Comments on 'On the use and significance of isentropic potential vorticity maps' by B. J. Hoskins, M. E. McIntyre and A. W. Robertson (October 1985, 111, 877–946)

By J. S. A. GREEN
University of East Anglia

The authors are to be congratulated on their excellent survey of the use of potential vorticity (PV). I guess that we are impressed that modern observations and analysis techniques are sufficiently accurate to determine PV reliably, and that it is so difficult to invert compared with (say) the quasi-geostrophic variety.

But one aspect of their story worries me. It concerns the nature of the boundary condition to be applied at a lid, particularly when this is treated by the device of a vorticity sheet. It seems to me that this conflicts with the way that Eady (1949) and Green (1960, 1970) and many others, have seen the atmosphere as having thermodynamical and dynamical aspects (vorticity, or one of its variants representing the dynamics, potential temperature, or one of its variants representing the thermodynamics) but both needed to describe the motion completely.

Bretherton (1966) showed that one could solve the problem of finding the flow field given only the potential vorticity, by supposing the boundary to be isentropic, but then allowing the potential vorticity to have unbounded magnitude over an infinitesimal height range just above the boundary, such that the vertical integral was finite.

This fictitious PV must be dominated by the term in $f \partial \theta / \partial z$ so when one actually performs this integration one discovers the value of the temperature just above the boundary. To integrate forward in time we must carry this boundary distribution of potential vorticity along with us just as we had to for potential temperature. Thus the layer merely behaves as a surrogate for potential temperature at the surface.

I had shown Green (1960) that for amplifying baroclinic waves, the isentropic boundary was a very peculiar state and therefore labelled this development as Bretherton's bogus boundary layer, or Bretherton's bogus boundary condition where appropriate, B3 for short.

Inclusion of this fictitious sheet in the system has several disadvantages. There is the basic physical implausibility of having an isentropic lower layer, just where convective exchange with the underlying surface makes it the least isentropic in the system in the sense that here isentropes are more likely to be vertical than horizontal. The convective boundary layer is already complicated enough without having B3 above/below/inside, as well.

Also, one can show (as indeed Bretherton did) that the fictitious layer has a finite flux of potential vorticity associated with it which, at least in some cases, exactly cancels the flux in the whole of the rest of the system. Thus the transport of potential vorticity shown on the present authors' maps might be exactly nullified by opposing transfer in the B3 at another level. As shown in Green (1970) this would make the height-integral of momentum flux, which drives the surface zonal-mean flow, vanish and give zero zonal-mean surface wind everywhere. It is therefore not a good device to use if you wanted to know the surface wind.

I suspect it follows that 'inversion' of IPV can be seen as a simple superposition only if the system is tolerably isentropic in its lower layers, as in the authors' Fig. 15 for example, or with no anomaly anywhere except at the B3, as in Fig. 16. Any realistic case will be a mixture of these two. Moreover the volume-IPV and the B3 contribution will not usually be independent.

This leads us into the main thing that bothers me. This is, what has happened to the potential energy? I have grown used to thinking of baroclinic motion as displacements designed to release potential energy from the mean flow, and being hampered in that task by constraints due to vorticity. The thermal wind inciting the flow to be along rather than across the isotherms, vertical velocities demanding the expenditure of kinetic energy in useless rotational motion through vortex-stretching, for example.

These constraints induce dynamical evasive action, like wave propagation at the 'Rossby speed' to neutralize the β effect, and dispersion to get rid of the kinetic energy before the system puts it back into local potential energy again. In this picture the dynamical repercussions of propagation and dispersions are secondary. Now if the authors' (and B3) are a complete specification of the problem, the potential energy seems to have disappeared except as a minor diagnostic detail. Indeed according to these authors even the classical Eady problem of releasing potential energy in the absence of the tiresome dynamical constraints, has now been reduced (again by the agency of B3) to aspects of essentially barotropic waves propagating along the top and bottom boundaries.

It seems to me that there are two possible solutions. One is that the bulk IPV and the B3 contribution must be highly correlated with each other in a way that re-introduces the notion of potential energy. This might imply that there is no sense in which the two can be considered separately. Rather, that they can but only if an unlikely previous history for the air motion is implied.

The other solution concerns the upside-downness of PV. It has been said many times that contours of PV look much nicer if it is the reciprocal of PV that is plotted. But this is the thickness (mainly the height) of the isentropic sheet divided by the rotation. This begins to look a bit like available potential energy, particularly when defined in the Lagrangian sense used in Green (1970), rather than in the Eulerian sense of Lorenz. It would also begin to echo my notion of the manipulation of potential energy with vorticity constraints. So maybe it is more readily available potential energy that is being invented.

REFERENCES

Bretherton, F. P.	1966	Critical layer instability in baroclinic flows. *Q. J. R. Meteorol. Soc.*, **92**, 325–334
Eady, E. T.	1949	Long waves and cyclone waves. *Tellus*, **1**, 25–52
Green, J. S. A.	1960	A problem in baroclinic stability, *Q. J. R. Meteorol. Soc.*, **86**, 237–251
	1970	Transfer properties of the large scale eddies and the general circulation of the atmosphere. *ibid.*, **96**, 157–185

Reply by B. J. HOSKINS, M. E. McINTYRE and A. W. ROBERTSON

We thank Dr Green for his comments on our review. We are unsure to what extent these are meant to suggest that 'dynamics' should be thought of separately from 'thermodynamics', but it does no harm to emphasize that, despite its name, the PV contains, broadly speaking, as much information about the 'thermodynamics' as the 'dynamics'. There is more to it than just vorticity.

We agree wholeheartedly with Dr Green that it is crucial to pay attention (in one way or another) to the potential temperature θ at or near the lower boundary, as well as to isentropic distributions of PV in the free atmosphere. Moreover, we certainly don't regard Bretherton's mathematical device for thinking about the lower boundary condition as mandatory for 'IPV thinking'. As stated on page 910 of our review, it is "just a conceptual device and it is probably best simply to think in terms of the surface θ anomalies and the rule that 'warm is cyclonic and cold is anticyclonic'".

However, we do find Bretherton's device useful in some contexts. For instance it leads to the unified view of the Rossby-wave propagation mechanism noted at the end of our section 6(*a*). Figure 17, appropriately interpreted, can be seen as applying equally well to topographic as to ordinary Rossby waves, and also to neutral Eady waves. In the first and last cases the wave propagation mechanism depends on the basic state having a gradient of potential temperature θ at the boundary, instead of an isentropic gradient of PV in the interior.

Unstable Eady waves fit into the picture quite easily, as we tried to show in section 6(*b*) and Fig. 18, as does the fact that an isentropic boundary is 'peculiar', as Dr Green has pointed out. Such phenomena can be described directly in terms of surface θ anomalies, if this is preferred to Bretherton's description. Both descriptions, it hardly needs adding, are demonstrably consistent with the energy-conservation principle. One is merely talking about different views of the same dynamical phenomenon—and it can, of course, be very useful to have more than one view.

Regarding the planetary boundary layer, this naturally concerned us also, since the boundary layer is one of the places where the assumption of balanced motion seems least secure. We concluded (again on p. 910) that in practice the theoretical 'surface θ' is "probably best interpreted to mean θ. . . . just above the planetary boundary layer".

On the idea of vorticity 'constraints' we are doubtful as to whether these should always be thought of as 'hampering' the release of potential energy by baroclinic instabilities. There is a sense in which, on the contrary, they help the potential-energy release. The reluctance of fluid columns to shrink or stretch vertically in a rapidly rotating fluid means that there is a tendency for parcel trajectories to be flatter (more nearly parallel to the lower boundary) than the isentropic surfaces in a baroclinic zone. It is this very fact that *enables* potential energy to be released in a simple baroclinic instability (e.g. McIntyre 1970, pp. 294–5). If parcels just followed isentropes there would be no buoyancy effect at all. The point is illustrated by the well-known thought-

experiment of sloping the lids to match the slope of the isentropes (making the boundary condition 'peculiar' at both lids). Sloping the lids in this way does not make much difference to the store of available potential energy but, as expected from the foregoing argument, it completely suppresses the instability.

The IPV viewpoint gives us an alternative way of looking at the same phenomenon and immediately confirms that the instability is thus suppressed, since elimination of the boundary θ gradients eliminates the boundary Rossby propagation mechanism, and hence the instability mechanism. The IPV viewpoint actually takes us a bit further, by making it obvious that sloping only *one* of the boundaries would be sufficient to suppress the Eady instability. That, too, is understandable in terms of our Fig. 18 and the discussion accompanying it. At least two Rossby waves must exist (and be capable of becoming phase-locked to each other) in order for the baroclinic instability mechanism to work. It is no accident that isentropic gradients of PV and surface gradients of θ (and not distributions of potential or kinetic energy) enter the relevant general theorems on sufficient conditions for stability.

Dr Green raises the question of the atmospheric general circulation and the mid-latitude surface westerlies. Although our paper was concerned mainly with synoptic-scale systems, we think that the IPV viewpoint can also make a useful contribution to general circulation studies, including certain nonlinear tropical–extratropical interactions (as was mentioned on our pp. 936–7). Dr Green is correct in saying that, in quasi-geostrophic theory, the globally integrated PV flux must vanish exactly when 'B3' boundary contributions are included (the PV flux being equal to a divergence, namely the divergence of the Eliassen–Palm flux). However, the vertically integrated PV flux, including the surface θ contribution, need not vanish at each latitude. Indeed the existence of surface westerlies in the real atmosphere can be related to the existence of PV fluxes concentrated near the subtropical tropopause (and uncompensated by surface θ fluxes underneath). A useful discussion may be found in Gill (1982, sections 13.9 and 13.10.2). These PV fluxes are associated with Rossby 'wave breaking' phenomena of the kind exemplified by our Figs. 2(a) and 20(b) and discussed in section 6(d) (see also McIntyre and Palmer 1986 and references), the Rossby waves having propagated upward and then equatorward along the tropopause, before breaking. The idea that equatorward Rossby-wave propagation is involved in the maintenance of the surface westerlies appears to go back to an unpublished suggestion by Eady (Green 1970, p. 160).

There are, on the other hand, problems, including general-circulation problems, for which it may be more convenient to consider the vorticity and thermodynamical equations separately. For instance this is often true of planetary-scale motions (horizontal scale $\gg$ Rossby radius of deformation), for reasons which seem worth a few words of explanation.

As has been noticed in some general circulation studies (e.g. Lau and Holopainen 1984), the transient eddy heat flux contribution to the interior quasi-geostrophic PV flux has a less important effect on the planetary-scale flow than its magnitude in the lower troposphere might suggest, when considered in isolation. In this case there is considerable cancellation with the effect of the lower boundary θ flux. Essentially the same cancellation manifests itself, also, in the paper by Pfeffer (1987). The cancellation is only approximate, not being the consequence of any exact theorem, but is to be expected from a scale analysis of the quasi-geostrophic inversion problem (e.g. our Eqs. (43)–(47)). One can also anticipate it by thinking about the typical structures of air masses of tropical and polar origin, especially air masses of large horizontal extent. The larger the horizontal scale, the greater the scope for cancellation between the wind fields induced by the surface θ and upper-air IPV anomalies (because the Rossby height is greater), and the greater the tendency for the interior and surface anomalies to have opposite senses as a result of the air mass moving bodily from its place of origin (e.g. the tendency for cold near-surface air to be advected from polar latitudes at the same time as cyclonic upper air PV, mentioned on p. 910 of our paper). The implication is that if one is interested in extratropical, planetary-scale, quasi-barotropic flow, one should get a simpler picture by integrating vertically and dealing directly with vorticity in the traditional way.

On the point about the principle of superposition for IPV inversion, this does not depend at all on the absence of surface θ (or interior IPV) anomalies. Strict applicability of the superposition principle does, on the other hand, depend on assuming that all anomalies, whether surface θ or interior IPV anomalies, are sufficiently 'weak', in the sense expressed by quasi-geostrophic theory. This was discussed in our section 5. Some more recent calculations by Thorpe (1986) have illustrated the extent to which superposition actually holds for realistic mixtures of surface θ and interior IPV anomalies.

We did give some consideration to using the reciprocal of the (Rossby–Ertel) PV; it certainly makes some of the equations neater, e.g. our Eq. (74b). But we are not convinced that it is always

the most convenient choice (nor that the PV or its reciprocal is the potential energy in disguise).

We take this opportunity to correct a sign error in Eqs. (72) and (73): the right-hand sides should be prefixed with plus and not minus signs. We should also like to record that generalizations of the balance concept (and therefore, by implication, of the invertibility concept) have been elegantly discussed, for a spectrally truncated model, in an important paper by Lorenz (1980) which we overlooked. Lorenz raises the intriguing possibility that temporally evolving states might exist satisfying a balance condition which is not merely more accurate than geostrophy, but is exact. He calls this 'superbalance'. However, the most recent mathematical evidence (e.g. Vautard and Legras 1986, Lorenz 1986, and references) suggests that, in realistic parameter ranges, the balance and invertibility concepts are more likely to be inherently approximate, as our discussion assumed, even for spectrally truncated models like Lorenz's, let alone the real atmosphere with its scale interactions and its capacity to develop fronts, strong jet streams, Kelvin–Helmholtz instabilities, etc. The question of the ultimate accuracy of the invertibility principle is still very much an unanswered one. Other key references we overlooked include the paper by Truesdell (1951), pointed out to us by Dr R. Hide, and the paper by Thorpe and Emanuel (1985). These draw explicit attention, respectively, to the frictional part, and to the full version including diabatic effects, of the important integral constraint expressed by our Eq. (70b). Haynes and McIntyre (1987) discuss some additional results associated with that constraint.

REFERENCES

Gill, A. E.	1982	*Atmosphere—ocean dynamics.* Academic Press
Green, J. S. A.	1970	Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Q. J. R. Meteorol. Soc.,* **96**, 157–185
Haynes, P. H. and McIntyre, M. E.	1987	On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci.,* **44**, (in press)
Lau, N.-C. and Holopainen, E. O.	1984	Transient eddy forcing of the time-mean flow as identified by geopotential tendencies. *ibid.,* **41**, 313–328
Lorenz, E. N.	1980	Attractor sets and quasi-geostrophic equilibrium. *ibid.,* **37**, 1685–1699
	1986	On the existence of a slow manifold. *ibid.,* **43**, 1547–1557
McIntyre, M. E.	1970	On the nonseparable baroclinic parallel-flow instability problem. *J. Fluid Mech.,* **40**, 273–306
McIntyre, M. E. and Palmer, T. N.	1986	A note on the general concept of wave breaking for Rossby and gravity waves. *Pageoph,* **123**, No. 6 964–975
Pfeffer, R. L.	1987	Comparison of conventional and transformed Eulerian diagnostics in the troposphere. *Q. J. R. Meteorol. Soc.,* **113**, 237–254
Thorpe, A. J.	1986	Synoptic scale disturbances with circular symmetry. *Mon. Weather Rev.,* **114**, 1384–1389
Thorpe, A. J. and Emanuel, K. A.	1985	Frontogenesis in the presence of small stability to slantwise convection. *J. Atmos. Sci.,* **42**, 1809–1824
Truesdell, C.	1951	Proof that Ertel's vorticity theorem holds in average for any medium suffering no tangential acceleration on the boundary. *PAGEOPH,* **19**, 167–169
Vautard, R. and Legras, B.	1986	Invariant manifolds, quasi-geostrophy and initialization. *J. Atmos. Sci.,* **43**, 565–584