized Eliassen-Palm and Charney-Drazin theorems for waves on axisymmetric flows in compressible atmospheres. *J. Atmos. Sci.*, **35**, 175-185.

———, 1978b: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.*, **89**, 609-646.

Bengtsson, L., M. Kanamitsu, P. Kållberg and S. Uppala, 1982: FGGE research activities at the European Centre for Medium-Range Weather Forecasts. *Bull. Amer. Meteorol. Soc.*, to appear under GARP Topics.

Benney, D. J., and R. F. Bergeron, 1969: A new class of nonlinear waves in parallel flows. *Studies in Appl. Math.* **48**, 181-204.

Blackmon, M. L., R. A. Madden, J. M. Wallace and D. S. Gutzler, 1979: Geographical variations in the vertical structure of geopotential height fluctuations. *J. Atmos. Sci.*, **36**, 2450-2466.

Boyd, J., 1976: The noninteraction of waves with the zonally-averaged flow on a spherical earth and the interrelationships of eddy fluxes of energy, heat and momentum. *J. Atmos. Sci.*, **33**, 2285-2291.

Bretherton, F. P. and Garrett, C. J. R., 1968: Wavetrains in inhomogeneous moving media. *Proc. Roy. Soc.*, **A 302**, 529-554.

Bridger, A.F. C. and Stevens, D. E., 1982: Numerical modelling of the stratospheric sudden warming: some sensitivity studies. *J. Atmos. Sci.*, to appear.

Butchart, N., S. A. Clough, T. N. Palmer, and P. J. Trevelyan, 1982: Simulations of an observed stratospheric warming with quasi-geostropic refractive index as a model diagnostic. *Quart. J. Roy. Meteorol. Soc.*, **108**, in press.

Chapman, W. A. and Miles, T., 1981: Planetary-scale wave guides in the troposphere and stratosphere. *Nature*, **293**, 108-112.

Charney, J. G. and DeVore, J. G., 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205-1216.

Charney, J. G. and Drazin, P. G., 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83-109.

Charney, J. G. and Stern, M. E., 1962: On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 159-172.

Coy, L., 1979a: An unusually large westerly amplitude of the quasi-biennial oscillation. *J. Atmos. Sci.*, **36**, 174-176.

———, 1979b: Corrigendum to Coy, 1979a: *J. Atmos. Sci.*, **37**, 912-913.

Clark, J. H. E., 1974: Atmospheric response to the quasi-resonant growth of forced planetary waves. *J. Meteorol. Soc. Japan*, **52**, 143-163.

Crighton, D. G., 1981: Acoustics as a branch of fluid mechanics. *J. Fluid Mech.* **106**, 261-298.

Davies, H. C., 1981: An interpretation of sudden warmings in terms of potential vorticity. *J. Atmos. Sci.*, **38**, 427-445.

Davis, R. E., 1969: On the high Reynolds number flow over a wavy boundary. *J. Fluid Mech.*, **36**, 337-346.

Dickinson, R. E., 1969: Theory of planetary wave-zonal flow interaction. *J. Atmos. Sci.*, **26**, 73-81.

———, 1970: Development of a Rossby wave critical level. *J. Atmos. Sci.*, **27**, 627-633.

Dole, R. M., 1981: Persistent anomalies of the extratropical northern hemisphere wintertime circulation. Ph.D. Thesis, Mass. Inst. of Technology.

Dunkerton, T. J., 1978: On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci.*, **35**, 2325-2333.

Dunkerton, T., Hsu, C.-P. F., and McIntyre, M. E., 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, **38**, 819-843.

Edmon, H. J., B. J. Hoskins, and M. E. McIntyre, 1980: Eliassen-Palm cross-sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600-2616. (See also Corrigendum, *J. Atmos. Sci.*, **38**, 1115, especially second last item.)

Eliassen, A., 1951: Slow thermally or frictionally controlled meridional circulation in a circular vortex. *Astrophys. Norvegica*, **5**, no. 2, 19-60.

Eliassen, A. and Palm, E., 1961: On the transfer of energy in stationary mountain waves. *Geofys.*

Publ., **22**, no. 3, 1-23.

Geisler, J. E., 1974: A numerical model of the sudden stratospheric warming mechanism. *J. Geophys. Res.*, **79**, 4989-4999.

Geisler, J. E. and R. E. Dickinson, 1974: Numerical study of an interacting Rossby wave and barotropic zonal flow near a critical level. *J. Atmos. Sci.*, **31**, 946-955.

Green, J. S. A., 1970: Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Quart. J. Roy. Meterol. Soc.*, **96**, 157-185.

Grimshaw, R., 1980: A general theory of critical level absorption and valve effects for linear wave propagation. *Geophys. Astrophys. Fluid Dyn.*, **14**, 303-326.

Grose, W. L. and K. V. Haggard, 1981: Numerical simulation of a sudden stratospheric warming with a three-dimensional, spectral, quasi-geostrophic model. *J. Atmos. Sci.*, **38**, 1480-1497.

Held, I. M., 1982: Stationary and quasi-stationary eddies in the extratropical troposphere: theory. In *Large-scale dynamical processes in the atmosphere*, R. P. Pearce and B. J. Hoskins, eds., Academic.

Hirota, I. and Y. Sato, 1969: Periodic variation of the winter stratospheric circulation and intermittent vertical propagation of planetary waves. *J. Meteorol. Soc. Japan*, **47**, 390-402.

Holton, J. R., 1975: *The dynamic meteorology of the stratosphere and mesosphere*, Boston, Massachusetts, American Meteorological Society (Meteorol. Monogr. no. 37), 218 pp.

———, 1976: A semi-spectral numerical model for wave, mean-flow interactions in the stratosphere: application to sudden stratospheric warmings. *J. Atmos. Sci.*, **33**, 1639-1649.

———, and Mass, C., 1976: Stratospheric vacillation cycles. *J. Atmos. Sci.*, **33**, 2218-2225.

———, and Dunkerton, T., 1978: On the role of wave transience and dissipation in stratospheric mean flow vacillations. *J. Atmos. Sci.*, **35**, 740-744.

———, and Tan, H.-C., 1982: The quasi-biennial oscillation in the northern hemisphere lower stratosphere. *J. Meteorol. Soc. Japan*, **60** (this issue).

Hoskins, B. J., 1982: Modelling of the transient eddies. In *Large-scale dynamical processes in the atmosphere*, R. P. Pearce and B. J. Hoskins, eds., Academic.

———, and D. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179-1196.

Horel, J. D. and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the southern oscillation. *Mon. Wea. Rev.*, **109**, 813-829.

Houghton, J. T., 1978: The stratosphere and the mesosphere. *Quart. J. Roy. Meteorol Soc.*, **104**, 1-29.

Hsu, C.-P. F., 1980: Air parcel motions during a numerically simulated sudden stratospheric warming. *J. Atmos. Sci.*, **37**, 2768-2792.

———, 1981: A numerical study of the role of wave-wave interactions during sudden stratospheric warmings. *J. Atmos. Sci.*, **38**, 189-214.

Kanzawa, H., 1980: The behavior of mean zonal wind and planetary-scale disturbances in the troposphere and stratosphere during the 1973 sudden warming. *J. Met. Soc. Japan*, **58**, 329-356.

———, and I. Hirota, 1981: The behavior of mean zonal winds and planetary waves during the 1973 sudden warming. In *Middle Atmosphere Program: Handbook for MAP, vol. 2:* Extended abstracts from International Symposium on Middle Atmosphere Dynamics and Transport, Urbana, Illinois. S. K. Avery, ed., 165-174. Available from SCOSTEP secretariat, University of Illinois, 1406 W. Green St., Urbana, Ill. 61801, U.S.A.

Karoly, D. and Hoskins, B. J., 1982: Three dimensional propagation of planetary waves. *J. Meteorol. Soc. Japan*, **60** (this issue).

Killworth, P. D. and M. E. McIntyre, 1982: Do Rossby-wave critical layers absorb, reflect or over-reflect? *J. Fluid Mech.*, to be submitted.

Koermer, J. P., A. Kasahara and S. K. Kao, 1982: Numerical studies of major and minor stratospheric warmings caused by orographic forcing. *J. Atmos. Sci.*, to appear.

Kuo, H.-L., 1979: Baroclinic instabilities of linear and jet profiles in the atmosphere. *J. Atmos. Sci.*, 2360-2378.

Labitzke, K., 1978: On the different behavior of the zonal harmonic height waves 1 and 2 during the winters 1970/71 and 1971/72. *Mon. Wea. Rev.*, **106**, 1704-1713.

———, 1981: The amplification of height wave 1 in January 1979: a characteristic precondition for the major warming in February. *Mon. Wea. Rev.*, **109**, 983-989.

———, 1982: On the interannual variability of the middle stratosphere during the northern winters. *J. Meteorol. Soc. Japan*, **60** (this issue).

Lau, N.-C., 1981: A diagnostic study of recurrent meteorological anomalies appearing in a 15-year simulation with a GFDL general circulation model. *Mon. Wea. Rev.*, **109**, in press.

Lin, Ben-Da, 1982: The behaviour of winter stationary planetary waves forced by topography and diabatic heating. *J. Atmos. Sci.*, to appear.

Lordi, N. J., A. Kasahara, and S. K. Kao, 1980: Numerical simulation of stratospheric sudden warmings with a primitive equation spectral model. *J. Atmos. Sci.*, **37**, 2746-2767.

Madden, R. A., 1975: Oscillations in the winter stratosphere: Part 2. The role of horizontal eddy heat transport and the interaction of transient

and stationary planetary-scale waves. *Mon. Wea. Rev.*, **103**, 717-729.

———, 1979: Observations of large-scale traveling Rossby waves. *J. Geophys. Res.*, **17**, 1935-1949.

———, and K. Labitzke, 1981: A free Rossby wave in the troposphere and stratosphere during January 1979. *J. Geophys. Res.*, **86**, 1247-1254.

Mahlman, J. D., 1969: Heat balance and mean meridional circulations in the polar stratosphere during the sudden warming of January 1958. *Mon. Wea. Rev.*, **97**, 534-540.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter northern hemisphere. *J. Atmos. Sci.*, **27**, 871-883.

———, 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28**, 1479-1494.

———, and K. Nakamura, 1979: The Eulerian- and Lagrangian-mean meridional circulations in the stratosphere at the time of a sudden warming. *J. Atmos. Sci.*, **36**, 640-654.

McInturff, R. M., ed., 1978: Stratospheric warmings: synoptic, dynamic and general-circulation aspects. *Nat. Aeronaut. Space Admin.*, Ref. Publ. 1017. 174 pp.

Murgatroyd, R. J. and A. O'Neill, 1980: Interaction between the troposphere and stratosphere. *Phil. Trans. Roy. Soc. London*, **296**, 87-102.

O'Neill, A., 1980: The dynamics of stratospheric warmings generated by a general circulation model of the troposphere and stratosphere. *Quart. J. Roy. Meteorol. Soc.*, **106**, 659-690.

———, and B. F. Taylor, 1979: A study of the major stratospheric warming of 1976/77. *Quart. J. Roy. Meteorol. Soc.*, **105**, 71-92.

———, and C. E. Youngblut, 1982: Stratospheric warmings diagnosed using the transformed Eulerian-mean equations and the effect of the mean state on wave propagation. *J. Atmos. Sci.*, to appear.

Paegle, J. N., 1979: The effect of topography on a Rossby wave. *J. Atmos. Sci.*, **36**, 2267-2271.

Palmer, T. N., 1981a: Diagnostic study of a wavenumber-2 stratospheric sudden warming in a transformed Eulerian-mean formalism. *J. Atmos. Sci.*, **38**, 844-855.

———, 1981b: Aspects of stratospheric sudden warmings studied from a transformed Eulerian-mean viewpoint. *J. Geophys. Res.*, **86**, 9679-9687.

Pick, D. R., 1979: Stratospheric charts for period 1 January-31 March 1979. Document MO19/SC/79/1/N, Meteorological Office, Bracknell, Berkshire, U.K.

Plumb, R. A., 1981a: Forced waves in a baroclinic shear flow, Part 2: Damped and undamped response to weak near-resonant forcing. *J. Atmos. Sci.*, **38**, 1856-1869.

———, 1981b: Instability of the distorted polar night vortex: a theory of stratospheric warmings. *J. Atmos. Sci.*, **38**, 2514-2531.

Quiroz, R. S., 1975: The stratospheric evolution of sudden warmings in 1969-74 determined from measured infrared radiation fields. *J. Atmos. Sci.*, **32**, 211-224.

———, 1979: Tropospheric-stratospheric interaction in the major warming event of January-February 1979. *Geophys. Res. Letters*, **6**, 645-648.

———, 1981: The tropospheric-stratospheric mean zonal flow in winter. *J. Geophys. Res.*, **86**, 7378-7384.

———, A. J. Miller and R. M. Nagatani, 1975: A comparison of observed and simulated properties of sudden stratospheric warmings. *J. Atmos. Sci.*, **32**, 1723-1736.

Ramanathan, V., 1977: Troposphere-stratosphere feedback mechanism: stratospheric warming and its effect on the polar energy budget and the tropospheric circulation. *J. Atmos. Sci.*, **34**, 439-447.

Rhines, P. B., 1979: Geostrophic turbulence. *Ann. Rev. Fluid Mech.*, **11**, 401-441.

———, and W. R. Holland, 1979: A theoretical discussion of eddy-driven mean flows. *Dyn. Atmos. Oceans*, **3**, 289-325.

Rhines, P. B. and W. R. Young, 1982: Homogenization of potential vorticity in planetary gyres. *J. Fluid Mech.*, to appear. See also T. Yamagata and T. Matsuura, 1981: A generalization of Prandtl-Batchelor theorem for planetary fluid flows in a closed geostrophic contour. *J. Meteorol. Soc. Japan*, **59**, 615-619.

Salby, M. L., 1981: Rossby normal modes in nonuniform background configurations. Part 2: Equinox and solstice conditions. *J. Atmos. Sci.*, **38**, 1827-1840.

Sato, Y., 1980: Observational estimates of Eliassen and Palm flux due to quasi-stationary planetary waves. *J. Meteorol. Soc. Japan*, **58**, 430-435.

Schoeberl, M. R., 1978: Stratospheric warmings: observations and theory. *Revs. Geophys. Space Phys.*, **16**, 521-538.

———, and J. H. E. Clark, 1980: Resonant planetary waves in a spherical atmosphere. *J. Atmos. Sci.*, **37**, 20-28.

———, and D. F. Strobel, 1980a: Numerical simulation of sudden stratospheric warmings. *J. Atmos. Sci.*, **37**, 214-236.

———, 1980b: Sudden stratospheric warmings forced by mountains. *Geophys. Res. Lett.*, **7**, 149-152.

Simmons, A. J., 1974: Planetary-scale disturbances in the polar winter stratosphere. *Quart. J. Roy. Meteorol. Soc.*, **100**, 76-108.

———, 1982a: The forcing of stationary wave motion by tropical diabatic heating. *Quart. J. Roy. Meteorol. Soc.*, **108**, to appear.

———, 1982b: Numerical forecasts of strato-

spheric warming events using a model with a hybrid vertical coordinate. In preparation.
Stewartson, K., 1978: The evolution of the critical layer of a Rossby wave. *Geophys. Astrophys. Fluid Dyn.*, **9**, 185-200.
Tung, K. K., 1979: A theory of stationary long waves. Part III: Quasi-normal modes in a singular waveguide. *Mon. Wea. Rev.*, **107**, 751-774.
———, and R. S. Lindzen, 1979a: A theory of stationary long waves. Part I: A simple theory of blocking. *Mon. Wea. Rev.*, **107**, 714-734.
———, 1979b: A theory of stationary long waves. Part II: Resonant Rossby waves in the presence of realistic vertical shears. *Mon. Wea. Rev.*, **107**, 735-750.
Wallace, J. M. and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Wea. Rev.*, **109**, 784-812.
Wallace, J. M. and F.-C. Chang, 1982: Interannual variability of the wintertime polar vortex in the northern hemispheric middle stratosphere. *J. Meteorol. Soc. Japan*, **60** (this issue).
Warn, T. and Warn, H., 1976: On the development of a Rossby wave critical level. *J. Atmos. Sci.*, **33**, 2021-2024.
———, 1978: The evolution of a nonlinear critical level. *Studies in Appl. Math.*, **59**, 37-71.

成層圏突然昇温の力学をどう理解すればよいか？

Michael E. McIntyre
Department of Applied Mathematics and Theoretical Physics
University of Cambridge, U. K.

成層圏突然昇温に関する Matsuno の先駆的な数値実験が成功して以来，この荘大な自然現象が力学的な原因に由来するものであることは疑いをはさむ余地のないところである。しかし，その理論的なモデル化や衛星観測に基づく諸研究は，昇温現象の詳細にわたる理解や適切な予測に関してある程度の見通しが得られた段階に到ったばかりである。本論文では，この現象に関する最近の研究の進展ぶりを自由に論じ，あわせて，数値モデル化に際して対流圏の運動を先駆的に与えることによって生ずる偽の共鳴を避ける方法など，将来の研究のあり方についても示唆を与える。