From M. Shapiro and S. Grønås (eds.), 1998: *The Life Cycles of Extratropical Cyclones.* Boston, Amer. Meteorol. Soc., preprint in press.

Numerical Weather Prediction: a Vision of the Future, Updated Still Further

Michael E. McIntyre
Centre for Atmospheric Science at the
Department of Applied Mathematics and Theoretical Physics,†
Silver Street, Cambridge CB3 9EW, UK.
M.E.McIntyre@damtp.cam.ac.uk; http://www.atmos-dynamics.damtp.cam.ac.uk/

1. Introduction

It is a great pleasure to be reminded of the Bergen Symposium, which I enjoyed so much, and to take the opportunity to update my own small, somewhat personal contribution (McIntyre 1994) culminating in a vision of the future of numerical weather prediction. One reason to update it and to reemphasize the points made at Bergen is that the most important, and perhaps the most controversial, of those points still seem to me likely to hold good at the future time envisaged — despite my being forced by today's politics to go a bit further into the future. More of that later; but let me just say for now that I shall go one decade further, twenty years into the Millennium.

The controversial points in question are the points about ergonomics. That is, they are the points about the potential for humans to interact usefully and efficiently with computer based systems, to whatever extent the latter become artificially intelligent in one sense or another — the potential for optimizing the person–machine synergies on which operational weather prediction and the responsible use of same will continue to depend (e.g., Tennekes 1988). It still seems to me, after sampling the most recent literature, that while seeking to exploit developments in artificial intelligence as fully as possible we should also be skeptical of the claims that tend to be made about such developments, most especially the claim that artificial intelligence will become, over the next several decades, absolutely superior to human intelligence (e.g., Warwick 1997). That claim seems to be based on a textbook model of the human brain (neuron–synapse networks as simple logic circuits) that is outdated today, and many orders of magnitude too simple, leading to gross underestimates of what would be involved in simulating or matching, let alone surpassing, human brain function in every respect. Indeed, hard evidence is now accumulating to confirm what the principles of molecular biology have long suggested, including what Jacques Monod (1971) called 'microscopic cybernetics', the arbitrary computational powers of protein molecules. Microscopic, more aptly nanometric, cybernetics suggests — and the new evidence is beginning to say it more directly — that textbook neurons and synapses represent only the tip of a hypermassively parallel computational iceberg of which we are very far indeed from possessing an adequate conceptual, let alone detailed computational, model.[1]

Of course electronic computers are much better than we are at certain things; and what is going to be done with various kinds of computers in the next few decades — electronic, photonic, and no doubt biomolecular — will surely be marvellous in itself. It would equally be a mistake to underestimate that. Indeed, my vision of the future assumes that computers and humans will together be doing some very marvellous things indeed.

So let me proceed as before. I believe that the points to be discussed are very fundamental, and I make no apology for repeating them. As I said at Bergen, they are fundamental from two quite different viewpoints. They are fundamental not only from a fluid dynamical viewpoint — and I shall mention some recent advances in understanding, for instance, Lighthill radiation, Hamiltonian balanced models, and potential vorticity inversion — but fundamental, also, from the viewpoint of how human perception works, and human cognition, especially as regards the visual–cognitive system. Notwithstanding what I just said about neurons, synapses, and protein molecules, we have considerable knowledge of at least the typical general properties of perception and cognition. That knowledge, which again points toward hypermassive parallelism and has far reaching human as well as technical importance, comes from general biological reasoning and from analogies with simpler biological systems whose molecular scale functioning is relatively well

[0] †The Centre for Atmospheric Science is a joint initiative of the Department of Chemistry and the Department of Applied Mathematics and Theoretical Physics. General enquiries, telephone (+44-1223-)336345 or web site http://www.atm.ch.cam.ac.uk/

[1] Keys to the burgeoning literature may be found at our URLs http://www.atmos-dynamics.damtp.cam.ac.uk/ and ftp://ftp.damtp.cam.ac.uk/pub/papers/mem/lucidity.ps

understood. It also comes from the wealth of perceptual phenomena, many of them counterintuitive yet easily observable but all pointing toward the same simple principles, and many of which have been studied in detail by psychologists, neurologists, and scientists working in related fields.[2]

For present purposes the implication of all this is still, it seems to me, an extraordinary potential for effective, efficient person–machine interaction in future weather prediction operations, and, one might add, in public liaison as well, including a significant contribution from human perceptual–cognitive skills that machines will not be able to match. It would be rash to predict exactly how the person–machine division and interface — the division of labor and responsibility — will be organized. But it is strongly arguable that, for the foreseeable future, there always will be some such division and that it will come much closer to optimal as the state of the art matures, a nontrivial process that is barely beginning. Any such progress will require not only that we continue to bring powerful computing and data handling technologies fully to bear on those processes that can be automated — including ensemble forecasting and four dimensional variational data assimilation, both of which now seem to be coming of age in an exciting way (see, e.g., within a burgeoning literature, Molteni et al. 1996, Palmer 1996, Thépaut et al. 1996) — but also that we continue to move toward exploiting fully, in particular, the potentialities of the human visual system, that most powerful of data interfaces between computers and humans. For clear biological reasons, connected with the survival of species, the human visual system has, for instance, a 'four dimensional intelligence' that even in terms of raw computing power still dwarfs, by many orders of magnitude, the power of today's largest electronic supercomputers.

And there must surely be great scope for exploiting what is perhaps the most remarkable human cognitive skill of all, forms of which we deploy every day with hardly a moment's thought. This is the ability, after appropriate learning, to interrelate intuitively, to simultaneously take account of, and to act swiftly on, very disparate kinds of information, if the information can be presented in an appropriate form and in a sufficiently familiar context. The successful development of this kind of cognitive skill is sometimes called 'getting a feel for the problem'. We all have some 'feel' for what happens, and why, if someone or something upsets a bucket of water on a table, despite the formidable complexity of the sequence of events when viewed in detail. We can make quick judgments of the consequences and their dependence on the configuration of neighboring objects, sometimes even quick enough to help with damage limitation. More specialized such skills, like flying aircraft safely, or playing tennis, with attention to a variety of sensory inputs, depend on other kinds of familiarity. These require more specialized learning on top of the range of childhood experiences. The potential for developing analogous skills on the human side of weather prediction is something that we surely tend to underestimate, if only because that potential, and the degree of skill attainable, and how to attain it, have hardly begun to be explored. This will need pioneering by talented individuals. Given the right background conditions — compare J. S. Bach's establishment of the equal tempered musical keyboard — there could yet be a Paganini, a Rachmaninov or a Charlie Parker of operational weather prediction or, if you prefer to switch metaphors, a Chuck Yeager or a Martina Navratilova.

[2] Keys to the literature, and some demonstrations of perceptual phenomena, may be found on the same web and ftp sites (footnote 1).

2. Research background

Several existing lines of research in the atmospheric sciences, some of them with a long history, have already given us some slight hint of the potential for person–machine efficiency. Much of this research has been oriented to synoptic and mesoscale tropospheric phenomena (e.g., within today's vast literature, Hibbard et al. 1989, Reed et al. 1992, 1994, Demirtas and Thorpe 1997, and many papers to the Bergen Symposium), and some of the underlying ideas go back all the way to the air mass concepts developed at Bergen and to the dynamical properties of potential vorticity discovered by Rossby (1936, 1940) and Ertel (1942). There has also been relevant research in, and crossfertilization from, adjacent fields. One such field is stratospheric meteorology and chemistry, where the research leading to our present understanding was stimulated first by the greater observational accessibility and dynamical simplicity of the stratosphere and second, more recently, by scientific and public concern about the ozone layer and its continuing depletion.

My own professional interest comes largely from trying to understand the fluid dynamics of the stratosphere, and dates back to attempts, in the early 1980s, to imagine what isentropic distributions of Rossby–Ertel potential vorticity (PV) might look like in the wintertime middle stratosphere and how they might help one to understand the preconditioning of the northern stratosphere for midwinter sudden warmings. This was followed by the intense excitement of being involved in seeing, and helping to make sense of, some of the earliest observational estimates of middle stratospheric PV distributions from satellite data. The results became an important landmark because the estimated PV distributions turned out to be of better quality than conventional wisdom said was possible with the satellite data. Although we were seeing only a "blurred view of reality" through "knobbly glass" (McIntyre and Palmer 1984), it was still a good enough view, when combined with theoretical insight — including insight from the set of idealized models labeled 'Rossby wave critical layer theory', and insight into what is now called 'PV invertibility' and its scale dependence — to catalyze important advances in our understanding of the real stratosphere both from a dynamical and, as it turned out, also from a chemical viewpoint.

Today, thanks to a wealth of new data on the stratosphere, the metaphorical glass has already become far less knobbly, as suggested by Figure 1a below. There is a wealth of new data, and clever new data analytic methods that exploit the kinematics of advection, vividly suggesting the potential for 'four dimensional intelligence' in data analysis and quality control. The research continues apace and I cannot do it justice here: a random and inadequate sample of the most closely relevant papers might include those by Schoeberl et al. (1992), Randel et al. (1993), Waters et al. (1993), Waugh (1993), Chen et al. (1994), Manney et al. (1994), Strahan et al. (1994), Sutton et al. (1994), Fisher and Lary (1995), Lahoz et al. (1996), Nakamura (1996), O'Sullivan et al. (1996), Lee et al. (1997), and Schoeberl et al. (1997), among many others. Not only cleverness with computing and data analysis, but also cleverness with remote sensing techniques and other observational techniques has been crucially important here.

I should mention also that, aside from the conceptual background of wave–mean interaction theory (going back to Lord Rayleigh), of which the Rossby wave critical layer theory is just a particular case, some of the dynamical ideas that adumbrated the relatively clear view we have today were discussed in a typically perceptive early contribution by Huw Davies (1981), on the nature of stratospheric warmings. More of the history is surveyed in my (1982) review and in a more recent review (1993a) written for an audience from the theoretical mechanics community, emphasizing how wave–mean and PV-related concepts have displaced what the admirable Doctor de Bono would call the 'intermediate impossible' or 'crazy ideas' stage of trying to understand atmospheric circulations, invoking things like 'negative viscosity' — now clearly understandable in terms of less crazy ideas, like Rossby wave propagation.

This stratospheric research, exciting though it has been for me personally, and continues to be, is only one of several lines of research that have reminded us of a key point about untapped potential in operational weather prediction. The point, which now seems to be widely appreciated — and indeed is coming close now to new forms of operational exploitation (A. Hollingsworth, personal communication 1997) — is that it makes good sense scientifically and data analytically to consider, either implicitly or explicitly, PV and chemical tracers together. Whether or not you find it insightful to think of the dynamics in terms of PV, the bottom line is that there is much more information in chemical and dynamical fields taken together than in either taken separately.

This point has cropped up again and again in atmospheric research, and has been another point of crossfertilization between our understanding of the troposphere and the stratosphere, including conditions near the tropopause. It emerges from many studies using data from remote sensors like the 6-micron water vapor imagers, and the famous Total Ozone Mapping Spectrometer, and goes back to much earlier work by pioneers like Starr and Neiburger (1940), Reed and Danielsen (1959), and Danielsen (1968), natural developments, in turn, of the Bergen air mass concepts and the advocacy of isentropic analysis by Napier Shaw (1930). Figures 1a–c, taken with kind permission from Appenzeller et al. (1996), beautifully illustrate the point with a three way snapshot of typical midlatitude, near tropopause phenomena. Figure 1a shows a hypothetical chemically inert tracer advected on the 320K isentropic surface over a 4-day period. The pattern was computed from analyzed winds by means of an accurate, sophisticated 'contour advection' technique, representing a significant 'technology transfer to the real world' from so called 'academic' theoretical fluid dynamical work (Dritschel 1989). This particular idea for technology transfer was a beautiful idea independently thought of and implemented by Norton (1994) on the one hand, and by Waugh and Plumb (1994) on the other. In Figure 1a (see caption for details), colors represent nominally stratospheric or ex-stratospheric air, and white tropospheric. Figures 1b,c show respectively a corresponding water vapor image and isentropic map of PV. The resemblance between the three is obvious at a glance, and tells us, I believe, something significant. The three maps show three quite different variables — yet human cognition, even without the help of animation, or forward radiation models, can instantly discern that the three pictures are all showing aspects of the same structure. Furthermore, we know that the resemblance is significant dynamically as well as morphologically, for the reasons to be recalled in sections 4–6.

In a recent effort to get an idea of the potential suggested by such cases, Demirtas and Thorpe (1997) have shown that other cases can be found that, by contrast, lack the kind of resemblance that is so conspicuous in Figures 1a–c, and that this can be symptomatic of highly significant analysis errors.

In the cases they present, a surprisingly crude and simple correction procedure turns out to be enough to improve some forecasts decisively. It was simply, in effect, to move PV contours — by implication, reshaping the ridge–trough structure of the tropopause — to give better qualitative agreement with a water vapor image as judged subjectively. Then PV inversion was used to generate the corresponding 'bogus data'. The subjective judgments involved were nontrivial. It seems plain that ozone can be used in a similar way; and there have even been first attempts, involving chemical modeling, at using information not only from relatively inert but also from relatively *reactive* chemicals, which are sensitive to solar zenith angle (Austin 1992, Fisher and Lary 1995). Much of this will no doubt be automated and made much more objective, reliable, and accurate, as soon as the state of the art of 4D variational assimilation becomes sufficiently developed. But my bet is that human vision will remain potentially very important for grasping how well any such assimilation is performing, for 'getting a feel for the problem' and hence insight into what to do or say if — or rather when — things go wrong, when the moisture field and all the other dynamical and chemical fields fail to fit together neatly in spacetime.

3. Cyclogenesis and Rossby wave propagation

There have been still other remarkable crossfertilizations between stratospheric and tropospheric research, yet again highlighting the dynamical relevance of isentropic distributions of PV and hence the dynamical information implicit in chemical tracer fields. To a dynamicist, of course, the stratosphere and troposphere are inseparable parts of the same system; and the extratropical tropopause, with its strong isentropic gradients of PV, is hardly a passive 'ceiling' but much more an erodable, highly deformable, dynamically active 'Rossby-elastic wall' (McIntyre and Palmer 1984, Holton et al. 1995), dynamically more like the edge of the stratospheric polar vortex. Its sideways 'Rossby-wave elasticity' obviously has an effect on downstream development, for instance, as well as on local development, once again bringing in the notions of advection and wave propagation simultaneously.

Indeed, to a dynamicist the phrase 'tropospheric' weather system sounds self contradictory. High PV stratospheric air is well known to be crucial to midlatitude cyclogenesis, as many papers in the Bergen Symposium reminded us. Nevertheless, it is mainly the lower stratosphere that has direct importance for cyclogenesis, the middle stratosphere being relatively remote from so called tropospheric weather and having stronger conceptual than direct relevance. Here, again speaking from my own experience, I am thinking especially of a line of research from which another — this time really surprising — insight emerged about tropospheric cyclogenesis (Edmon et al. 1980; see also Dunkerton et al. 1981, Hoskins et al. 1985, Held and Hoskins 1985, Thorncroft et al. 1993, Magnusdottir and Haynes 1996).

The surprise was that, contrary to what the textbooks still say, the major part of the growth of eddy kinetic energy seen in the typical Simmons–Hoskins baroclinic wave life cycle is due to upward Rossby wave propagation and Doppler shifting, not to baroclinic instability.

The upward Rossby wave propagation is dynamically indistinguishable from the Charney–Drazin type of propagation traditionally thought of as relevant, on a larger scale, to the wintertime middle stratosphere. Less than half of the growth of eddy kinetic energy in the life cycle could be attributed to baroclinic instability. The initial baroclinic instability saturates early in the cycle (e.g., Thorncroft et al. 1993, Figure 4); and the subsequent peaking of the eddy kinetic energy is associated with a more or less distinct second saturation event involving an extreme, irreversible deformation of the Rossby-elastic tropopause, and dynamically much the same as the breaking of planetary-scale Rossby waves in the wintertime middle stratosphere (McIntyre and Palmer 1983–85) and indeed significantly like the idealized form of breaking described by the Rossby wave critical layer theory. When the importance of upward Rossby wave propagation in the baroclinic life cycle was first discovered by Edmon et al., the surprise was total. I don't think anyone had the slightest inkling that stratospheric planetary-scale Rossby wave propagation had anything to teach us about such a highly nonlinear phenomenon as tropospheric cyclogenesis.

4. Why PV?

This question is still asked. I think it is prompted by the mathematical truism that one dependent variable is as good as another if they carry the same information. Let us, therefore, recall the fluid dynamical reasons why the PV calls for special attention, or, more precisely, summarize why it is useful to make a particular choice of dependent variables, to focus prognostic attention on near-surface distributions of potential temperature (PT) and interior isentropic distributions of PV — which latter, once upon a time, I rashly called 'IPV distributions'. The presentation follows that in McIntyre (1993b) and builds on lecture material I have used for many years. (Why was 'IPV distributions' rash? I have to take the blame here. I thought I had come up with a neat mnemonic acronym but failed to remember something else about human perception, the phenomenon of 'stray adjective' or 'disabled toilet syndrome'. This is a potentially confusing manifestation of our power of lateral thinking. The 'I' for ‘isentropic’ was meant to apply to ‘distributions’, ‘maps’, ‘gradients’, ‘fluxes’ etc., and not to ‘PV’.)

The important idea, going back at least as far as Charney (1948) and Kleinschmidt (1950–1), is to recognize that certain aspects of the PV–PT fields can be regarded as controlling, in a certain sense, the dynamical evolution. More precisely, there is an ‘invertibility principle’ — so familiar to theoreticians that it is hardly ever 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 inertia–gravity oscillations that may be present including equatorial Kelvin waves. This diagnostic process is what we now call PV inversion. A precise statement will be given in the next section. 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, since there is no escape from considering the advective nonlinearity somehow. As every meteorologist knows, advection is fundamental to practically all weather developments, whatever else is going on.

(b) The PV–PT viewpoint recognizes, makes explicit, and keeps conceptually separate those aspects of the dynamics that are intrinsically nonlocal. The nonlocal or action-at-a-distance aspects 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) nondivergent barotropic vortex dynamics.[3] Barotropic dynamics is included as a particular case characterized by a particular inversion operator, a simple inverse Laplacian.

Thus the PV–PT viewpoint shows, for instance, why certain classical aerodynamical phenomena such as 2D shear instability, vorticity shedding, vortex rollup, merging, and vortex core isolation all seem to have baroclinic, ‘layerwise-2D’ counterparts in synoptic and mesoscale atmosphere–ocean dynamics, even down to the vortex streets behind oceanic islands and, of more concern to weather forecasters, the larger scale shedding of PV anomalies from, for instance, the European Alps. According to the recent work of Aebischer and Schär (1997), this last contributes significantly to Alpine lee cyclogenesis. Again, vortex core isolation, as in smoke rings, is a central theme in today’s stratospheric research because of its importance to ozone hole chemistry (references cited in section 2 above). Vortex core isolation is related to Rossby-wave elasticity — another phenomenon most easily understood from the PV–PT viewpoint — as well as to the horizontal shear just outside the vortex edge (Juckes and McIntyre 1987). It is probably important also, I suspect, to tropical cyclone maintenance (McIntyre 1993b) because it can protect the moist updraft from drier incursions, keeping the tropical cyclone thermodynamically efficient. Whether anything similar happens in extratropical explosive marine cyclogenesis is an open question, but an interesting one.

Point (a) is of central importance 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. Then features in the PV–PT fields often become sharp edged and frontlike 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, as in so called ‘chaotic advection’ — the advected quantities rapidly developing spatial scales far smaller than those of the advecting wind fields, and doing so, in most cases, exponentially rapidly. Figure 1a well illustrates this ubiquitously important process, all the fine scales and steep gradients shown there having developed advectively over just 4 days. The figure also reminds us of one of the forecaster’s nightmares: a high PV air mass drawn out into a long sheet or filament, and eluding operational analyses, can sometimes, if large scale deformation rates slacken off, undergo vortex rollup and turn into a compact cyclone like, for instance, the cyclone over the Balkan Peninsula near the right hand edge of the figure.

Synopticians have long been familiar, of course, with the zoology of all the various advectively produced structures, including jets, shear lines, tropical upper tropospheric troughs, and so on. The PV–PT viewpoint when linked to the classical insights about advection and deformation, including those of the Bergen school about surface frontogenesis, provide by far the simplest explanation of why such structures occur so commonly. In a nutshell, one has the steepening of low-altitude temperature gradients and of high-altitude PV gradients.

The PV–PT viewpoint enables one to regard advection as a quasi-horizontal, quasi-2D process. This is another powerful conceptual simplification — and, incidentally, not at all dependent on the familiar but crude approximations of quasi-geostrophic theory, a point that will be underlined by Figure 2 below. Advection for this purpose can be considered quasi-2D despite its really be-

[3] See also the supplement to HMR (Quart. J. Roy. Meteorol. Soc. **113**, 402), regarding ‘equivalent barotropic’ dynamics.

ing 3D. It is quasi-horizontal and quasi-2D in the sense of referring to horizontally projected motion along isentropic, that is to say constant PT, surfaces, and to horizontally projected motion near the Earth's surface — meaning, in practice, just above the boundary layer (Hoskins et al. 1985, hereafter HMR) — with no need to refer explicitly to vertical motion despite the latter's dynamical importance in many cases.

When material PV and PT conservation fails, as happens for instance in moist convection, the evolution of the PV–PT field can still be described as resulting from dynamically local effects only: advective, diabatic, frictional, and so on. Moreover — and very surprisingly — the quasi-two-dimensionality persists, in a certain sense, even when diabatic heating, for instance latent heating, is taking air parcels across isentropic surfaces. To see this, one need only adopt a viewpoint that is arguably a natural consequence of taking seriously the analogy between PV and chemicals (section 5, item **7**).

The resulting view of the dynamical evolution is fundamentally simpler than any view that refers directly to the primitive equations. The primitive equations intimately combine the local frictional and diabatic effects with *three dimensional* advection and dynamically *nonlocal* interactions. The nonlocal interactions are mediated by the pressure and buoyancy fields, constrained by the requirements of mass conservation, hydrostatic balance, and other forms of balance (see section 6 below). 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, balanced and unbalanced, local and nonlocal, prognostic and diagnostic.

The importance of dynamically nonlocal interactions, incidentally, is one of the reasons 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 totally misleading impression of causal linkages and the workings of the dynamics (e.g., Holton et al. 1995).

5. PV and its invertibility

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

$$P = (f + \hat{\mathbf{z}} \cdot \boldsymbol{\nabla}_\theta \times \mathbf{V})\,(-g\,\partial\theta/\partial p) \qquad (1)$$

is essentially that proposed by Rossby in 1940 as a natural generalization of the single layer 'shallow water' PV he had proposed in 1936 (Rossby 1936, 1940). In the expression on the right, f is the Coriolis parameter, $\hat{\mathbf{z}}$ is a unit vertical vector, $\mathbf{V}$ is the horizontal wind vector, θ is the PT or any function of PT alone, such as specific entropy, p is pressure altitude, and g is the gravity acceleration. The subscript θ indicates differentiation along an isentropic surface, so that although the quantity $\hat{\mathbf{z}} \cdot \boldsymbol{\nabla}_\theta \times \mathbf{V}$ looks like relative vorticity, it is quite different, in principle, from the ordinary relative vorticity. It is more appropriately called Rossby's 'isentropic vorticity' to keep the distinction clear, as explained in HMR. As is well known, Ertel's (1942) more general formula reduces to (1) in the hydrostatic case.

A convenient SI unit for the PV is 10^{-6} m^2 s^{-1}K kg^{-1}; it could perhaps be called the 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, with the convention of using the PT as the materially conserved thermodynamic variable, θ in the formula (1), PV values $P < 1$ usually imply that we are looking at tropospheric air, and $P > 2$ extratropical stratospheric air. One of the exceptions to this is the case of a mature tropical cyclone, where lower tropospheric PV values well over 10 PVU may be encountered, in dramatic contrast with typical surrounding values, of order 10^{-1} PVU. Extratropical marine cyclogenesis is another exception, again with substantial contrasts in lower tropospheric PV values (e.g., Reed et al. 1992, Davis et al. 1993).

One way of stating the invertibility principle is that proposed in 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 the PV on each isentropic surface, and of the PT at the lower boundary, one can deduce, diagnostically, all the other dynamical fields such as winds, temperatures, geopotential heights, local static stabilities, and the adiabatic part of the vertical motion, to the extent that, and to the accuracy with which, the motion can be regarded as balanced (again see section 6 below).

The diagnostic nature of PV inversion means, of course, that the prognostic aspects of the problem have been entirely confined to describing the advective and nonadvective evolution of the PV and PT. This is part of the conceptual separation referred to in section 4. The following additional points can be made:

1. Just as in the simpler case the inverse Laplacian, in order to carry out an inversion we must solve the diagnostic problem globally, with proper attention to boundary conditions. This is an inescapable consequence of 'nonlocalness'. The inversion problem is an elliptic problem partially analogous to membrane deformation problems or problems in electrostatics (as discussed in HMR and in my more recent reviews), with the geopotential qualitatively like the electrostatic potential and the wind vector approximately at right angles to the electrostatic field.

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 and other phenomena involving balanced motion. In the electrostatic analogy, the interior PV and the surface PT are on the same footing: they have comparable roles, and comparable status as, so to speak, free electric charges subject to advection.

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

4. The principle works only for 'dry inversion', e.g., when the PV and PT fields refer to the dry and not to the moist equivalent PT, say θ_e. The latter, and its associated moist PV, say P_e, are more useful for the purposes of parameterizing upright and slantwise moist convection, for instance by setting $P_\mathrm{e} = 0$ in the eye wall of a tropical cyclone (e.g., Emanuel 1991) or in the updraft of an intense extratropical marine cyclone. The P_e and θ_e fields are not useful for inversion, for mathematical reasons (lack of ellipticity).

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 — an outlook already being vindicated, as it turns out, at least for shallow water models as illustrated in Figure 2 below.

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', was given in HMR (see also the Appendix of HMR); it shows how upper air PV advection can induce large scale, adiabatic upward motion just ahead of it, in practice often triggering moist convection in the manner familiar to synopticians and forecasters studying what is called Pettersen type B cyclogenesis.

7. PV distributions and their possible evolution, and the associated transports and budgets of PV — more precisely, of the transportable 'charge' or signed quasi-chemical 'substance', PVS say, whose mixing ratio or 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. They are simple, almost trivial, to prove by standard vector calculus (e.g., Haynes and McIntyre 1990), but have nontrivial significance. Specifying a PV distribution that violated the implied constraints would presumably lead to failure of any attempt at inversion; see also the caveat in HMR section 3, equations (17b)ff. The first theorem says that PVS, considered as an additive, conservable, transportable quasi-charge or quasi-substance, is *indestructible*, like real 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 theorem says that PV distributions behave as if isentropic surfaces were *impermeable* to the 'charged particles' of PVS, even in the presence of diabatic heating. The mathematics permits us to think of an isentropic surface in a stably stratified atmosphere as acting like a *semipermeable membrane*, allowing mass to cross it but not the notional particles of PVS.

It is this 'impermeability' or 'semipermeability', then, that allows the quasi-2D representation of advection to be retained, very surprisingly, even when mass and real chemical substances are crossing isentropic surfaces diabatically. The PVS particles are to be imagined as moving strictly on the isentropic surfaces, regardless of how those surfaces themselves move. In this picture, values of the PV, which have the nature of chemical mixing ratios, can of course change; but they can change only through the PVS being transported, diluted, or concentrated in various ways.[4] Both the 'indestructibility theorem' and the 'impermeability theorem' are straightforward consequences of the way in

[4]At this point I should remind the reader that there has been an unfortunate and dangerous confusion of terminology in the literature. Words like 'source' have been used in different senses for PV and chemicals — even, sometimes, in the same discussion with PV and chemicals considered together. If the mixing ratio of an inert chemical like helium were to change through dilution, I don't think we would want to speak of a 'sink' of helium. Yet the parallel situation for PVS is often spoken of in terms of a 'sink for PV'. Further discussions of the various incompatible meanings of words like 'source', 'sink' — and 'transport' as well — are given in my 1992 review, section 11, and under STRAY ADJECTIVE in the file lucidity-supplem.tex on the web site already noted. (So powerful can be the straying of adjectives that, very confusingly, through a kind of linguistic–cognitive domino effect, the word 'substance' has become, in the present context, almost a synonym for 'per unit volume'. More precisely, you will find the term 'PV substance' used to mean the PV times the mass density ρ, even though 'water substance', for instance, is never used to mean the corresponding thing, water mixing ratio times ρ. The domino effect, well illustrating the danger of dismissing all such problems as 'mere semantics', is traced in the STRAY ADJECTIVE section of the abovementioned file lucidity-supplem.tex.)

which the PV is constructed mathematically, equation (1) above and its nonhydrostatic generalization due to Ertel. This has the significant further consequence that the theorems apply not only to the exact Rossby–Ertel PV that could be constructed from exact wind and PT fields if one knew them (which is never the case in practice) but also, exactly, to any 'coarse grain PV' constructed from observational datasets for the wind and PT fields (Haynes and McIntyre 1990, Keyser and Rotunno 1990).

In the interesting example of explosive marine cyclogenesis, the highly concentrated low altitude PV anomalies that are characteristic of the phenomenon are 'concentrated' in a literal sense, if one describes what goes on in terms of PVS. According to that description, one can picture such PV anomalies as having resulted from a strongly convergent inflow of PVS particles trapped on dry-isentropic surfaces and hence, inevitably, concentrated near the convergence maximum. The PVS is concentrated there in just the way a real chemical can't be (because the molecules of the real chemical aren't, of course, trapped on the dry isentropes but are carried across them in the cyclone's upward vertical motion).

6. Limitations on balance and PV invertibility

Over the past decade or so there has been interesting work on PV inversion at accuracies comparable to that of the gradient wind and Bolin–Charney balance approximations, a level of accuracy that is often, in practice, comparable to data errors and data assimilation errors (e.g., Thorpe 1985, Raymond 1992, Davis 1992, Davis et al. 1993). I shall not go into this except to express my continuing interest and to say that I believe that the development of robust and efficient inversion algorithms having this kind of accuracy will sooner or later be seen to be of the greatest practical importance — not, of course, as a means of timestepping the numerical weather models themselves, which are most efficiently based on the primitive equations, but rather as a means of generating 'bogus data' in the manner of Demirtas and Thorpe (1997), and generally for 'getting a feel for the problem' in research mode. Because inversion poses a nonlinear elliptic problem, massively parallel computation should help future developments here.

I would like to say just a little, however, about some recent and theoretically important advances in our understanding of PV invertibility and its *ultimate* limitations — the limitations on how accurate it can be in principle, the limitations beyond which we cannot go no matter how clever we are computationally. These limitations are, of course, bound up with the ultimate limitations on the concepts of 'balance' and 'slow manifold'.

Some of these limitations have become clearer from studies of a phenomenon characteristic of the deep tropics, asymmetric inertial instability (Ciesielski et al. 1989, O'Sullivan and Hitchman 1992, Dunkerton 1993, Clark and Haynes 1994). If we add moist processes (Bennetts and Hoskins 1979) we similarly encounter what might have to be called 'conditional *asymmetric symmetric* instability' (just as we have to speak nowadays of the '*variable* solar *constant*'). Perhaps it should be 'asymmetric slantwise conditional instability'. In any case, its importance in intense extratropical as well as tropical cyclogenesis seems to be in little doubt, and its velocity fields are, of course, another kind of unbalanced motion inaccessible to PV inversion as usually understood.

When and where such dry or moist instabilities are unimportant, there remains another, entirely different, more subtle, and ever present limitation on PV invertibility, which has no necessary connection with any instability and which might be summarized in the phrase 'fuzziness of the slow quasi-manifold'. It manifests itself in various subtle ways, including

(a) the characteristic 'schizophrenia' or 'velocity splitting' of filtered equation models, (usually in the form of two coexisting, unequal velocity fields, there often being no such thing as 'the' velocity field of the model; this has recently been shown to be generic to a large and important class of Hamiltonian balanced models, whether constructed by filtering or other means (Allen and Holm 1996, McIntyre and Roulstone 1997)),

(b) the mathematical phenomenon called Poincaré's 'homoclinic tangle', and

(c) the physical phenomenon described by the Lighthill theory of aerodynamic sound emission generalized to the rotating frame.

This last physical phenomenon might be described as the weak but nonvanishing 'spontaneous-adjustment emission' of inertia–gravity waves by unsteady, 2D or layerwise-2D vortical motion. The wave emission mechanism is sometimes referred to, illogically, as 'geostrophic' adjustment even though — as a consideration of circular vortices immediately suggests — it need not take the system toward geostrophic balance. (This has long been a source of confusion, I suspect, about the significance of 'ageostrophic' winds.)

The phrase *spontaneous* adjustment is also intended to distinguish the phenomenon in question from the quite different phenomenon of Rossby or initial condition adjustment. The latter is simply due to the initial conditions being out of balance. The essential character of spontaneous-adjustment emission is its remarkable weakness. It is often far weaker, in circumstances of interest, than you would deduce from any simple order-of-magnitude or scaling analysis, from the values of Froude and Rossby numbers and so on. It is this remarkable weakness that gives rise to a correspondingly remarkable accuracy in PV inversion, or, rather, possible accuracy, a possibility that can be realized if one is clever enough computationally.

It was James Lighthill's great contribution, in the simpler but fundamentally similar aeroacoustic or sound-emission problem (Lighthill 1952), to show why the emission is often far weaker than you would think from a standard order-of-magnitude or scaling analysis. He achieved the required conceptual clarification despite having no clues from numerical modeling, and despite the dauntingly nonlinear yet delicate nature of the problem. The Lighthill theory tells us, in effect, that spontaneous-adjustment emission is often weak because destructive interference among the emitted waves is often important. This is not something that you can notice from scaling analysis. Recent work by my former students Dr Warwick Norton and Dr Rupert Ford, and by other colleagues including Drs Jim McWilliams and Lorenzo Polvani, has added to our detailed knowledge of how all this works in a rotating system and, in particular, has given us some detailed quantitative checks on the Lighthill theory and its range of validity, which for order-of-magnitude purposes turns out to be astonishingly wide (Ford 1994, Ford et al. 1997).

In a rotating system, Coriolis effects tend to enhance still further the possibilities for accurate balance and PV invertibility, because they tend to weaken still further the spontaneous-adjustment emission. One can say, following Lighthill, that in the acoustic or pure gravity wave problem one has a mismatch of spatial scales at small Mach or Froude number. With Coriolis effects, one has in addition a frequency mismatch at small Rossby number. Both things, in combination with the destructive interference, weaken the spontaneous-adjustment emission. The weakening due to the frequency mismatch at small Rossby number was already observed and noted in numerical experiments by Errico (1982).

Even more remarkably, when one goes to larger Froude and Rossby numbers the spontaneous-adjustment emission tends to remain weak by any practical criterion, and *weaker* than the Lighthill theory might appear to predict without detailed consideration of the vortical motion (Ford 1994). The reason is that the unsteadiness of the vortical motion, necessary to excite spontaneous adjustment, tends to become weaker as a result of the same parameter changes. This in turn is because inversion operators have a *short range* character under these conditions, or 'small deformation radius', related to the slow propagation speeds of the inertia–gravity waves that mediate the apparent action-at-a-distance (McIntyre 1992, p. 345).

If the spontaneous-adjustment emission were zero, then we could have a true slow manifold within phase space, in the strict mathematical sense of the term 'manifold'. States on this (invariant) manifold would correspond to what Edward Lorenz once called 'superbalance', implying exact and unique PV invertibility and exactly zero inertia–gravity wave activity, given suitable initialization. The abovementioned and other evidence is overwhelming, however, in favor of there being no such thing as a strict slow manifold, as originally argued via another approach by Warn (1997, q.v.) and by Warn and Ménard (1986). Spontaneous-adjustment emission turns what might otherwise be a manifold into a much fuzzier mathematical object in phase space. That object almost certainly has the character of a thin chaotic or 'stochastic' layer, hence the term 'slow quasi-manifold'. This means that PV inversion itself must have a corresponding fuzziness, or inherent approximateness, representing an ultimate limitation on its possible accuracy.

Though rigorous proof seems beyond present day mathematical resources, for realistic fluid models with their infinite dimensional phase spaces, it is worth recalling that fuzzy mathematical objects of the kind in question have long been known to dynamical systems experts studying low order Hamiltonian systems like perturbed pendulums and other weakly coupled oscillators. Experts on these topics, who have studied multitudes of cases and developed some general theoretical perspectives, assure me that stochastic layers seem to be generic — that is to say typical of nearly all such systems — as expected intuitively from Liouville's theorem and considerations of the infinite time available to disturb what is called a homoclinic orbit. This is the relevance of Poincaré's 'homoclinic tangle'; for some relevant recent work see Bokhove and Shepherd (1996 & refs.). But the good news is that, under the usual parameter conditions, the atmosphere's slow quasi-manifold, though fuzzy, appears likely to be thin enough to be, for practical purposes, almost a manifold, at least the parts of it corresponding to motions whose Froude number is not too large (vertical scale not too small). The upshot is that PV inversion, despite being inherently approximate and inherently nonunique, can nevertheless, with enough computational ingenuity, be made astonishingly accurate — again provided that the vertical scale is not too small — and certainly far more accurate than any simple filtered theory or scaling analysis could ever suggest.

Figure 2 presents an example, taken from McIntyre and Norton (1990), showing direct evidence for this astonishing accuracy, far greater than any of the standard filtered theories would suggest and, by implication, confirming the weakness of spontaneous-adjustment emission in at least some circumstances. Like most of the fundamental theoretical work to date, the computations have been done, so far, only for the simplest dynamical system for which the balance and inversion concepts are nontrivial, namely, a shallow water system. This example is from the PhD thesis work of my former student Dr Warwick Norton (1988), full details of which we are hoping to publish soon in a long delayed joint paper (McIntyre and Norton 1997). The computation was done with a hemispherical pseudospectral shallow water model at fairly high resolution, triangular truncation T106. A carefully initialized primitive equation integration produced the geopotential height, wind, and divergence fields shown in the top two panels, representing a complicated, unsteady vortical motion somewhat like an exaggerated 300-hPa flow in the real atmosphere, with a strongly meandering midlatitude jet. There is a gigantic blocking anticyclone, and three major troughs. Froude and Rossby numbers reach values of the order of 0.5 in parts of the jet; Rossby numbers are of course infinite at the equator. Such values can hardly be thought of as small, and we may expect that the standard filtered theories of balanced motion will not be accurate. Nevertheless, in this and a range of similar cases, it proved possible to reconstruct the height, wind, and divergence fields by PV inversion — i.e., knowing only the mass of the fluid layer and the PV distribution, to reconstruct all the other fields from that information alone — and to do so with astonishing accuracy. Here of course the relevant PV is the shallow water PV of Rossby (1936), namely absolute vorticity divided by layer depth. In this example, the PV distribution is shown by contours and grayscale in the middle two panels; and the reconstruction of the other fields from it, the output of an accurate PV inversion algorithm, is shown in the bottom two panels of Figure 2. In comparing the original primitive equation fields in the top two panels to the reconstructions in the bottom two panels, you have to look carefully to see the differences: by implication, the invertibility principle is at least this good, in this particular example.

To get results with this sort of accuracy, a rather elaborate inversion algorithm had to be used; the equations are formidable looking and take about half a page to write out, and are non-trivial to solve numerically (Norton 1988). The computations are like a high order extension of nonlinear initialization in the manner of Hinkelmann (1969). In other words, we are dealing with a fairly complicated nonlinear elliptic problem, about which little is known in any mathematically rigorous sense; however, a powerful check on the numerical procedures, and on the accuracy of the invertibility principle itself, came from tests of cumulative accuracy in 10 day experiments (not shown here) running the primitive equations in parallel with the balanced model defined by advecting the PV and using the inversion algorithm. These 10 day experiments again showed astonishing accuracy. Similar results, indeed somewhat more accurate still, as it turned out, were obtained using the nonlinear normal mode approach. All these results are in stark contrast with those that would be obtained using standard quasigeostrophic inversion, such inversion being hopelessly *inaccurate* under the parameter conditions of Figure 2. In the case of Figure 2, a quasigeostrophic inversion (not shown) produced wind speeds in error by factors of order 2 even well away from the equator.

7. Implications for weather prediction

It has been said that "there is nothing so practical as a good theory." One could say that the theoretical developments I have been discussing underline the fact that the dynamics of weather developments has as strongly advective a character as does the moist and dry thermodynamics. One has moist advection, warm advection, and PV advection. This does not capture everything about real weather developments, but it surely captures an important part of them. To say that the dynamics has a strongly advective character is perfectly compatible, incidentally, with classic perceptions about the role of Rossby wave propagation, as I have already remarked in connection with downstream and local development and the 'Rossby-elastic tropopause'. The Rossby wave propagation mechanism is, indeed, partly advective — one could say 'sideways advective' — as explained for instance in the reviews cited, including HMR section 6. Ordinary textbook plane polarized waves on stretched strings are 'sideways advective' in a partly analogous sense: the material of the string is carried sideways, back and forth, whenever we excite any wave motion. Of course, to complete the intuitive picture for Rossby waves, and to include for instance such things as group velocity (HMR section 6c), one must obviously, as always, consider advection and PV inversion together.

If we now bring this view of the dynamics and thermodynamics together with an appreciation of the nature of human perception, especially human vision — the '4D intelligence' I have referred to — then the practical implications for weather prediction become conspicuous. What follows is a slight update of a discussion (1988) that I published some years ago in the Royal Meteorological Society's popular magazine 'Weather', in response to an article by Tennekes (1988), which also appeared in *Weather* and which I read with great interest.

Professor Tennekes gave what I saw as a thoughtful, provocative and timely warning of what may be at stake; and I wanted to add something to his discussion of that vexed and critical question, the future role of the human forecaster. My remarks were based on earlier lectures I had given to the ECMWF Seminar of 1987. I suspect that the same remarks — Tennekes' remarks as well as mine — are still worth making, and that the issues are going to stay alive for quite a few years to come.

Professor Tennekes suggested that the automation of analysis and forecasting has been taking us into a situation where already, for some purposes, "the added value of human skills has become marginal," and that commercial pressures, and commercial thinking, will lead to pressure to "bypass the forecasters altogether." Are we, he asks, "going to leave it to private enterprise to redefine the added value of human skills?" Against this, Professor Tennekes pleaded that "only human professionals can bear the responsibility for life and property in emergencies."

I agree; and I believe that the case can be further strengthened. The peculiar way in which the atmosphere works argues strongly in itself, as I have been hinting, for an enhanced and not a diminished role, sooner or later, for human professional skills.

Consider the most basic aspect of the forecaster's responsibility, the quality control, and the quality *assessment*, of the analysis–forecast process on which everything else depends. In view of the massive data throughput and the requirement to "work at lightning speed," as Tennekes put it, the tasks involved are daunting. Clearly they do indeed call for as much automation as possible, wherever appropriate. However, there are some things that humans can do faster, and better, than any existing or foreseeable machine. It may be useful to step sideways and look at these first in another context.

A motorist driving in city traffic is continually keeping track of, and extrapolating the positions of, a variety of independently moving objects most of which are rapidly changing their apparent size and shape, some in very complicated ways. Some of the objects intermittently obscure others from the field of view, and dirt on the windscreen may impose further complications. A skilled driver seeing a pedestrian about to emerge from behind a parked vehicle some way down the street may 'instinctively' begin to slacken speed within a few tenths of a second. It is possible that only the pedestrian's feet were visible at first, their characteristic motion showing beneath the parked vehicle. "So what?" you may say. "These are familiar facts, but what is the relevance to weather prediction?" Be patient a moment longer.

In performing these data processing tasks that we blithely call 'instinctive' or 'intuitive', the human brain and its visual peripherals are performing almost unimaginably prodigious computational feats, involving massive data throughput and much more, all taking place at more than "lightning speed" by the standards of any electronic computer. Furthermore, the data in my example, carried by the light entering the driver's eyes, on which safe driving depends, are gappy, noisy data. These are among the reasons why, for instance, computers don't yet drive taxis. The fact that we are not conscious of all this computational activity does not mean that it is not taking place. It may even be quite literally unimaginable, in any complete and detailed sense. As Marvin Minsky has emphasized, an ability to grasp intuitively how one's own brain works is biologically speaking an unlikely sort of mental ability. We quite naturally tend to underestimate what is involved. In watching young video games addicts at play, we tend to be impressed by the fast moving action on the colour displays, though what is truly remarkable is the human, not the electronic, side of the person–machine interface.

The actual computing power that the human brain deploys on visual data processing and interpretation is hardly a well known quantity. But I well recall hearing in 1979 a fascinating radio interview in which John Maddox talked to the late David Marr, then an outstanding young thinker among the new generation of researchers on vision and artificial intelligence. Marr made the point that even the earliest (and evolutionarily most ancient) computational processes in the human visual system — the sort of thing we have in common with frogs and other creatures we call primitive — already appear to need several orders of magnitude more computing power than a "fast general purpose computer". When one considers the vastly more complicated functioning of the entire visual cortex, there seems little doubt that the total equivalent computing power must be staggering. The advantage of three dimensional miniaturization down to cellular and molecular scales overwhelms the disadvantage of what might loosely be called a slower 'cycle time'. There is also the robust and highly developed 'programming' — here the distinction between 'software' and 'hardware' would be pushing the analogy too far — the cumulative result of hundreds of millions of years of natural selection.

It is dangerous to guess what human technology might look like even a decade or two from now, but my own guess is that the human eye–brain system will continue to make even our biggest supercomputers, our most sophisticated artificial-intelligence software, and our most massive experiments on artificial neural networks look feeble for some decades to come, when it comes to processing, interpreting, and interacting at high speed with appropriate kinds of visual data, and doing so with what we call 'understanding', or cognitive grasp. On the other hand, independent commercial pressures to develop electronic computers to do what they are good at, including animated graphics all the way through to what is charmingly called 'virtual reality' (as with 'asymmetric symmetric instability' and the 'variable solar constant') will produce mindboggling possibilities for the person–machine interface.

You will have no difficulty in seeing the point I am coming to. Provided that the interface with the forecaster is appropriately designed, the prodigious computing power — and cognitive–perceptual power — of the human brain and its peripherals could be harnessed far more efficiently to some very considerable electronic or photonic computing power and through this to the quality control and quality assessment of weather forecasts under operational pressures — just as it is already harnessed to other high speed, 'four dimensional' problems under operational pressures such as playing world-class football, flying military aircraft, or for that matter driving in heavy traffic. It is in such situations, involving the visual detection of coherent motion, that the eye–brain's '4D intelligence' is most conspicuously powerful. One of the computational tasks to which our visual systems are most exquisitely adapted is that of following moving, coherent features, even when the features continually change their shape and intensity, and even in the presence of noise, and of data gaps — as my traffic example was meant to suggest. This prompts the question, why should we not use, among other things, this '4D intelligence' in weather prediction?

In order to exploit the implied potential to the full, it will be crucial to present information to the duty forecaster's eye–brain system in an appropriate form. To use '4D intelligence' to the full we need to use animation, and we need to find, if possible, ways of presenting the evolution of the atmosphere in terms of moving, visibly coherent features. It is an extraordinary piece of good fortune, therefore, that so much of the dynamical and thermodynamical behavior of the atmosphere can be described in terms of quantities that often behave advectively. And when they do not behave advectively, as with the low altitude PV changes involved in explosive marine cyclogenesis or tropical cyclogenesis, then this fact calls for the forecaster's attention, the more so since it means that the numerical prediction model is invoking so called 'physics' parametrizations that may be less soundly based than the model's dynamics. We therefore need animated displays that make it instantly obvious whether the behavior is advective or not. Such displays are possible: if a moving vehicle hits a brick wall, or suddenly doubles in size, or suddenly appears from thin air, so to speak, it is a very visible thing.

One of the promising graphical choices for the near future seems to be the tropopause 'isostrophic PT maps' used at Reading, of which Thorncroft et al. (1993) give examples both from the real atmosphere and from life cycle experiments. These are maps of PT on the 2 PVU surface, approximately marking the extratropical tropopause and reflecting its character as an 'erodable, highly deformable, dynamically active 'Rossby-elastic wall'. The tropopause could be marked even more accurately if we were to redefine the PV using some suitable function of PT in place of the PT itself, θ in the Rossby–Ertel formula (1); it is merely a lucky accident that the conventional definition of the PV, using θ rather than an alternative thermodynamic function such as specific entropy, or some other function of θ, happens to be a good approximate extratropical tropopause marker. A single tropopause map has the advantage, with present day graphical methods, of summarizing most of the upper air information that would otherwise require a whole stack of isentropic maps of PV.

It is possible of course that, some years from now, the 'virtual reality' techniques will get so clever that it becomes perfectly easy to look at a whole stack of maps anyway, or to fly oneself very quickly, supermanlike, around some kind of 3D representation of significant features such as the tropopause and other regions of steep isentropic gradients of PV. Whatever visualization methods turn out to be best, it seems to me that experience with the typical patterns of behavior of such features, once they are made highly visible, cannot fail to lead to sharpened insights into the problems of data analysis and quality control; and as the automated system itself becomes more consistently 'four dimensional' and thence 'advectively intelligent' there will be an increasing potential, as hinted at by Figure 1a, for improvement in the accuracy and clarity of any graphical representation of the PV–PT fields, greatly enhancing their visual intelligibility as well as the accuracy of the numerical forecasts that use them.

8. A vision, or dream, of the future

I should like to conclude, then, by sharing, in a lighthearted vein, the vision — or perhaps I should not risk sounding too pretentious, and just say *dream*, or just plain *fantasy* — about future weather prediction operations. I hope it will be taken in the spirit of the real visionaries, the great pioneers Vilhelm Bjerknes and Lewis Fry Richardson. The original dream was something that Dr Glenn Shutts and I indulged in during a brief conversation at the UK Meteorological Office, probably sometime in 1986 or 1987. Today's politico-economic conditions make the whole thing even more dreamlike, in some ways, than it did in the 1980s; and, as I said earlier, I am going to push the time of the dream a further decade into the future. But, wearing my optimist's hat, I shall also dare to make it part of a larger dream in which, through practical necessity, the chaotic pendulum of politics and economics has swung toward a new attractor basin representing a degree of sanity for the longer term, including saner uses of our new 'psychological nuclear energy', the power of market forces (von Weizsäcker et al. 1997). Though there is hardly room for complacency I actually think that, despite everything, there is a tenable optimist's theory of the future, meaning the future of many things besides weather prediction.

The year is 2020. The international demand for good one week deterministic–stochastic weather forecasts has led to the establishment of a greatly enhanced global observing network, and to an unprecedented concentration of electrophotonic and biomolecular computing power and human skills at the great new International Centers for weather and climate prediction. The observing network includes not only sophisticated remote sensors but also a large fleet of ultra-light, ultra-cheap, remotely controlled miniature aircraft or 'aerosondes' that can be flexibly deployed in forecast-sensitive regions as necessary, including for instance the centers of tropical cyclones.[5]

An added incentive has been the recent acceleration in the excess greenhouse warming, exacerbated by a slight increase in solar irradiance. These trends were accompanied by an increase in the frequency and severity of moisture related weather phenomena, not just tropical cyclones of increasing power but also destructive storms in higher latitudes, due to maritime explosive cyclogenesis — the latter one of the fastest evolving forms of hazardous weather and one of the trickiest challenges to the weather forecaster. The trends, moreover, had all been predicted fairly convincingly, and a few years in advance of their occurrence, by climate models, solar physics models, and mesoscale models. All of these models had reached a state of refinement undreamt of, by scientists at least, before the Millennium, and the climate models had had notable successes in reproducing Pleistocene as well as present climate regimes. Development of the solar models had been greatly stimulated by the improved helioseismic data that began to be available in the 1990s. In terms of their perceived impact on human societies, weather and climate prediction had become increasingly important everywhere, including the low lying coastal areas where the rise in sea level had become unequivocal and where further population growth, only just coming under control, had been accommodated by new amphibian dwelling-house technologies.

Weather prediction had been made far more socially effective. This was not only through new communication skills and technologies but also through one of the spinoffs from those skills and technologies — the new cooperation and mutual understanding between scientists, social scientists, and the intelligent lay public that was built up in the aftermath of the damaging 'Science Wars' of the 1990s and the accompanying 'Little Dark Age' of commercially amplified superstitious belief. The effort to ward off a greater Dark Age had begun with attention to what was called the problem of the public understanding of science, followed by recognition that public respect for the scientific ideal, for rational thinking and respect for evidence — and a wider appreciation that such questions are distinct from questions about personal faith and spiritual health — was even more important than understanding particular bits of science. This in turn had led scientists to develop an improved, publicly declared code of professional conduct that helped to reduce scientific arrogance and theologizing. Although these various efforts had seemed quite futile at first, they began to have a big effect when, to everyone's astonishment, the international mass newsmedia swung behind them.

This happened when, after fully establishing their supranational power base, the international newsmedia and internet moguls had begun to realize that their practice of promoting scientifically irresponsible mass journalism, though profitable in the short term, was a bit like playing with matches in a planetary-scale explosives factory. They saw that they were playing games, on a global scale, with exceedingly powerful human instincts — which instincts and their bioclimatological origins were by then better understood scientifically — and they saw that such games could lead in the end to catastrophic social destabilization and widespread totalitarian repression. This in turn, they saw, would cause the destruction of the newsmedia's own power base, dependent as it was, and is, on the continued existence of the free market democracies and internet. The growth of technologically sophisticated kamikaze terrorism and apocalyptic balkanization or sectarianization, not only in the poorer countries but also, during the 1990s, in the affluent USA, was one of the signs to which the international newsmedia eventually paid heed, it being recognized that, with biotechnological advances, the terrorism might even become quite literally a global scale phenomenon.

Scientists, for their part, having seen more clearly the damage that had been done to science through its deliberate or inadvertent promotion as the An-

[5] There are already precursors to such a development: see, e.g., Lighthill (1993) on the Aerosonde Project for the improvement of tropical cyclone forecasts: current designs call for a 12 kg, 3 meter wingspan aircraft capable of in situ and surface pressure measurements using radar altimetry. See also, e.g., Palmer et al. (1997) on singular vectors as a guide to deployment of observations.

swer to Everything and the Way to the Mind of God, had developed as part of their professional code a publicly declared 'humility principle', in turn leading to a better public understanding of the nature of scientific uncertainty. The upshot had been a renewed respect for science and a more widespread acceptance that good science, rather than astrology or witchcraft, would be our best eyes and ears on an uncertain future, and an indispensable aid to managing medical, environmental, and societal risk.

The same professional principle of humility and recognition of uncertainty — indeed the development of an improved, simple *language* of uncertainty — had helped to increase respect for weather forecasting in particular. Public weather bulletins, early in the Millennium, were already routinely incorporating well explained, semiquantitative estimates of forecast uncertainty. It was widely believed that these largely successful, and widely publicized, estimates of forecast uncertainty had been the single most important influence toward building public confidence in the scientific respectability of weather prediction — helped of course by improvements in actual forecasting skill, and by the increasingly informative animated-graphical techniques used in public presentations, in broadcasts, and on the Internet. (Some early versions of those techniques had actually been seen, from time to time, in the British Broadcasting Corporation's television weather bulletins as long ago as the 1980s, some of which dared to be as informative as possible, going against the then prevailing 'look of the thing' or 'fool the public' culture. Even such simple but effective devices as cutting straight from an animated rainfall radar map into an animation, at the same speed, of the predicted forecast rainfall pattern — much fuzzier of course — instantly conveyed to an interested lay person a useful intuitive idea of predictable and unpredictable spatial scales, and some inkling of the overall complexity of the problem.)

Today, in the year 2020, the global observing network has long been largely automated — including even aerosonde flight plans, whose bureaucracy had been computerized — and the network performs far more consistently than in the 1980s. But data assimilation and forecasting procedures have deliberately not been quite fully automated. Artificial intelligence systems, though very powerful, are still far from being directly competitive with the human brain in certain respects — particularly the visual–cognitive system of a gifted individual who has had the full childhood educational and virtual-reality games experience, and who has been given the incentive of a highly respected professional status after surviving rigorous training and selection procedures. This continuing supremacy of certain aspects of human intelligence is now better understood, even though it had surprised some artificial-intelligence researchers who had based their predictions of absolute machine supremacy on textbook models of neurons and synapses that were, in fact, outdated even before the Millennium.

The data and forecasting quality control systems at the new International Centers are among the many twenty-first-century information systems in which the efficiency of person–machine interaction has been fully developed and exploited, with a finesse undreamt of before the Millennium. The basic mode of operation is, of course, simple in principle: the visual inspection and manipulation of animated thermodynamical and dynamical fields — not only to inform the forecasters about the evolution of the system's model atmosphere and its response to data assimilation, but also to facilitate extremely rapid interactive repairs to the model fields. Techniques such as variable speed animation within virtual reality environments have long been taken for granted: the duty forecaster is, in a sense, thoroughly immersed in the four dimensional weather. Also taken for granted is the succinctness and high visibility with which advected quantities, especially PV, PT, water vapor and other chemical mixing ratios, convey large amounts of dynamical and thermodynamical information in an intuitively assimilable form, while fully recognizing, of course, the importance of large scale wave propagation effects.

On some occasions, a forecaster carries out the repairs directly on the PV–PT and moisture fields, for instance moulding, by direct hyper-electroencephalography, the shapes of missing sharp edges or shear lines whose likely presence in an upper air PV–PT distribution might be a good guess from experience. The system maintains the appropriate integral constraints on the potential vorticity field (respecting the indestructibility of 'PV substance'), and a system of audio signals instantly informs the forecaster whether, and in what way, he or she is reducing or increasing the stress between the 4D analysis and the observational constraints. The latter include satellite brightness temperatures in various spectral regions, giving information not only on temperatures but also on advected quantities such as water vapor, ozone and many other trace chemicals. There is also a warning signal when the limitations of the balance and potential vorticity invertibility concepts are approached, and an ability to diagnose any spontaneous-adjustment emission of inertia–gravity waves that results, as well as emission from other sources such as cumulonimbus activity, and assess the consequences for mesoscale developments and for aviation fuel savings in the lower stratosphere.

Even in the year 2020, there are still fairly severe limitations on the size of the ensemble of initial conditions that can be used to help assess forecast uncertainty. So although a basic ensemble of a few thousand forecasts is always run automatically, for which the initial conditions are varied objectively using forms of singular vector analysis, there is also provision for a practically unlimited number of special forecast runs, based on the forecasters' subjective assessment of the most sensitive locations for varying the initial conditions, helped by the locations of features in displays of advected quantities and an array of other diagnostics. This subjective or semisubjective assessment of sensitive locations is regarded as one of the most important parts of the duty forecaster's responsibility, if only because it is the smallest scales that are most conspicuous to human perception but least well handled by the automatic analysis–forecast process. Human awareness of sensitive locations is considered to be an essential part of the background to the judgments that have to be made when, as happens increasingly often these days, there is a risk of hazardous weather.

As a typical manifestation of the newly enlightened uses of market forces, forecasters are provided with strong incentives to develop their subjective skills — even though there will always be some element of luck to add to the excitement. For instance, one member of the ensemble of forecasts always takes the fully automated objective analysis as initial condition. After verification the system logs the improvement in skill, if any, in the final forecast, resulting from the forecaster's repair work, if any, and subjective assessment of ensemble properties. This is added to the forecaster's personal score and thence bonus payment. A particularly high score is earned by finding initial conditions for a special forecast run that proves to have a higher rate of divergence from the main run than a typical member of the automatic ensemble, for a given root mean square initial difference. It is becoming something of a legend that a select few in the forecasting profession have begun to develop an almost uncanny level of skill, like the legendary 'top guns' or ace fighter pilots; and such individuals are coming to be known as *ace forecasters*. A few of these individuals have, in newsmedia interviews, described their subjective experience as something like merging with, or becoming part of, the weather, or not just being in the eye of the storm but becoming the eye of the storm, as one of them put it. And the urge to aspire to such skill and to join the ranks of internationally famous ace forecasters is intense, even though only a select few have so far earned that status. It is just one small but significant part of the drama that has contributed to the new prestige, glamour, and indeed human interest of the atmospheric sciences.

Acknowledgements Many colleagues have kindly shared their knowledge, ideas, historical recollections, and unpublished work over the years, including John Allen, Rainer Bleck, Onno Bokhove, Lance Bosart, Keith Browning, Oliver Bühler, Jule Charney, Mike Cullen, Ed Danielsen, Chris Davis, Huw Davies, David Dritschel, Franco Einaudi, Arnt Eliassen, Kerry Emanuel, Mike Fisher, Rupert Ford, Bill Grose, Peter Haynes, Raymond Hide, Tony Hollingsworth, Darryl Holm, Jim Holton, Brian Hoskins, Ian James, Martin Juckes, Steve Koch, Dan Keyser, Ed Lorenz, Bob Lunnon, Robert MacKay, Jerry Mahlman, Taroh Matsuno, Jim McWilliams, Geoff Monk, Mike Montgomery, Phil Morrison, John Norbury, Warwick Norton, Alan O'Neill, Tim Palmer, Anders Persson, Norman Phillips, Ray Pierrehumbert, Alan Plumb, Dave Raymond, Dick Reed, Peter Rhines, Rich Rotunno, Ian Roulstone, Rick Salmon, Prashant Sardeshmukh, Wayne Schubert, Mike Sewell, Mel Shapiro, Ted Shepherd, Glenn Shutts, Adrian Simmons, Chris Snyder, Susan Solomon, George Sutyrin, Henk Tennekes, Alan Thorpe, Jürgen Theiss, Joe Tribbia, Adrian Tuck, Louis Uccellini, Tom Warn, Darryn Waugh, Jeff Whitaker, Geoff Vallis, Martin Young, and Vladimir Zeitlin. 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 Programme, from the Science and Engineering Research Council, subsequently Engineering and Physical Sciences Research Council, through research grants and through their generous award of a Senior Research Fellowship.

REFERENCES

Allen, J. S., and Holm, D. D, 1996: Extended-geostrophic Hamiltonian models for rotating shallow water motion. Physica D, 98, 229–248.

Appenzeller, C., Davies, H. C., Norton, W. A., 1996: Fragmentation of stratospheric intrusions. J. Geophys. Res., 101, 1435–1456.

Austin, J., 1992: Towards the four-dimensional assimilation of stratospheric chemical constituents. J. Geophys. Res., 97, 2569–2588.

Bennetts, D. A., Hoskins, B. J., 1979: Conditional symmetric instability — a possible explanation for frontal rainbands. Quart. J. Roy. Meteorol. Soc., 105, 945–962.

Bokhove, O., Shepherd, T. G., 1996: On Hamiltonian balanced dynamics and the slowest invariant manifold. J. Atmos. Sci., 53, 276–297.

Buizza, R., Palmer, T. N., 1993: The singular-vector structure of the atmospheric global circulation. J. Atmos. Sci., 52, 1434–1456.

Charney, J. G., 1948: On the scale of atmospheric motions. Geofysiske Publ., 17(2), 3–17.

Chen, P., Holton, J. R., O'Neill, A., Swinbank, R., 1994: Isentropic mass exchange between the Tropics and Extratropics in the stratosphere. J. Atmos. Sci., 51, 3006–3018.

Ciesielski, P. E., Stevens, D. E., Johnson, R. H., Dean, K. R., 1989: Observational evidence for asymmetric inertial instability. J. Atmos. Sci., 46, 817–831.

Clark, P. D., Haynes, P. H., 1994: Inertial instability of an asymmetric low-latitude flow. Quart. J. Roy. Meteorol. Soc., submitted.

Danielsen, E. F., 1968: Stratospheric-tropospheric exchange based on radioactivity, ozone and potential vorticity. J. Atmos. Sci., 25, 502–518.

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

Davis, C. A., 1992: Piecewise potential vorticity inversion. J. Atmos. Sci., 49, 1397–1411.

Davis, C. A., Stoelinga, M. T., Kuo, Y.-H., 1993: The integrated effect of condensation in numerical simulations of extratropical cyclogenesis. Mon. Wea. Rev., 121, 2309–2330.

Demirtas, M., Thorpe, A. J., 1997: Sensitivity of short-range weather forecasts to local potential-vorticity modifications. Mon. Wea. Rev., submitted.

Dritschel, D. G., 1989: Contour dynamics and contour surgery: numerical algorithms for extended, high-resolution modelling of vortex dynamics in two-dimensional, inviscid, incompressible flows. Computer Phys. Rep., 10, 78–146.

Dunkerton, T. J., 1993: Inertial instability of non-parallel flow on an equatorial beta-plane. J. Atmos. Sci., 50, 2744–2758.

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

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

Emanuel, K. A., 1991: The theory of hurricanes. Ann. Rev. Fluid Mech., 23, 179–196.

Errico, R. M., 1982: Normal mode initialization and the generation of gravity waves by quasi-geostrophic forcing. J. Atmos. Sci., 39, 573–586.

Ertel, H., 1942: Ein Neuer hydrodynamischer Wirbelsatz. Met. Z., 59, 271–281.

Fisher, M., Lary, D. J., 1995: Lagrangian four-dimensional variational data assimilation of chemical species. Q. J. Roy. Meteorol. Soc., 121, 1681–1704.

Ford, R., 1994: Gravity wave radiation from vortex trains in rotating shallow water. J. Fluid Mech., 281, 81–118.

Ford, R., McIntyre, M. E., Norton, W. A., 1997: Balance and the slow quasi-manifold: some explicit results. J. Atmos. Sci., to be submitted.

Haynes, P. H., McIntyre, M. E., 1990: On the conservation and impermeability theorems for potential vorticity. J. Atmos. Sci., 47, 2021–2031.

Hibbard, W. et al., 1989: Application of the 4-D McIDAS to a model diagnostic study of the Presidents' Day cyclone. Bull. Amer. Meteorol. Soc., 70, 1394–1403.

Hinkelmann, K. H., 1969: Primitive equations. In: *Lectures in Numerical Short-range Weather Prediction. Regional Training Seminar, Moscow.* World Met. Org. No. 297, pp. 306–375.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., Pfister, L., 1995: Stratosphere–troposphere exchange. Revs. Geophys., 33, 403–439.

Hoskins, B. J., McIntyre, M. E., Robertson, A. W., 1985 [HMR]: On the use and significance of isentropic potential-vorticity maps. Quart. J. Roy. Meteorol. Soc., 111, 877–946. Also **113**, 402–404. [*Note:* The title should have been 'On the use and significance of isentropic maps of potential vorticity.']

Juckes, M. N., McIntyre, M. E., 1987: A high resolution, one-layer model of breaking planetary waves in the stratosphere. Nature, 328, 590–596.

Keyser, D., Rotunno, R., 1990: On the formation of potential-vorticity anomalies in upper-level jet-front systems. Mon. Wea. Rev., 118, 1914–1921.

Kleinschmidt, E., 1950–1: Über Aufbau und Entstehung von Zyklonen (1–3 Teil). Met. Runds., 3, 1–6; 3, 54–61; 4, 89–96.

Lahoz, W. A., O'Neill, A., Heaps, A., Pope, V. D., Swinbank, R., Harwood, R. S., Froidevaux, L., Read, W. G., Waters, J. W., Peckham, G. E., 1996: Vortex dynamics and the evolution of water vapour in the stratosphere of the southern hemisphere. Quart. J. Roy. Meteorol. Soc., 122, 423–450.

Lee, A. M., Carver, G. D., Chipperfield, M. P., Pyle, J. A., 1997: Three-dimensional chemical forecasting: a methodology. J. Geophys. Res., 102, 3905–3919. [Special issue on the ASHOE airborne experiment.]

Lighthill, M. J., 1952: On sound generated aerodynamically. Proc. Roy. Soc. Lond., A 211, 564–587. Also *Collected Papers*, Vol. III, ed. M. Y. Hussaini. Oxford, University Press.

Lighthill, J., 1993: Final recommendations of the Symposium. In: Proc. ICSU/WMO Internat. Symp. on Tropical Cyclone Disasters, Beijing, ed. J. Lighthill, Z. Zheng, G. Holland, K. Emanuel; Beijing, Peking University Press, 582–587.

Magnusdottir, Gudrun., Haynes, P. H., 1996: Application of wave-activity diagnostics to baroclinic-wave life cycles. J. Atmos. Sci., 53, 2317–2353.

Manney, G. L., Zurek, R. W., O'Neill, A., Swinbank, R., 1994: On the motion of air through the stratospheric polar vortex. J. Atmos. Sci., 51, 2973–2994.

McIntyre, M. E., 1982: How well do we understand the dynamics of stratospheric warmings? J. Meteorol. Soc. Japan, Special Centennial Issue, 60, 37–65. [*Note:* "latitude" should be "altitude" on page 39a, line 2.]

McIntyre, M. E., 1988: Numerical weather prediction: a vision of the future. Weather (Roy. Meteorol. Soc.), 43, 294–298.

McIntyre, M. E., 1994: Numerical weather prediction: a updated vision of the future. In: *The Life Cycles of Extratropical Cyclones*, Proceedings of an International Symposium at the University of Bergen, Norway, 27 June – 1 July 1994, ed. S. Grønås and M. Shapiro, ISBN 82–419–0144–5. Vol. I, Invited Papers, 275–286.

McIntyre, M. E., 1992: Atmospheric dynamics: some fundamentals, with observational implications. In: Proc. Internat. School Phys. "Enrico Fermi", CXV Course, ed. J. C. Gille, G. Visconti (ISBN 0-444-89896-4); Amsterdam, Oxford, New York, Toronto, North-Holland, 313–386. [A list of updates and minor corrigenda is available via ftp://ftp.damtp.cam.ac.uk/pub/papers/mem; get mcintyre.ps. Also accessible via http://www.atmos-dynamics.damtp.cam.ac.uk/.]

McIntyre, M. E., 1993a: On the role of wave propagation and wave breaking in atmosphere–ocean dynamics. In: Theoretical and Applied Mechanics 1992, ed. S. Bodner, J. Singer, A. Solan, and Z. Hashin. (Sectional Lecture, Proc. XVIII Int. Congr. Theor. Appl. Mech., Haifa.) Amsterdam, New York; Elsevier, 281–304.

McIntyre, M. E., 1993b: Isentropic distributions of potential vorticity and their relevance to tropical cyclone dynamics. In: Proc. ICSU/WMO Internat. Symp. on Tropical Cyclone Disasters, Beijing, ed. J. Lighthill, Z. Zheng, G. Holland, K. Emanuel; Beijing, Peking University Press, 143–156.

McIntyre, M. E., Norton, W. A., 1990: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. J. Fluid Mech., 212, 403–435 (G. K. Batchelor Festschrift Issue). *Corrigendum* 220, 693.

McIntyre, M. E., Norton, W. A., 1997: Potential-vorticity inversion on a hemisphere. J. Atmos. Sci.,, resubmitted.

McIntyre, M. E., Palmer, T. N., 1983: Breaking planetary waves in the stratosphere. Nature, 305, 593–600.

McIntyre, M. E., Palmer, T. N., 1984: The "surf zone" in the stratosphere. J. Atm. Terr. Phys., 46, 825–849.

McIntyre, M. E., Palmer, T. N., 1985: A note on the general concept of wave breaking for Rossby and gravity waves. Pure Appl. Geophys., 123, 964–975.

McIntyre, M. E., Roulstone, I., 1997: Hamiltonian balanced models: slow manifolds, constraints and velocity splitting. J. Fluid Mech., in revision.

Molteni, F., Buizza, R., Palmer, T. N., Petroliagis, T., 1996: The ECMWF ensemble prediction system: Methodology and validation. Quart. J. Roy. Meteorol. Soc., 122, 73–119.

Monod, J., 1971: *Chance and Necessity*, transl. A. Wainhouse. Glasgow, Collins, 187 pp.

Nakamura, N., 1996: Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate. J. Atmos. Sci., 53, 1524–1537.

Norton, W. A., 1988: *Balance and potential vorticity inversion in atmospheric dynamics.* University of Cambridge, PhD Thesis, 167 pp.

Norton, W. A., 1994: Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following coordinate system and a technique for advecting material contours. J. Atmos. Sci., 51, 654–673.

O'Sullivan, D. J., Hitchman, M. H., 1992: Inertial instability and Rossby wave breaking in a numerical model. J. Atmos. Sci., 49, 991–1002.

O'Sullivan, D., Chen, P., 1996: Modeling the QBO's influence on isentropic tracer transport in the subtropics. J. Geophys. Res., 101, 6811–6821.

Palmer, T. N., 1996: Predictability of the atmosphere and oceans: from days to decades. In: *Decadal Climate Variability* (NATO Advanced Study Institute, Les Houches, 1995) ed D. Anderson and J. Willebrand. Series 1, Global Envirnmental Change, Vol. 44. Heidelberg, Springer.

Palmer, T. N., Gelaro, R., Barkmeijer, J., Buizza, R., 1997: Singular vectors, metrics, and adaptive observations. J. Atmos. Sci., in press.

Polvani, L. M., McWilliams, J. C., Spall, M. A., Ford, R., 1994: The coherent structures of shallow water turbulence: deformation-radius effects, cyclone/anticyclone asymmetry and gravity-wave generation. Chaos, 4, 177–186 and 427–430. [Special Volume on geophysical flows; Pp. 427–430 are colour plates.]

Rabier, F., Klinker, E., Courtier, P., Hollingsworth, A., 1996: Sensitivity of forecast errors to initial conditions. Q. J. Roy. Meteorol. Soc., 122, 121–150.

Randel, W. J., Gille, J. C., Roche, A. E., Kumer, J. B., Mergenthaler, J. L., Waters, J. W., Fishbein, E. F., Lahoz, W. A., 1993: Stratospheric transport from the tropics to middle latitudes by planetary-wave mixing. Nature, 365, 533–535.

Raymond, D. J., 1992: Nonlinear balance and potential-vorticity thinking at large Rossby number. Quart. J. Roy. Meteorol. Soc., 118, 987–1015.

Reed, R. J., Danielsen, E. F., 1959: Fronts in the vicinity of the tropopause. Arch. Met. Geophys. Biokl., A 11, 1–17.

Reed, R. J., Stoelinga, M. T., Kuo, Y.-H., 1992: A model-aided study of the origin and evolution of the anomalously high potential vorticity in the inner region of a rapidly deepening marine cyclone. Mon. Wea. Rev., 120, 893–913.

Reed, R. J., Kuo, Y.-H., Low-Nam, S., 1994: An adiabatic simulation of the ERICA IOP 4 storm: an example of quasi-ideal frontal cyclone development. Mon. Wea. Rev., 122, 2688–2708.

Rossby, C. G., 1936: Dynamics of steady ocean currents in the light of experimental fluid mechanics. Mass. Inst. of Technology and Woods Hole Oc. Instn. Papers in Physical Oceanography and Meteorology, 5(1), 1–43.

Rossby, C. G., 1940: Planetary flow patterns in the atmosphere. Quart. J. Roy. Meteorol. Soc., 66(Suppl.), 68–97.

Aebischer, U., Schär, C., 1997: Low-level potential vorticity and cyclogenesis to the lee of the Alps. J. Atmos. Sci., in press.

Schoeberl, M. R., Lait, L. R., Newman, P. A., Rosenfield, J. E., 1992: The structure of the polar vortex. J. Geophys. Res., 97, 7859–7882. [Polar Ozone Special Issue, no. D8]

Schoeberl, M. R., Roche, A. E., Russell III, J. M., Ortland, D., Hays, P. B., Waters, J. W., 1997: An estimation of the dynamical isolation of the tropical lower stratosphere using UARS wind and trace gas observations of the quasi-biennial oscillation. Geophys. Res. Lett., 24, 53–56.

Shaw, Sir Napier, 1930: *Manual of Meteorology. Vol. III: The Physical Processes of Weather.* Cambridge, University Press.

Starr, V. P., Neiburger, M., 1940: Potential vorticity as a conservative property. J. Marine Res., 3, 202–210.

Strahan, S. E., Mahlman, J. D., 1994: Evaluation of the SKYHI general circulation model using aircraft N_2O measurements. 1. Polar winter stratospheric meteorology and tracer morphology. J. Geophys. Res., 99, 10305–10318.

Sutton, R. T., Maclean, H., Swinbank, R., O'Neill, A., Taylor, F. W., 1994: High-resolution stratospheric tracer fields estimated from satellite observations using Lagrangian trajectory calculations. J. Atmos. Sci., 51, 2995–3005.

Tennekes, H., 1988: Numerical weather prediction: Illusions of security, tales of imperfection. Weather (Roy. Meteorol. Soc.), 43(4), 165–170.

Thépaut, J.-N., Courtier, P., Belaud, G., Lemaître, G., 1996: Dynamical structure functions in a four-dimensional variational assimilation. A case study. Quart. J. Roy. Meteorol. Soc., 122, 535–561.

Thorncroft, C. D., Hoskins, B. J., McIntyre, M. E., 1993: Two paradigms of baroclinic-wave life-cycle behaviour. Quart. J. Roy. Meteorol. Soc., 119, 17–55.

Thorpe, A. J., 1985: Diagnosis of balanced vortex structure using potential vorticity. J. Atmos. Sci., 42, 397–406.

Warn, T., 1997: Nonlinear balance and quasi-geostrophic sets. Atmos.–Ocean, 35, 135–145. [*Note:* This pioneering paper was written in 1983 but rejected by the journal to which it was submitted. It seems to have been the first to recognize that the 'slow quasi-manifold' of the parent dynamics is in some sense fuzzy.]

Warn, T., Ménard, R., 1986: Nonlinear balance and gravity–inertial wave saturation in a simple atmospheric model. Tellus, 38A, 285–294.

Warwick, K., 1997: *March of the Machines: Why the New Race of Robots will Rule the World.* London, Random House (Century Books), 263 pp.

Waters, J. W., Froidevaux, L., Read, W. G., Manney, G. L., Elson, L. S., Flower, D. A., Jarnot, R. F., Harwood, R. S., 1993: Stratospheric ClO and ozone from the Microwave Limb Sounder on the

Upper Atmosphere Research Satellite. Nature, 362, 597–602.

Waugh, D. W., 1993: Subtropical stratospheric mixing linked to disturbances in the polar vortices. Nature, 365, 535–537.

Waugh, D. W., Plumb, R. A., 1994: Contour advection with surgery: a technique for investigating fine scale structure in tracer transport. J. Atmos. Sci., 51, 530–540.

von Weizsäcker, E., Lovins, A. B., Lovins, L. H., 1997: *Factor Four: Doubling Wealth, Halving Resource Use — The New Report to the Club of Rome.* London, Earthscan Publications, 322 pp.

FIGURE CAPTIONS

Figure 1. **(a)** Hypothetical passive tracer on the 320 K isentropic surface for 14 May 1992 at 1200Z, calculated by the *contour advection* technique (Norton 1994, Waugh and Plumb 1994) using winds from the operational analyses of the European Centre for Medium Range Weather Forecasts (ECMWF), initialized four days earlier to coincide with the PV (potential vorticity) contours whose values are 1, 2, 3, 4, and 5, in the standard PV units of $10^{-6}\mathrm{m}^2\mathrm{s}^{-1}\,\mathrm{K\,kg}^{-1}$, of a smoothed isentropic map of PV from the same operational analyses. Values from 1 unit upwards are colored rainbow-wise from dark blue to red and represent nominally stratospheric air. **(b)** Meteosat 5.7–7.1μm water vapor image at approximately the same time as in **(a)**. **(c)** Isentropic map of PV for the same time as in **(a)** and from the same operational analyses. The dotted contours near the tropopause have values 0.5, 1.0, 1,5 PV units and the solid contours 2, 3, 4,... PV units.

Figure 2. Demonstration of highly accurate PV inversion in a shallow water system, the simplest dynamical system for which the balance (slow quasi-manifold) and inversion concepts are nontrivial. Quasi-geostrophic inversion would be grossly inaccurate in this example, if only because there are substantial layer depth variations, with fractional departures 0.4 or more of the area-mean depth, 2 km. The domain is the hemisphere and the map projection is polar stereographic. Solid contours show positive values, dashed contours negative values, and dotted contours zero. **(a)** wind field from the primitive equation solution, arrows scaled by the $100\,\mathrm{m\,s}^{-1}$ standard arrow shown, and geopotential height, contoured at intervals of 0.1 km, defined as departure of layer depth from its area mean (2 km). **(b)** (horizontal) divergence from the primitive equation solution, contoured at intervals of $0.6 \times 10^{-6}\mathrm{s}^{-1}$. **(c,d)** grayscale and contour maps of the shallow water PV from the primitive equation solution, i.e., absolute vorticity divided by local layer depth; the contour interval is $1{\times}10^{-8}\mathrm{m}^{-1}\mathrm{s}^{-1}$ and the hatching in the contour plot highlights values lying between 4 and 6 of these units. The grayscale representation of the same information is monotonic from light to dark, from zero to a maximum value of $1\times10^{-7}\mathrm{m}^{-1}\mathrm{s}^{-1}$. **(e,f)** As in **(a,b)** but reconstructed from the PV alone using an accurate nonlinear PV inversion algorithm. From McIntyre and Norton (1990).