THE DYNAMICAL SIGNIFICANCE OF ISENTROPIC DISTRIBUTIONS OF POTENTIAL VORTICITY AND LOW-LEVEL DISTRIBUTIONS OF POTENTIAL TEMPERATURE

Michael McIntyre
Department of Applied Mathematics and Theoretical Physics
Silver St, Cambridge CB3 9EW, U.K.

1. INTRODUCTION

I want to talk about a way of thinking about dynamical processes that is useful both in synoptic and in theoretical meteorology, and that greatly illuminates the connection between the two. It has given theoreticians like myself an enormously sharpened appreciation of insights traditionally reserved for synopticians; and it gives the synoptician and the forecaster direct access to concepts traditionally reserved for those more inclined to abstract mathematical thought. It applies to almost all large-scale dynamical processes of meteorological interest, and to some mesoscale processes as well. More precisely, it applies to any dynamical process that can be considered to be balanced in the sense that inertio-gravity oscillations are absent, or balanced after averaging out any distinguishable such oscillations. This means for instance that it applies to every dynamical process describable by filtered equations such as the quasigeostrophic and semigeostrophic equations, and all their higher-order elaborations. It is an excellent way of looking at synoptic-scale processes such as cyclogenesis and blocking, and it promises to make a contribution to practically important problems such as the quality control of numerical weather analysis and prediction, and the operational, real-time assessment of forecast uncertainty.

The key idea, which goes back to Charney (1948) and Kleinschmidt (1950a,b, 1951), is to recognize that certain aspects of the fields of Rossby-Ertel potential vorticity (PV) and potential temperature (PT) can be regarded as controlling the dynamical evolution. There is an

"invertibility principle" -- so familiar to theoreticians that it is not always mentioned explicitly -- that says that a certain subset of the information in the PV/PT fields can be used to diagnose everything about the other dynamical fields, apart from any inertio-gravity oscillations (including Kelvin waves) that may be present. A precise statement will be given in §3 below. The resulting viewpoint has advantages that become particularly clear at the interface between the observation-analysis-prediction system and the human brain.

For practical purposes, that interface is our visual system; and it is a very remarkable interface. The visual cortex and its peripherals deploy prodigious computational power on data transfer and processing, estimated to be many thousands of Crays' worth even for the earliest stages of the data processing. Keeping in mind the biological purposes for which the system came into being, we might expect it to be particularly powerful when used on suitable visualizations of the temporally evolving PV/PT fields. Why is this? Many significant weather developments depend on fast upper-air motions that are to a first approximation adiabatic and "frictionless", implying that both PV and PT are materially conserved. Features in the PV/PT fields therefore tend to be carried along with the air motion, and often become sharp-edged and front-like because of the strong deformation rates in the large-scale wind field, with their well known tendency to create steep gradients in the distributions of materially conserved quantities. Synopticians have long been familiar with the associated structures: jets, shear lines, and so on (e.g. Palmén and Newton 1969). One of the functions to which the eye-brain system is exquisitely well adapted is the near-instantaneous recognition of moving, sharp-edged features. When helped by suitable graphical techniques, it can lock on to, and follow, several such features at once. It can perform this computational feat far better than any artificial intelligence system yet in prospect, and can do it to a remarkable extent even when the data are noisy or gappy.

Practical ways of viewing the moving features in the PV/PT fields, and relating them to material air motions, include the time-honoured method of looking at the evolution of PV distributions on constant-PT, i.e. isentropic, surfaces. Some historical notes are given in the recent review paper by Hoskins et al. (1985, hereafter HMR). Isentropic distributions, maps, etc., of PV will sometimes be referred to for brevity as "IPV"

distributions, maps, etc. The "I" is intended to emphasize the importance of viewing the PV on an isentropic surface, this being a natural way of visualizing material conservation. Most isentropic surfaces are mappable, because most of the atmosphere is stably stratified; moreover IPV maps provide a useful way of formulating the invertibility principle (§4). A complementary approach is to use isostrophic maps of PT, i.e. maps of PT on constant-PV surfaces. These will be referred to for brevity as "IPT" maps. They have less general utility since the atmosphere is not monotonically stratified in PV, but recent work at Reading has shown that they can be extremely useful when the PV surface is taken at a PV value characteristic of the tropopause (B.J. Hoskins, personal communication). The typical structure of the upper troposphere, lower stratosphere system -- our knowledge of which goes back to the pioneering work of Danielsen, Kleinschmidt, Platzman, Reed, and Sanders -- gives rise to the fortunate circumstance that a single tropopause IPT map can convey nearly as much upper-air dynamical information as a whole stack of IPV maps. The only difficulty is the graphical difficulty of depicting tropopause folds. Still other important aspects of the PV/PT fields are those depicted by low-level maps of PT, particularly just above the planetary boundary layer. We shall see that such low-level PT maps can actually be regarded, in a certain sense, as another special case of IPT maps. Together with tropopause IPT maps they contain information of crucial importance to many cases of cyclogenesis.

The fact that the evolution of the PV/PT fields is well suited to viewing by the human eye, in various ways, would have only limited significance were it not for the invertibility principle. The latter therefore needs to be stated more precisely, and its theoretical status and régime of applicability discussed -- a matter of some mathematical subtlety being actively researched today. Before getting into such questions, therefore, I should like to talk about a simpler dynamical system that has a surprising amount in common with the real atmosphere and that illustrates, in a relatively straightforward way, much of what is involved.

2. THE SIMPLEST PARADIGM: NONDIVERGENT BAROTROPIC VORTICITY DYNAMICS

Despite their simplicity, the equations of nondivergent barotropic dynamics exhibit a number of counterparts to real-atmospheric phenomena, including several already mentioned, as we shall see in the next section. This is

understandable from the viewpoint of "PV thinking", since from that viewpoint the system has the same two basic characteristics -- a dynamical variable, the absolute vorticity q, that is materially conserved in the absence of friction, and an invertibility principle saying that all the other dynamical variables can be diagnosed from a knowledge of q. The system is of course a very well understood one, long studied in the context of classical aerodynamics as well as that of classical dynamical meteorology; and indeed one of the advantages of PV thinking is that it enables us to make use of many well developed and well understood ideas such as, for instance, "vortex rollup" (to be illustrated below). The equations may be written as follows:

$$Dq/Dt = \text{frictional terms}, \tag{2.1a}$$

$$\psi = \nabla^{-2}(q - f), \tag{2.1b}$$

where $q = \nabla^2\psi + f$, the absolute vorticity, f is the Coriolis parameter, ψ the streamfunction, and D/Dt the two-dimensional material derivative. The latter is defined by

$$D/Dt = \partial/\partial t + u\,\partial/\partial x + v\,\partial/\partial y \tag{2.2}$$

where

$$u = -\partial\psi/\partial y, \quad v = \partial\psi/\partial x; \tag{2.3}$$

(x, y) are eastward and northward coordinates and (u, v) the corresponding eastward and northward components of the wind vector $\underset{\sim}{v}$.

If friction is neglected in equation (2.1a), the equation states that q is materially conserved. And if we were to watch a moving picture of the q field we would be following everything else about the dynamics since, by (2.1b), knowledge of q implies knowledge of ψ and hence, by (2.3), of the wind field. This is the relevant invertibility principle, and three points about it should be noted for later reference:

1. Local knowledge of q does not imply local knowledge of ψ; the inversion is very much a global process. In particular, the principle depends on specifying suitable boundary conditions to make (2.1b) unambiguous.

2. The notion of balance, on which the inversion depends, corresponds in this system to the absence of sound or external gravity waves. These have been filtered out simply by the assumption

of incompressible, nondivergent motion.

3. There is a scale effect, whereby small-scale features in the q field have a relatively weak effect on the ψ and $\underset{\sim}{v}$ fields whereas large-scale features have a relatively strong effect. In particular, ψ and $\underset{\sim}{v}$ are to varying degrees insensitive to fine-grain structure in the q field. The inverse Laplacian operator in (2.1b) is a smoothing operator, and some of the smoothing survives even when followed by the single differentiations in (2.3).

The non-localness of the inversion operator and the implied "action at a distance" emphasized in point 1 may seem strange at first sight. The apparent "action at a distance" summarizes, with remarkable succinctness, the peculiar way in which fluid elements push each other around. The resulting gain in economy of thinking has long been made use of by aerodynamicists as well as by theoretical meteorologists. Indeed it is built into classical aerodynamical language, in such phrases as the velocity field "induced by" the vorticity field, and the idea that a vortex can roll "itself" up. It may be useful to note from (2.1b) that diagnosing the ψ field from the q field is more or less the same thing, mathematically, as calculating the electrostatic potential from the charge distribution, or the static displacement of a stretched membrane from the pressure distribution on it.

Point 2 prompts a question that is both theoretically intriguing and, it will appear, practically relevant. To what extent can more refined versions of (2.1) can be constructed in which slight compressibility is taken into account? We know that (2.1a) remains exactly valid, provided that q is replaced by the relevant potential vorticity $Q = q/\rho$ or q/h, where ρ is the relevant density or h the relevant layer depth; but what happens to the inversion (2.1b)? One can imagine correcting (2.1b) iteratively, thus modelling the balanced motion more accurately. One might even wonder whether such a process can converge, i.e. whether it is dynamically possible for the motion of a slightly compressible fluid to be exactly free of sound waves, and to have an exact invertibility principle, to which (2.1b) is merely a first approximation. Essentially the same questions arise for a baroclinic atmosphere, regarding acoustic-inertio-gravity waves; and they have been much discussed in the recent literature under the heading of the existence or nonexistence of a "slow manifold". It may be of interest to note that there is one case, namely the present case with $f = 0$, in which the answer is well established and well known. It is a clear "no". Unsteady, vortical motions in a slightly

compressible fluid always emit some sound or external gravity waves, albeit very weak ones if the Mach or Froude number F is small. The acoustic or external gravity wave power propagating to the far field scales as a high power of F (F^7 in two dimensions and F^8 in three, the latter being the celebrated Lighthill eighth-power law -- for relevant review material see Crighton 1975, 1981, and Lighthill 1978, pp. 21, 64). One consequence is that any process of iteratively refining (2.1) can have at best asymptotic, not convergent, behaviour. But for moderate values of F the refinements may still lead to balance conditions and inversion operators that are far more accurate than those expressed by (2.1), even though not exact in principle. I shall present some specific evidence for this in §6, in the rotating case $f \neq 0$.

On point 3, the classic illustration is simple Rossby wave propagation and its scale dependence. Indeed the Rossby wave propagation mechanism nicely illustrates all the foregoing ideas, including the idea of the "induced" velocity field. Imagine a basic state of relative rest ($\psi = 0$ everywhere) in which f and therefore q has a large-scale y-gradient $\beta = df/dy > 0$. If friction is neglected in (2.1a), the contours of constant q are also material contours. If a disturbance makes these contours undulate as shown in Fig. 1, the right hand side of (2.1b) will be

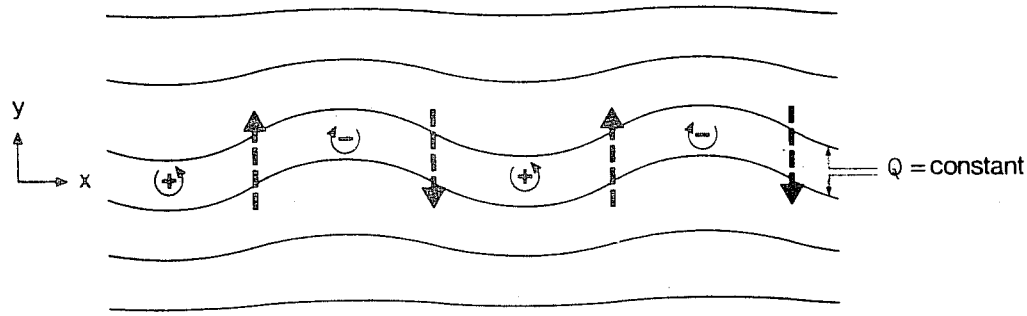

Fig. 1 Sketch of the q or Q contours and anomaly pattern, and the induced velocity field, in a simple Rossby wave.

alternately positive and negative as shown by the plus and minus signs. To see what the induced ψ field must look like, one can solve (2.1b), or simply picture the contours of ψ as the equipotentials of the electrostatic field due to a pattern of alternating positive and negative charges, or as the topographical contours giving the diplacement of a stretched elastic membrane that is pushed and pulled alternately in the same pattern. It can be seen that ψ will have hills and valleys centred respectively on the minus and the plus signs, and that the strongest

north-south winds will therefore occur at intermediate positions, a quarter wavelength out of phase with the displacement, and in the sense shown by the heavy, dashed arrows in the figure. If you now make a moving picture in your mind's eye of what this induced velocity field will do to the contours, you can see that the undulations will propagate westward. This is the celebrated Rossby-wave propagation mechanism: whenever material parcels get displaced back and forth across a large-scale northward q gradient, the resulting q anomalies immediately induce a velocity field that, for small displacements (more precisely, small sideways contour slopes), always tends to make the undulations propagate westward -- relative, of course, to any mean flow that may be present.

The scale effect, point 3 above, implies that the westward phase speed c increases with wavelength. This is another celebrated property of Rossby waves. It is easy to check explicitly in the textbook case where β is constant. We then have simple solutions of the type $\psi \propto \cos \ell y \cos\{k(x-ct)\}$, which can be seen to satisfy (2.1a,b) (in this case without linearization, as it happens) provided that the phase speed c and the wavenumbers k, ℓ satisfy the Rossby dispersion relation

$$c = -\beta / (k^2 + \ell^2) \ . \qquad (2.4)$$

The scale-dependent factor $-(k^2+\ell^2)^{-1}$ comes from the inverse Laplacian operator, explicitly showing the scale effect upon phase speed. Further discussion of Rossby-wave propagation (elucidating the mechanisms of lateral and vertical group propagation and of Rossby-wave critical-layer absorption and reflection) may be found in §6c of HMR (q.v., Fig. 19).

3. A NONLINEAR BAROTROPIC THOUGHT-EXPERIMENT

When nonlinearity becomes important, simple Rossby-wave undulations tend to break down in interesting, and synoptically recognizable, ways. Figs. 2-4 show a sequence of daily q and $\underset{\sim}{v}$ fields for a very high resolution, T159, nearly frictionless integration of (2.1) on a hemisphere, recently carried out by Martin Juckes at Cambridge. For full details, see the recent paper by Juckes and McIntyre (1987). The q fields are depicted in monotonic grayscale shading, with light shades denoting low q values, and dark shades high. The original motivation was to do a thought-experiment to model certain aspects of the wintertime middle and upper stratosphere,

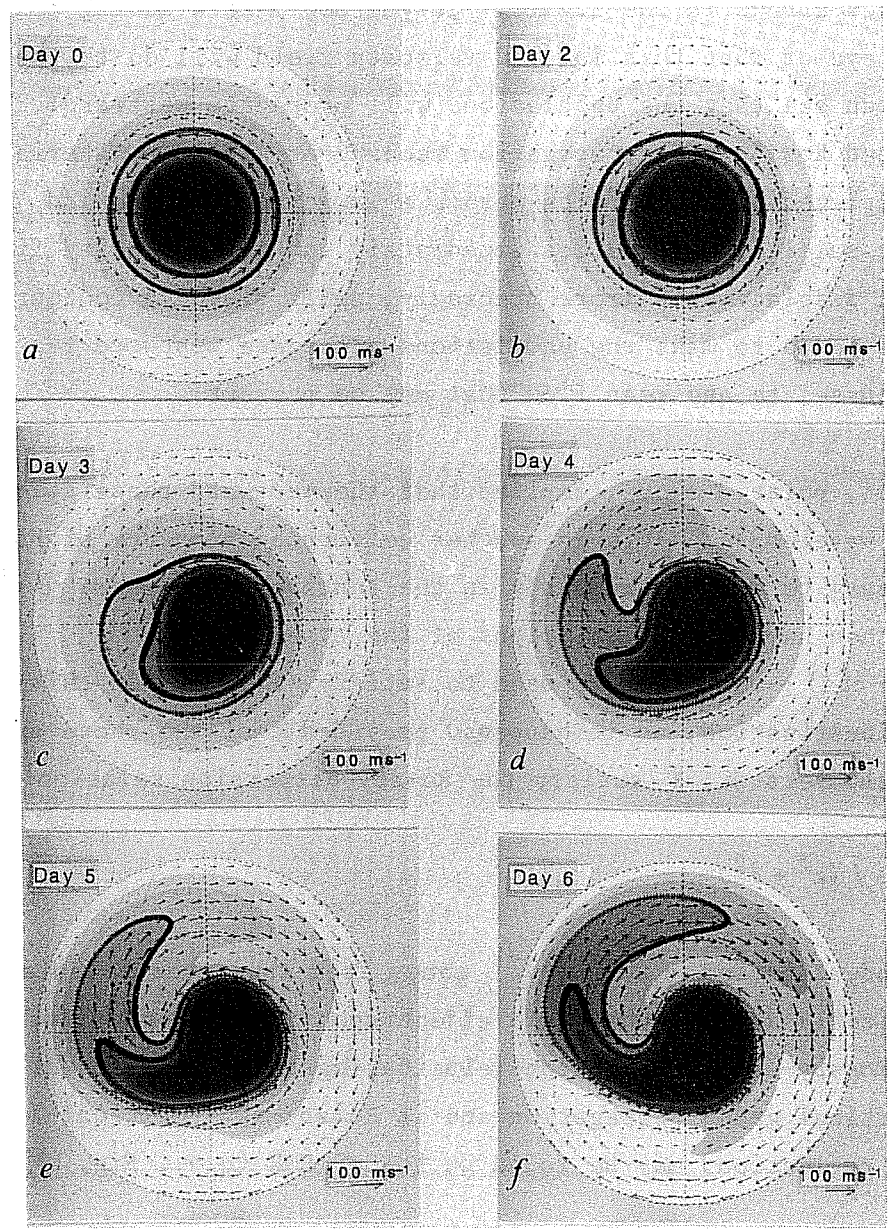

Fig. 2 High resolution (T159, Reading spectral code), nearly-frictionless simulation of a breaking Rossby wave in a barotropic, hemispheric model stratosphere having a strong polar-night vortex, from Juckes and McIntyre (1987). Grayscale shading denotes the q field, shown monotonically over nine intervals from zero (lightest) to $1.36\, f_{max}$ (darkest), where f_{max} is the maximum planetary vorticity, $1.458 \times 10^{-4} s^{-1}$. About half the initial q range is concentrated into a fairly steep gradient near the edge (jet-core) of the model's polar vortex, just inside the two material contours initially marking the fourth interval. Arrows show the wind field. Maps are polar stereographic and show the whole hemisphere.

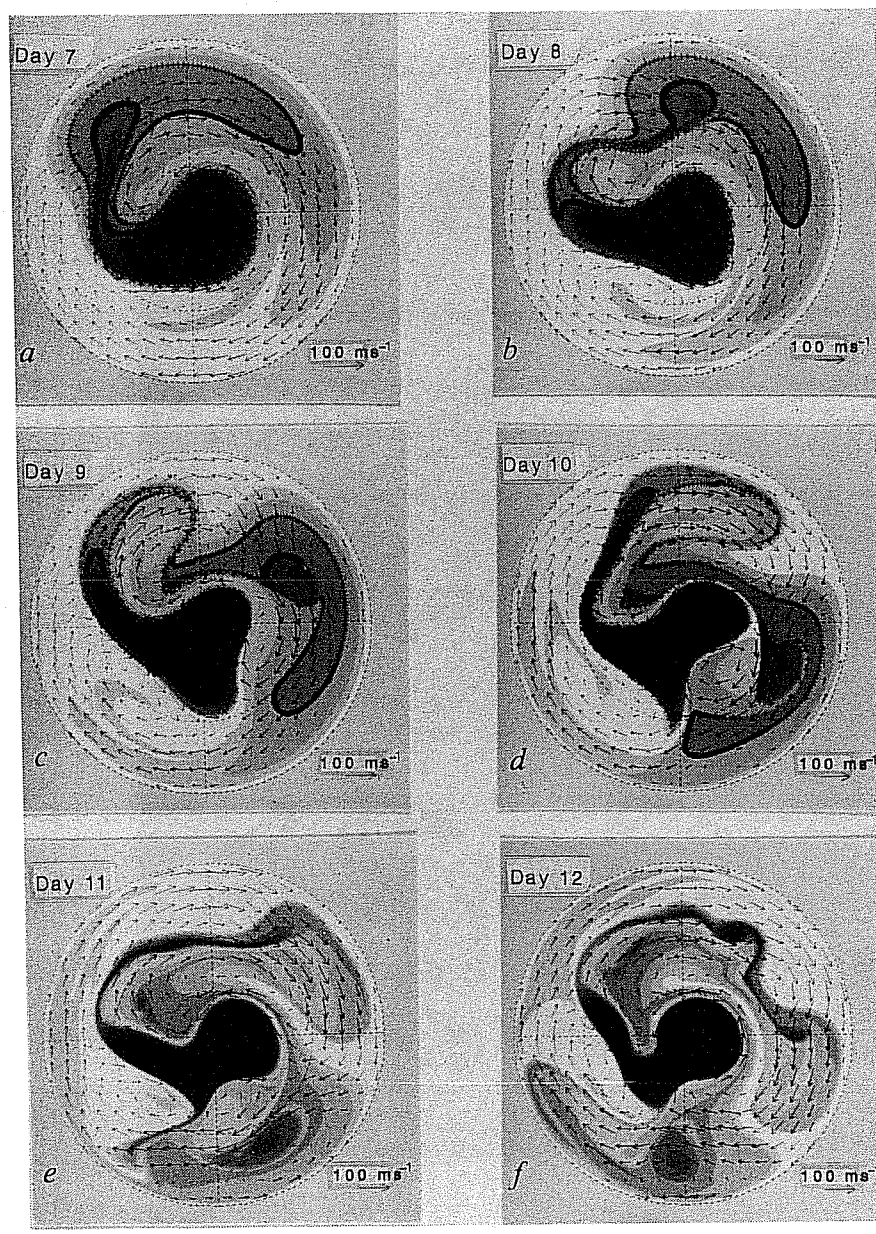

Fig. 3 Continuation of Fig. 2. The q gradients at the edge of the model's polar vortex are rapidly steepening to values limited by the model's hyperdiffusion.

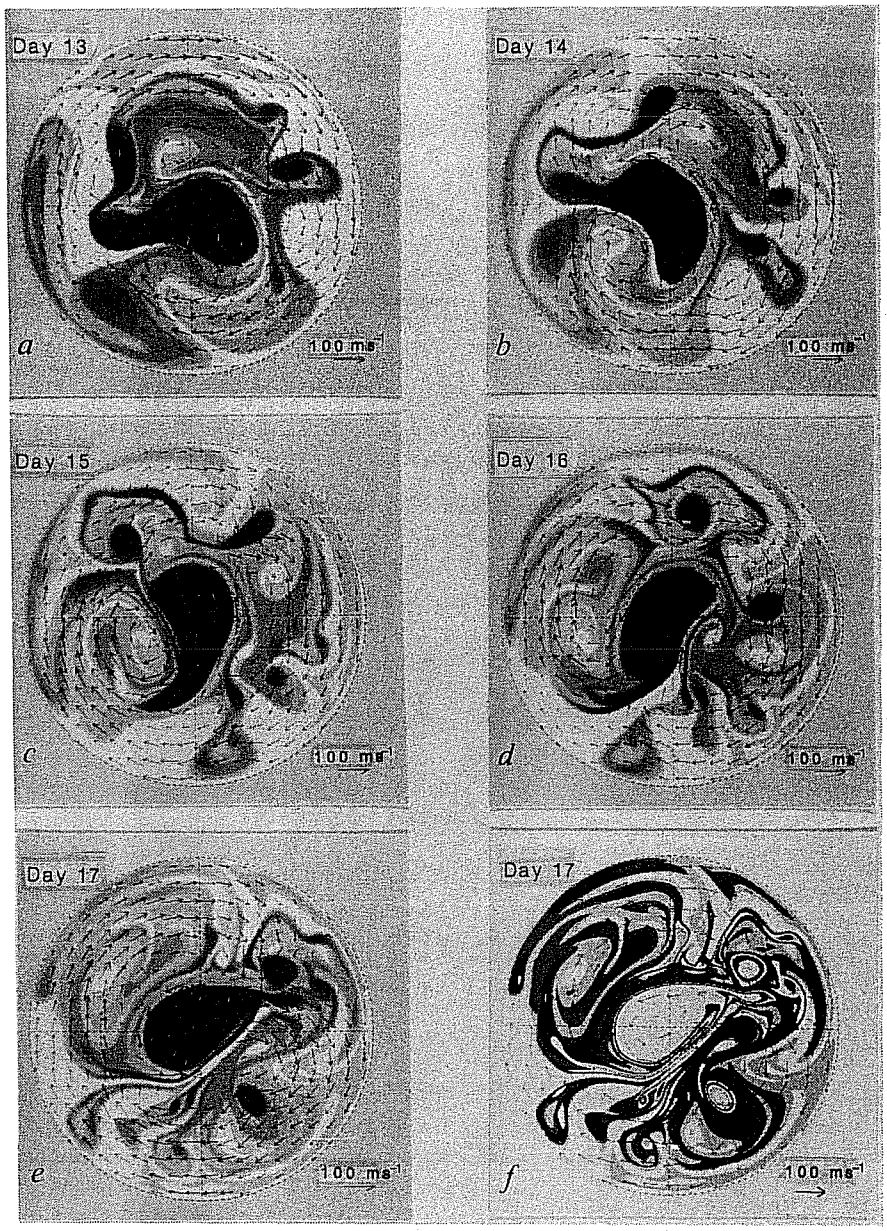

Fig. 4 Continuation of Fig. 3. Day 17 is shown twice, the second time with a different grayscale sequence, white-gray-black, white-gray-black, white to cover the range of increasing q values. Notice how two of the cutoff cyclones, those at top right in (a), can be traced back to identifiable barotropic shear instabilities (visible in Fig. 3e,f), while the others cannot.

starting from a hypothetical, zonally symmetric state on day zero that is subsequently disturbed in high latitudes by a wave-one planetary-scale Rossby wave arriving from below. This disturbance was modelled in the simplest possible way that preserves the basic mathematical structure of (2.1), by redefining q as $\nabla^2\psi + f$ plus a prescribed, smooth function of space and time having the planetary-scale structure of the incident disturbance. This function peaks at about 65°N, is relatively small in the tropics, and its zonal dependence contains wave 1 only. Its time dependence was taken as a smooth buildup from zero during days 0-4, then constant to day 12, then a smooth decrease to zero by day 16. Two material contours (heavy circles in Fig. 2a) were independently followed up to day 10, as a check both on the numerical accuracy and on the smallness of the right-hand side of (2.1a). The latter was taken to be a triharmonic hyperdiffusion whose effects were, indeed, mostly insignificant for present purposes except as noted in the caption to Fig. 3.

Linear Rossby-wave theory breaks down between days 3 and 4 (Figs. 2c,d), when material and q contours begin to deform irreversibly (in striking contrast to the simple undulations shown in Fig. 1), leading subsequently to a variety of sizes and shapes of small-scale features in the q field, most of them far smaller in scale than the incident planetary-scale disturbance. This is an inherently nonlinear, advective process to which Rossby waves are often subject, and which is fundamentally analogous to the breaking of other kinds of waves such as gravity waves; and it has significance for questions about tropospheric as well as stratospheric dynamics that range from matters of synoptic detail to the shaping of the entire general circulation. Four points will be singled out for explicit mention here:

1. Note the general tendency to form sharp gradients in the q field, which is particularly evident in Figs. 3 and 4 and which occurs for the reasons mentioned in §1. One example is the long, dark (cyclonic) strip that is conspicuous in the top half of Fig. 3e, day 11. This is the model's analogue of an upper-tropospheric shear line, albeit on a rather larger scale. Another is the dramatic steepening of the gradient at the edge of the model's main, cyclonic polar-night vortex (the large, central dark region), the edge corresponding more or less to the location of the main polar-night jet core. This tendency for the q field to develop sharp gradients is a characteristic feature that occurs again and again in

this and in many other, similar numerical experiments that have been carried out.

2. Also of interest, in the context of these lectures, are the synoptic-scale cutoff cyclones (dark) and anticyclones (light) that appear during the later stages shown in Fig. 4. Two of the cyclones, namely the pair at top right in Fig. 4a, day 13, have their origins in a recognizable barotropic shear instability. Inspection of the two previous frames shows very clearly that they originate from an instability of the abovementioned shear line. The later stages also provide a good example of the previously-mentioned mechanism of "vortex rollup", which is germane to the cutting-off process. Because of the existence of a small-amplitude instability stage, their genesis could be said to be a barotropic counterpart to Petterssen's "type A cyclogenesis". Others, like the two in the top right quadrant of Fig. 4d, never go through any recognisable linear-instability phase. They arise directly as finite amplitude disturbances ("Petterssen type B"), and indeed from the other experiments this seems to be the commonest way in which cyclones and anticyclones arise in the model -- and possibly, also, dare I say it, in the real stratosphere and troposphere. I shall return to this point in the second lecture.

3. Another point of great interest, both synoptically and also from the viewpoint of two-dimensional turbulence theory, is the way in which the smaller-scale q anomalies exhibit a mixture of dynamically active and passive-tracer behaviour. Their induced velocity fields are locally significant in the former case, and relatively insignificant in the latter: some anomalies roll themselves up, and some do not. In particular, many of the long, thin q filaments or shear lines that appear in Fig. 4 -- and in many similar experiments not shown here -- do <u>not</u> go locally unstable, despite the fact that like the shear line in Fig. 3e their q distributions locally satisfy the usual Fjørtoft-Arnol'd conditions for shear instability. It now seems likely from recent work at Cambridge (D. Dritschel, M.N. Juckes, P.H. Haynes, T.G. Shepherd, personal communication) that a sufficient reason is the stabilizing effect of larger-scale strain and shear fields induced from some distance away, which are the reason why the anomalies got long and thin in the first place and to which local shear instabilities appear to be quite sensitive. Such a stabilizing effect may help us understand not only the common occurrence of passive-tracer-like, filamental features such as those evident in Fig. 4, but also the

occurrence of shear lines in the real atmosphere (C. Thorncroft, personal communication).

4. These nearly frictionless numerical experiments provide particularly clear illustrations of the way in which a breaking Rossby wave can irreversibly transfer negative angular momentum, into the tropics in this case. The effect of the transfer appears simply as a permanent change in the tropical zonally averaged zonal flow, which becomes more easterly by, typically, several tens of metres per second. The theory of wave, mean-flow interaction implies that this phenomenon is closely analogous to the way in which breaking ocean waves induce longshore currents on beaches. The process appears to be fundamental to understanding certain tropical, extratropical interactions both in the stratosphere and in the troposphere, particularly the "negative viscosity" effect in the angular momentum budget of the general circulation and its relation to large-scale tracer transport. Full and careful discussions of these points have been given elsewhere (McIntyre and Palmer 1984,5; Haynes and McIntyre 1987a,b, & refs.).

5. On the question of graphical technique, note that it would be difficult, if not impossible, to plot contour maps of the last few q fields in Fig. 4 that would make much sense to the eye. For viewing materially advected fields like q, with their tendency to develop fine structure and sharp-edged features, "photographic" techniques like the grayscale shading used here are more suitable. Animated versions making the continuity of evolution visible can be especially powerful; and demonstrations of this have recently been given by John Marshall (personal communication) using the image-processing facility at Imperial College.

4. THE PV AND THE INVERTIBILITY PRINCIPLE FOR A BAROCLINIC ATMOSPHERE

The standard meteorological definition of PV for a three-dimensional, baroclinic, hydrostatic atmosphere is essentially that proposed by Rossby in 1940:

$$Q = (f + \hat{\underset{\sim}{z}}.\nabla_\theta \times \underset{\sim}{v})(-g\partial\theta/\partial p) \ , \qquad (4.1)$$

where $\hat{\underset{\sim}{z}}$ is a unit vertical vector, θ is the PT, subscript θ indicates differentiation along an isentropic surface, $\underset{\sim}{v}$ is the horizontal wind

vector as before, and p is hydrostatic pressure. Ertel's independently discovered and more general formula reduces to (4.1) in the hydrostatic case. A convenient SI unit for Q is $10^{-6}m^2s^{-1}K\,kg^{-1}$; it could perhaps be called the μR (microrossby), but since Ertel has an independent claim I shall settle for calling it the PV unit or PVU, as was done in HMR. It is a convenient unit since values < 1 usually imply that we are looking at tropospheric air, and > 2, stratospheric air except near the equator. Most other units in common use are numerically more or less the same thing times some power of ten, depending on whether SI or cgs is used and on whether the factor g is dropped.

One way of stating the invertibility principle for the Q and θ fields is that proposed by HMR. It begins by assuming that the mass under each isentropic surface is specified, or some equivalent information giving the static stability of a suitable reference state, just as is done in the theory of available potential energy. The principle asserts that given this information, together with the global distribution of Q on each isentropic surface, and of θ at the lower boundary, one can deduce, diagnostically, all the other dynamical fields such as winds, temperatures, geopotential heights, local static stabilities, and vertical motion, to the extent that, and to the accuracy with which, the motion can be regarded as balanced.

The following points may be noted:

1. Just as in the simpler case of §2, we must solve the diagnostic problem globally, with proper attention to boundary conditions; the same remarks about the apparent "action at a distance" apply.

2. The principle, in the form just stated, helps to explain why IPV gradients and surface PT gradients keep on turning up as key factors in theoretical studies of barotropic and baroclinic instabilities, large-scale waves, vortices, and other phenomena involving balanced motion. Some examples will be given later.

3. For practical purposes the phrases "surface" and "at the lower boundary", in connection with PT distributions and gradients, will usually mean just above the planetary boundary layer, as already hinted in §1.

4. The invertibility principle, as stated here, carefully avoids any prior commitment as to the best balance condition under which to carry out the inversion. Indeed a strong reason for elevating it to the status of a "principle" is to focus attention on the idea, already hinted at in §2, that the balance and invertibility concepts need not be tied to any particular set of approximations, filtered equations, or explicit formulae, and to leave open the possibility that more accurate ways of quantifying balance and invertibility may yet be found. Such ways, in fact, do now seem to be in prospect; some preliminary evidence will be presented in §6. The ultimate limitations are as yet unknown, although it is obvious enough that limitations must exist, within known requirements such as those of static and inertial stability, Froude and Rossby numbers not being large, and so on.

5. The statement that vertical motion can be deduced is related of course to the omega-equation principle. A simple illustration, the "vacuum-cleaner effect", will be given in the next lecture. The more accurate inversions require the vertical motion to be found as part of the inversion procedure, and so the more accurate ideas of "balance" have some dependence on information about frictional and diabatic effects. Further discussion is given in HMR (§4 and appendix), and below in §6.

6. IPV distributions and their possible evolution, and the associated transports and budgets of PV, are constrained by two exact, general theorems that hold even in the presence of diabatic heating and frictional or external forces. Specifying an IPV distribution that violated the associated constraints would presumably lead to failure of any attempt at inversion (see also HMR §3, eq. 17b et seq.). The first is that PV, considered as a convected and transported tracer, is <u>indestructible</u>, except where isentropic surfaces intersect a boundary such as the earth's surface. The second is that isentropic surfaces are <u>impermeable</u> to PV -- even in the presence of diabatic heating. An isentropic surface in a stably stratified atmosphere acts like a semi-permeable membrane, allowing mass to cross it but not PV.

Thus although values of Q, which have the nature of tracer mixing ratios, can change, they can change only by the tracer being transported, diluted, or concentrated. Both theorems are direct consequences of the way in which the PV is constructed mathematically. This, incidentally, has the

significant further consequence that the theorems apply not only to the exact PV constructed from exact $\underset{\sim}{v}$ and θ fields, but also, exactly, to any "coarse-grain PV" constructed from coarse-grain observational datasets.†

7. Recalling point 3 of §2, we might expect that the inversion operators delivering geopotential heights and winds from PV distributions are broadly speaking not too sensitive to fine-grain detail in the PV field. Eq. (4.1) involves differentiation and so we might naively expect the inversion operator to involve integration and to be a smoothing operator, somewhat like the operator ∇^{-2} in (2.1b). This indeed is manifestly the case for the crudest, quasigeostrophic inversions (e.g. HMR §5b). However, the theory of semigeostrophic inversion (op. cit., §5c) indicates that the more accurate forms of the inversion operator, which are nonlinear, may not be quite such good smoothers under some conditions, e.g. in certain locations within strong fronts and jet streams.

8. PV inversion in the presence of topography ill-understood, but would worth trying to understand better in connection with problems such as Alpine lee cyclogenesis (e.g. Bleck and Mattocks 1984, 1986). In some cases involving shallow, smooth topography, the invertibility principle appears to hold exactly as stated above, although the precise nature of the solution is known only within the context of quasigeostrophic theory. The generalization beyond that theory presents conceptual as well as technical difficulties, even for smoothed topography, as can be seen from the subtleties encountered in the analysis of a related problem by Eliassen (1980). There is also the obvious fact that real topography commonly generates unbalanced motions in the form of mountain waves, signalling one very definite limitation to the balance and invertibility concepts.

5. CYCLONIC PV ANOMALIES IN A BAROCLINIC ATMOSPHERE

It is one thing to say that, in principle, one can deduce everything from IPV and surface PT distributions (in a wide, albeit not yet precisely

†The history of these theorems is somewhat unclear, to me at least. Some aspects of the first theorem were noted by Truesdell (1951) for the adiabatic case, and generalized by Thorpe and Emanuel (1985) to the (much more interesting and significant) diabatic case. As far as I can tell from extensive correspondence, the complete picture including the impermeability theorem has apparently been noted only very recently, by Haynes & McIntyre (1987).

defined, set of circumstances). It is another to know what the answers look like. A useful approach is to consider the structures induced by isolated PV/PT anomalies. Not surprisingly, isolated cyclonic anomalies tend to induce cyclonic wind structures, and anticyclonic anomalies anticyclonic wind structures. Each exhibits characteristic features in the associated fields of vorticity and static stability, the two factors appearing in (4.1). These features and the physical and mathematical reasons for them were discussed in some detail in HMR §3, following the pioneering work of Kleinschmidt and drawing on some specific examples from the work of Thorpe (1985, 1986).

Figs. 5 and 6 reproduce two of Thorpe's examples. Both are structures

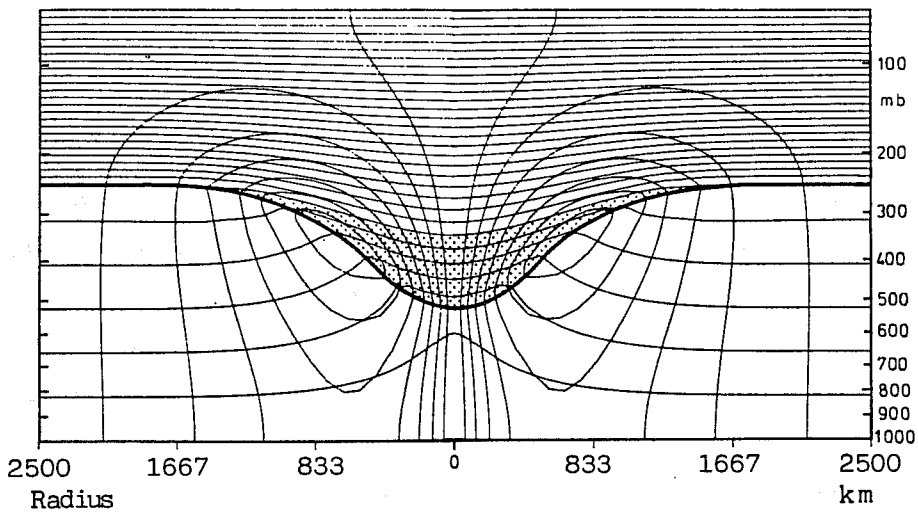

Fig. 5 Cyclonic structure induced by a simple, circularly symmetric, upper-air PV anomaly, from the work of Thorpe (op. cit.). θ surfaces are shown at 5K intervals, and isotachs at $3\,ms^{-1}$ intervals. The domain shown has a radius of 2500km. The heavy curve represents the tropopause, across which the PV jumps by a factor 6.

induced by single, isolated, circularly symmetric cyclonic anomalies under gradient-wind balance, the first by an upper-air IPV anomaly and the second by a surface PT anomaly. The resemblance to cyclonic structures that have often been observed in the real atmosphere (e.g. Palmén and Newton 1969) is immediately striking. Indeed it is remarkable how much in the way of realistic-looking features appear, illustrating the economy of description that can result from "IPV thinking", which says that the whole structure, in each of these cases, follows from the presence of a single IPV anomaly.

In the case of Fig. 5 the region containing the anomaly is shown stippled.

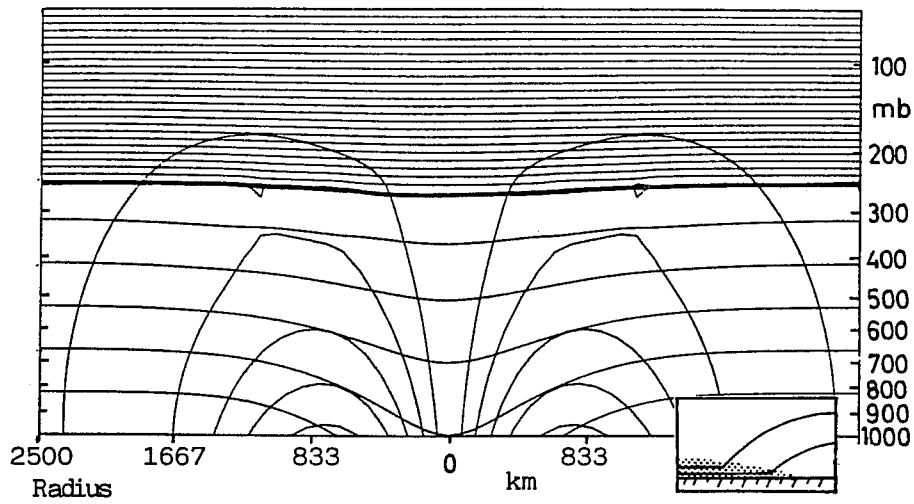

Fig. 6 As in Fig. 5 except that the anomaly is located at the surface.

Note that the phrase "IPV anomaly" is intended to denote, as indicated by the presence of the letter "I", values that are low or high relative to surrounding values on a given isentropic surface. For instance the value of Q is about $\frac{1}{2}$PVU at the edge of Fig. 5 near 400mb, and if we follow an isentrope inwards from that point, Q jumps to about 3 PVU upon crossing the tropopause. On a tropopause IPT map, the amplitude of the θ anomaly would be -24K. The absolute vorticity at the tropopause reaches the very large value 2.7f, implying that a quasigeostrophic inversion would be exceedingly inaccurate. Down on the surface, the maximum wind speed is just over 15 m/s and the surface pressure anomaly is -41mb. For more detail see HMR §3, and Thorpe (op. cit.).

In Fig. 6 the surface PT anomaly has amplitude +10K. The inset suggests how the warm surface PT anomaly can be thought of as a limiting case of a cyclonic IPV anomaly. Alternatively, one can think of a surface PT map as a special case of an IPT map: the solid earth acts as a kind of nether-region "stratosphere" having infinite static stability and PV. In this case the wind induced at the surface peaks at about 16 m/s and the surface pressure anomaly is -31mb.

Upper-air PV and surface PT anomalies of this magnitude occur in the real atmosphere, and may on occasion be very rapidly advected into a given location. This is one way of seeing why upper-air PV advection and surface warm advection should be so important for cyclogenesis, as synopticians have long known. The second lecture will indicate how these two basic

advective processes can amplify each other, and how this coupling can be enhanced and the whole process still further amplified, sometimes explosively, by latent-heat release. It will be convenient to postpone discussion of the associated vertical motion field until then.

An important practical point evident from Fig. 5 is that a tropopause IPT map can be used to give a direct indication of the vertical extent of an anomaly like the stippled region, or of its opposite-signed, anticyclonic counterpart, which induces a raised tropopause. In both cases a single such IPT map will show directly how deep and therefore how powerful a major anomaly is, whereas a stack of several IPV maps would be required to make the same thing clear. It should by now be evident why I suggested in §1 that IPT and surface PT maps together "contain information of crucial importance to cyclogenesis".

6. HOW ACCURATE IS THE INVERTIBILITY PRINCIPLE IN GENERAL?

In examples like those given in Figs. 5 and 6, the balance condition is simply gradient-wind balance; and balance and invertibility hold exactly since the flows are circular, steady, frictionless, and adiabatic. Such examples leave open the question of how accurately, or inaccurately, the concepts of balance and invertibility might apply to more realistic situations. This question lies at the research frontier today.

Point 2 of §2, together with the discussion below it, has already warned us to expect an ultimate, bedrock limitation on the accuracy that can be obtained, even if Froude and Rossby numbers are small. The same point has been cogently argued for rotating, baroclinic atmospheres by Warn and Ménard (1985), Vautard and Legras (1986), and Lorenz and Krishnamurthy (1987). The implication is that "balance", "slow manifold", and "invertibility" are all inherently approximate concepts, although, it seems likely, often far more accurate than one might think from experience with the simpler systems of filtered equations. The success of initialized numerical weather forecasts is one encouragement towards this latter belief.

The question of how to define balance and PV inversion to the highest possible accuracy involves, then, some mathematical subtlety. One difficulty is that the nonexistence of a strictly defined slow manifold

means that inversion operators are not only inherently approximate, but also necessarily non-unique. The operators are nonlinear and can be defined only iteratively (as far as I know); and for any given proposed operator there is a very large function space to explore.

One can narrow down the possibilities by imposing some requirements on the inversion operator that seem prima facie sensible. In order to do this, and in any case to define the operator itself, it is convenient to consider the associated balanced dynamical model, having a structure analogous to (2.1). That is, the evolution of this notional model is defined by time-stepping the exact equations governing IPV distributions and surface PT distributions [collectively analogous to (2.1a)], followed by inversion to get the new wind field [analogous to (2.1b)], then another time step, and so on, for sufficiently small time steps.

It seems natural to require that this notional balanced model conserve mass exactly. A model atmosphere in which mass can appear or disappear in an arbitrary way is too unphysical for most of us. It follows in turn (since the exact PV and PT evolution equations are used, implying material conservation of PV and PT in frictionless, adiabatic conditions) that the vorticity equation is also satisfied exactly and that the only scope for approximation is then in the divergence equation. We may also wish to require that the inversion operator be Galilean invariant: the result of applying it should be physically the same when computed in reference frames rotating at different angular velocities.

I am not, however, going to suggest that the associated balanced model should exactly conserve energy and momentum. Since this may seem heretical at first sight it may be worth digressing briefly to explain why. We want the balanced model to get as close to reality as it can. In the corresponding real motion we expect that spontaneous emission of inertio-gravity waves is likely to be going on nearly all the time, albeit usually very weakly, just as with sound waves in §2 (and probably even weaker than when $f = 0$ because of the inertia-frequency cutoff). In reality, many of these waves probably end up being dissipated in distant places, for example the mesosphere, and this may change the energy and momentum of the region from which they were emitted (albeit possibly by only tiny amounts). A balanced model that underwent similar changes in energy and momentum, at sites from which gravity would have been emitted in reality, might be more

realistic than one that conserved energy and momentum.

The foregoing considerations can be made the basis of various algorithms for iteratively defining PV inversion, by successively reducing the error in the divergence equation. Tests to date have been confined to the shallow-water equations on a hemisphere, almost the simplest system for which the questions involved are nontrivial, but already very nontrivial since the tropics are included.

For this system a number of iterative definitions of PV inversion have been devised and implemented, with considerable ingenuity, by Warwick Norton at Cambridge. Some are Galilean-invariant and some are not (for example a version using the techniques of nonlinear normal-mode initialization); in practice the differences have so far been found to be small, consistent with what we think is the likely degree of accuracy of the invertibility concept itself. Figs. 7-9 show one example (Norton, personal

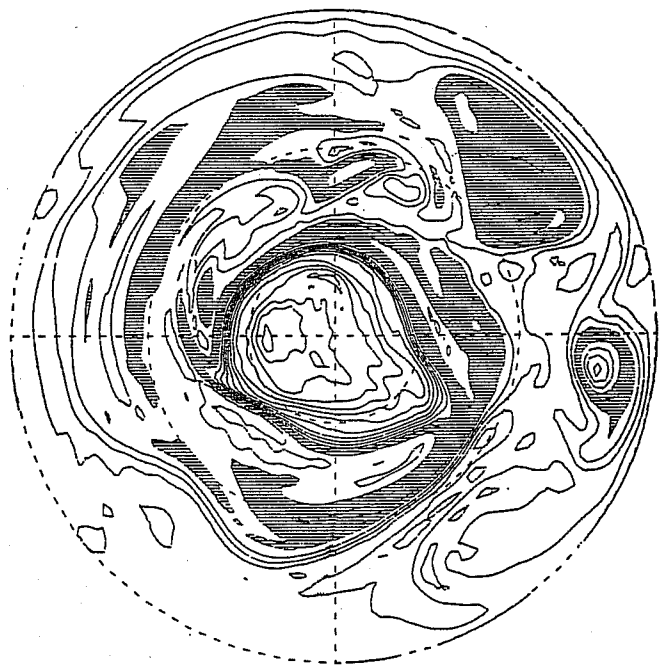

Fig. 7 (a) PV distribution from a shallow-water, primitive-equation integration on a hemisphere, originally intended as a model stratosphere along the lines of Figs. 2-4. Map is polar stereographic and shows the whole hemisphere. Resolution T63, Reading spectral code; equivalent depth is 8 km. Contour interval is $0.2 \times 10^{-8} s^{-1} m^{-1}$; shading marks values from $0.8 \times 10^{-8} s^{-1} m^{-1}$ to $1.2 \times 10^{-8} s^{-1} m^{-1}$. Higher values are mainly confined to the main polar vortex and the prominent secondary vortex on the right, deep inside the tropics.

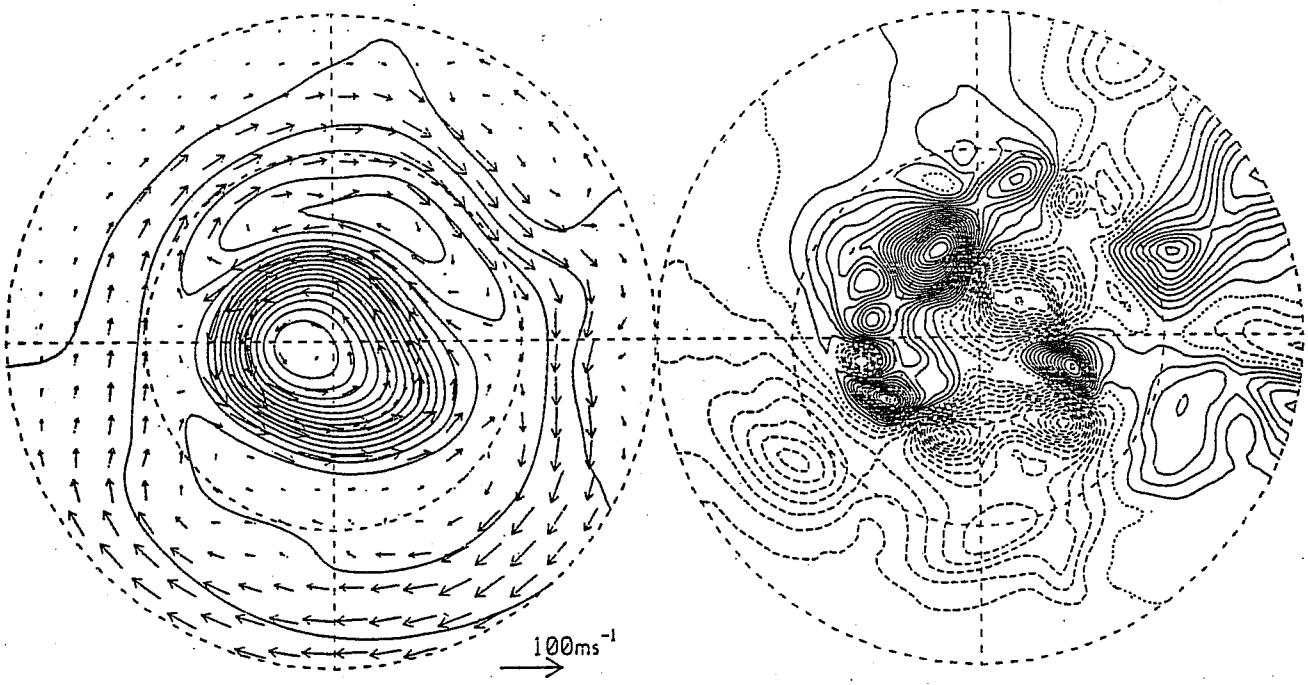

Fig. 8 (a) height and wind, and (b) divergence, from the same shallow-water, primitive-equation integration as in Fig. 7. Contour intervals 200m and $0.8 \times 10^{-7} s^{-1}$ respectively. Zero contour dotted, negative values dashed, in (b).

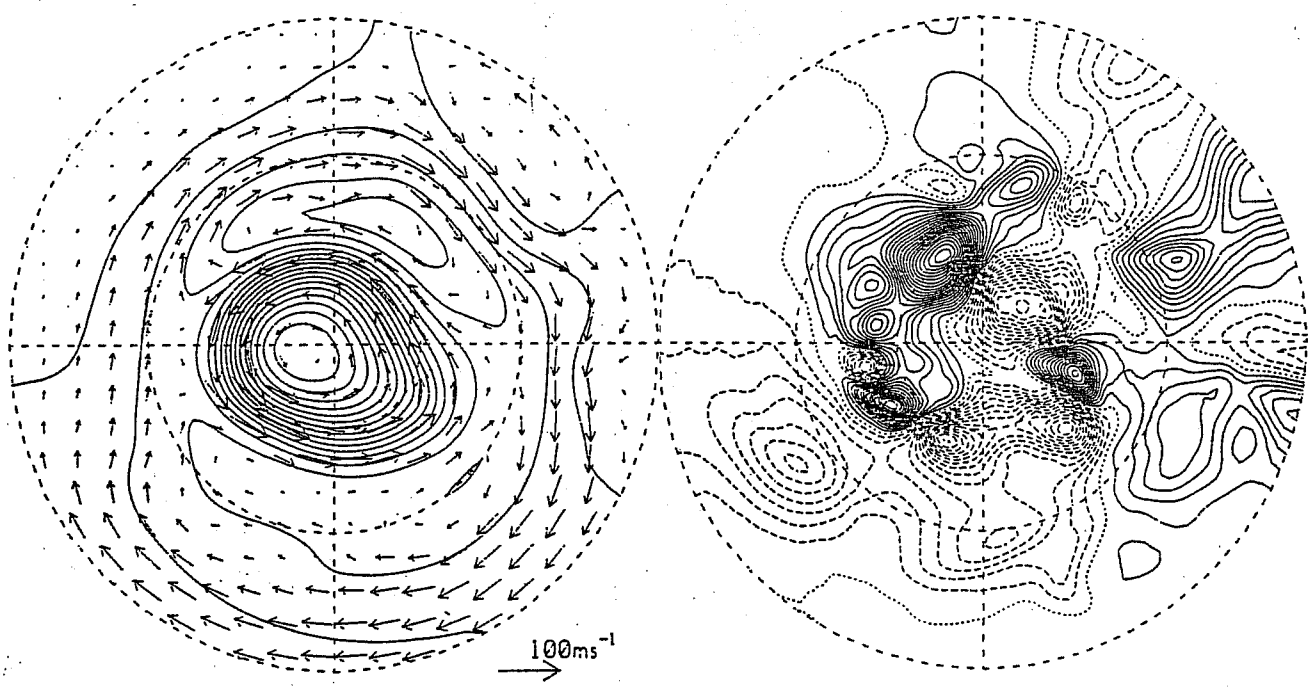

Fig. 9 Same fields as in Fig. 8, but recovered by a high-order inversion from the data in Fig. 7 alone. (Calculation by W.A. Norton.)

communication). Fig. 7 depicts the PV field from a shallow-water, primitive-equation integration rather like that of Figs. 2-4. Fig. 8 shows the corresponding height, wind and divergence fields from the same primitive-equation integration. Fig. 9 shows the height, wind and divergence fields recovered by one of Norton's inversion algorithms. You have to look quite closely to see the differences, even in the tropics. A difference map of the divergence fields using the same contour interval is blank apart from the zero contour. The same thing happens for several different inversion algorithms: all give almost exactly the same result.

The equivalent depth for this case is 8km. Other cases down to 0.5km have been tried, some with Froude numbers well above unity in the circumpolar jet. This begins to show more differences between the inversion algorithms -- for instance, there is a Galilean-invariant algorithm that does better than the nonlinear-normal-mode one -- but it is a remarkable fact that invertibility, at accuracies not far short of that shown in Fig. 9, still holds.

****

References and acknowledgements may be found at the end of the second lecture, "The use of PV and low-level temperature/moisture to understand extratropical cyclogenesis".