Some Fundamental Aspects of Atmospheric Dynamics, with a Solar Spinoff

Michael E. McIntyre

Centre for Atmospheric Science at the Department of Applied Mathematics and Theoretical Physics, Cambridge, UK
http://www.atm.damtp.cam.ac.uk/people/mem/

Global-scale atmospheric circulations illustrate with outstanding clarity one of the grand themes in physics: the organization of chaotic fluctuations in complex dynamical systems, with nontrivial, persistent mean effects. This is a theme recognized as important not just for classical textbook cases like molecular gas kinetics but also for an increasingly wide range of dynamical systems with large phase spaces, not in simple statistical–mechanical equilibrium or near-equilibrium. In the case of the atmosphere, the organization of fluctuations is principally due to three wave-propagation mechanisms or quasi-elasticities: gravity/buoyancy, inertia/Coriolis, and Rossby/vortical. These enter the fluctuation dynamics along with various forms of highly inhomogeneous turbulence, giving rise to a "wave–turbulence jigsaw puzzle" in which the spatial inhomogeneity—characteristic of "wave-breaking" understood in a suitably generalized sense—exhibits phase coherence and is an essential, leading-order feature as illustrated, also, by the visible surf zones near ocean beaches. Such inhomogeneity is outside the scope of classical turbulence theory, which assumes spatial statistical homogeneity or small departures therefrom. The wave mechanisms induce systematic correlations between fluctuating fields, giving rise to mean fluxes or transports of momentum quite different from those found in gas kinetics or in classical turbulence theory. Wave-induced momentum transport is a long-range process, effective over distances far exceeding the fluctuating material displacements or "mixing lengths" characteristic of the fluid motion itself. Such momentum transport, moreover, often has an "anti-frictional" character, driving the system away from solid rotation, not toward it, as underlined most plainly by the existence of the stratospheric quasi-biennial oscillation (QBO) and its laboratory couunterparts.

Wave-induced momentum transport drives the Coriolis-mediated "gyroscopic pumping" of meridional circulations against radiative relaxation, as illustrated by the Murgatroyd–Singleton mesospheric circulation and the Brewer–Dobson (misnamed Hadley) stratospheric circulation. The Brewer–Dobson can be contrasted with the convectively driven tropospheric Hadley circulation and with the oceans' thermohaline circulation. The ocean is an opposite extreme case in the sense that radiative relaxation has no counterpart. If the stratosphere is compared to a tape recorder whose motor is the gyroscopic pump and whose recording head is the tropical tropopause with its seasonal water-vapour signal, then the distribution of radiative-equilibrium temperatures across the tropics may be compared to the guidance wheels influencing the upward path of the tape. In other words, solar heating does not *drive* the tropical stratospheric upwelling, but does *influence* the way in which the upwelling mass flux demanded by the wave-driven pumping is distributed across the tropics. Because the tropical mean circulation problem is nonlinear, one cannot think of the tropical part of the circulation as a linear superposition of thermally and mechanically driven "contributions".

Insights into the dynamics of atmospheric circulations have recently led to a breakthrough in an astrophysical problem, that of understanding the differential rotation, meridional circulation, and helium distribution within the sun, with strong implications for helioseismic inversion and the so-called lithium and beryllium problems. This is an example of meteorological understanding informing solar and stellar physics.

A. INTRODUCTION

I want to revisit the fluid dynamics of global-scale atmospheric circulations in a way that emphasizes physical fundamentals, and to show how this has led to new insights into the workings of the sun's interior and other stellar interiors. In all these problems we are dealing with chaotic fluid motion at vast Reynolds numbers, involving various kinds of turbulence. But a leitmotif running through the whole story is the inapplicability of the classical turbulence paradigm, not only quantitatively but also qualitatively. Classical turbulence theory was historically important and is still implicitly relied on both in the meteorological literature and in the

Meteorology at the Millennium
ISBN 0-12-548035-0

astrophysical literature, whenever the concept of "eddy viscosity" is used in one or another of its versions. The chaotic fluctuations that are most important in shaping global-scale circulations are organized, by contrast, in a very nonclassical way, outside the scope of classical turbulence theory but now recognized as essential even to a qualitative understanding of those circulations.

Our understanding is most secure in the case of the middle atmosphere, the region consisting of the stratosphere below 50 km and mesosphere above. One reason is that we now have a wealth of Lagrangian information from long-lived chemical tracers, thanks to the effort to understand ozone-layer chemistry. The middle atmosphere is more accessible to remote sensing than the troposphere, because of its chemical diluteness and lack of moist convection. Ozone, water vapour, and methane mixing ratios are measured in parts per 10^6, and most other trace chemicals in parts per 10^9 or less. Clouds and aerosols are relatively speaking thin or nonexistent, though crucial to the chemistry when present (Albritton *et al.*, 1998). These circumstances, taken advantage of by an array of clever and sophisticated observing methods, have made the middle atmosphere into a wonderful outdoor laboratory for the study of basic processes in stratified, rotating fluid motion, such as must also exist beneath the convection zones of sun-like stars.

In the fluid dynamics of the middle atmosphere, the Coriolis and buoyancy effects associated with rotation and stable stratification are of course immensely strong, in relation to mean circulations and their chemical consequences. Typical Coriolis timescales, of the order of hours over a wide range of latitudes, and buoyancy timescales, of the order of minutes, are several orders of magnitude less than the timescales for mean circulations and chemical transports. Coarse-grain gradient Richardson numbers are large. Thus mean circulations feel an enormous stiffness from the Coriolis and buoyancy effects. This implies stiffness of the fluid-dynamical equations, in the numerical-analytical sense—the same stiffness that famously spoiled L. F. Richardson's pioneering attempt at numerical weather prediction. It also reminds us of the possible importance of wave propagation mechanisms—the reason why the classical turbulence paradigm fails.

B. THERMAL AND CHEMICAL EVIDENCE

Before saying more about the circulation dynamics, let us recall two well-known facts that constrain any theory. One is that, at solstice, the summer pole is the sunniest place on Earth. Diurnally averaged insolation—the solar energy arriving per unit time per unit horizontal area—is a maximum there, because of the substantial tilt of the earth's axis, 23.5° or 0.41 radian. The other fact is that, at altitudes between about 80 km and 90 km, the summer pole has been observed again and again to be the coldest place on earth (excluding laboratories) despite being the sunniest.

This is the famous cold summer polar mesopause. Even in a heavily averaged climatological picture (Fleming *et al.*, 1990), temperatures dip below about 150 K—colder even than the high Antarctic plateau in midwinter, which seldom if ever gets below 170 K. And there are large fluctuations about climatological values: rocket-launched falling sphere measurements have recorded, on at least one occasion (Lübken, 1999), temperatures as low as 105 K. Partly because of these extremes of cold, the summer polar mesopause—more precisely, the region a few kilometres below it—is the region of formation of, by far, the world's highest clouds, the so-called "polar mesospheric", "noctilucent", or "night shining" clouds (Gadsden and Schröder, 1989; Thomas *et al.*, 1989; Schröder, 1999a). They occur at altitudes usually within the range 80–90 km and have often been detected by space-based instruments, including visible-wavelength imagers on geosynchronous meteorological satellites (Gadsden, 2000). The clouds are also seen, less often, by ground-based observers between latitudes ~50° to 60°, around midnight on some clear summer nights. They have been noticed ever since the "Krakatau twilights" of 1883, according to Schröder (1999). I have seen them myself, looking north around 2 a.m. from a low hill near Cambridge, which at 52.2°N is near the equatorward limit of observability.

Now noctilucent clouds are believed on good evidence, including *in situ* rocket measurements, to be made of ice crystals (G. Witt and E. Kopp, personal communication). Their formation therefore depends not only on the low temperatures but also, crucially, on a supply of water vapour from below (Thomas *et al.*, 1989). Water vapour is photochemically destroyed on timescales of days by the very hard solar ultraviolet that can penetrate to the upper mesosphere (Brasseur and Solomon, 1984, Fig. 4.40). The necessary supply of water vapour depends on the existence of a systematic rising motion over the summer polar cap, as suggested schematically by the heavy dashed curve in Fig. 1. Air and water vapour are thus brought up from the less fiercely irradiated, and slightly less dry, lower layers. Water vapour mixing ratios near 50 km altitude, for instance, are maintained relatively stably at just over 6 ppmv (parts per million by volume), partly by direct supply of water and partly by the oxidation of methane, both coming from the troposphere far below, via the freeze-drying tropical tropopause.

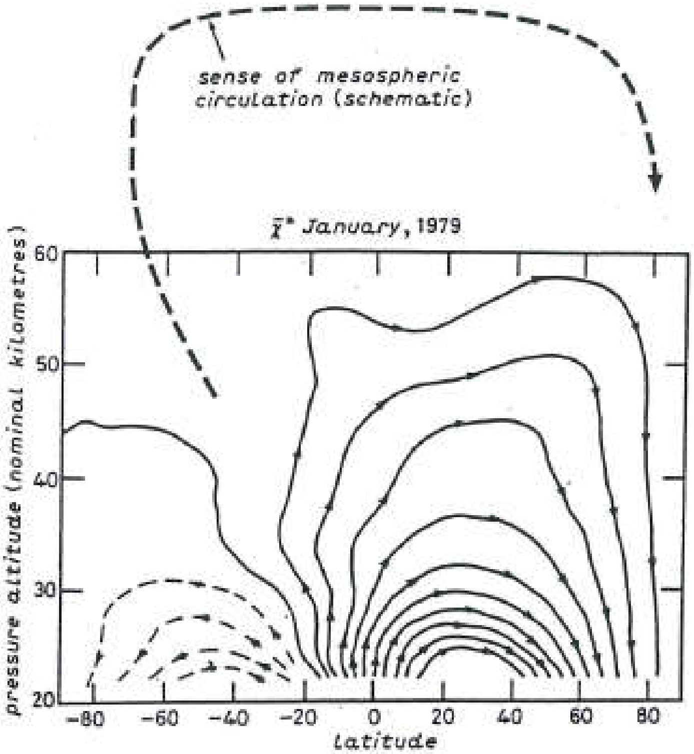

FIGURE 1 Mass transport streamlines of the global-scale mean circulation for January 1979, solid curves, from Solomon *et al.* (1986). This was estimated from LIMS satellite radiometer data in conjunction with detailed calculations of radiative heating and cooling rates. The picture gives the typical latitude and height dependence of the zonal and time averaged mass circulation of the stratosphere, defined in a quasi-Lagrangian sense (Section H) giving a simplified, but roughly correct, indication of the vertical advective transport of chemical constituents (Section H). The heavy dashed streamline indicates the qualitative sense of the mesospheric or Murgatroyd-Singleton branch of the circulation, deduced from other observational and theoretical evidence (Andrews *et al.*, 1987) The lower pair of circulation cells is often referred to as the Brewer–Dobson circulation and the mesosopheric branch as the Murgatroyd–Singleton circulation. The altitude scale is nominal log-pressure altitude, based on a nominal pressure scale height of 7 km. The right-hand cell of the circulation is a little stronger than usual, in this picture; the wintertime stratosphere was dynamically very active in January 1979. The northward mean velocities at top right (within the frame, not counting the heavy dashed streamline) are of the order of 2 or 3 m s^{-1}. On the heavy dashed streamline they would be more like 10 m s^{-1}.

The rising motion in the summer polar mesosphere is part of a self-consistent picture that checks out very well in other ways. For instance the rising motion is an essential part of *why* summer polar mesopause temperatures are so low. There is a refrigerating action due to the expansion of the air as it rises through the stable stratification. The consequent adiabatic cooling opposes the net radiative heating, which has positive contributions both from the sun and from infrared radiation emitted by warmer layers near the stratopause below (the infrared photons having long free paths at these altitudes, many of them escaping directly to space, but a few being reabsorbed on their way up). The result is that temperatures around 80–90 km are pulled down, by the refrigerating action, by as much as ~100 K below radiative equilibrium, though quantitative estimates are difficult because of uncertainties about the detailed photochemistry, and about departures from local thermodynamic equilibrium (e.g. Andrews *et al.*, 1987; Ward and Fomichev, 1996).

Temperatures in certain other parts of the middle atmosphere also depart significantly from radiative equilibrium, for instance over the winter polar cap. There, air is compressed, and its temperature pushed above radiative equilibrium, by the systematic descending motion seen on the right of Fig. 1. This contrasts with the summer stratopause, the temperature maximum at pressure altitudes near 1 hPa, or about 50 km or 7 pressure scale heights, which is relatively close to radiative equilibrium, probably within 10 K or so, the circulation being relatively weak. The high temperatures at the summer stratopause are well explained by the absorption, by ozone, of solar ultraviolet at wavelengths ~200–300 nm, which can penetrate down to 50 km or lower (Brasseur and Solomon, 1984, Fig. 4.40; Andrews *et al.*, 1987). Implicit in all the foregoing is, of course, the notion of a radiative equilibrium temperature $T_{\rm rad}$ toward which global-scale temperatures, in the absence of any circulation, would tend to relax. This notion is well justified by detailed studies of radiative transfer (e.g. Andrews *et al.*, 1987; Fels, 1985), relaxation times being of the order of days to weeks.

A circulation like that in Fig. 1 and its seasonal variants is consistent not only with observed temperatures and the observed behaviour water vapour and methane but also with that of a number of other chemical constituents (nearly all of which, incidentally—having molecules larger than diatomic—are greenhouse gases, with quasi-bending modes of vibration at infrared frequencies).

For instance, man-made chlorofluorocarbons (CFCs) are chemically inert under everyday conditions. Together with their low water-solubility and small rate of uptake by the oceans, this means that, as has been checked and rechecked by countless observations (Albritton *et al.*, 1998), they are mixed fairly uniformly throughout the troposphere. The troposphere can therefore be thought of as a big tank or reservoir whose contents are recirculated through the middle atmosphere at rates of the order suggested by Fig. 1, or a fraction less. This accounts rather well for the observed destruction rates of CFCs, which correspond to *e*-folding atmospheric lifetimes of order a century—actually somewhat shorter, by a factor 2 or so, for CFC11 ($CFCl_3$) (Albritton *et al.*, 1998), and somewhat longer for CFC12 (CF_2Cl_2), the former being a bit less stable because of its extra chlorine atom. It is mainly at altitudes ≳25 km that the CFCs are destroyed, by ultraviolet photolysis; and so rates of CFC destruction are closely tied to rates of circulation. In terms of mass flow that reaches altitudes ≳25 km, these rates amount to a few percent of the tropospheric mass per year. This is consistent with circulation times of the order of several years (Hall *et al.*, 1999) and CFC *e*-folding lifetimes of the order of a century because of the falloff of mass density with altitude, roughly in proportion to pressure, by a factor e^{-1} for each pressure scale height, i.e. for each 7 km or so. Thus mass densities at, e.g., 30 km, are about 1% of those at sea level.

It is easy to check (Holton *et al.*, 1995) that the orders of magnitude in this picture require upward velocities at, say, 20 km in the tropical lower stratosphere (bottom centre of Fig. 1) of the order of $0.2\,{\rm mm\,s^{-1}}$ or 6 km per year, about one pressure scale height per year. This sort of velocity is far too small to observe directly—or so it was thought until a few years ago. Figure 2, from Mote *et al.* (1998; see also Mote *et al.*, 1996)), comes close to qualifying as a direct observation, and gives us a powerful independent consistency check on the whole picture.

Figure 2 was derived from satellite data obtained quite recently, long after results like that of Fig. 1 had been obtained, indirectly, with the help of an earlier

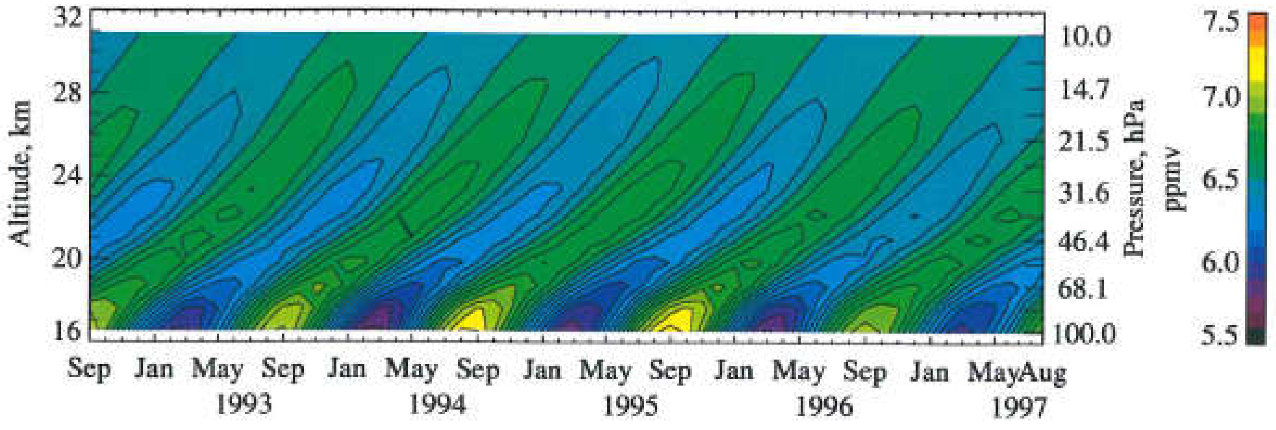

FIGURE 2 The tropical stratosphere as a "tape recorder" (from Mote *et al.*, 1998; *q.v.* for a careful discussion of the data analysis and interpretation). The time–altitude plot is from HALOE, a solar occultation limb sounder on the Upper Atmosphere Research Satellite (UARS), showing the signal from the annual cycle in water vapour being advected upward by the large-scale mean velocity $\bar{w}^*$. Near 20 km, $\bar{w}^* \approx 0.2\,{\rm mm\,s^{-1}}$. Soundings between 14°N and 14°S were used. The quantity plotted is "total hydrogen", water vapour plus twice methane, in parts per million by volume. Because of methane oxidation chemistry, this is to excellent approximation a passive tracer in the altitude range shown. The raw data have been fitted to a suitable set of extended empirical orthogonal functions (EEOFs) to extract the annual cycle in an objective way; the plot shows the contributions from the dominant two EEOFs, accounting for 68% of the variance. But the tape-recorder signal is still easily visible without any such processing (see Plate 1 of Mote *et al.*, 1998)—and in data from other instruments, notably the passive Microwave Limb Sounder on UARS (Plate 1a of Mote *et al.*, 1996).

generation of satellites. It shows the water vapour signature of the annual cycle being carried up in the rising flow just like a signal recorded on a moving magnetic tape. At altitudes around 20 km, the velocity apparent in the picture is very close to 0.2 mm s^{-1}. Mote *et al.* (1998) follow a careful model-fitting procedure to verify that upward advection at this speed is indeed the main effect near 20 km, as distinct from, e.g., the vertical eddy diffusion due to small-scale turbulence or infrared random walking (Sparling *et al.*, 1997).

But what, then, is driving the global-scale circulation illustrated in Fig. 1? As with ordinary mechanical refrigerators or compressors, something has to pump the air around. It will not, by itself, move vertically against the strong stable stratification. To make noctilucent clouds you have to do work against buoyancy forces and and pull air upward. To warm the winter pole you have to push air downward. So what is doing the pulling and pushing?

The answer, as hinted earlier, is *wave-induced momentum transport*—more precisely, the irreversible transport of momentum and angular momentum by wave motions generated in one place (mostly in the more massive troposphere below) and dissipated in another (mostly higher up), the dissipation mechanisms including various kinds of *wave breaking*, in a sense to be explained in Section I, p. 294. This, in combination with Coriolis effects, acts as a global-scale mechanical pump, as we shall be reminded in Section H. One implication, confirmed by an analysis of MSU (Microwave Sounding Unit) channel 4 satellite data (Yulaeva *et al.*, 1994), is that the seasonal variation of the total upwelling mass flux in the tropical stratosphere is tied to seasonal variations in extratropical rather than in tropical conditions.

The way the momentum transport works has features that may seem surprising, and historically did seem surprising, especially from a statistical–mechanical or turbulence–theoretical viewpoint. Such surprise was expressed, for instance, in the closing pages of E. N. Lorenz's (1967) landmark review on the general circulation. For if we regard the wave fields as examples of chaotic fluctuations, what a plasma physicist might call "weak turbulence"—and the real wave fields are often, indeed, more or less chaotic—then we find on analysing their properties that they have a distinct tendency to behave "anti-frictionally" in the sense of driving the mean state of the atmosphere not toward, but away from, solid rotation.

This is the opposite to what is predicted by classical turbulence theory, which says that chaotic fluctuations about the mean should give rise to a viscosity, in much the same way as in gas kinetics. Hence the notion of "eddy viscosity". Of course there was never the slightest reason why fluctuations should necessarily behave in this classical way. Chaotic though they may be, many of the fluctuating motions strongly feel one or more of the wave propagation mechanisms (Rossby, gravity, inertia–gravity) that are neglected in classical turbulence theory but are characteristic of stratified rotating fluid systems. By its nature, any wave propagation mechanism tends to promote systematic correlations between the different fluctuating fields, hence systematic mean stresses. In other words the fluctuations, however chaotic or otherwise they may be, tend to be organized dynamically in such a way that momentum is systematically transported through the fluid—over distances limited only by the distances over which waves can propagate.

This is the fundamental point. Wave-induced momentum transport is, by its nature, a long-range process, acting over distances that can be vastly greater than the "mixing lengths", or length scales for typical material displacements, that characterize local turbulent eddy motion. By contrast, the notion of "eddy viscosity" makes sense only when the fluctuation-induced momentum transport is at an opposite extreme—a short-range process dominated by mixing length scales that are smaller than length scales characterizing the mean state, as with mean free paths in gas kinetics.

C. WAVE-INDUCED MOMENTUM TRANSPORT: SOME SIMPLE EXAMPLES

Wave-induced momentum transport has been well known for two centuries or more as a generic phenomenon, characteristic of many types of waves in a vacuum or in fluid media. Photons in a vacuum provide the simplest example, with wave amplitude a proportional to a typical electric or magnetic field strength: we can think of photons as being just like bullets having a certain momentum and carrying that momentum from one place to another, at rates $O(a^2)$. Waves in fluid media are often described in the same way, but it is important to understand that the picture is not literally true any more. The analogy with photons in a vacuum is only an analogy, indeed only a partial analogy (McIntyre, 1981, 1993; Bühler and McIntyre, 2001). This is because the presence of the material medium not only vitiates Lorenz invariance of the wave problem—there is now a preferred frame of reference—but also significantly complicates wave-mean momentum budgets correct to $O(a^2)$. A material medium can support $O(a^2)$ mean stresses in ways that a vacuum (in classical physics) cannot.

Careful analyses of problems of this kind correct to

$O(a^2)$, with attention to boundary conditions as well as to the nonlinearities in the fluid-dynamical equations, show that the notion of a wave packet possessing a definite momentum no longer makes sense except in some very restricted circumstances. It makes no sense to talk for instance about the waves in material media depositing "their momentum", or exchanging "it" with the mean flow. But the notion of a wave-induced momentum *flux* (a) does generally make sense, (b) is more often than not correctly quantified by the photon analogy, though not always, and (c) is in any case the relevant thing for our purpose. It is in momentum fluxes, i.e. transport rates, that we are interested. To make the literature on atmospheric waves make sense, one sometimes has to insert the word "flux" or "transport" after the word "momentum".[*]

The underpinning mathematical theory is conceptually and technically tricky, especially in its most general, three-dimensional form. It is still under development today; among recent contributions an especially important one is that of Bühler (2000). However, the theory for the simplest and, for present purposes, most important case, that of waves on a zonally symmetric mean flow, is relatively straightforward and is available in detail in the standard literature (e.g. Andrews *et al.*, 1987). Some key points will be recalled briefly in Section F onwards.

A simple experiment can be used to shortcut the theoretical issues and illustrate the physical reality, and robustness, of wave-induced momentum transport. Figure 3a shows a version convenient for lecture demonstrations, in which a small cylinder is oscillated vertically and radiates capillary–gravity waves anisotropically to either side. If chalk dust or other floating powder is sprinkled on the surface, where the fluctuations are concentrated, one sees that the water flows away from the wavemaker in a persistent and conspicuous mean motion that sweeps the surface dust along with it. The experiment can be done on an overhead projector, and always works robustly. I have demonstrated it countless times when lecturing on this topic.

The observed mean flow depends on strong dissipation of the waves. This results in an irreversible flux or transport of momentum from the wavemaker to the fluid on either side of it, where the waves are dissipating. The wave dissipation is enhanced by the dirty water surface, which gives rise to a strong viscous boundary layer just under the surface, which may be intermittently turbulent. That is, the waves can be *breaking* even though no air entrainment takes place. The fact that irreversible momentum transport is involved can be seen by carefully stopping the wavemaker, and observing that the mean flow persists much longer than does the outgoing wave field. There is a significant contribution to the mean flow that is not a mere Stokes drift or other reversible mean-flow contribution dependent on the continued presence of the outgoing waves (and on the choice of averaging operator to define "mean flow"). Such reversible contributions return to zero as soon as the waves propagate out of the region of interest, or become directionally randomized or finally dissipated.

The fact that the flow is due to wave-induced momentum transport and not to the oscillatory boundary layer on the wavemaker itself is also clear in the experiment, if one uses vertical oscillations of the wavemaker. The Rayleigh–Schlichting boundary-layer streaming generated at the surface of the wavemaker is then directed toward the wavemaker, i.e., in a sense opposite to that observed. So the observed mean flow is due, rather, to the wave-induced momentum transport. Moreover, with a larger tank, one can do versions of the experiment in which the waves are generated by a larger, concave wavemaker that focuses them at a more distant spot, as suggested in Fig. 3b. This larger-scale demonstration is set up every year at the Cambridge Summer School in Geophysical and Environmental Fluid Dynamics. Sure enough, the momentum is transported, then, over the greater distance. The strongest mean flow is seen where the waves are focused.

---

[*] These points, including point (b), were in essence made long ago by Léon Brillouin. Following him, I have discussed them more fully in earlier publications. See McIntyre (1981, 1993), and their bibliographies, tracing the problem back to the "Abraham–Minkowski controversy" about light waves in a refractive medium, and to a rare mistake by Lord Rayleigh. There is a perpetual source of confusion insofar as two distinct conservable quantities enter the theory, *momentum* on the one hand and *quasimomentum* or *pseudomomentum*, often misleadingly called "wave momentum", on the other. They are the same in a vacuum but differ when a material medium is present. Pseudomomentum is an $O(a^2)$ wave property, calculable from linearized wave theory alone. Momentum is calculable only if the state of the material medium is known correct to $O(a^2)$. Both conservable quantities are associated with a translational symmetry, via Noether's theorem, hence look superficially similar within the mathematical formalisms of Hamiltonian and Lagrangian dynamics, a simple example being Hamilton's equations for ray tracing. But the two translational symmetries are physically different when a material medium is present. One of them is translational invariance of the *physics*, giving rise to conservation of momentum. The other is translational invariance, i.e., homogeneity, of the *medium*, giving rise to conservation of pseudomomentum. When ray tracing is a valid approximation, one can usually take the wave-induced momentum flux to be *equal* to the pseudomomentum flux (which in turn can be equated to the outer product of the group velocity and the pseudomomentum density) without incurring serious errors in calculating wave-induced forces. This is point (b) above, sometimes called the "pseudomomentum rule". It is a consequence of Kelvin's circulation theorem (Andrews and McIntyre, 1978; McIntyre, 2000a; Bühler, 2000). For the most recent update see Bühler and McIntyre (2001).

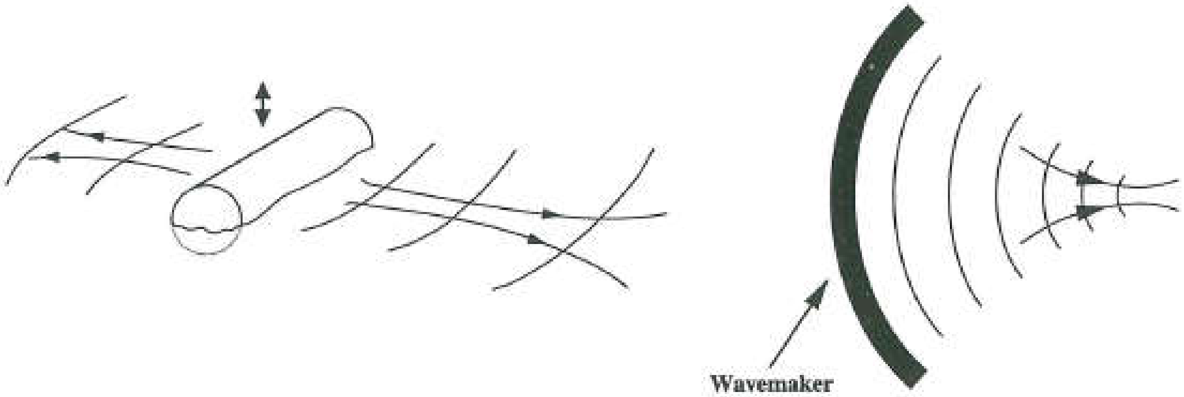

FIGURE 3 Simple experiments with water waves, illustrating the wave-induced momentum transport (here manifested by the appearance of a strong mean flow directly driven by it) that results from the generation of waves in one place and their dissipation in another (Section C). The mean flow can be made visible by sprinkling a little powder such as chalk dust on the surface of the water. In experiment (a), a cylinder about 10 cm long and 4 cm in diameter is oscillated vertically. Experiment (b) uses a curved wavemaker about 60 cm in radius and arc length. Good results are easily obtained with capillary–gravity waves having frequencies ~5 Hz. (From McIntyre and Norton, 1990.)

A spectacular case of wave-induced momentum transport over still greater distances was demonstrated by Walter Munk and colleagues (Snodgrass *et al.*, 1966) in a famous observational study of surface gravity waves propagating from the Southern Ocean across the entire Pacific, all the way to beaches in Alaska, where the waves, breaking in the near-shore surf zone, generate longshore mean currents in fundamentally the same way as in Fig. 3. In other words, mean forces exerted on the Southern Ocean, by winds in the latitudes of the "roaring forties" and "fifties", were shown to drive mean currents off the coast of Alaska.* Even more than Fig. 3b, this underlines the point that wave-induced momentum transport is a long-range process.

D. THE QUASI-BIENNIAL OSCILLATION (QBO)

Fundamentally similar effects occur with, for example, internal gravity waves in a stably stratified fluid. Figure 4 shows schematically the apparatus for the celebrated Plumb–McEwan experiment (Plumb and McEwan, 1978), in which the effects manifest themselves in an especially interesting form, which I think must surprise anyone encountering them for the first time. A flexible membrane is oscillated in a standing wave, at the bottom of an annular container filled with a stably stratified salt solution. The imposed conditions are close to mirror-symmetric. There is no externally-imposed difference between the clockwise and anticlockwise directions around the annulus; in this experiment, Coriolis effects are negligible. When the bottom boundary is oscillated, with a period exceeding $2\pi/N$, the buoyancy or Brunt–Väisälä period of the stable stratification, internal gravity waves are generated. They can be regarded as a superposition of two progressive modes, clockwise and anticlockwise. They propagate upward and dissipate mainly by viscosity, salt diffusivity being negligible for this purpose. In a larger-scale apparatus they would dissipate by wave breaking, but to my knowledge such an experiment has not been done.

Naive symmetry arguments suggest that nothing interesting will happen. But, when the experiment is run at sufficient wave amplitude, the mirror symmetry is broken spontaneously and a conspicuous mean flow around the annulus is generated. This goes robustly into a quasi-periodic sequence of states, in which the mean flow keeps reversing, and only the long-time-average state is approximately mirror-symmetric. Sufficient amplitude means sufficient to overcome viscous drag on the mean flow. The experiment has been repeated at the University of Kyoto, and the results are available on the web along with advice on experimental technique; see caption to Fig. 4.

The mechanism operating in the experiment is simple and well understood (Plumb, 1977). It involves irreversible wave-induced transport of angular momentum, dependent on the viscous dissipation of the waves, and on wave refraction and Doppler shifting by the

---

* For a careful discussion of why the photon analogy is still only an analogy, see especially McIntyre (1993).

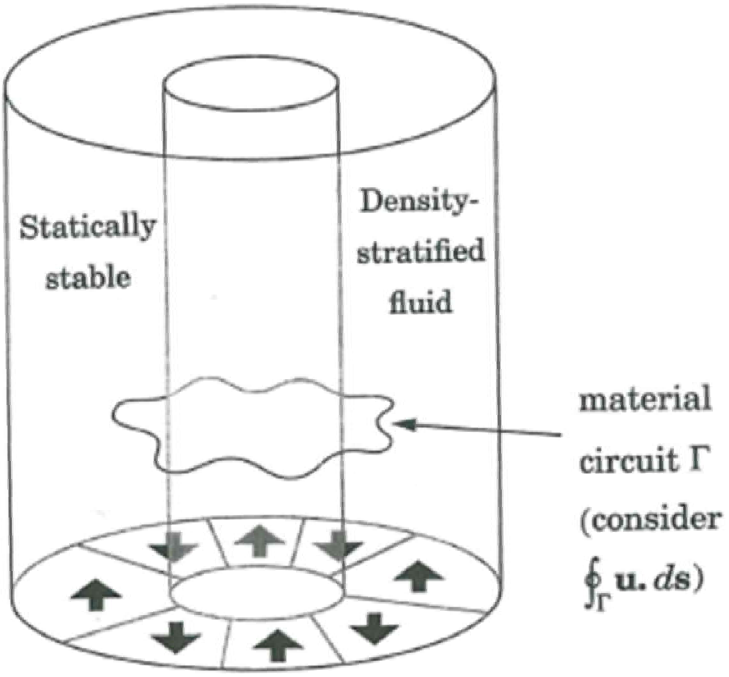

FIGURE 4 Schematic of the Plumb–McEwan experiment (Plumb and McEwan, 1978), showing how a reversing mean flow, first clockwise then anticlockwise around the annulus, can be generated by nothing more than standing oscillations of a flexible membrane at the bottom of an annular container filled with a stably stratified salt solution. The standing oscillations generate internal gravity waves at periods around 20 s $> 2\pi/N$, the buoyancy or Brunt–Väisälä period of the stable stratification. The annular gap width in the original apparatus was 12 cm, which is smaller than sketched here, relative to the outer radius, 30 cm. Also, the original apparatus had twice the number of flexible segments, giving 8 full wavelengths around the annulus. When gap widths are made significantly less than 12 cm, with water as the working fluid, the experiment fails because of viscous drag on the mean flow. For a useful set of notes on how the experiment was successfully repeated at Kyoto University, see the "Inside Stories" and "Tech Tips" at http://www.gfd-dennou.org/library/gfd_exp/exp_e/ under the heading "QBO", where a movie of the experiment running (.avi file format) can also be seen.

resulting shear flow, operating in a feedback loop. The Doppler shifting leads to enhanced dissipation, reduced vertical group velocity, and reduced vertical penetration—or *filtering*, as it is often called—of the clockwise and anticlockwise progressive wave components alternately. This gives rise to a characteristic space–time signature in which features such as zero crossings in the mean velocity profile propagate inexorably downward. The behaviour is clearly anti-frictional: if one calculates an eddy viscosity as momentum flux divided by mean shear, then one gets nonsensical negative and infinite values. The fluctuations drive the system away from, not toward, solid rotation.

It is now recognized as no accident that a similar space–time signature is observed in the real atmosphere, in the form of the famous quasi-biennial oscillation (QBO) of the zonal winds in the equatorial lower to middle stratosphere. The overwhelming weight of evidence, from observation and modelling, says that this flow in the real atmosphere is a wave-driven mean flow, like the laboratory experiment, even though there are still, even today, large uncertainties as to exactly which wave types are significant (e.g. Baldwin *et al.*, 2001), and large uncertainties as to the nature and strength of their sources and as to whether some of them propagate horizontally as well as vertically (Andrews and McIntyre, 1976; Dunkerton, 1983).

By a strange coincidence, the whole oscillation, with features descending at variable rates of the order of one or two scale heights per year, takes place in air *rising* at nearly the same rate, as illustrated in Fig. 2. Wave-induced angular momentum fluxes have to be almost twice what would otherwise be required (Gray and Pyle, 1989; Dunkerton, 1991). This point seems to have been missed in early work on the QBO, which also tended to consider too small a set of wave types. The important wave types include, but are not restricted to, the equatorial Kelvin wave and other vertically propagating waves trapped in an equatorial waveguide (Matsuno, 1966; Gill, 1982) by the "potential well" in the Coriolis parameter $f = 2|\Omega|\sin(\text{latitude})$, twice the vertical component of the earth's angular velocity Ω, which is picked out by the atmosphere's stable stratification $N^2 \gg 4|\Omega|^2$.

E. THE MICHELSON–MORLEY PRINCIPLE

So how, then, despite the uncertainties about wave types and wave sources, can we be so confident that the real QBO is a wave-driven mean flow? In my view an important part of the answer goes back to the seminal paper of Wallace and Holton (1968). It points up the importance of negative results in science—one might call it the Michelson–Morley principle—a principle neglected today, but neglected at our peril.

I have discussed this more fully elsewhere (McIntyre, 1994, 2000a); but the essence of the matter is that strenuous efforts to model QBO dynamics without wave-induced momentum transport—by scientists who well knew what they were doing and were well aware of observational constraints—failed in a significant way. That failure greatly adds to our confidence that wave-induced momentum transport is essential. Such transport had not yet been thought of at the time; and Wallace and Holton were led to a surprising, and no doubt disconcerting, conclusion—that it was impossible to explain the QBO without postulating a mysterious zonal force field that descended as time progressed and displayed an anti-frictional character. Despite being unable to guess what physical processes might give rise to such a force field, Wallace and Holton rather courageously published their negative result. It was only then recognized, increasingly clearly, through the work of Lindzen and others (Lindzen and Holton, 1968; Wallace and Kousky, 1968; Holton and Lindzen, 1972; Plumb, 1977), that there is just one kind of physical process in the atmosphere, wave-induced momentum transport, that can and should produce the otherwise mysterious force through filtering by wave refraction and Doppler shifting. With a suitable spread of phase speeds (Saravanan, 1990), one easily gets anti-frictional behaviour and a space–time signature of just the kind observed.

The essential behaviour does not, incidentally, depend on the existence of "critical levels" (an extreme case of wave filtering), nor on influence from above, by the semi-annual oscillation for instance, as was once thought (Plumb, 1977). These points are underlined by the results of the laboratory experiment. In the experiment, there are no critical levels, nor any phenomenon resembling the semi-annual oscillation. The experiment also illustrates anti-frictional behaviour in its purest form: the system is driven away from solid rotation by literally nothing other than the fluctuations generated by the oscillations of the lower boundary.

F. THE NONACCELERATION CONSTRAINT

Theoretical modelling of the feedback loop or wave–mean interaction giving rise to the anti-frictional behaviour is usually based on zonal averaging. The most convenient formulation uses the so-called TEM (transformed Eulerian-mean) equations, in which the rate of wave-induced momentum transport is quantified by an accurate standard diagnostic, the Eliassen–Palm flux (Andrews *et al.*, 1987; Andrews and McIntyre, 1976). According to the TEM equations, one may picture the dynamics in terms of a zonally symmetric "mean fluid motion" in which notional rings of fluid are pushed eastward or westward by the wave-induced force, $\bar{F}$ say. The force is given quantitatively by the divergence of the Eliassen–Palm flux, which vanishes under so-called "nonacceleration conditions".

The most important ways for such conditions to be violated—this is sometimes called "breaking the nonacceleration constraint" (McIntyre, 1980; McIntyre and Palmer, 1985)—are for the waves to be breaking or otherwise dissipating. The nonacceleration constraint can be regarded as coming from Kelvin's circulation theorem, which is applicable to closed material contours or circuits undulated by the wave motion, such as the contour Γ sketched in Fig. 4. For instance the mean-flow changes in the Plumb–McEwan experiment are associated with changes in the Kelvin circulation, $\oint_\Gamma \mathbf{u}\cdot d\mathbf{s}$, caused by the viscous forces associated with wave dissipation. The Kelvin circulation would otherwise be invariant. Here $\mathbf{u}$ is the three-dimensional velocity field and $d\mathbf{s}$ the element of arc length. The role of material contours like Γ is also, incidentally, the reason why Lagrangian means arise naturally in the theory of wave–mean interaction (Andrews and McIntyre, 1978; Bühler and McIntyre, 1998; McIntyre, 2000a; Bühler, 2000).

G. EXTRATROPICAL LATITUDES

With wave-induced angular momentum transport so conspicuously important near the equator, as illustrated by the real QBO, it is perhaps not surprising that such transport should turn out to have a role in higher latitudes as well—though just how central a role was not recognized until recently, outside small communities of specialists who gradually developed the needed fluid-dynamical understanding. With hindsight, those developments can be traced back to (a) the work of Kelvin, Hough, and Rossby on what are now called Rossby waves or vorticity waves, which turn out to be the principal wave type involved; (b) the discovery by Rossby (1936, 1940) and Ertel (1942) of the concept of potential vorticity (PV), measuring the Kelvin circulation $\oint_{\Gamma'} \mathbf{u} \cdot d\mathbf{s}$, for small circuits Γ' lying on an isentropic surface; (c) the recognition by Charney (1948) and Kleinschmidt (1950) of what is now called PV invertibility, allowing Rossby-wave dynamics to be understood intuitively, beyond idealized cases (Hoskins *et al.*, 1985); (d) the recognition of large-scale anti-frictional behaviour, principally by Jeffreys in the 1920s and by Starr and co-workers in the 1950s; and (e) the attempts by Taylor, Rossby, and others (echoed for instance in Eady's 1950 RMS Centenary paper) to develop a "vorticity-transfer" view of turbulence stemming from Taylor's classic paper on eddy motion in the atmosphere (Taylor, 1915). The TEM theory, providing important conceptual simplifications within a coherent mathematical framework, began to take shape in the work of Eliassen and Palm (1961), Charney and Drazin (1961) and Dickinson (1969). Dickinson's seminal and perceptive work gave strong hints, moreover, that it would be fruitful to consider the waves and turbulence together—as complementary, interacting pieces of a single jigsaw puzzle. The first fully consistent theoretical model to capture all the essential aspects of such a jigsaw was the "nonlinear Rossby wave critical-layer theory" of Stewartson, Warn and Warn (Killworth and McIntyre, 1985), hereafter "SWW theory", which used matched asymptotic expansions to represent the dynamics of the interacting pieces for certain idealized cases of breaking Rossby waves. We return briefly to the SWW theory in Section J.

H. GYROSCOPIC PUMPING

Away from the equator the response to the zonal force $\bar{F}$ depends on Coriolis effects, which, as noted earlier, are immensely strong for present purposes. The equator is exceptional because everything is dominated there by the still stronger stable stratification, as reflected in the sine-latitude dependence of the Coriolis parameter f and in the large values of Prandtl's ratio

$$N/|2\Omega| \approx 10^2 \qquad (1)$$

with N the buoyancy frequency as before, making the horizontal component of Ω unimportant. Haynes (1998) has recently clarified what "away from the equator" means for present purposes, through scale-analytical estimates and numerical experiments with given zonally symmetrical zonal forces $\bar{F}$. When typical parameters are substituted into his estimates, including height scales, say $\gtrsim 10\,\text{km}$, it is found that "away from the equator" means further than about 15° latitude. This number is insensitive to the precise choice of parameters because of a quarter-power dependence on time and height scales and on radiative relaxation times. An interesting corollary of Haynes' results is that there is no need for the QBO to depend exclusively on forces $\bar{F}$ confined to the tropics.

Away from the equator, the most important part of the response to a given $\bar{F}$ is what is now called the "wave driving", or "gyroscopic pumping", of the mean meridional circulation. If a ring of fluid feels a persistent westward push, for instance, as suggested in Fig. 5, then Coriolis effects tend to deflect it persistently poleward—a mechanical pumping action that depends on the earth's rotation. This is the essence of what is happening at most altitudes and latitudes in Fig. 1. At top left, above the frame of Fig. 1, near the summer mesopause, the situation is similar apart from sign: because of gravity-wave filtering, the wave-driven push is eastward and the air is pumped equatorward. This in turn pulls air upward in the polar cap, producing the strong refrigeration already noted.

We can learn more about the fluid dynamics involved through thought experiments and numerical experiments with given forces $\bar{F}$, applied to a zonally symmetric model atmosphere away from the equator. Because of the tendency toward anti-frictional behaviour of the real wave-induced forces, this is arguably a more useful experiment than the kind in which $\bar{F}$ is taken to be a friction force. The mean circulation that arises in response to the given $\bar{F}$ can be identified with the "residual mean" meridional and vertical velocities $(\bar{v}^*, \bar{w}^*)$ that appear in the TEM equations.

The vertical velocity $\bar{w}^*$ is not, incidentally, the same as a Lagrangian mean. But it has in common with Lagrangian means—in the restricted circumstances in which the latter are well defined (Sparling *et al.*, 1997; Thuburn and McIntyre, 1997)—the property of pushing or pulling temperatures above or below radiative equilibrium temperatures T_{rad}, as assumed in Section B. This property can be seen from the TEM equations, in which eddy heat fluxes can be neglected under typical parameter conditions (Andrews *et al.*, 1987). It is this same property that was used in deriving Fig. 1 from observed temperatures, with the help of a sophisticated radiation scheme. The $\bar{w}^*$ field and the associated meridional velocity $\bar{v}^*$ required to produce the observed departures of zonal-mean temperatures from T_{rad} are what were actually calculated, with appropriate corrections to ensure mass conservation (Shine, 1989), then converted into the corresponding stream function. For

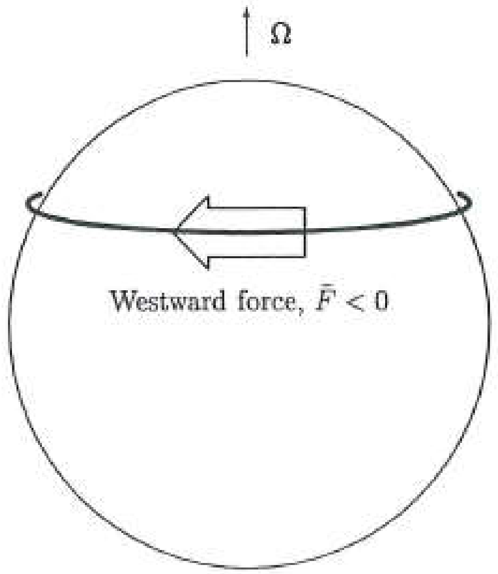

FIGURE 5 The gyroscopic pumping mechanism, in the case $F < 0$ (appropriate to Rossby waves everywhere and to internal gravity waves in the winter mesosphere). When a ring of air located away from the equator is pushed westwards, Coriolis effects try to deflect it poleward. This is the mechanical pumping action that drives the mean circulation against radiative relaxation. The picture reminds us that the old idea of a sun-driven "Hadley" circulation in the middle atmosphere cannot, for instance, explain why the Brewer–Dobson cells look nothing like the Murgatroyd-Singleton cell at solstice.

related reasons, $\bar{w}^*$ is also a better measure of the mean vertical advection of chemical tracers than is the Eulerian mean, $\bar{w}$.

What happens when one starts the gyroscopic pump? More precisely, what do the equations say will happen in a thought experiment in which one starts exerting a given, zonally symmetrical, steady westward force, $\bar{F} < 0$, in some region occupying finite ranges of altitude and latitude away from the equator? To a first approximation the answer is that one sees a time-dependent adjustment in which the motion gradually settles toward steady poleward flow $\bar{v}^*$ through the forcing region, whose Coriolis force $f\bar{v}^*$ comes into balance with the force $\bar{F}$ as the steady state is approached. Away from the equator, this behaviour is robust.

In the final steady state, the mass streamlines close entirely downward (Haynes *et al.*, 1991, 1996). This last aspect of the response pattern is sometimes called the "downward control" property. The word "control" does not mean anything absolute, of course, but simply refers to what happens in this particular thought experiment, or numerical experiment, in which $\bar{F}$ is given and in which attention is focused on the circulation pattern $(\bar{v}^*, \bar{w}^*)$ that arises in response to the given $\bar{F}$.

The steady-state response can be pictured as resembling the lower part of Fig. 1, in a typical experiment, if we choose an $\bar{F}$ distribution with suitable negative values. If, for instance, we choose $\bar{F}$ to be negative within, but zero beneath, the frame of the figure then, during the adjustment toward the steady state, the lower part of the streamline pattern gradually burrows downward until it reaches the earth's surface, where it closes off in a frictional boundary layer. This downward burrowing is an aspect of the downward control property that will be important for our discussion of the solar interior.

The downward burrowing is fastest, and the steady state is approached fastest, for broad, deep, global-scale $\bar{F}$ distributions covering a large range of latitudes, having low-order zonally symmetrical Hough function structure. The approach to steadiness is then usually, for practical purposes, somewhat faster than seasonal timescales in the middle and upper stratosphere at least. Subglobal scales, on the other hand, represented by higher-order Hough functions, take much longer to adjust. Haynes *et al.* give a careful analysis and discussion of the finer points of the adjustment process and of its timescales and their relation to latitudinal structure and to f^{-1}, N^{-1}, and radiative relaxation times. There is some further discussion of the time-dependent behaviour in Holton *et al.* (1995) and in Haynes (1998).

In the most accurate form of the theory, the steady-state balance $\bar{F} = -f\bar{v}^*$ is replaced by $\bar{F} = -f_{\bar{M}}\bar{v}^*$ where $f_{\bar{M}}$ is the latitudinal gradient of $\bar{M}$, the zonal-mean specific angular momentum, divided by the perpendicular distance to the earth's axis. One can then have steady-state "sideways control" (Dunkerton, 1991), to the extent that the isopleths of $\bar{M}$ slope sideways as must happen most significantly in parts of the tropics, where f is relatively small.

The downward control property is related to the finiteness of the pressure and density scale heights, i.e., to the exponential falloff of pressure and density with altitude. Again glossing over some technicalities (Haynes *et al.*, 1991), we have a situation in which the finite mass of atmosphere above the forcing region is incapable of absorbing infinite amounts of angular momentum. In the parameter regimes of interest the system solves this problem, so to speak, by approaching a steady state in which all the angular momentum being supplied by $\bar{F}$ is sent downward into the solid earth, whose moment of inertia is regarded as infinite for this purpose. Of course the ability to approach a steady state at all depends on the relaxational character of infrared radiation.

If one gets too close to the equator, then not only can the $\bar{M}$ isopleths slope sideways, but their pattern can be substantially reshaped as part of the response to $\bar{F}$. Advection of the relative contribution to $\bar{M}$, as distinct from the Coriolis contribution, makes the problem nonlinear. A possible consequence is that the response to the given $\bar{F}$ may no longer be robust, or even unique, though mass conservation still demands that air must be pulled up from somewhere not too far from the equator. This is at the research frontier of atmospheric dynamics and is not yet well understood, though progress is being made (Haynes, 1998; Scott and Haynes, 1998; Sankey, 1998; Plumb and Eluszkiewicz, 1999). There are simplified (shallow-water) models that do exhibit nonuniqueness in the form of hysteretical behaviour—multiple equilibria, regime flipping (Sankey, 1998)—in the upward path taken by the tropical or subtropical branch of the circulation that arises in response to a given $\bar{F}$, though we do not yet have a clear view of the significance for stratified flow.

The path taken is also influenced, within the tropics, by the distribution of radiative equilibrium temperatures T_{rad} and therefore of solar heating across the tropics. In terms of the tape-recorder analogy, changing the solar declination is a bit like moving the guidance wheels of the tape recorder, again affecting the path taken by the rising air within the tropics but not, to a first approximation, the total mass flux demanded by the gyroscopic pumping that is extracting the air from the tropics. Such a picture is consistent with the observational evidence from MSU channel 4, mentioned in Section B, on the seasonal variation in the upwelling mass flux (Yulaeva *et al.*, 1994).

In the analogy we may think of the gyroscopic pumping action as the tape recorder's "motor". Of course from an old-time audio enthusiast's viewpoint the tape recorder is a poor quality one, with no "capstan" and with much "wow and flutter" in the form of seasonal, intraseasonal, and interannual variability. And the nonlinearity of the tropical upwelling problem, and its possible hysteretical behaviour, mean that the guidance wheels are sloppy and allow lots of "play" in the path taken.

I. ROSSBY AND GRAVITY WAVE BREAKING

Which wave types are most important for the global-scale circulation in the real extratropical middle atmosphere? Here, as hinted earlier, we think we know more than in the case of the tropics and the QBO. Below 50 km or so in the winter hemisphere, and below 25 km or so in the summer hemisphere, it seems clear from observational and modelling studies that the circulation is driven mainly by large-scale Rossby waves propagating or diffracting upwards.

Because of their strong chirality or handedness, tied to the sense of the earth's rotation, Rossby waves have ratchet-like properties that tend to produce persistently westward, or retrograde, mean forces whenever and wherever the waves are dissipated. This explains the persistently poleward mean motion in the Brewer–Dobson cells illustrated in Fig. 1. The two principal ways in which Rossby waves are dissipated are by infrared radiative damping and by wave breaking. Usually, in the real atmosphere, there is an intimate combination of both. For Rossby waves, as for all the other wave types of interest, "breaking" is most aptly defined as the rapid and irreversible deformation of those material contours that would otherwise undulate reversibly in the manner described by linearized wave theory (McIntyre and Palmer, 1984, 1985), as with the contour Γ sketched in Fig. 4. "Rapid" means on timescales comparable to an intrinsic wave period or less.

This definition of wave breaking is very widely applicable. It applies to the surface capillary–gravity waves of Fig. 3 and to breakers on Alaska beaches, as well as to Rossby waves and to the internal gravity waves of Fig. 4. It applies regardless of whether wave breaking is mediated by dynamical instabilities, as is usual with internal gravity waves, or not, as with plunging ocean-beach breakers. In ocean-beach breakers there is no identifiable instability process. What is common to all cases is that the irreversible deformation of material contours like Γ is another way to "break the nonacceleration constraint". When Rossby waves break, the resulting motion can often be described as layerwise two-dimensional turbulence, constrained, like the waves themselves, by the strong stratification. The typical consequence is again to drive the system away from, not toward, solid rotation as we shall see shortly.

Observational and model studies quantitatively diagnosing the Eliassen–Palm flux show that the forces due to breaking Rossby waves are indeed sufficient to drive the poleward flows in Fig. 1, in the Brewer–Dobson cells (Rosenlof and Holton, 1993; Holton *et al.*, 1995), though internal gravity waves may make a contribution that could also be significant. Gravity waves may be especially significant in the summertime upper stratosphere where Rossby-wave activity is weak and $\bar{v}^*$ is small (Rosenlof, 1996; Alexander and Rosenlof, 1996), and perhaps also in the subtropical stratosphere where a relatively small $\bar{F}$ can produce significant gyroscopic pumping, because of the relatively small values of $\bar{M}$ gradients. Above 50 km or so, internal gravity waves take over as the most important wave type (e.g. Lindzen, 1981; Holton, 1982; Hamilton, 1997). Unlike Rossby waves, internal gravity waves can of course produce mean forces that are either prograde or retrograde, depending on wave filtering, as in the Plumb–McEwan experiment. Observations of internal gravity waves in the mesosphere, mostly by radar and lidar techniques, supplemented by rocket soundings, suggest that in order-of-magnitude terms, at least, such waves are well able to drive a Murgatroyd–Singleton circulation qualitatively like the heavy dashed curve in Fig. 1 and strong enough to make noctilucent clouds. The waves are observed to dissipate mainly by breaking due to unstable convective overturning, of heavy air over light, as clearly seen in the formation of adiabatic layers (e.g. Fritts *et al.*, 1988) and in the mixing of certain chemical tracers such as atomic oxygen (C. R. Philbrick, personal communication).

J. THE JIGSAW PUZZLE: BAROTROPIC MODELS

Rossby-wave dynamics, including propagation, breaking, and the associated momentum and angular-momentum transport—the whole wave–turbulence jigsaw puzzle—can be illustrated by the simple textbook case of barotropic nondivergent vortex dynamics in Rossby's "beta plane" or "approximately flat earth" model. In place of the PV we have the absolute vorticity $Q(x,y,t)$, say. Invertibility holds exactly: a stream function $\psi(x,y,t)$ describing the incompressible velocity field can be derived by from the Q field by inverting a Poisson equation, given suitable boundary conditions such as evanescence at infinity:

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

Here $f(y)$ is again the Coriolis parameter, and

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

where (x,y) are eastward and northward Cartesian coordinates and (u,v) the corresponding components of the velocity vector $\mathbf{u}(x,y,t)$. There is just one evolution equation

$$DQ/Dt = 0 \tag{4}$$

where D/Dt is the two-dimensional material derivative, defined by

$$D/Dt = \partial/\partial t + \mathbf{u}\cdot\nabla = \partial/\partial t + u\partial/\partial x + v\partial/\partial y \tag{5}$$

If desired, one can model dissipation by adding terms to the right of equation (4). Note incidentally that invertibility, as always, entails nonlocality, i.e. instantaneous action-at-a-distance—perfectly accurate in this case because all fast waves, such as acoustic waves and gravity waves, have been suppressed by giving them infinite propagation speeds. In this implicit sense, the system is infinitely stiff: there is no delay in the pushing of one fluid element by another, all the way to infinity. Note also the single time derivative, in the single evolution equation (4), pointing to the chirality and ratchet-like properties of Rossby waves.

We may regard the system (2)–(5) as describing motion on an isentropic or constant-θ surface in a strongly stratified atmosphere, in the limit of infinite buoyancy frequency N, where θ is potential temperature. Figure 6 shows the characteristic pattern of a Rossby-wave disturbance, in which constant-Q contours, Γ_{Q0}, say, lying on the isentropic surface—which according to (4) are also material contours—undulate in the ±y-direction, corresponding to north–south on the real earth and implying a pattern of Q anomalies from which inversion, (2), (3), yields a velocity field qualitatively as shown by the solid arrows (Hoskins *et al.*, 1985). If one now makes a mental movie of the resulting motion, one sees that the undulations will propagate westward because the velocity field is a quarter wavelength out of phase with the displacement field. This is the Rossby-wave propagation mechanism, sometimes called "quasi-elasticity" to emphasize that it is a restoring effect, tending to return parcels to their equilibrium latitudes, but perhaps better called "semi-elasticity" because of the chirality.

The standard Rossby wave theory—linearize equation (4) about rest, take $\beta = df/dy$ = constant, look for

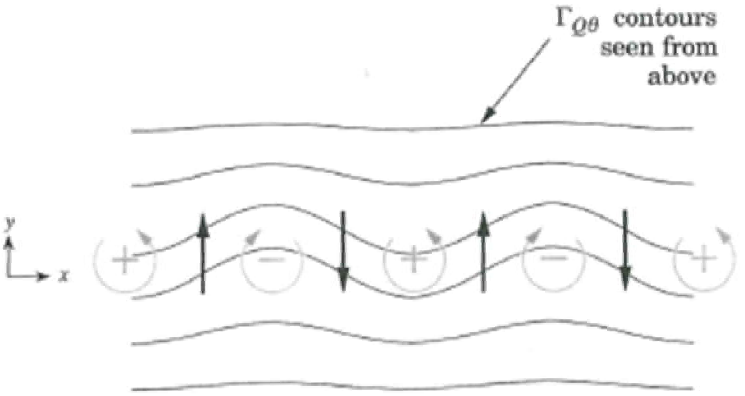

FIGURE 6 Plan view of the chiral or one-way restoring mechanism or "quasi-elasticity" to which Rossby waves owe their existence (diagram taken from Hoskins *et al.*, 1985). The mechanism has a central role in nearly all large-scale dynamical processes in the atmosphere and oceans, including barotropic and baroclinic instabilities, vortex coherence (e.g. blocking), and vortex interactions.

solutions $\propto \exp(ikx + ily - i\omega t)$ so that $\nabla^{-2} = -(k^2 + l^2)^{-1}$, and deduce $\omega = -\beta k(k^2 + l^2)^{-1}$—presumes that the sideways displacements and slopes of the undulating contours $\Gamma_{Q\theta}$ are small and that the undulations are reversible. Rossby-wave breaking is the opposite extreme situation, in which this condition of reversible undulation is violated, and the contours $\Gamma_{Q\theta}$ deform irreversibly; the relevant arrow-of-time "paradox" is discussed in McIntyre and Palmer (1984).

Now if the Rossby wave undulations take place in a mean or background shear flow $\bar{\mathbf{u}} = (\bar{u}(y), 0)$, and have a monochromatic zonal (x) phase speed that coincides with the background flow velocity at some value of y—by convention called a critical line—then there is always some surrounding region in which Rossby-wave breaking, in the foregoing sense, is important. This region is sometimes called a "Rossby-wave surf zone". Such zones can be broad or narrow according as the wave amplitude is large or small. In the narrow case, the SWW theory and its further development in Killworth and McIntyre (1985) and in Haynes (1989) has given us a comprehensive understanding of how the wave field and surf zone interact, when modelled using equations (2)–(5). The results illustrate in detail how the nonacceleration constraint can be broken even in a dissipationless fluid system such as (2)–(5).

The idealized problems considered in the SWW theory and its further development are problems in which a small-amplitude Rossby wave, excited by an undulating side boundary, encounters a critical line in a basic flow initially having constant shear. A narrow surf zone forms in which the contours $\Gamma_{Q\theta}$ deform irreversibly in a recirculating flow, a so-called Kelvin cat's-eye pattern, straddling the critical line. As the contours wrap around, the Q distribution changes and, partly through the action-at-a-distance implied by the inversion, i.e. by the ∇^{-2} operator in equation (2), induces a change in the flow outside the surf zone. This affects phase gradients with respect to y in the outer flow, hence trough tilts and Reynolds stresses $\overline{u'v'}$, in such a way that the surf zone appears as an absorber to the Rossby waves outside it during the early stages of wave breaking, but later becomes a reflector because of the finite scope for rearrangement of the Q distribution within a surf zone of finite width. The way in which the surf zone interacts with its surroundings is precisely expressed by the use of matched asympotic expansions. Full details of the analysis are too lengthy to reproduce here but can be found in the review material of Killworth and McIntyre (1985), which also establishes a general theorem implying that the absorption–reflection behaviour is generic.

The essence of what happens is, however, very simple, and is summarized in Fig. 7. The sketch on the left, Fig. 7a, shows the background $\bar{Q}$ profile, $\bar{Q}$ = const. + βy, superposed on the result of rearranging it in such as way as to homogenize it over some finite y-interval. This mimics in simplified form what happens in the surf zone. Figure 7(b) is the zonally (x-)averaged $\bar{Q}$ profile from an actual solution, in a case where the flow in the surf zone is chaotic because of the onset of secondary instabilities (Haynes, 1989, and personal communication). To get the corresponding zonal mean momentum change, the inversion implied by (2)–(3) is trivial: one merely has to integrate with respect to y the mean change $\delta\bar{Q}$ in $\bar{Q}$:

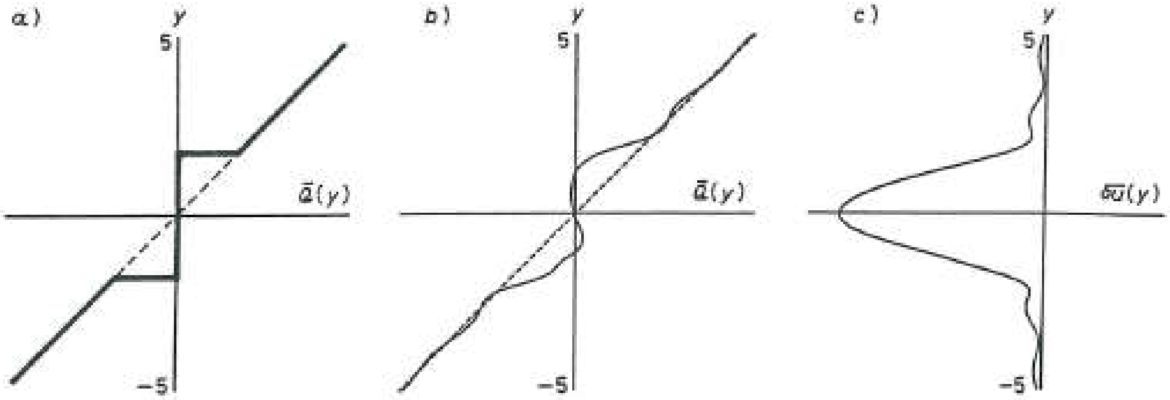

FIGURE 7 The relation between mean wave-induced force and Q-rearrangement by a breaking Rossby wave, in the simplest relevant model system, the dynamical system (2)–(5). (Courtesy P. H. Haynes; for mathematical details see Killworth and McIntyre (1985) and Haynes (1989).) Plot (a) shows idealized $\bar{Q}$ distributions before and after mixing Q in some y-interval or latitude band; (b) shows the x-averaged $\bar{Q}$ distribution in an actual model simulation using equations (2)–(5); (c) shows the resulting mean momentum deficit, equation (6), whose profile would take a simple parabolic shape in the idealized case corresponding to (a). Notice that the momentum has been transported against its own mean gradient, because the wave source is at positive y and the background shear is positive.

$$\delta\bar{u}(y) = \int_y^\infty \delta\bar{Q}(\tilde{y})d\tilde{y} \tag{6}$$

The result is shown in Fig. 7c. The momentum change corresponding to the simplified, left-hand $\bar{Q}(y)$ profile in Fig. 7a is a parabolic shape, not shown, qualitatively similar to the right-hand graph. The parabolic shape results from integrating the linear $\delta\bar{Q}$ profile. More generally, it is plain from (6) that any Q rearrangement that creates a surf-zone-like feature with weakened mean Q gradient will give rise to a momentum deficit. This is another clearcut example of anti-frictional behaviour, because $\delta\bar{u}(y)$ has to be added to a background flow $\bar{u}(y) \propto y$ with positive shear $\partial\bar{u}(y)/\partial y > 0$ (not shown in the figure), and the wave source is located at positive y, outside the domain of the figure. Thus the momentum has been transported against its own mean gradient.

The mean momentum tendency is given by

$$\frac{\partial\bar{u}}{\partial t} = -\frac{\partial}{\partial y}\overline{(u'v')} = \overline{v'Q'} \tag{7}$$

where the overbars denote the Eulerian zonal mean and primes denote fluctuations about that mean. This makes explicit the relation between the mean momentum tendency, the wave-induced Reynolds stress convergence, and the eddy flux of Q measuring the rate of rearrangement of the $\bar{Q}$ distribution in the surf zone. The second equality, a corollary of (3) and the relation $Q' = \partial^2\psi'/\partial x^2 + \partial^2\psi'/\partial y^2$ from (2), was given in Taylor's (1915) classic paper and is often called the "Taylor identity".

The essential point, then, is that whenever the Q distribution is rearranged by Rossby-wave breaking, a momentum deficit appears—signalling the presence of a retrograde force due to irreversible wave-induced momentum transport. This illustrates not only anti-frictional behaviour but also the ratchet-like character of the whole process, ultimately a consequence of the single time derivative in (4). The momentum change where Rossby waves dissipate by breaking is always retrograde, i.e., against the background rotation, westward in the case of the earth. If there is some forcing effect that tends to restore the background Q gradient (requiring a nonzero term on the right of (4)), then the Q rearrangement and the irreversible momentum transport can persist.

A feature that is robust, and appears to carry over to larger-amplitude cases, is the strong spatial inhomogeneity of the "jigsaw puzzle". The flow is more "wave-like" in some places (the outer flow) and more "turbulent" in others (the surf zone), where "turbulent" means that the material contours deform irreversibly. Each region strongly affects the others dynamically. More realistic cases with larger wave amplitude typically show the same generic features. Figure 8 shows some material tracer fields for such a case, mimicking conditions in the real lower to middle stratosphere by going to a more realistic spherical geometry and to divergent (shallow-water) barotropic flow (see caption and Norton, 1994; McIntyre and Norton, 2000). The inhomogeneity is very striking indeed. The midlatitude surf zone is strongly turbulent, while its borders are relatively wave-like. This is especially so at the inner

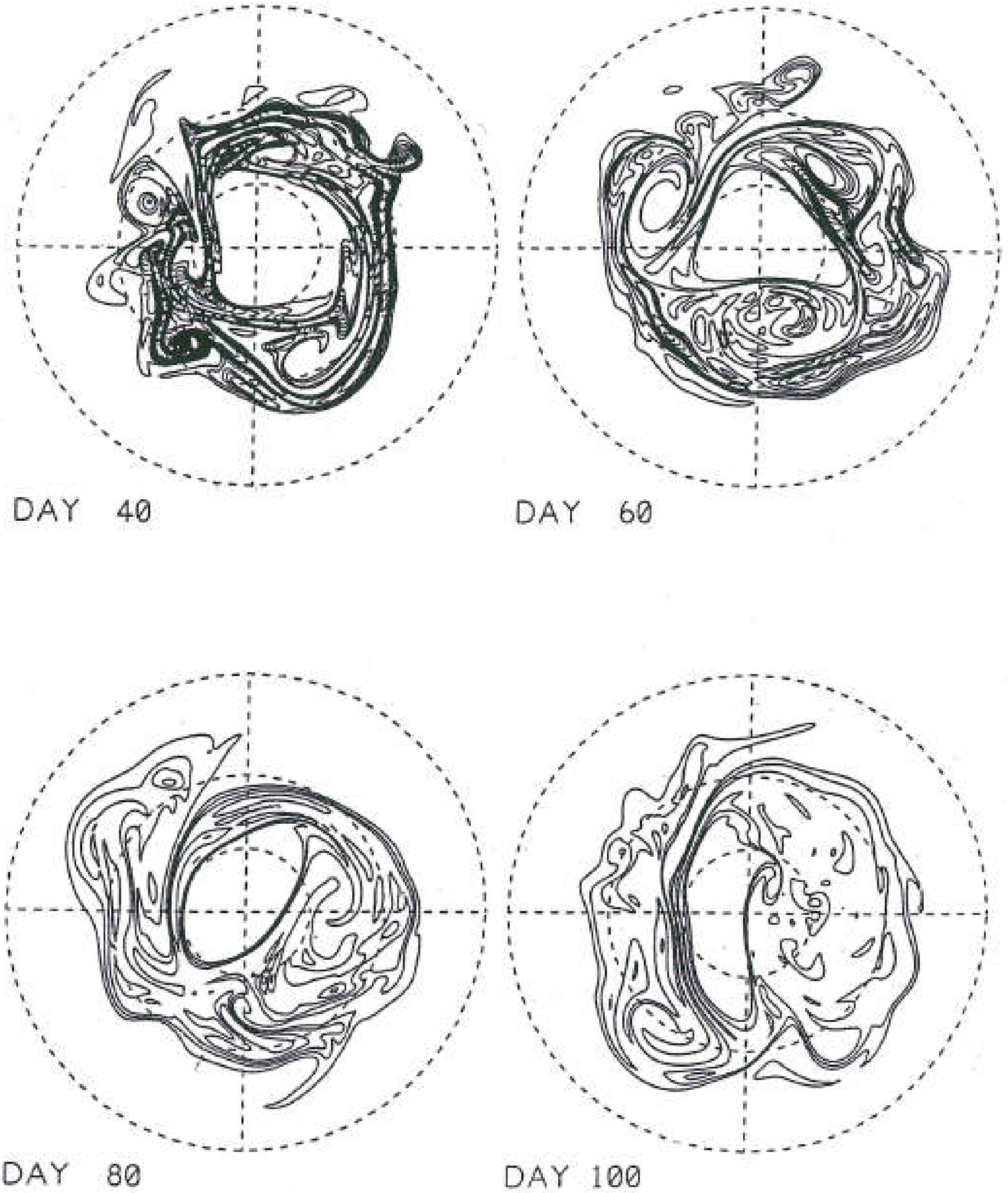

FIGURE 8 Model flow on the sphere, closely resembling flow in the real wintertime stratosphere at altitudes around 25 or 30 km. The map projection is conformal (polar stereographic), with the equator and the 30°N and 60°N latitude circles shown dashed. The flow is visualized by passive tracer released as a compact blob into the midlatitude stratospheric surf zone, clearly showing the fast two-dimensional turbulent mixing in that region, despite which the stratospheric polar vortex remains almost completely isolated from its surroundings, and likewise, to a lesser extent, the tropics (helping to explain what is seen in Fig. 2). The isolation of the (core of the) polar vortex recalls classic smoke rings and is of great importance to stratospheric polar chemistry, including the Antarctic ozone hole and its (so far less severe) Arctic counterpart. The isolation is due to the combined effects of the Rossby-wave restoring mechanism and the strong shear just outside the edge. (Courtesy of Dr W. A. Norton, from whom an animated video of the model run is available (Dept of Atmospheric, Oceanic, and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, UK).) Details of the model and the model run are given in Norton (1994). A "shallow-water" model is used, with mean depth 4 km, giving behaviour qualitatively close to that of (2) except that the inversion operator is replaced by one with shorter range (Rossby length ~2000 km in middle latitudes); for justification of this last assertion, see McIntyre and Norton (2000).

border of the surf zone, marking the edge of the polar vortex—whose resilence makes it appear to behave, in animated versions of Fig. 8, almost like an elastic band. This is a clear example of what is now called an "eddy-transport barrier". The barrier effect results from the combined effects of the Rossby-wave restoring mechanism and strong shear just outside the polar vortex (Juckes and McIntyre, 1987). All this has implications not only for chemistry, in a well-known way, but also, for instance, for the dynamics of the much-discussed Holton–Tan effect, as will be noted shortly.

K. GENERALIZATION TO REALISTIC STRATIFIED FLOW

The whole picture just summarized generalizes to layerwise two-dimensional motion in a finitely stratified atmosphere provided that we reinterpret Q as the Rossby–Ertel PV

$$Q = \rho^{-1}(2\Omega + \nabla \times \mathbf{u}) \cdot \nabla\theta \tag{8}$$

where ρ is mass density, and replace ∇^{-2} in equation (2) by a more complicated, three-dimensional inversion operator (quasi-geostrophic or higher-order (Hoskins *et al.*, 1985; Davis, 1992)). Qualitatively, everything is much the same as before: equation (4) is unchanged, and its single time derivative implies the same chirality as before; Rossby quasi-elasticity still works as shown in Fig. 6; and the rearrangement of PV on isentropic surfaces is still robustly associated with an angular momentum deficit (Robinson, 1988), and with ratchet-like, irreversible angular momentum transport in the sense required to drive the Brewer–Dobson cells. The Taylor identity still holds except that the Reynolds stress divergence is replaced, to a first approximation, by a form of the Eliassen–Palm flux divergence.

The wave–turbulence inhomogeneity illustrated in Figs 7 and 8 is also found to be typical of the real, stratified atmosphere. One reason for the generic character of the inhomogeneity is its tendency to be self-reinforcing. Where Q contours are crowded together, typically at the edges of surf zones—idealized as the "corners" in the $\bar{Q}$ profile in Fig. 7a—one has a strengthening of the Rossby quasi-elasticity. Even more importantly, there is a weakening within surf zones. Thus mixing becomes easier in surf zones once they begin to be established, and *vice versa*. Other things being equal, then, the inhomogeneity tends to perpetuate itself. The structure can of course be changed by sufficiently drastic changes in circumstance, such as the introduction of strong Rossby-wave motion with a different phase speed from the original one, tending to carve out a surf zone near a different y value in the background shear (Bowman, 1996). Likewise, the QBO-related changes in tropical winds can change relative or intrinsic phase speeds and hence modulate the tropical and subtropical structure, a fact that, together with the absorption–reflection behaviour of surf zones, must change the dynamics of the winter-stratospheric "Rossby-wave cavity" (Matsuno, 1970). Such a mechanism of modulation by the QBO—a highly nonlinear mechanism—probably underlies the celebrated Holton–Tan effect (Baldwin *et al.*, 2001).

Another reason to expect the wave–turbulence inhomogeneity is, as I have argued more carefully elsewhere (McIntyre, 1994, 2000a), the fundamental integral constraint on PV rearrangement on a (topologically spherical) isentropic surface S enclosing the earth. Using the definition (8) together with Stokes' theorem on the closed isentropic surface S, we have

$$\iint_S bQ\,dA = 0 \tag{9}$$

where dA is the area element and b is defined such that $b\,dA\,d\theta$ is the mass element. That is, $b = \rho/|\nabla\theta|$. Because b is positive definite, (9) tells us at once that the only way to mix Q to homogeneity on the surface S is to make Q zero everywhere—a fantastically improbable state on a planet as rapidly rotating as the earth. Since real Rossby waves do break, and do mix Q, they must do so imperfectly and produce spatial inhomogeneity and anti-frictional tendencies.

To be sure, one can imagine a thought experiment in which the air on and near the isentropic surface S begins by rotating solidly with the earth and then has its angular velocity uniformly reduced by some incident Rossby-wave field, in such a way as to give a uniformly reduced pole-to-pole latitudinal profile of $\bar{Q}$. But the tailoring of a Rossby-wave field to do this would be a more delicate affair than standing a pencil on its tip; and the natural occurrence of such a wave field would be another fantastically improbable thing. To summarize, then, these fluid systems naturally exhibit a strong wave–turbulence inhomogeneity, and with it a marked tendency for the associated wave-induced momentum transport to behave anti-frictionally, in the sense of driving the system away from, not toward, solid rotation.

L. THE SUN'S RADIATIVE INTERIOR

1. Inevitability of a Magnetic Field

When put together with new helioseismic data, our understanding of stratospheric circulations has recently enabled us to learn something new and significant about the sun's deep interior. Here I mean the radiative interior, at radii $< 5 \times 10^5$ km, as distinct from the overlying convection zone, whose upper surface is the visible surface at about 7×10^5 km. We can now answer in the affirmative, with high confidence, an age-old question: does a dynamically significant magnetic field pervade the interior? The mere existence of such a field has immediate and far-reaching consequences, as will be explained, for our ability to sharpen theoretical hypotheses about other aspects of the sun's interior and to make inferences from helioseismic inversion.

Of course the possibility that there *could* be such an interior field has long been recognized, if only because magnetic diffusion times ($\sim 10^{10}$ y) are comparable to, or somewhat longer than, the age of the sun (Cowling, 1945), the latter currently being estimated as close to 4×10^9 y. Thus the interior needs no dynamo action: it could contain a magnetic field left over from the sun's formative years. Indeed, when the sun began to form as a protostar, the interstellar magnetic field must have been greatly concentrated as material collapsed inward (e.g., Gough, 1990, and references therein), dragging the field with it and spinning up to high rotation rates. On the other hand, it is also possible, and has often been hypothesized, that any such interior field was expelled or annihilated long ago—perhaps by some kind of turbulent motion within the early sun—leaving only the rapidly oscillating field associated with the convection zone and its 22-year cycle, for which dynamo action is a plausible explanation. Be that as it may, models of solar spindown—the sun's evolution from its early, rapidly rotating state, exporting angular momentum through the solar wind—have tended to assume that the interior was free of dynamically significant interior magnetic fields for most of the relevant timespan of 4×10^9 y (see, for instance, Spiegel and Zahn, 1992, and references therein).

The argument now to be sketched (for more detail, see McIntyre, 1994 and Gough and McIntyre, 1998) depends crucially on recent helioseismic results concerning the sun's differential rotation (e.g. Thompson *et al.*, 1996; Kosovichev *et al.*, 1997). These strongly indicate, for one thing, that classical laminar spindown models will not fit observations. The convection zone is in quite strong differential rotation, as suggested schematically in Fig. 9 and its caption. But the radiative interior, by contrast, turns out to be approximately in solid body rotation except possibly near the poles. (The helioseismic technique, which depends on observing the rotational splitting of acoustic vibrational modes of the sun, is insensitive to details near the rotation axis.) Now this near-solid interior rotation, of which the first indications began to emerge about a decade ago, is incompatible not only with classical laminar spindown but also with practically any other purely fluid-dynamical picture. How can anyone claim such a thing? The basis is our observational and theoretical knowledge of terrestrial middle-atmospheric dynamics.

The sun's interior is strongly stratified, and, for present purposes, is a fluid-dynamical system very like the terrestrial stratosphere, with short buoyancy and Coriolis timescales. Even the ratio of those timescales is similar: throughout most of the interior, Prandtl's ratio $N/2|\Omega| \approx 10^2$. So, in a laminar spindown model for instance, the dynamics is qualitatively like that in a zonally symmetrical terrestrial model stratosphere. The main difference is one of detail: because photon mean free paths are small, radiative heat transfer is diffusive, and also rather weak for fluid-dynamical purposes (diffusivities $\sim 10^3\,\mathrm{m^2 s^{-1}}$).

Now if laminar spindown were the only thing happening, it would drive the interior away from solid rotation toward some differentially rotating state that matches the differential rotation of the convection zone. If, on the other hand, layerwise two-dimensional turbulence were present in the interior, perhaps through the breaking of Rossby waves emitted from the convection zone, then the effect would again be to drive the interior away from solid rotation in, probably, some other way (McIntyre, 1994), subject to the same kinds of dynamical constraints as those with which we are familiar in the terrestrial stratosphere, not least the integral constraint (9). Broadband gravity waves emitted from the convection zone would tend to do the same thing, drive the interior away from solid rotation, in yet another way, through anti-frictional behaviour as in the terrestrial QBO. In summary, one can get all kinds of forms of departures from solid rotation, but one cannot get solid rotation out of the fluid dynamics except by some fantastically improbable accident. Recall again, for instance, the spontaneous symmetry-breaking in the Plumb–McEwan experiment.

I am claiming, therefore, that there appears to be no purely fluid-dynamical process that can make the interior rotate nearly solidly as observed. This is the basis for concluding that there must be an interior poloidal magnetic field. Such a field, a few lines of which are sketched in the right half of Fig. 9, can do the job very easily, through Alfvénic torques. Moreover, it is the only known way to do the job—to stop the interior differential rotation that fluid-dynamical processes would otherwise bring about.

Recent refinements to the helioseismic results, from the MDI (Michelson Doppler Interferometer) instrument on the SOHO spacecraft (Solar and Heliospheric Observatory) have strongly confirmed the near solid interior rotation and have furthermore begun to re-

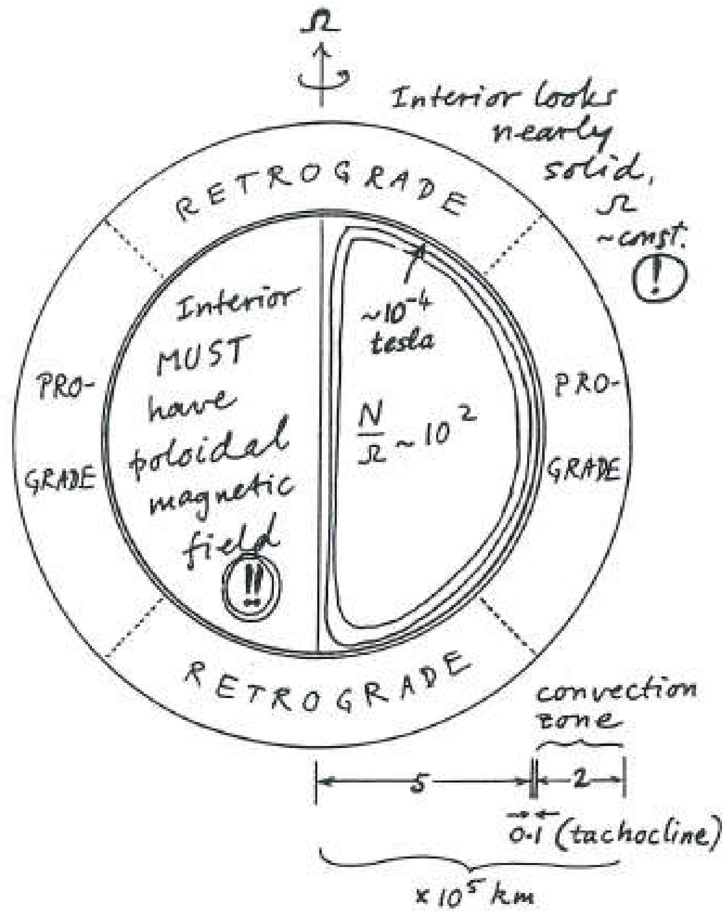

FIGURE 9 Schematic of proposed new model for the sun's interior (Gough and McIntyre, 1998). The strong differential rotation of the convection zone, now well known from helioseismic inversion (Thompson *et al.*, 1996; Kosovichev *et al.*,1997), is roughly depth-independent and is indicated schematically by the words "PROGRADE" and "RETROGRADE". Angular velocities range over about ±10% of an intermediate value, $|\Omega| \approx 2.7 \times 10^{-6}\,\mathrm{s}^{-1}$, close to the typical interior value. Interior rotation rates are much more nearly uniform. The poloidal magnetic field occupying the radiative interior is a necessary dynamical ingredient, crucial to explaining the observed depth of the tachocline, the thin shear layer separating the base of the convection zone from the interior, whose depth Δ is roughly of the order 0.1×10^5 km; cf. the sun's radius, 7×10^5 km. The estimate of magnetic field strength $|B| \approx 10^{-4}$ T (tesla) is an extremely rough, preliminary estimate, refinement of which will depend on intricate nonlinear modelling not yet done, and on the fullest possible use of helioseismic data now being accumulated. Because of the expected approximate scaling (Gough and McIntyre, 1998), which turns out to involve a ninth power law, $|B| \propto \Delta^{-9}$, it will be especially crucial to work toward refined estimates of Δ. The latest attempt at a quantitative estimate (Elliott and Gough, 1999) tentatively replaces 0.1×10^5 km by 0.13×10^5 km. The next refinement will be to make Δ a function of latitude.

solve the detailed vertical structure at the top of the interior, just under the base of the convection zone. There is strong vertical shear in a layer of small but now just-resolvable thickness, $\lesssim 2\%$ of the solar radius; but the magnitudes involved (Richardson numbers $N^2/|\text{vertical shear}|^2 \gg 1$) imply overwhelmingly strong control by the stable stratification, leading to layerwise two-dimensional motion. Older ideas that the region might be shear-unstable, and three-dimensionally turbulent, appear to be definitely ruled out. Consequently, the shear layer, which solar physicists call the "tachocline", must have a relatively quiescent dynamical character like the rest of the interior, i.e., at worst layerwise two-dimensionally turbulent like the wintertime terrestrial stratosphere, or possibly as quiescent, even, as the summertime upper stratosphere.

On this basis, and from the helioseismic results, one can estimate the circulation of the tachocline in much the same way as the circulation shown in Fig. 1. It turns out that the circulation or ventilation time is about 10^6

times longer: not several years but more like several million years (Gough and McIntyre, 1998). But this is a mere instant in comparison with 4×10^9 y, the age of the sun and the timescale of its spindown. Rates of change due to spindown can therefore be neglected in theories of tachocline structure and dynamics: the dynamics can be treated as quasi-steady. Furthermore, wave-induced forces, which are hard to estimate but are generally thought to be significant, if at all, on longer timescales $\sim 10^8$ y to 10^9 y (see Garcia López and Spruit, 1991, and references therein) can likewise be neglected, almost certainly, in the tachocline dynamics with its 10^6 y circulation timescale.

The only remotely plausible candidate for driving a circulation that is so fast, relative to the other timescales involved, is gyroscopic pumping by Reynolds stresses within the dynamically vigorous convection zone above. As in the terrestrial stratosphere, this must produce a downward-burrowing circulation. The burrowing has to be stopped, however, not by a frictional boundary layer at a solid surface but by a magnetic diffusion layer capping the interior magnetic field. If the burrowing is not stopped somehow, then we are back to a classical laminar spindown scenario, which would produce a tachocline much deeper than observed (Haynes *et al.*, 1991; Spiegel and Zahn, 1992; Gough and McIntyre, 1998).

2. The Tachopause and the Lithium and Beryllium Problems

An idealized boundary-layer theory for the magnetic diffusion layer (Gough and McIntyre, 1998) confirms that the magnetic field is, indeed, capable of stopping the burrowing, and furthermore that the boundary layer is thin, more than an order of magnitude thinner than the tachocline itself. This is turn implies that the tachocline must end in a relatively sharp "tachopause". Such information can be fed back into helioseismic inversion models. In particular, the fast tachocline circulation implies that there has to be a near-discontinuity in helium abundance, hence sound speed, at the tachopause; such a structure has recently been shown to fit the helioseismic data very well and to lead to a refined estimate of the tachocline thickness (Elliott and Gough, 1999); see caption to Fig. 9.

The boundary layer theory also gives us a first, rather crude, estimate, *ca.* 10^{-4}T, of typical magnetic field strengths just below the tachopause. Stronger fields would imply smaller tachocline depths Δ, and vice versa. However, this estimate is likely to need revision, partly because it turns out that Δ is only weakly dependent on field strength (caption to Fig. 9), and partly because a quantitative estimate will require solution of a rather complicated, and strongly nonlinear, magnetohydrodynamic problem for the upwelling branch of the tachocline circulation. This is work in progress. In the upwelling branch, which has to be in middle latitudes to be consistent with the observed pattern of differential rotation, we expect the interior field to be skimmed up into the convection zone, where it will be subject to reconnection. There will probably be significant zonal Lorenz forces in this upwelling region.

The implications are, however, already far-reaching in at least three senses. First, if the interior is constrained by the simplest possible magnetic field, well diffused from its primordial state and aligned with the rotation axis, hence in the form of a poloidal field with field lines configured as in Fig. 9, as if wrapped around a torus, or doughnut, then what is known as *Ferraro's law of isorotation* applies. This says that the angular velocity on the surface of each torus containing closed field lines must be constant, which is equivalent to saying that no Alfvénic torsional oscillations are excited. (There is an argument saying that such torsional oscillations will be damped rather quickly relative to the sun's age, essentially because of the near-degeneracy of the problem, with singular eigenfunctions in the form of a delta function confined to the surface of one torus (Hollweg, 1978; McIntyre, 1994.)) Because some of the toroidal surfaces penetrate deep into the interior, this offers for the first time a chance of obtaining information about differential rotation near the sun's core—information that is unobtainable directly from helioseismic inversion.

The second implication is a better understanding of spindown itself. The spindown problem under the constraint of Ferraro's law (with or without wave-induced forces) becomes very different from the classic spindown problem, again bringing new understanding to what we observe of the present differential rotation.

The third is a new possibility for solving the notorious lithium-burning problem (e.g. Garcia López and Spruit, 1991; S. Vauclair, personel communication). To explain observed lithium abundances at the surfaces of various populations of stars, one needs constituent-transporting mixing mechanisms or circulations that reach below the tachocline. The strong gyroscopic pumping by the convection zone might have such a role. On the simplest picture sketched in Fig. 9, there are just two places where the circulation might be able to burrow to sufficient depth for the purpose: they are the two poles, where the magnetic field must vanish, if confined by the convection zone. The downward-burrowing circulation might therefore be able to "dig" a kind of "polar pit" in the magnetic field, at each pole (Gough and McIntyre, 1998), penetrating much more deeply than elsewhere. It can probably afford to take its time over this, say $\lesssim 10^9$ y. This could still be enough to burn the lithium. Even if the interior magnetic field were more complicated, the hairy-sphere theorem tells us that weak points vulnerable to "pit digging" must exist. Detailed predictions will again depend on numerical solution of a highly nonlinear magnetohydrodynamic problem, yet to be attempted. There is a similar problem with beryllium, sharpening the observational constraints on the numerical modelling in the same way as with multiple chemical species in the earth's middle atmosphere.

I shall end on a much more speculative note. It is conceivable that, despite their generally high damping rates, interior Alfvénic torsional oscillations might nevertheless be significant especially if the field somehow configures itself to minimize the damping. Could there be, one wonders, a QBO-like torsional oscillation in the sun's interior excited by broadband gravity waves from the convection zone, despite the likely weakness of such waves? This would be a bit like the balance wheel of a watch. My present feeling is that it is less likely than when I argued for it in 1994 (McIntyre, 1994)—but then again, people who study isotopes in palaeoclimatic records tell us that there is evidence for solar oscillations on many timescales, centuries to millennia, far slower than the sunspot cycle. Perhaps all these fluctuations on various timescales originate, without exception, in the convection zone. That is entirely possible, in such a strongly nonlinear dynamical subsystem. But then again, perhaps some of these time scales come from the radiative interior. They are compatible with the range of possible interior field strengths.

ACKNOWLEDGEMENTS

This chapter draws on material from invited lectures given previously to the 12th Yukawa Memorial Symposium held in Nishinomiya, Japan (McIntyre, 1998), to the European Space Agency (McIntyre, 1999), and to the 1999 IUGG Symposium on solar variability (McIntyre, 2000b). The Yukawa Symposium was especially important in reminding me of problems in other branches of physics, engineering, and biology, such as molecular motors, that depend in interesting ways on the dynamical organization of fluctuations. I thank Joan Alexander, David Andrews, Dave Broutman, Oliver Bühler, Keith Browning, David Dritschel, Patrick Espy, Michael Gadsden, Mark Hallworth, Björn Haßler, Peter Haynes, Tony Hollingsworth, Jim Holton, Brian Kerridge, Ernst Kopp, Adrian Lee, Philip Mote, Warwick Norton, Tim Palmer, Russ Philbrick, Alan Plumb, John Pyle, Clive Rodgers, Karen Rosenlof, David Sankey, Richard Scott, Adrian Simmons, Keith Shine, Claude Souprayen, Chris Warner, Darryn Waugh, Georg Witt, Michio Yamada and Shigeo Yoden for valuable help, stimulation, comments and correspondence. My research has received generous support from the UK Natural Environment Research Council through the UK Universities' Global Atmospheric Modelling Programme, from the Isaac Newton Institute for Mathematical Sciences, and from a SERC/EPSRC Senior Research Fellowship.

References

Albritton, D. L. *et al.*, 1998: *Scientific Assessment of Ozone Depletion 1998* (WMO Global Ozone Research and Monitoring Project Report No. 44). Geneva, World Meteorol. Org. Available from WMO, GO$_3$OS, P.O. Box 2300, 1211–Geneva–2, Switzerland.

Alexander, M. J. and J. R. Holton, 1997: A model study of zonal forcing in the equatorial stratosphere by convectively induced gravity waves. *J. Atmos. Sci.*, **54**, 408–419.

Alexander, M. J. and K. H. Rosenlof, 1996: Nonstationary gravity wave forcing of the stratospheric zonal mean wind. *J. Geophys. Res.*, **101**, 23465–23474.

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

Andrews, D. G. and M. E. McIntyre, 1978: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.*, **89**, 609–646.

Andrews, D. G., J. R. Holton and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, London, New York.

Baldwin, M. P. *et al.*, 2001: The quasi-biennial oscillation. *Revs. Geophys.*, **39**, 179–229.

Bowman, K. P., 1996: Rossby wave phase speeds and mixing barriers in the stratosphere. Part I: Observations. *J. Atmos. Sci.*, **53**, 905–916.

Brasseur, S. and S. Solomon, 1984: *Aeronomy of the Middle Atmosphere*. Reidel, Dordrecht.

Bretherton, F. P., 1969: Waves and turbulence in stably stratified fluids. *Rad. Sci.*, **4**, 1279–1287.

Bühler, O., 2000: On the vorticity transport due to dissipating or breaking waves in shallow-water flow. *J. Fluid Mech.*, **407**, 235–263.

Bühler, O. and M. E. McIntyre, 1998: On non-dissipative wave-mean interactions in the atmosphere or oceans. *J. Fluid Mech.*, **354**, 301–343.

Bühler, O. and M. E. McIntyre, 2001: The recoil from acoustic or gravity waves refracted by a vortex: a new wave–mean interaction effect. In preparation.

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

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

Cowling, T. G., 1945: On the Sun's general magnetic field. *Mon. Not. R. Astronom. Soc.*, **105**, 166–174.

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

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

Dunkerton, T. J., 1983: Laterally-propagating Rossby waves in the easterly acceleration phase of the quasi-biennial oscillation. *Atmos.-Ocean*, **21**, 55–68.

Dunkerton, T. J., 1991: Nonlinear propagation of zonal winds in an atmosphere with Newtonian cooling and equatorial wavedriving. *J. Atmos. Sci.*, **48**, 236–263.

Dunkerton, T. J., 1997: The role of gravity waves in the quasi-biennial oscillation. *J. Geophys. Res.*, **102**, 26053–26076.

Dunkerton, T. J. *et al.*, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.*, **38**, 819–843.

Eady, E. T., 1950: The cause of the general circulation of the atmosphere. In Centenary Proceedings Royal Meteorological Society (P. A. Sheppard, J. K. Bannon, C. R. Burgess, G. Manley, G. D. Robinson and R. C. Sutcliffe, Eds), Reading, UK, Royal Meteorological Society, 156–172.

Eliassen, A. and E. Palm, 1961: On the transfer of energy in stationary mountain waves. *Geofysiske Publ.*, **22**(3), 1–23.

Elliott, J. R. and D. O. Gough, 1999: Calibration of the thickness of the solar tachocline. *Astrophys. J.*, **516**, 475–481.

Ertel, H., 1942: Ein Neuer hydrodynamischer Wirbelsatz. *Met. Z.*, **59**, 271–281. English translation in W. Schröder (Ed.) 1991: *Geophysical Hydrodynamics and Ertel's Potential Vorticity: Selected Papers of Hans Ertel*, Bremen-Rönnebeck, Interdivisional Commission of History of IAGA, Newsletter No. **12**.

Fels, S. B., 1985: Radiative-dynamical interactions in the middle atmosphere. In *Advances in Geophysics: Issues in Atmospheric and Oceanic modeling* (**28A**), S. Manabe, Ed., Academic Press, Orlando, FL, pp. 277–300.

Fleming, E. L. *et al.*, 1990: Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. COSPAR International Reference Atmosphere 1986. Part II: Middle atmosphere models. *Adv. Space Res.*, **10**(12), 11–62.

Fritts, D. C. *et al.*, 1988: Evidence of gravity wave saturation and local turbulence production in the summer meosphere and lower thermosphere during the STATE experiment. *J. Geophys. Res.*, **93**, 7015–7025.

Gadsden, M., 2000: Polar mesospheric clouds seen from geostationary orbit. *J. Atmos. Solar-Terrest. Phys.*, **62**, 31–36.

Gadsden, M. and W. Schröder, 1989: *Noctilucent Clouds*. Springer, Heidelberg.

Garcia, R. R., 1989: Dynamics, radiation, and photochemistry in the mesosphere: implications for the formation of noctilucent clouds. *J. Geophys. Res.*, **94**, 14605–14615 (Special Issue for the International Workshop on Noctilucent Clouds).

García López, R. J. and H. C. Spruit, 1991: Li depletion in F stars by internal gravity waves. *Astrophys. J.*, **377**, 268–277.

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics*. Academic Press, London and New York.

Gough, D. O., 1990: On possible origins of relatively short-term variations in the solar structure. *Phil. Trans. R. Soc. Lond.*, **A330**, 627–640.

Gough, D. O. and M. E. McIntyre, 1998: Inevitability of a magnetic field in the Sun's radiative interior. *Nature*, **394**, 755–757.

Gray, L. J. and Pyle, J. A., 1989: A two-dimensional model of the quasi-biennial oscillation of ozone. *J. Atmos. Sci.*, **46**, 203–220.

Hall, T. M. *et al.*, 1999: Evaluation of transport in stratospheric models. *J. Geophys. Res.*, **104**, 18815–18839.

Hamilton, K., (Ed.), 1997: *Gravity Wave Processes: Their Parameterization in Global Climate Models*. Springer-Verlag, Heidelberg.

Haynes, P. H., 1989: The effect of barotropic instability on the nonlinear evolution of a Rossby wave critical layer. *J. Fluid Mech.*, **207**, 231–266.

Haynes, P. H., 1998: The latitudinal structure of the quasi-biennial oscillation. *Q. J. R. Meteorol. Soc.*, **124**, 2645–2670.

Haynes, P. H. *et al.*, 1991: On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.*, **48**, 651–678; see also **53**, 2105–2107.

Haynes, P. H. *et al.*, 1996: Reply to comments by J. Egger on 'On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces'. *J. Atmos. Sci.*, **53**, 2105–2107.

Hines, C. O., 1972: Gravity waves in the atmosphere. *Nature*, **239**, 73–78.

Hollweg, J. V., 1978: Some physical processes in the solar wind. *Rev. Geophys. Space Phys.*, **16**, 689–720.

Holton, J. R., 1982: The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *J. Atmos. Sci.*, **39**, 791–799.

Holton, J. R. and R. S. Lindzen, 1972: An updated theory of the quasi-biennial cycle of the tropical stratosphere. *J. Atmos. Sci.*, **29**, 1076–1080.

Holton, J. R. *et al.*, 1995: Stratosphere–troposphere exchange. *Rev. Geophys.*, **33**, 403–439.

Hoskins, B. J., M. E. McIntyre and A. W. Robertson, 1985: On the use and significance of isentropic potential-vorticity maps. *Q. J. R. Meteorol. Soc.*, **111**, 877–946. Also *Corrigendum*, etc., **113** (1987), 402–404.

Houghton, J. T., 1978: The stratosphere and the mesosphere. *Q. J. R. Meteorol. Soc.*, **104**, 1–29.

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

Killworth, P. D. and M. E. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect or over-reflect? *J. Fluid Mech.*, **161**, 449–492.

Kleinschmidt, E., 1950: Über Aufbau und Entstehung von Zyklonen (1,2 Teil). *Met. Runds.*, **3**, 1–6, 54–61.

Kosovichev, A. G. *et al.*, 1997: Structure and rotation of the solar interior: initial results from the MDI medium-l program. *Sol. Phys.*, **170**, 43–61.

Lindzen, R. S., 1981: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.*, **86**, 9707–9714.

Lindzen, R. S. and J. R. Holton, 1968: A theory of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 1095–1107.

Lorenz, E. N., 1967: *The Nature and Theory of the General Circulation of the Atmosphere.* World Meteorol. Org., Geneva.

Lübken, F.-J., 1999: Thermal structure of the Arctic summer mesosphere. *J. Geophys. Res.*, **104**, 9135–9149.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Met. Soc. Jpn.*, **44**, 25–43.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.*, **27**, 871–883.

McIntyre, M. E., 1980: Towards a Lagrangian-mean description of stratospheric circulations and chemical transports. *Phil. Trans. R. Soc. Lond.*, **A296**, 129–148.

McIntyre, M. E., 1981: On the "wave momentum" myth. *J. Fluid Mech.*, **106**, 331–347.

McIntyre, M. E., 1989: On dynamics and transport near the polar mesopause in summer. *J. Geophys. Res.*, **94**, 14617–14628 (Special Issue for the International Workshop on Noctilucent Clouds).

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

McIntyre, M. E., 1994: The quasi-biennial oscillation (QBO): some points about the terrestrial QBO and the possibility of related phenomena in the solar interior. In *The Solar Engine and its Influence on the Terrestrial Atmosphere and Climate* (Vol. 25 of

NATO ASI Subseries I, Global Environmental Change) (E. Nesme-Ribes, Ed.), Springer-Verlag, Heidelberg, pp. 293–320.

McIntyre, M. E., 1998: Breaking waves and global-scale chemical transport in the Earth's atmosphere, with spinoffs for the Sun's interior. *Progr. Theor. Phys. (Kyoto) Suppl.*, **130**, 137–166 (invited review for 12th Nishinomiya Symposium in memory of Hideki Yukawa). See also an important Corrigendum, *Progr. Theor. Phys. (Kyoto)*, **101** (1999), 189.

McIntyre, M. E., 1999: How far have we come in understanding the dynamics of the middle atmosphere? Invited Symposium Lecture, *Proceedings 14th European Space Agency Symposium on European Rocket and Balloon Programmes and Related Research*, Potsdam, Germany, 31 May–3 June 1999, ESA SP–437 (B. Kaldeich-Schürmann, Ed.), ESA Publications Division, ESTEC, Noordwijk, The Netherlands.

McIntyre, M. E., 2000a: On global-scale atmospheric circulations. In *Perspectives in Fluid Dynamics: A Collective Introduction to Current Research* (G. K. Batchelor, H. K. Moffatt and M. G. Worster, Eds), Cambridge University Press, pp. 557–624.

McIntyre, M. E., 2000b: The Earth's middle atmosphere and the Sun's interior. In *Long and Short Term Variability of the Sun's History and Global Change* (Proceedings IUGG Symposium GA6.01 held in Birmingham, July 1999) (W. Schröder, Ed.), Interdivisional Commission on History, IAGA/IUGG, Newsletter **39**, pp. 13–38. Science Edition, D–28777 Bremen-Rönnebeck, Hechelstrasse 8, Germany.

McIntyre, M. E. and W. A. Norton, 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** (1990) 693. (A further, important correction is described in R. Mo, O. Bühler and M. E. McIntyre, 1998: Permeability of the stratospheric vortex edge: On the mean mass flux due to thermally dissipating, steady, non-breaking Rossby waves. *Q. J. R. Meteorol. Soc.*, **124**, 2129–2148.

McIntyre, M. E. and W. A. Norton, 2000: Potential-vorticity inversion on a hemisphere. *J. Atmos. Sci.*, **57**, 1214–1235. Corrigendum: *J. Atmos. Sci.*, **58**, 949.

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

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

Mote, P. W. *et al.*, 1996: An atmospheric tape recorder: the imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res.*, **101**, 3989–4006.

Mote, P. W. *et al.*, 1998: Vertical velocity, vertical diffusion, and dilution by midlatitude air in the tropical lower stratosphere. *J. Geophys. Res.*, **103**, 8651–8666.

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.

Plumb, R. A., 1977: The interaction of two internal waves with the mean flow: implications for the theory of the quasi-biennial oscillation. *J. Atmos. Sci.*, **34**, 1847–1858.

Plumb, R. A. and J. Eluszkiewicz, 1999: The Brewer–Dobson circulation: dynamics of the tropical upwelling. *J. Atmos. Sci.*, **56**, 868–890.

Plumb, R. A. and A. D. McEwan, 1978: The instability of a forced standing wave in a viscous stratified fluid: a laboratory analogue of the quasi-biennial oscillation. *J. Atmos. Sci.*, **35**, 1827–1839.

Robinson, W. A., 1988: Analysis of LIMS data by potential vorticity inversion. *J. Atmos. Sci.*, **45**, 2319–2342.

Rosenlof, K., 1996: Summer hemisphere differences in temperature and transport in the lower stratosphere. *J. Geophys. Res.*, **101**, 19129–19136.

Rosenlof, K. and J. R. Holton, 1993: Estimates of the stratospheric residual circulation using the downward control principle. *J. Geophys. Res.*, **98**, 10465–10479.

Rossby, C.-G., 1936: Dynamics of steady ocean currents in the light of experimental fluid mechanics. *Mass. Inst. Technol. Woods Hole Oc. Inst. Pap. Phys. Oceanogr. Meteorol.*, **5**(1), 1–43. See Eq. (75).

Rossby, C.-G., 1940: Planetary flow patterns in the atmosphere. *Q. J. R. Meteorol. Soc.*, **66**(Suppl.), 68–97.

Sankey, D., 1998: Dynamics of upwelling in the equatorial lower stratosphere. Ph.D. thesis, University of Cambridge.

Saravanan, R., 1990: A multiwave model of the quasi-biennial oscillation. *J. Atmos. Sci.*, **47**, 2465–2474.

Schröder, W., 1999: Were noctilucent clouds caused by the Krakatoa eruption? A case study of the research problems before 1885. *Bull. Am. Meteorol. Soc.*, **80**, 2081–2085.

Scott, R. K. and Haynes, P. H., 1998: Internal interannual variability of the extratropical stratospheric circulation: the low-latitude flywheel. *Q. J. R. Meteorol. Soc.*, **124**, 2149–2173.

Shine, K. P., 1989: Sources and sinks of zonal momentum in the middle atmosphere diagnosed using the diabatic circulation. *Q. J. Roy. Meteorol. Soc.*, **115**, 265–292.

Snodgrass, F. E. *et al.*, 1966: Propagation of ocean swell across the Pacific. *Phil. Trans. R. Soc. Lond.*, **A259**, 431–497.

Solomon, S., J. T. Kiehl, R. R. Garcia and W. L. Grose, 1986: Tracer transport by the diabatic circulation deduced from satellite observations. *J. Atmos. Sci.*, **43**, 1603–1617.

Sparling, L. S. *et al.*, 1997: Diabatic cross-isentropic dispersion in the lower stratosphere. *J. Geophys. Res.*, **102**, 25817–25829.

Spiegel, E. A. and J.-P. Zahn, 1992: The solar tacholine. *Astron. Astrophys.*, **265**, 106–114.

Taylor, G. I., 1915: Eddy motion in the atmosphere. *Phil. Trans. Soc. Lond.*, **A215**, 1–23.

Thomas, G. E. *et al.*, 1989: Relation between increasing methane and the presence of ice clouds at the mesopause. *Nature*, **338**, 490–492.

Thompson, M. J. *et al.*, 1996: Differential rotation and dynamics of the solar interior. *Science*, **272**, 1300–1305.

Thuburn, J. and M. E. McIntyre, 1997: Numerical advection schemes, cross-isentropic random walks, and correlations between chemical species. *J. Geophys. Res.*, **102**, 6775–6797.

Wallace, J. M. and J. R. Holton, 1968: A diagnostic numerical model of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 280–292.

Wallace, J. M. and V. E. Kousky, 1968: Observational evidence of Kelvin waves in the tropical stratosphere. *J. Atmos. Sci.*, **25**, 900–907.

Ward, W. E. and V. I. Fomichev, 1996: Consequences of CO_2–O collisional relaxation on the energy budget of the mesosphere and lower thermosphere. *Adv. Space Res.*, **17**(11), 127–137.

Yulaeva, E., J. R. Holton and J. M. Wallace, 1994: On the cause of the annual cycle in tropical lower stratospheric temperatures. *J. Atmos. Sci.*, **51**, 169–174.