Isentropic distributions of potential vorticity and their relevance to tropical cyclone dynamics

Michael E. McIntyre

Proc. ICSU/WMO International Symposium on
Tropical Cyclone Disasters,
ed. J. Lighthill, Z. Zheng, G. Holland, K. Emanuel.
Beijing, Peking University Press (1993), pp.143–156.

Proc. ICSU/WMO International Symposium on
Tropical Cyclone Disasters, ed. J. Lighthill, Z. Zheng.
Beijing University Press (1993), pp.143–156.

ISENTROPIC DISTRIBUTIONS OF POTENTIAL VORTICITY AND THEIR RELEVANCE TO TROPICAL CYCLONE DYNAMICS

Michael E. McIntyre

(Dept. Appl. Math. Theor. Phys., Univ. of Cambridge, Silver St CB3 9EW, U.K.)

ABSTRACT This paper concerns 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 synoptic meteorologists; and it gives the synoptician and the forecaster direct access to concepts traditionally reserved for those more inclined to abstract mathematical thought. It is equally relevant to oceanic dynamics. A brief review of the fundamental concepts is given here, with some remarks on the elastic 'eye-wall PV barrier' and other points of special relevance to tropical cyclones (TCs).

I. INTRODUCTION

I cannot speak as a TC expert, but only as a fluid dynamicist interested in the workings of the atmosphere. Our theoretical understanding of atmosphere–ocean dynamics is today reflected in many contributions scattered throughout a vast literature, but a major part of that understanding — concerned with the kind of stratified, rotating fluid motion that underlies most weather developments — can be most succinctly expressed in terms of the fields of dry potential temperature (PT) or specific entropy, or other thermodynamical equivalent, and the associated *Rossby–Ertel potential vorticity* (PV). Section II will recall why this is so. To take just one example, the interesting 'wind surges' whose association with TC genesis has been suggested by Professor W. M. Gray, in his paper to this Proceedings[1], are stratified, rotating fluid motions. It seems likely that we shall not fully understand their role in initiating TC development until we understand their role in advecting air with different PV values from one place to another, in the stably-stratified tropical tropospheric environment.*

We also need to understand the PV–PT changes due to any concurrent or subsequent moist thermodynamic processes, and how those changes, in turn, modify the subsequent fluid motion. A mature TC probably owes its survival to the strong, quasi-elastic 'PV barrier' surrounding its core, tending to isolate the saturated eye-wall upflow from large nonaxisymmetric, horizontal incursions of drier environmental air. In this regard, the core presumably resembles a small segment of the core of a turbulent smoke ring, whose smoke-filled interior is only slowly diluted by its turbulent surroundings[3]. An even closer analogy is the wintertime stratospheric polar vortex since, although the horizontal scale is much larger, it is another case of stratified, rotating fluid motion of

*The recent work of Ritchie and Holland[2] is of interest in this connection.

a fundamentally similar kind. The polar vortex has a quasi-elasticity or resilience that tends to isolate its core from zonally asymmetric horizontal incursions of the surrounding air, permitting special 'ozone hole chemistry' to take place (§IV below). Strong dry-isentropic gradients of PV, near the jet maximum surrounding the vortex core, are essential to this quasi-elasticity. In the case of TCs, PV advection by the abovementioned surges cannot, of course, account for the extremely strong PV gradients that must exist near the eye wall and wind maximum in a mature TC — it is here that moist processes must be crucial — but such advection could well be significant in the early stages by helping to initiate a rotationally isolated moist updraft.

It seems to me that we have yet to attain a detailed understanding of how such rotationally isolated updrafts might be initiated and remain stable — among other things it will surely require some very careful three-dimensional mesoscale numerical modelling — but a brief survey of some relevant dynamical fundamentals may be not inappropriate to this Symposium. The following presentation is based on ideas and material developed over the years in previous articles and lectures, including lectures given to the ECMWF Seminar of September 1987, to the International Fermi School of July 1990[4], and to the more recently inaugurated UK Summer School in Geophysical and Environmental Fluid Dynamics.

II. PV FUNDAMENTALS

The key idea, which goes back to Charney[5] and Kleinschmidt[6], is to recognize that certain aspects of the PV–PT fields can be regarded as *controlling* the dynamical evolution. There is an 'invertibility principle' — so familiar to theoreticians that it is not always mentioned explicitly — saying that a certain subset of the information in the PV–PT field can be used to diagnose everything about the other dynamical fields, apart from any inertio-gravity oscillations that may be present, including equatorial Kelvin waves. This diagnostic process may be called PV 'inversion'. A precise statement will be given in §V below. The advantages of this viewpoint include the following:

(a) The evolution of the PV–PT field incorporates the *effects of advection* in the conceptually simplest way possible. This is a powerful advantage because advection is almost always a crucial process: there is no escape from considering the advective nonlinearity somehow.

(b) The PV–PT viewpoint recognizes, makes explicit, and keeps conceptually separate the *non-local* aspects of the dynamics. These are all incorporated into the idea of PV 'inversion'.

(c) The PV–PT viewpoint makes precise the basis, extent, and limitations of the partial analogy with two-dimensional (2D) barotropic vortex dynamics. This is included as a special case, characterized by a particular inversion operator.

Thus the PV–PT viewpoint shows, for instance, why certain classical-aerodynamical phenomena (such as 2D shear instability, vortex rollup, merging, and vortex-core isolation) all seem to have baroclinic, 'layerwise-2D' counterparts in synoptic-scale and mesoscale atmosphere–ocean dynamics.

Point (a) is significant whether or not the motion is 'frictionless' and adiabatic. Of course many significant weather developments do depend on fast upper-air motions that are, to a first approximation, frictionless and adiabatic, so that both PV and PT are materially conserved, i.e. simply advected. Features in the PV–PT fields are then simply carried along with the air motion. They 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, TUTTs and so on[7] — and the PV–PT viewpoint provides the simplest explanation of why such structures occur so commonly.

Furthermore, advection can from this viewpoint be considered as a quasi-horizontal, *quasi-two-dimensional* process. This is another powerful simplification. Advection is quasi-2D in the sense of referring to horizontally-projected motion along isentropic, i.e. constant-PT, surfaces, and to horizontally-projected motion near the Earth's surface (meaning, in practice, just above the boundary layer)[8], with no need to refer explicitly to vertical motion despite the latter's dynamical importance.

When material conservation fails, as happens for instance in moist convection, the evolution of the PV–PT field can still be described as resulting from local effects only: advective, diabatic and frictional. Moreover, and very surprisingly (§V, note 7), one can still retain the quasi-two-dimensionality, even when diabatic heating, for instance latent heating, is taking air parcels across isentropic surfaces.

Such a view of the evolution is fundamentally simpler than referring directly to the primitive equations. The primitive equations intimately combine the local frictional and diabatic effects with *three-dimensional* advection and *non-local* interactions. The non-local interactions are mediated by the pressure and buoyancy fields, constrained by the requirements of mass conservation, hydrostatic balance, and so on. So although the primitive equations are useful in showing how Newton's second law of motion is satisfied, they impede understanding by intertwining all the different aspects of the stratified, rotating fluid dynamics — 3D and quasi-2D, prognostic and diagnostic, balanced and unbalanced, local and non-local. The importance of non-local interactions, incidentally, is an especially strong reason why studying the local balance of terms in a single equation, while sometimes useful when seen as part of a wider picture, can often give a misleading impression of causal linkages and the workings of the dynamics.

III. THE GENERAL FORM OF THE PROBLEM, AND ITS SIMPLEST PARADIGM

The PV–PT viewpoint applies most directly to any dynamical process that can be considered 'balanced' in the sense that inertio–gravity oscillations are absent, or balanced after averaging out any distinguishable such oscillations. The general form of the resulting problem is essentially that of its simplest paradigm, 2D nondivergent barotropic dynamics on an f-plane or β-plane, whose equations can be written

$$DQ/Dt = \text{frictional terms}\ , \tag{1a}$$

$$\psi = \nabla^{-2}(Q - f)\ , \tag{1b}$$

where f is the Coriolis parameter and D/Dt the two-dimensional material derivative, defined by

$$D/Dt = \partial/\partial t + \mathbf{v}\cdot\nabla = \partial/\partial t + u\partial/\partial x + v\partial/\partial y\ , \tag{2}$$

and where, in this simplest case, the wind field is strictly nondivergent:

$$\mathbf{v} = (u, v) , \qquad u = -\partial\psi/\partial y , \qquad v = \partial\psi/\partial x . \tag{3}$$

If the right-hand side of (1a) is zero, the equation states that Q (which in this case is simply the absolute vorticity, $Q = f + \nabla^2\psi$) is materially conserved, i.e. advected, by the two-dimensional velocity field $\mathbf{v}$. And if we were to watch a moving picture of the Q field (whether or not it is materially conserved), then we would be following everything about the dynamics since, by (1b), knowledge of Q implies knowledge of ψ and hence, by (3), of the wind field. This is the relevant invertibility principle, and three points about it should be noted:

(i) Local knowledge of Q does not imply local knowledge of ψ or $\mathbf{v}$; the inversion operator defined by (1b) and (3) is a global and not a local operator. In particular, it depends on specifying suitable boundary conditions to make the inverse Laplacian ∇^{-2} unambiguous.

(ii) Invertibility always depends on some kind of balance condition; and in this example the balance condition corresponds simply to an absence of sound and external gravity waves. They have been filtered out by the assumption of incompressible, nondivergent motion expressed by (3).

(iii) There is a scale effect, whereby small-scale features in the Q field have a relatively weak effect on the ψ and $\mathbf{v}$ fields, while large-scale features have a relatively strong effect. In particular, ψ and $\mathbf{v}$ are to varying degrees insensitive to fine structure in the Q field. The inverse Laplacian ∇^{-2} in (1b) is a smoothing operator, and some of the smoothing survives even when followed by the single differentiations in (3).

Equations (1a) and (1b) summarize, with remarkable succinctness, the peculiar way in which air parcels push each other around. The nonlocalness of the inversion operator and the implied action at a distance, (i), are related of course to the balance condition, (ii). The invertibility principle holds exactly in this case, because the waves representing departures from balance have been assumed to have infinitely stiff 'elasticities' or restoring mechanisms, and hence to propagate infinitely fast. It may be useful to note from (1b) that diagnosing the ψ field from the Q field is almost the same thing, mathematically, as calculating the static displacement of a stretched membrane induced by a given pressure distribution on it, or calculating the electrostatic potential induced by a given charge distribution; note that in this second analogy $\mathbf{v}$ is at right angles to the electric field. Thus strong local anomalies in Q tend to induce strong circulations around them, in the corresponding sense. Such a Q anomaly, together with its induced velocity field, is nothing other than the coherent structure that fluid dynamicists call a 'vortex'. The corresponding layerwise-2D coherent structures in a stratified, rotating flow have long been familiar to meteorologists as 'cyclones' and 'anticyclones', e.g. Fig. 1, from the work of Thorpe[9]; further interesting examples have been given, for instance, by Raymond[10] and Davis[11]. The term 'vortex' may be used to mean either the entire coherent structure, or simply the PV anomaly that induces it.

The succinctness of (1a, b) is to be compared with what is involved in thinking directly in terms of Newton's second law of motion. Newton's law applied to every

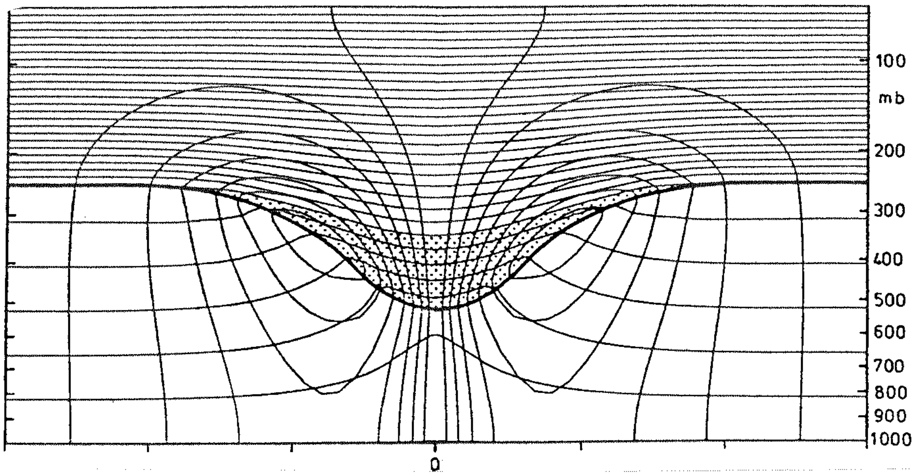

Figure 1. Section across the circularly symmetric structure induced by an isolated, circularly symmetric, cyclonic potential-vorticity anomaly near a model tropopause (heavy curve) across which the Rossby–Ertel potential vorticity Q has a strong discontinuity, by a factor of 6. The anomaly (meaning a PV contrast on isentropic surfaces, ref. [8], Eq. 29), is shown by stippling. The family of thin curves some of which are closed are isopleths of tangential velocity, at $3\,\mathrm{ms}^{-1}$intervals — the greatest velocities $> 21\,\mathrm{ms}^{-1}$ being at the tropopause — and the other family of thin curves, more nearly horizontal, are the isentropic surfaces, plotted at 5 K intervals in PT. The induced surface-pressure minimum is 41 hPa below ambient. The structure is typical for middle latitudes; the Coriolis parameter is $10^{-4}\,\mathrm{s}^{-1}$(as at latitude 43.3°N). The domain shown has a radius of 2500km.

air parcel is mathematically equivalent to (1), but requires us to think explicitly about subtle, non-local aspects of the pressure field.* The power of the viewpoint represented by (1) has long been recognized, and made use of, by aerodynamicists as well as by theoretical meteorologists. Indeed the idea of invertibility, as exemplified by (1*b*), is built into classical low-Mach-number aerodynamical language, in such phrases as the velocity field 'induced by' a given vorticity field, and in such ideas as the idea that a strong vorticity anomaly can roll 'itself' up into a nearly circular vortex.

In stratified, rotating flow, the waves representing departures from balance are inertio–gravity waves with finite phase and group speeds. It can be shown[12] that, not surprisingly from this viewpoint, the balance (slow-'manifold') and PV-invertibility concepts are then inherently approximate, except in a few very special cases including exactly circular vortices in gradient-wind balance, such as that of Fig. 1. However, balance conditions and corresponding inversion operators can be defined that are surprisingly accurate (at the price of being computationally more elaborate, and to some extent nonlinear), even in cases where inertio-gravity-wave phase speeds c are not at all large numerically in comparison with typical wind speeds $|\mathbf{v}|$. We possess examples in which local Froude numbers $|\mathbf{v}|/c$ reach values as high as 0.7, while the accuracy

---

*As suggested in ref. [4], the formulation (1) represents an economy of thinking analogous to, but greater than, the economy that results from treating normal reaction forces as constraints when discussing the dynamics of a roller coaster on a nearly-rigid track. Such a treatment reduces the three-dimensional roller-coaster problem to a one-dimensional problem in an obvious way, making the problem easier than explicitly using the two normal components of Newton's second law. In the fluid-dynamical system, the mass-conservation and boundary conditions likewise have the nature of constraints, felt through the pressure field: to this extent the pressure field has a role like that of the normal reaction forces in the roller-coaster problem.

remains much better than typical observational and numerical forecast accuracies[12]. This is very surprising indeed. The explanation seems to be the weakness of the nonlinear coupling from vortical motions to gravity waves, analogous to the well known weakness of 'aerodynamical sound generation' [13] pointed out by Lighthill in the 1950s. Coriolis effects tend to make this coupling weaker still. It is this same nonlinear coupling that underlies the fact that balance and invertibility are indeed inherently approximate, albeit often remarkably accurate, concepts, the inherent approximateness being also related to the mathematical concepts of 'homoclinic tangle' and 'stochastic layer' explained in books on 'chaos'. PV inversion operators in stratified, rotating flow are qualitatively not unlike 3D versions of the simple inverse Laplacian (1*b*), with a shorter-range character related to the finiteness of c.

On point (iii), the classic illustration is simple Rossby wave propagation* and its scale dependence. The Rossby wave propagation mechanism is part of what underlies the phenomenon of 'PV-barrier elasticity' or, as it could well be called, 'Rossby-wave elasticity'. The theory is standard, but gives a nice illustration of all the foregoing ideas, including the idea of the 'induced' velocity field implied by whatever inversion operator is relevant. In brief, the theory expresses the fact that if you have a gradient of Q — and, in the baroclinic case, what is relevant is an *isentropic* gradient of (Rossby–Ertel) PV, or gradient of PV along isentropic surfaces, what ref. [8] incautiously called an 'IPV gradient' — then the moment you undulate the Q contours sideways you create a row of Q anomalies with alternating signs. The induced velocity field given by the inversion operator is then a quarter wavelength out of phase with the displacement field describing the original undulations, causing the displacement of each air parcel to oscillate in time. The undulations propagate in the direction along the Q contours that has the higher Q values on the right. Further discussion is given in refs. [4] and [8].

When speaking of this 'PV-barrier elasticity' or 'Rossby-wave elasticity' — which, incidentally, is basic to understanding such phenomena as baroclinic instability, and many other large-scale motions of meteorological interest[8] — one usually has synoptic or planetary scales in mind, as in the stratospheric polar vortex or ozone-hole problem. But if isentropic gradients of PV are strong enough, then the resulting Rossby-wave elasticity can be effective on much smaller scales, particularly, for instance, in the case of the very strong gradients near the wind maximum surrounding the core of a mature TC. Indeed, to a first approximation the phenomenon of vortex-core isolation is scale-independent, as the smoke-ring example of §I reminds us.

IV. ROSSBY-WAVE ELASTICITY IN THE STRATOSPHERE

It is worth taking note of the remarkable effectiveness of Rossby-wave elasticity in isolating the core of the stratospheric polar vortex, both in a fairly realistic model of the winter stratosphere[14] and, it can strongly be argued, in the real stratosphere as well.[4,15,16] The model results are illustrated in Fig. 2, of which an animated version was shown in a video at the Symposium.

The model dynamical system comprises the shallow-water equations (the simplest relevant dynamical system with finite c) solved numerically on a sphere by a high-

---

*Rossby waves (historically, 'Kelvin–Kirchhoff–Rayleigh–Rossby waves') may also be called 'vorticity waves' or, more generally, 'potential-vorticity waves'. I am following today's established usage.

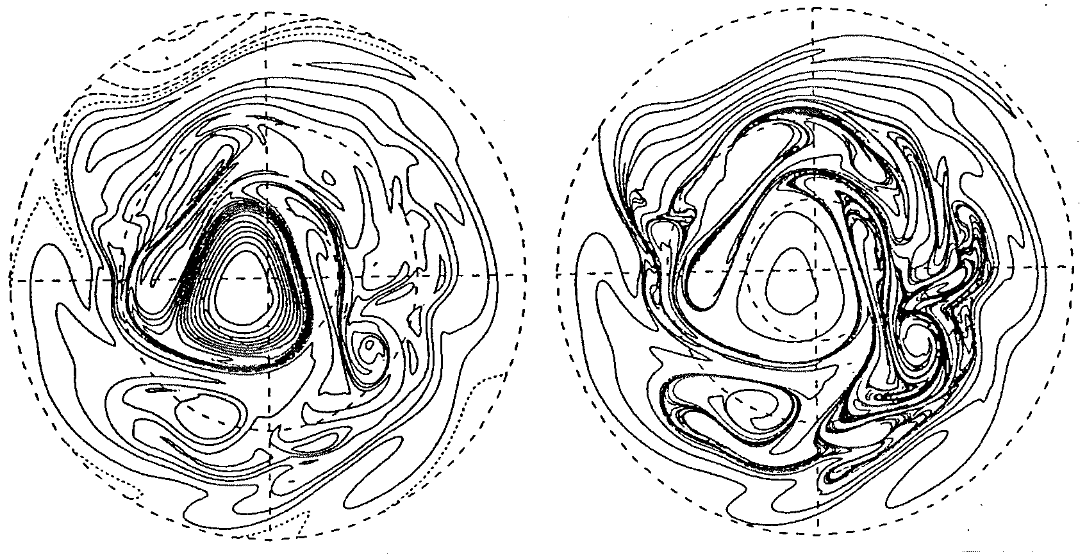

Fig. 2. Winter hemisphere in a high-resolution, shallow-water numerical model of the stratosphere, from ref. [14]. The mean depth is 4 km, and the numerical resolution (triangular truncation at total wavenumber 127) corresponds to a mesh size roughly 1° latitude. *Left panel*: shallow-water potential vorticity Q, contour interval $4 \times 10^{-9}\,\mathrm{m}^{-1}\mathrm{s}^{-1}$, zero contour dotted. *Right panel*: isopleths of an advected tracer field, initially axisymmetric and coincident with the Q contours, showing the fluid-dynamical irreversibility characteristic of Rossby-wave 'breaking'[4] and two-dimensional 'turbulence', also the complete isolation of the vortex core from zonally asymmetric horizontal incursions of the surrounding air. This was computed with near-perfect accuracy using a high-precision 'contour advection' technique, introduced independently by Norton[14] and by Waugh and Plumb[17] using an algorithm developed in another context by D. G. Dritschel[18].

resolution pseudospectral method. The projection is polar stereographic, and the winter northern hemisphere is shown. An axisymmetric initial state is disturbed by smoothly distorting the lower boundary in a large-scale pattern so as to imitate the effect, on the real stratosphere, of planetary-scale Rossby waves propagating up from the much denser troposphere below[19]. The result looks remarkably similar to what is seen in the real winter stratosphere at altitudes of the order of 25 to 50 km.

The left panel of Fig. 2 shows Q, now the shallow-water potential vorticity defined as absolute vorticity over local layer depth. The model problem still has the generic form (1) to excellent approximation, but with a shorter-range potential-vorticity inversion operator* because of the finiteness of c. The central region is the model's stratospheric polar vortex, where material contours that were initially on latitude circles remain closely coincident with the Q contours, and undulate nearly reversibly. Here the 'Rossby elasticity' is the dominant effect. This is further illustrated by comparison with the right panel, which shows the behaviour of some material contours computed very accurately using a high-precision 'contour advection' technique adapted from the work of Dritschel[18]. Outside the polar vortex, in middle latitudes, is a region in which the waves are 'breaking' vigorously. There, the initially-latitudinal material contours

---

*To rough approximation, this inversion operator is given by (1*b*) with ∇^{-2} replaced by $(\nabla^2 - \kappa^2)^{-1}$, corresponding to an elastic membrane tethered by local springs, somewhat like a spring mattress, with a latitude-dependent e-folding scale κ^{-1} (the Rossby radius) of about 1400 km at the pole, 2000 km at 45°N, and 3000 km in the tropics. Ref. [12] gives examples of much more accurate, albeit much more elaborate, potential-vorticity inversion operators for shallow-water models.

are deformed rapidly and irreversibly, and mixed into a broad, 2D-turbulent 'Rossby-wave surf zone'; e-folding times for contour lengthening were estimated by Norton to be about 4 days.

The contours in the right-hand panel of Fig. 2 can be regarded as isopleths of an advected passive tracer field, the advection being very accurately simulated, with no artificial diffusion. The simulation shows that, at least in the model, chemical substances in the turbulent midlatitude 'surf zone' do not penetrate past the region of strong Rossby elasticity concentrated in the steep Q gradient near the vortex edge. This is believed to be important for ozone-hole chemistry. The contour-advection technique, conceived of as a benchmark numerical tracer advection algorithm, was introduced independently by Norton[14] and by Waugh and Plumb[17]. Such numerical experiments, and laboratory counterparts[20] are important in showing the likely reality, for a nonlinearly disturbed vortex, of the vortex-core isolation suggested by the linear concept of 'Rossby elasticity'. In the nonlinear reality, it turns out that the strong shear just outside the vortex has an important role as well.[15,21,22]

V. THE PV INVERTIBILITY PRINCIPLE FOR A BAROCLINIC ATMOSPHERE

The standard meteorological definition of PV for a three-dimensional, baroclinic, hydrostatic atmosphere,

$$Q = (f + \hat{\mathbf{z}} \cdot \nabla_\theta \times \mathbf{v})(-g\partial\theta/\partial p) \, , \tag{4}$$

is essentially that proposed by Rossby in 1940[23] as a natural development of the single-layer 'shallow-water' PV he proposed in 1936[24]. In (4), $\hat{\mathbf{z}}$ is a unit vertical vector, θ is the PT, or any of its thermodynamical equivalents, such as specific entropy, subscript θ indicates differentiation along an isentropic surface, $\mathbf{v}$ is the horizontal wind vector as before, and p is hydrostatic pressure. Ertel's more general formula published in 1942[25] reduces to (4) in the hydrostatic case.

A convenient SI unit for Q is 10^{-6} m^2 s^{-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 ref. [8]. It is a convenient unit since, with the convention of using θ as the thermodynamic variable, PV values < 1 usually imply that we are looking at tropospheric air, and > 2, extratropical stratospheric air. A notable exception to this is the case of a mature TC, where lower-tropospheric PV values well over 10 PVU may be encountered, in dramatic contrast with typical surrounding values ($\sim 10^{-1}$ PVU).

One way of stating the invertibility principle for the Q and θ fields is that proposed in ref. [8]. 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 *diagnostic* nature of PV inversion means that the *prognostic* aspects of the problem have been confined to describing the adiabatic or diabatic evolution of Q and θ; this is part of the conceptual separation referred to in §II.

The following additional points should be noted:

1. Just as in the simpler case (1*b*), to carry out an inversion we must solve the diagnostic problem globally, with proper attention to boundary conditions; the same remarks about 'action at a distance' apply. This is an inescapable consequence of 'non-localness'.

2. The principle, in the form just stated, helps to explain why isentropic gradients of PV and surface gradients of PT keep on turning up as key factors in theoretical studies of barotropic and baroclinic instabilities, large-scale waves, vortices, so-called 'geostrophic' turbulence (Fig. 2) and other phenomena involving balanced motion.

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 §II.

4. The principle works only for 'dry inversion', i.e. when the Q and θ fields refer to the dry and not the moist equivalent PT, say θ_e. The latter, and its associated moist PV, say Q_e, are more useful for the purposes of parameterizing upright and slanting moist convection, for instance by setting $Q_e = 0$ in the eye wall of a TC[26]. The Q_e and θ_e fields are not useful for inversion, for mathematical reasons; so the Q and θ fields need to be kept track of in moist-convective situations.

5. 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 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.

6. 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 §VI; it shows how PV advection can trigger moist convection. 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 ref. [8] (§4 and appendix).

7. PV distributions and their possible evolution, and the associated transports and budgets of PV — more precisely, of the transportable 'substance' PVS, say, whose amount per unit mass is the PV — are constrained by two exact, general theorems that hold even in the presence of diabatic heating and frictional or external forces. Specifying a PV distribution that violated the associated constraints would presumably lead to failure of any attempt at inversion (see also ref. [8] §3, Eq. 17b *et seq.*). The first is that PVS, considered as a conservable, transportable quasi-substance, is *indestructible*, like electric charge (one can have pair production and mutual annihilation, but no net charge creation), except where isentropic surfaces intersect a boundary such as the

earth's surface. The second is that isentropic surfaces act as if they are *impermeable* to PVS — 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. It is this that allows the quasi-two-dimensional representation of advection to be retained, very surprisingly, even when mass and chemical substances are crossing isentropic surfaces; further discussion may be found in ref. [27], and [4] §11.

Thus, although values of Q, which have the nature of chemical mixing ratios, can change, they can change only by the notional 'substance' involved being transported, diluted, or concentrated in various ways. Both the indestructibility theorem and the impermeability theorem are direct consequences of the way in which the PV is constructed mathematically. This has the significant further consequence that the theorems apply not only to the exact PV constructed from exact $\mathbf{v}$ and θ fields, but also, exactly, to any 'coarse-grain PV' constructed from coarse-grain observational datasets.

VI. THE SIMPLEST KIND OF CYCLOGENESIS, AND THE 'VACUUM-CLEANER' EFFECT

This is worth a quick reminder here, if only because of its connection with possible contributions to TC forcing by tropical upper-tropospheric troughs (TUTTs).[28,29] The simplest conceivable kind of cyclogenesis occurs when a single upper-air isentropic anomaly of PV is advected into the region of interest, from some other location where its surroundings make it less 'anomalous'. The invertibility principle, as applied to isolated anomalies like that of Fig. 1, tells us that the arrival of a single such anomaly has to be accompanied by the arrival also of its induced wind, vorticity and static-stability structure — for example, in the case of an upper-air anomaly and no low-level anomaly, by a structure like that shown in Fig. 1. Bleck[30] gives a three-dimensional graphical depiction of this process, following Kleinschmidt.

Fig. 3 shows some coarse-grain PV maps from operational analyses, taken from ref. [8], in a real case giving a fairly clear example of essentially this kind of cyclogenesis (20-25 September 1982). Low-level PT anomalies seemed to be relatively unimportant in this case. An upper-air anomaly, consisting of high-PV stratospheric air, was advected from near Hudson Strait across the Atlantic towards Europe, and appeared to roll itself up into a large cutoff cyclone (at 18°W on 24 Sept.) having a structure like that of Fig. 1. The resulting surface cyclone is prominently visible near the centre of the top left panel of Fig. 4, the ECMWF surface analysis for 24 September 1982.

The vertical motion field has to fit in with these events. Consider an air parcel in the lower troposphere ahead of the cyclone — imagine it, say, near 800mb at the edge of Fig. 1 — and suppose that it subsequently gets caught up in the advancing cyclone. If diabatic and frictional contributions are negligible, the parcel must clearly ascend, since it has to ride some way up the relevant isentrope as it enters the cyclone. The uphill slope of the isentrope is a qualitatively robust part of the structure induced by the moving upper anomaly; any other shape compatible with the PV distributions would be impossibly far out of balance — the apparent 'action at a distance' again. The same ascent stretches vortex lines below the upper anomaly, and spins up the lower part of the cyclone.

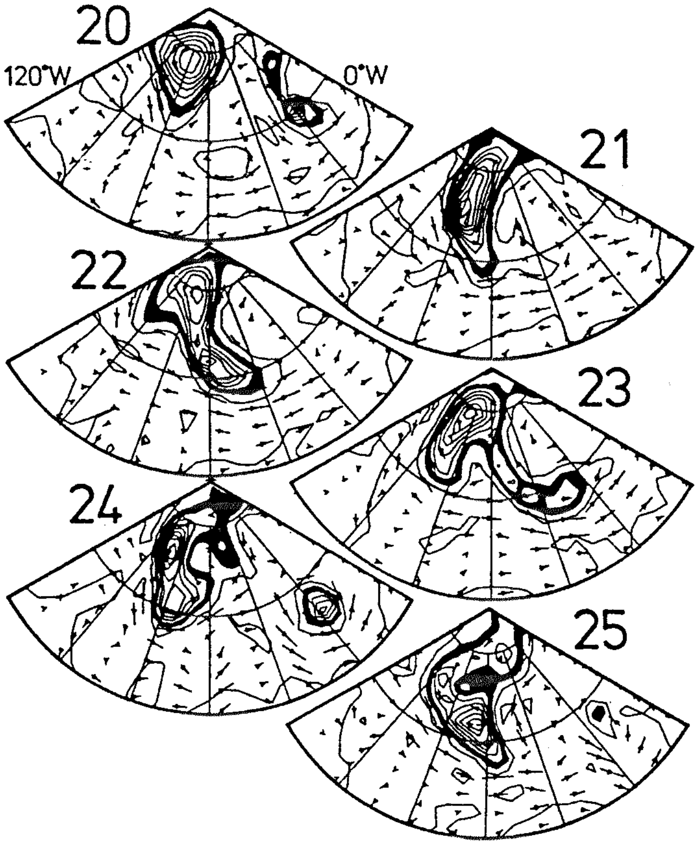

Fig. 3. 300 K isentropic maps of PV for 20-25 September 1982, September 1982 from ECMWF operational analyses. Latitude circles are 40°N, 60°N and 80°N; Greenwich meridian on the right, 120W on the left. Contour interval is 0.5 PVU, with 1.5 to 2 PVUblacked in, locating the tropopause. Wind arrows on the 300 K surface are scaled such that an arrow drawn northward from 40°N to 60°N would indicate a speed of 100 m s^{-1}. From ref. [8].

In summary, other contributions aside (see ref. [8], appendix), it is as if a cyclonic upper anomaly, advancing relative to the atmosphere below it, is "acting on the underlying layers of the atmosphere somewhat like a rather broad, very gentle 'vacuum cleaner', sucking air upwards towards its leading portion" (ref. [8] §4), in just such a way as to give rise (so to speak) to the vorticity and temperature anomalies that comprise the induced flow structure in those underlying layers. The vertical-motion signature produced by this 'vacuum-cleaner effect' is well known in synoptic meteorology. It is the basic reason why cloud and precipitation tend to form more or less on the advancing side of a cyclone; some case studies are cited in ref. [8] §4. The strength of the effect is evidently related to the PV anomaly strength and to the speed of upper-air advection relative to the lower layers. It also increases with increasing f, and so in the context of TCs will tend to be most effective on the fringes of the tropics.

The 'vacuum-cleaner' effect just described is, of course, nothing other than what is sometimes called the quasigeostrophic 'forcing' of vertical ascent and surface-cyclone development by an advancing cyclonic structure, another of whose characteristics is positive upper-air vorticity advection. Despite the terminology the effect is not de-

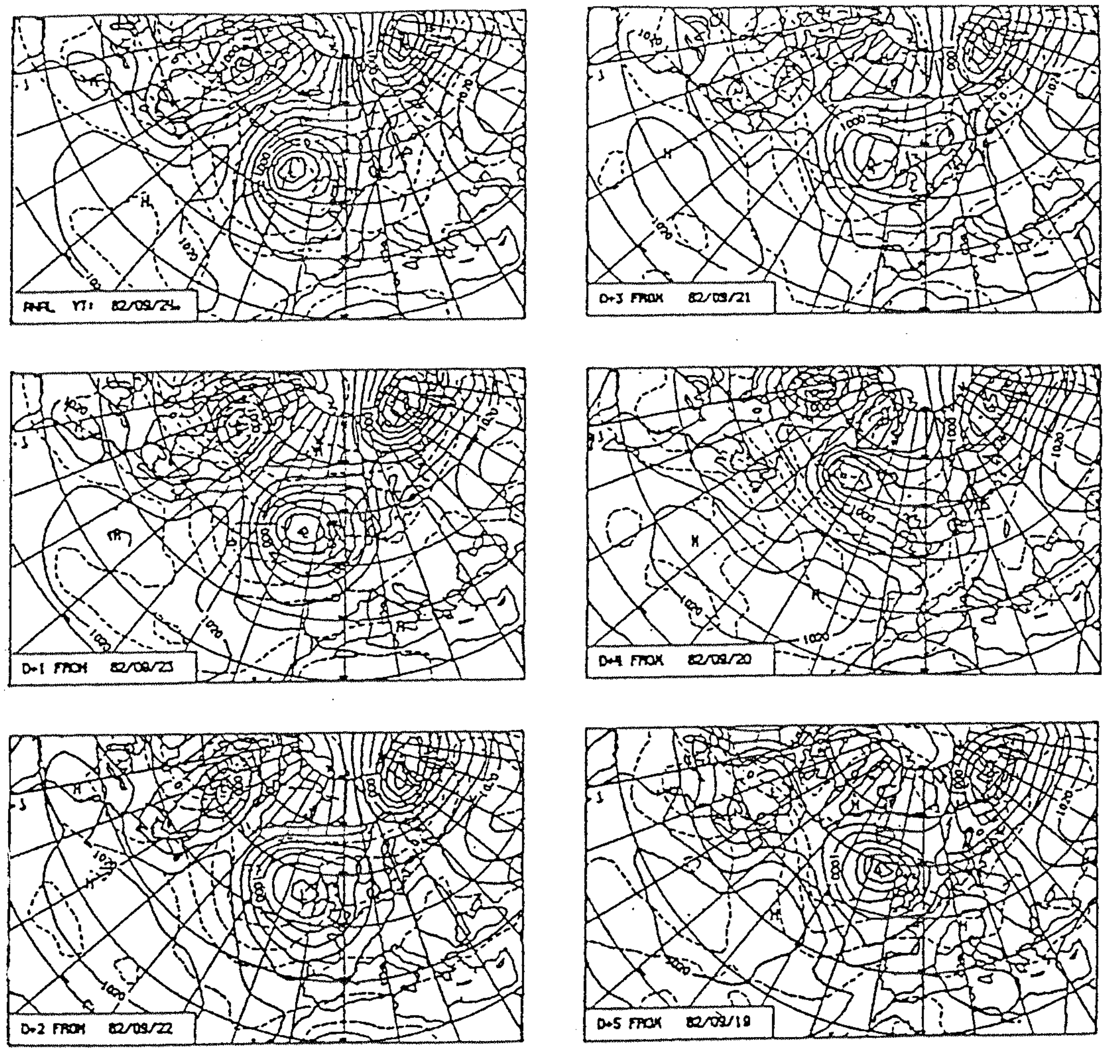

Fig. 4. ECMWF sea-level pressure (solid contours) and 850mb temperature (dashed contours) analysis for 24 September 1982 (top left), and five forecasts for that day from 23, 22, 21, 20 and 19 September. Contour intervals 5 hPa and 5 K.

pendent, of course, on the quasigeostrophic approximations. In principle it can be computed much more accurately.

VII. AN INTERESTING CASE OF A FAILED FORECAST

Fig. 3 suggests something interesting about the ECMWF operational forecasts shown in the remaining panels of Fig. 4. It can be seen that the east-Atlantic surface cyclone discussed above was reasonably well forecast from 1, 2, 3 and 5 days before 24 September, whereas the 4-day forecast was much worse (middle right panel). The initial condition for the 4-day forecast corresponds to the top left PV map in Fig. 3. We may hypothesize that a slight error in, say, the length of the left-hand 'trough' or cyclonic feature in that map could easily have been made in the analysis, and could easily have accounted for the bad forecast. It seems obvious that such an error could critically affect the precise amount of high-PV air subsequently advected from the polar stratosphere, to form the PV anomaly that induced the cutoff cyclone. It would be of great interest to run suitable transplant experiments[31], concentrating on the region surrounding the tip of the PV trough, in order to test this hypothesis.

VIII. CONCLUDING REMARKS

It is sometimes said that 'ordinary vorticity is enough in the tropics', and that there is no need to consider the PV. This is true to some extent of the prognostic or evolutionary aspect of the problem, but not of the diagnostic or 'inversion' aspect. Consideration of this latter aspect reminds us that the tropics is very far from being a 2D barotropic system, as indeed was emphasized long ago, in a different way, by Charney in his papers on large-scale motions in the tropics.[32,33]

A parallel has been drawn between possible TC core behaviour and stratospheric polar vortex core behaviour. It is intriguing to speculate that the shrinking of the eye wall radius observed in 'eye-wall cycles' might be not unrelated to the vortex-erosion process seen in model stratospheric polar vortices. Erosion must also presumably take place in TCs, although how it might interact with diabatic effects, including moist symmetric (zero-Q_e) neutral eye-wall flow[26], and whether it would add significantly to the simple effect of symmetric radial inflow is an open question — one of the questions that call among other things for careful mesoscale numerical modelling. The resolution required to represent correctly any such process might be quite severe.[18,21,22,34]

Many other questions suggest themselves, all beyond the scope of this brief discussion. To what extent are mature TCs like isolated, barotropic vortices? It might be thought that since they are largely created by diabatic heating, they must be like vertical vortex-pairs (what oceanographers call 'heatons'). However, there can be interesting effects in which much of the upper anticyclone is blown away by upper-tropospheric and lower-stratospheric winds, depending on vertical wind shears[35], giving interesting baroclinic steering effects that are easily comprehensible in terms of the PV–theta viewpoint.

Acknowledgements I am grateful to Dr Warwick Norton for supplying Fig. 2, and to the European Centre for Medium Range Weather Forecasts for supplying, and allowing me to reproduce, Fig. 4. Many colleagues have kindly shared their knowledge, ideas, historical recollections, and unpublished work over the years, including John Allen, Rainer Bleck, Lance Bosart, Keith Browning, Jule Charney, Mike Cullen, Ed Danielsen, Chris Davis, David Dritschel, Franco Einaudi, Arnt Eliassen, Kerry Emanuel, Bill Grose, Peter Haynes, Raymond Hide, Tony Hollingsworth, Jim Holton, Brian Hoskins, Ian James, Martin Juckes, Steve Koch, Dan Keyser, Ed Lorenz, Bob Lunnon, Jerry Mahlman, Taroh Matsuno, Jim McWilliams, Geoff Monk, Warwick Norton, Tim Palmer, Norman Phillips, Ray Pierrehumbert, Alan Plumb, Dave Raymond, Dick Reed, Peter Rhines, Rich Rotunno, Rick Salmon, Prashant Sardeshmukh, Wayne Schubert, Ted Shepherd, Glenn Shutts, Adrian Simmons, Chris Snyder, Susan Solomon, George Sutyrin, Alan Thorpe, Joe Tribbia, Adrian Tuck, Louis Uccellini, Tom Warn, Jeff Whitaker, Geoff Vallis and Martin Young. Work at Cambridge received support in part from the Natural Environment Research Council, through the British Antarctic Survey and through the UK Universities' Global Atmospheric Modelling Project, from the Innovative Science and Technology Program funded through the US Naval Research Laboratory, and from the Science and Engineering Research Council through research grants and through the award of a Senior Research Fellowship. Finally, I am grateful to Sir James Lighthill and Professor Zhemin Zheng, and to the sponsors of the Beijing Symposium, without whom this paper would never have been written.

REFERENCES

[1] Gray WM. In: Proc. ICSU/WMO International Symp. on Tropical Cyclone Disasters, ed Lighthill MJ, Zheng Z. Beijing: University Press, 1993

[2] Ritchie EA, Holland GJ. Q J Roy Meteorol Soc, 1993, ~~in revision.~~ 119, 1363

[3] e.g., Vladimirov VA, Lugovtsov BA, Tarasov VF. Zh Prikl Mekhan Tekhn Fiziki, 1980, 5:69~76 (Soviet Physics English translation).

[4] McIntyre ME. In: Proc Internat School Phys "Enrico Fermi", CXV Course, ed Gille JC, Visconti G. Amsterdam: North-Holland, 1992, 313~386, & refs. (A list of updates and corrections is available from the author, on paper or by email, ~~mem2@phx~~.cam.ac.uk.) mem@damtp.cam.ac.uk

[5] Charney JG. Geofysiske Publ, 1948, 17(2):3~17.

[6] Kleinschmidt E. Met Runds, 1950–1, 3:1~6, 3:54~61, 4:89~96.

[7] e.g., Palmén, E, Newton C. Atmospheric Circulation Systems. New York: Academic Press, 1969, 603pp.

[8] Hoskins BJ, McIntyre ME, Robertson AW. Q J Roy Meteorol Soc, 1985, 111:877~946. Also 113:402~404.

[9] Thorpe AJ. J Atmos Sci, 1985, 42:397~406.

[10] Raymond DJ. Q J Roy Meteorol Soc, 1992, 118:987~1015.

[11] Davis CA. J Atmos Sci, 1992, 49:1397~1411.

[12] McIntyre ME, Norton WA. J Atmos Sci, ~~1993, to appear~~ (& refs. therein). 57, 1214 (2000); corrig. 58, 949

[13] e.g., Crighton DG. J Fluid Mech, 1981, 106:261~298.

[14] Norton WA. J Atmos Sci, ~~1993, submitted.~~ 51, 654 (1994).

[15] McIntyre ME. J Atmos Terrest Phys, 1989, 51:29~43, & refs.

[16] Jones RL *et al.*Nature, 1993, to be submitted.

[17] Waugh DW, Plumb RA. J Atmos Sci, ~~1993, submitted.~~ 51, 530 (1994)

[18] Dritschel DG. J Comput Phys, 1988, 77:240~266.

[19] Charney JG, Drazin PG. J Geophys Res, 1961, 66:83~109.

[20] Sommeria, J., Meyers, S. D., Swinney, H. L., 1989: Laboratory model of a planetary eastward jet. Nature, 337, 58~61.

[21] Juckes MN, McIntyre ME. Nature, 1987, 328:590~596.

[22] Dritschel DG. J Fluid Mech, 1988, 194:511~547 (See p.516).

[23] Rossby CG. Q J Roy Meteorol Soc, 1940, 66(Suppl.):68~97.

[24] Rossby CG. Mass Inst of Technology and Woods Hole Oc Instn Papers in Physical Oceanography and Meteorology, 1936, 5(1):1~43.

[25] Ertel H. Met Z, 1942, 59:271~281.

[26] e.g., Emanuel KA. Ann Rev Fluid Mech, 1991, 23:179~196.

[27] Haynes PH, McIntyre ME. J Atmos Sci, 1987, 44:828~841. Also 47:2021~2031.

[28] Pfeffer RL, Challa M. J Atmos Sci, 1992, 49:1051~1059.

[29] Montgomery MT, Farrell BF. J Atmos Sci, 1993, 50:285~310.

[30] Bleck R. Mon Wea Rev, 1974, 102:813~829 (See Fig. 2).

[31] Hollingsworth A, Lorenc AC, Tracton MS, Arpe K, Cats G, Uppala S, Kållberg P. Q J Roy Meteorol Soc, 1985, 111:1~66.

[32] Charney JG. J Atmos Sci, 1963, 20:607~609.

[33] Charney JG. J Atmos Sci, 1969, 26:182~185.

[34] McIntyre ME, Palmer TN. Nature, 1983, 305:593~600. Also J Atm Terr Phys, 1984, 46:825~849, Pure Appl Geophys, 1985, 123:964~975.

[35] Wu C-C, Emanuel KA. J Atmos Sci, 1993, 50:62~76.