Breaking planetary waves in the stratosphere

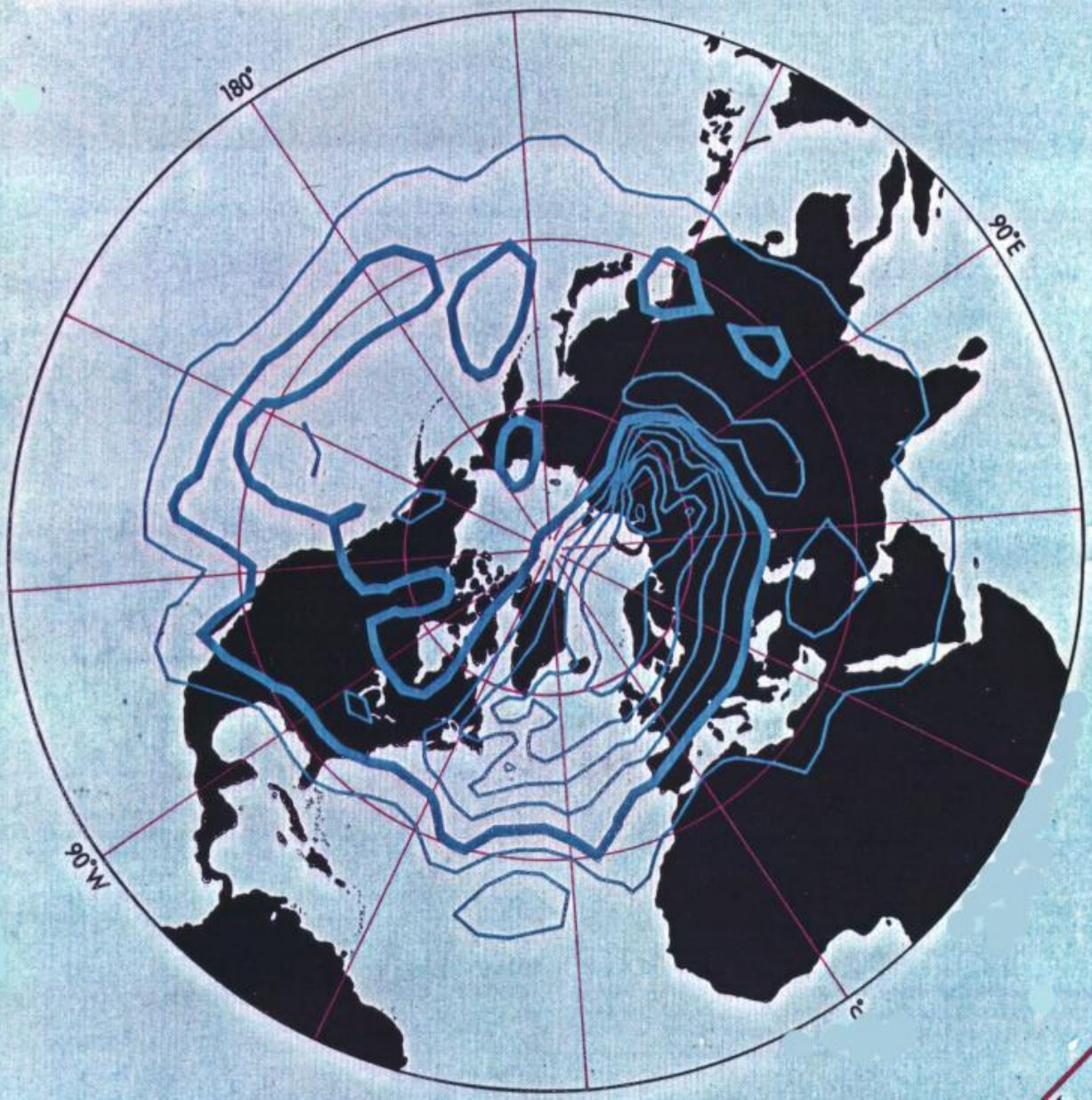
M. E. McIntyre and T. N. Palmer

Nature., **305**, 593–600 (1983)

nature

INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Volume 305 No 5935 13-19 October 1983 £1.80 $4.50



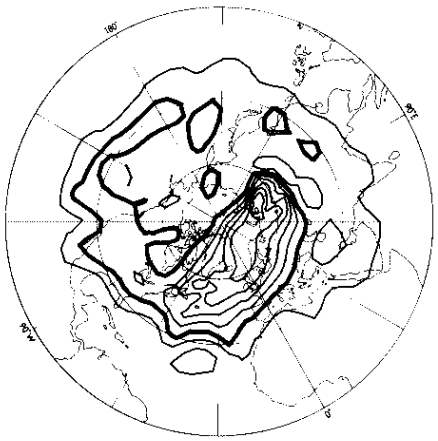
PLANETARY WAVES IN THE STRATOSPHERE

BIOCHEMICAL INSTRUMENTS
product review

nature

NATURE VOL. 305 13 OCTOBER 1983

COVER



A breaking Rossby wave in the stratosphere, at an altitude of about 30 km. The contours roughly represent material contours. A long tongue of polar air is being pulled into the subtropics, while tropical air carrying ozone and pollutants is mixed polewards. Satellite-borne IR radiometers have made this global-scale process, and some of its consequences, directly visible for the first time (see p.593).

OPINION

Poor outlook for arms control 561
Animal rights nonsense 562
Triumph for unreason

NEWS

US national laboratories 563
London merger
Genetic engineering 564
Swiss science policy
US science budget 565
US risk assessment
Star wars 566
UK weapons tests
Australian veterinary science 567
Polish agriculture
Chinese irrigation 568
French research
Genetics Inst. deal
British industry 569
Gene diagnosis
***Nature* index of biotechnology stocks**

CORRESPONDENCE

Animal experimentation/Thanks from Poland/Coolidge effect/etc. 570

COMMENTARY

Creation versus evolution: no middle way
G M Marsden 571

NEWS AND VIEWS

Nobel Prize to Barbara McClintock
John Maddox 575
Trophic factors and postsynaptic activity in synapse formation
T Lømo 576
Pathophysiology of schizophrenia — causal role for dopamine or noradrenaline
L L Iversen, G P Reynolds & S H Snyder, Oleh Hornykiewicz replies 577
IRAS circular 3
G Neugebauer & H Habing 578
Do metals conduct colour change?
Frank Close 579
Tricks for lasers and atom beams
Anne Thorne 580
Dynamic proteins
Roger H Pain 581
Magnetism and evolution of the terrestrial planets
J A Jacobs 582

BOOK REVIEWS

A Prehistory of Australia, New Guinea and Sahul *By Peter J White and James F O'Connell* and **Archaeology of the Dreamtime** *By Josephine Flood*
J V S Megaw 647
Naked Emperors *By Garrett Hardin*
Robert Ubell 649
Animal Mechanics *By R McNeill Alexander*
K E Machin
Fish Locomotion *By R W Blake*
John J Videler 650
Fish Stock Assessment *By J A Gulland*
D H Cushing
Handbook of Polycyclic Aromatic Hydrocarbons *Edited by Alf Bjørseth*
Peter Brookes

REVIEW ARTICLE

Protein phosphorylation in the brain
E J Nestler & P Greengard 583

ARTICLES

Frequencies, amplitudes and linewidths of solar oscillations from total irradiance observations
M Woodard & H S Hudson 589

Breaking planetary waves in the stratosphere
M E McIntyre & T N Palmer 593

Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3′ consensus sequence AAUAAA
C Montell, E F Fisher, M H Caruthers & A J Berk 600

Structural and immunobiological similarities between simian sarcoma virus gene product(s) and human platelet-derived growth factor
K C Robbins, H N Antoniades, S G Devare, M W Hunkapiller & S A Aaronson 605

LETTERS TO NATURE

Cosmic γ rays and cosmic-ray particles
J Wdowczyk & A W Wolfendale 609

Measurements of ^{129}I in meteorites and lunar rock by tandem accelerator mass spectrometry
K Nishiizumi, D Elmore, M Honda, J R Arnold & H E Gove 611

An unusual interplanetary event: encounter with a comet?
C T Russell, J G Luhmann, A Barnes, J D Mihalov & R C Elphic 612

Efficient and stable solar cell by interfacial film formation
S Menezes, H-J Lewerenz & K J Bachmann 615

Viscous magnetization
D Walton 616

A stress province boundary and tractions on the North American plate
D I Gough, C K Fordjor & J S Bell 619

Why large earthquakes do not nucleate at shallow depths
S Das & C H Scholz 621

Abnormal floral development of a tobacco mutant with elevated polyamine levels
R L Malmberg & J McIndoo 623

Long-term culture of murine erythropoietic progenitor cells in the absence of an adherent layer
F Wendling, M M Shreeve, D L McLeod & A A Axelrad 625

Contents continued overleaf

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Journals Ltd and includes the Annual Index (mailed in February). Annual subscription for USA and Canada US $220 (for subscription prices elsewhere, see next page). USA and Canadian orders only to: *Nature*, Subscription Dept PO Box 1018, Manasquan, New Jersey 08736, USA. All other orders to *Nature*, Brunel Road Basingstoke, Hants RG21 2XS, UK. Second class postage paid at New York, NY10010 and additional mailing offices. US Postmaster send form for 3579 to: *Nature*, 15 East 26 Street, New York, NY 10010. Published in Japan by Macmillan Shuppan K.K., 5th Floor Eikow Building, 1-10-9 Hongo, Bunkyo-ku, Tokyo, Japan.

Conclusions

The most straightforward interpretation of the power spectrum of total irradiance data is that the low-l p-modes behave like independent, randomly excited harmonic oscillators. The damping time $\tau \simeq 2$ days of the multiplet substates, inferred directly from the line widths of the main $l = 0$ modes and indirectly from the scatter in the power of the modes, is incompatible with the previous observation of rotationally split components. The composite width of the unresolved $l = 1$ peaks in our spectrum is consistent with either the amount of splitting (0.75 μHz per m-state) claimed by Claverie *et al.*, which we take as an upper bound, or with the splitting expected from uniform rotation at a rate of ~0.43 μHz. We conclude that the precise rotation rate of the solar interior is still an open question.

We thank W. A. Coles and P. Delache for useful discussions and R. C. Willson for providing both reduced ACRIM data and continuing assistance. This work was supported by NASA under grant NSG 7161.

Received 30 June; accepted 10 August 1983.

1. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **282,** 591–594 (1979).
2. Ulrich, R. K. & Rhodes, E. J. Jr *Astrophys. J.* **265,** 551–563 (1983).
3. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **293,** 443–445 (1981).
4. Ledoux, P. *Astrophys. J.* **114,** 373–384 (1951).
5. Isaak, G. R. *Nature* **296,** 130–131 (1982).
6. Dicke, R. H. *Nature* **300,** 693–697 (1982).
7. Gough, D. O. *Nature* **298,** 350–354 (1982).
8. Grec, G., Fossat, E. & Pomerantz, M. *Solar Phys.* **82,** 55–66 (1983).
9. Willson, R. C. *Appl. Opt.* **18,** 179–188 (1979).
10. Willson, R. C. & Hudson, H. S. *Astrophys. J. Lett.* **244,** L185–189 (1981).
11. Woodard, M. & Hudson, H. *Solar Phys.* **82,** 67–73 (1983).
12. Woodard, M., Hudson, H. & Willson, R. in *Pulsations in Classical and Cataclysmic Variable Stars* (eds Cox, J. P. & Hansen, C. J.) 152–156 (JILA, Boulder, 1982).
13. Gough, D. O. in *Nonradial and Nonlinear Stellar Pulsation* (eds Hill, H. A. & Dziembowski, W. A.) 273 (Springer, Berlin, 1980).
14. Christensen-Dalsgaard, J. & Frandsen, S. *Solar Phys.* **82,** 469–486 (1983).
15. Goldreich, P. & Keeley, D. A. *Astrophys. J.* **212,** 243–251 (1977).
16. Duvall, T. L. Jr & Harvey, J. W. *Nature* **302,** 24–27 (1983).
17. Shibahashi, H. & Osaki, Y. *Publ. astr. Soc. Japan* **33,** 713–719 (1981).
18. Gabriel, M., Scuflaire, R. & Noels, A. *Astr. Astrophys.* **110,** 50–53 (1982).
19. Gough, D. O. in *Pulsations in Classical and Cataclysmic Variable Stars* (eds Cox, J. P. & Hansen, C. J.) 117–137 (JILA, Boulder, 1982).
20. Christensen-Dalsgaard, J. & Frandsen, S. *Solar Phys.* **82,** 165–204 (1983).
21. Delache, P. & Scherrer, P. *Nature* (in the press).

Breaking planetary waves in the stratosphere

M. E. McIntyre* & T. N. Palmer†

* Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW, UK
† Meteorological Office, Bracknell, Berks RG12 2SZ, UK

Satellite-borne IR radiometers are turning the Earth's stratosphere into one of the best available outdoor laboratories for observing the large-scale dynamics of a rotating, heterogeneous fluid under gravity. New insight is being gained not only into stratospheric dynamics as such, with its implications for pollutant behaviour and the ozone layer, but also indirectly into the dynamics of the troposphere, with its implications for weather forecasting. Similar dynamical regimes occur in the oceans and in stellar interiors. A key development has been the construction of coarse-grain maps of the large-scale distribution of potential vorticity in the middle stratosphere. Potential vorticity is a conservable quantity which has a central role in the dynamical theory, but is difficult to calculate accurately from observational data. We present the first mid-stratospheric potential vorticity maps which appear good enough to make visible the 'breaking' of planetary or Rossby waves, a phenomenon ubiquitous in nature and arguably one of the most important dynamical processes affecting the stratosphere as a whole.

THE flow in the northern wintertime stratosphere is often thought of as consisting of an anticlockwise vortex centred on the cold polar cap, disturbed by a pattern of 'planetary waves'. Two examples are shown in Fig. 1*a*, *b*, which are conventional maps showing the height of the 10-mbar isobaric surface on 17 and 27 January 1979. This surface lies in the middle stratosphere, at heights ranging between about 28 and 32 km. Because there is approximate geostrophic balance between horizontal pressure gradient and Coriolis acceleration, the air flow is nearly parallel to the height contours, as suggested by heavy arrows in Fig. 1. Velocities in the fastest parts of the stream may reach about 75 m s^{-1} (~90° longitude per day at 50° N). The pattern in Fig. 1*a* contains a wave-3 disturbance, consisting of three pairs of ridges and troughs around a latitude circle, while Fig. 1*b* is dominated by a wave-1 disturbance, with a single trough over Europe. This wave-1 disturbance was observed to be nearly stationary with respect to the Earth's surface, and had an unusually large amplitude. It was the first of a sequence of events leading several weeks later to a major sudden warming, during which the circumpolar vortex broke up completely and temperatures in the polar cap rose by tens of degrees in just a few days.

The most prominent wave patterns at these altitudes in winter tend to be more or less stationary and to be dominated by the largest spatial scales, associated with waves 1 and 2 and to a lesser extent with wave 3. As is well known, this fact is in qualitative agreement with the predictions of linear wave theory, if one assumes that the primary source of the waves is in the much denser troposphere below and involves processes tied to large-scale geographical features, favouring the generation of stationary waves. Linear theory predicts that those stationary waves which have the largest horizontal scales penetrate highest.

The dynamical restoring mechanism, whereby the waves can propagate westward relative to the stream and remain stationary on the polar vortex, depends on the existence of a cross-stream gradient of Ertel's potential vorticity in isentropic surfaces. Part of this cross-stream gradient is due to the varying direction of gravity relative to the Earth's rotation axis, the so-called planetary vorticity gradient or 'beta effect', and there are further important contributions dependent on the velocity profile of the airstream. Ertel's potential vorticity Q is defined by

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

where $\boldsymbol{\Omega}$ is the Earth's angular velocity, $\mathbf{u}$ is air velocity relative to the Earth, ρ is air density and θ specific entropy, the gradient of which is nearly vertical in the stratosphere because the isentropic surfaces are nearly horizontal. One may equally well take θ to be potential temperature, or any other function of specific entropy. The equations of motion imply that, if the motion is adiabatic, that is, if

$$D\theta/Dt = 0 \qquad (2)$$

0028-0836/83/410593—08$01.00

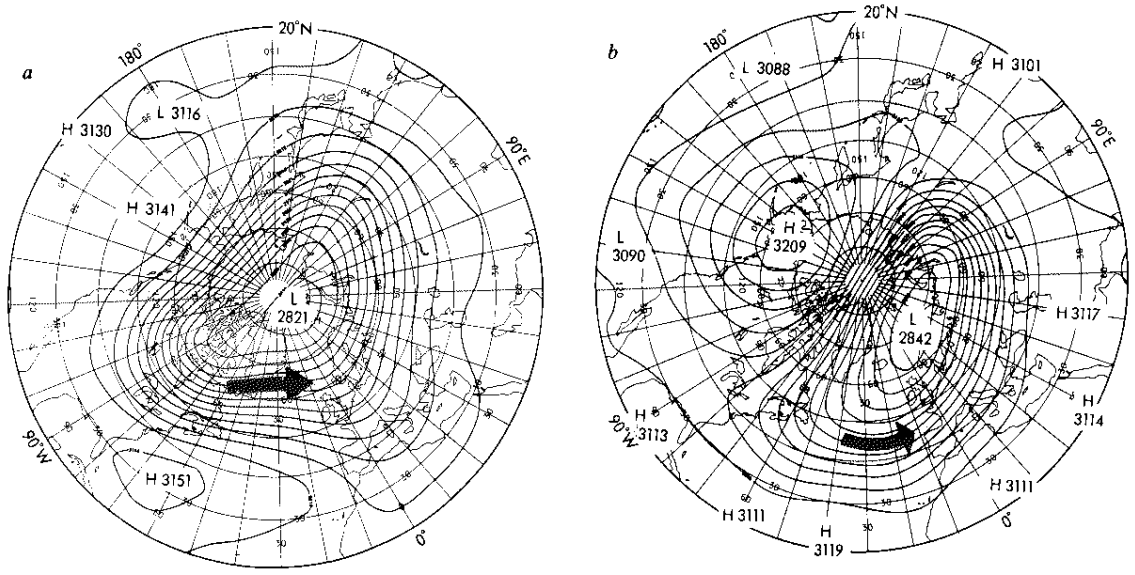


Fig. 1 Geopotential height in decametres (numerically nearly the geometric altitude above sea level) of the 10-mbar isobaric surface on 17 (*a*) and 27 (*b*) January 1979, at 00 h GMT. Contour interval is 24 decametres. Map projection is polar stereographic. The southernmost latitude circle shown is 20° N.

then, provided also that the motion is frictionless,

$$DQ/Dt = 0 \qquad (3)$$

(Ertel's theorem). $D/Dt = \partial/\partial t + \mathbf{u}\cdot\nabla$ is the rate of change seen by an individual air parcel; thus, Q and θ are constant following the motion. For present purposes this is a qualitatively reasonable approximation, for time scales up to a week or so. Over longer time scales, and in any case for quantitative work, the effects of radiative heating or cooling would have to be included, especially at the highest levels in the stratosphere. Equations (2) and (3) imply that when material contours are displaced northward and southward in a wavy pattern, the contours of potential vorticity in an isentropic surface are similarly displaced. The velocity field associated with the resulting potential vorticity disturbance lags the displacement pattern by a quarter wavelength, implying that the whole pattern propagates westward relative to the stream. This wave propagation mechanism was identified in 1939 by C.-G. Rossby, who recognized its importance for large-scale flow in the troposphere. It is an essentially linear mechanism. Of course, equation (3) itself applies equally well to nonlinear as to linear phenomena. A time sequence of isentropic maps of Q provides not only a good way of seeing when and where the Rossby-wave propagation mechanism is effective (and hence of judging the possible relevance of associated theoretical concepts like group velocity, refractive index, and so on), but also provides a very direct insight into some of the more important consequences of nonlinearity.

Data and approximations

Contour maps of Q on the 850 K isentropic surface in the Northern Hemisphere, at daily intervals during January–February 1979, were constructed at the UK Meteorological Office. 850 K is the potential temperature defined as the temperature of an air parcel brought adiabatically to a nominal sea-level pressure of 1,000 mbar. The data used were 50- and 100-mbar height fields from conventional meteorological objective analyses (NMC, Washington), together with IR spectroscopic radiances from channels 25 and 26 of the prototype Stratospheric Sounding Unit on board the polar orbiting satellite Tiros-N. Channels 25 and 26 receive radiation mainly from the region above 100 mbar, and their weighting functions peak at about 16 and 6 mbar, respectively[1]. The 850 K isentropic surface was chosen because it lies mainly between the peaks of the weighting functions and is one of the most suitable isentropic surfaces, in terms of signal-to-noise ratio, on which to estimate the derivatives required for calculating Q. It is also close to the 10-mbar isobaric surface depicted in Fig. 1, and is centrally placed for observing stratospheric planetary-wave phenomena.

To estimate Q, smoothed height fields on the 20, 10, 5, 2 and 1 mbar isobaric surfaces were derived from the hydrostatic relation and the IR radiances, using standard techniques summarized in Fig. 3 of ref. 1; this is also how the maps in Fig. 1 were obtained. The 1- and 2-mbar fields were inaccurate through failure of the third IR channel, 27, but we believe that this did not materially affect the analysis near 10 mbar. The so-called gradient wind approximation, which assumes geostrophic balance plus a correction allowing for the local curvature of air parcel trajectories, was used to estimate the vertical component ζ of the vorticity vector, $\nabla\times\mathbf{u}$, on each isobaric surface. A simple centred-difference formula having second-order accuracy was used, the height fields having been interpolated onto a square grid on a polar stereographic projection, the grid size being about 580 km at 50° N. The pressure p on the 850 K isentropic surface was then found at each grid point by a cubic spline interpolation procedure, and ζ was similarly interpolated on to that surface. Q was then found from equation (1), taking θ as potential temperature relative to 1,000 mbar and neglecting the small horizontal component of $\nabla\theta$. In estimating the vertical component $\partial\theta/\partial z$ from the data, the hydrostatic relation $\partial p/\partial z = -\rho g$ was again used, giving

$$\rho^{-1}\partial\theta/\partial z = -g\partial\theta/\partial p = -gp^{-1}\partial\theta/\partial(\ln p) \qquad (4)$$

where z is the altitude and g the acceleration due to gravity; $\partial\theta/\partial(\ln p)$ was estimated as the local slope of the curve defined by the appropriate cubic spline. All the vertical interpolations were carried out with respect to $\ln p$, roughly equivalent to using the true altitude z. Further details are given in ref. 2.

It is emphasized that the satellite data can at best represent smeared-out versions of the temperature and motion fields. The horizontal resolution is limited, especially in the tropics, by the number of orbits, about 14 polar orbits per day. The vertical resolution is limited by the half-widths of the weighting functions, just under 10 km. *A fortiori*, the data cannot resolve the fine-grain structure in the θ and Q fields which is to be expected for various reasons, and evidence of which has frequently been

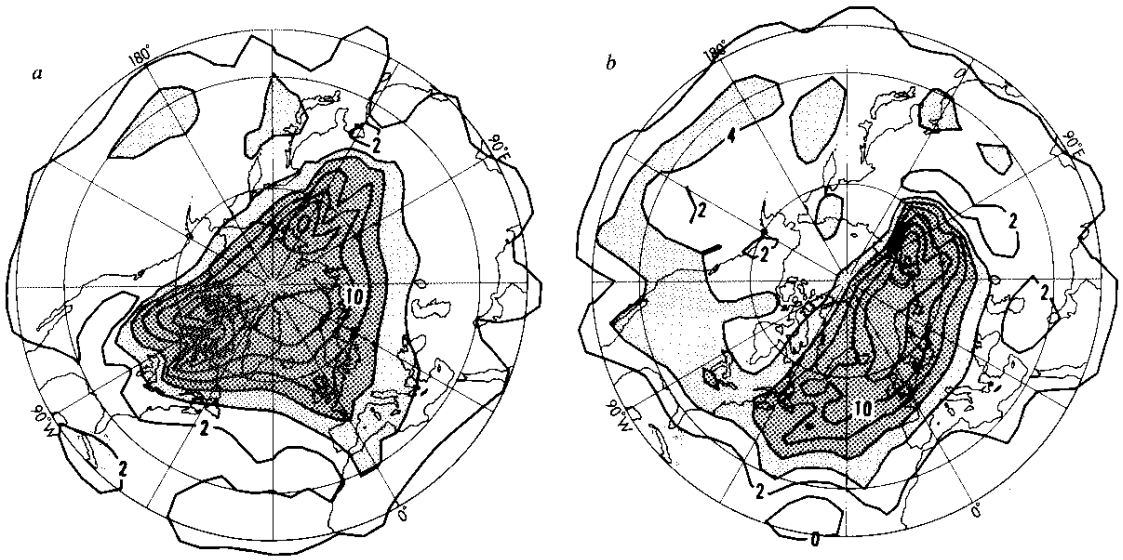


Fig. 2 Coarse-grain estimates of Ertel's potential vorticity Q on the 850 K isentropic surface (near the 10-mbar isobaric surface) on 17 (*a*) and 27 (*b*) January 1979, at 00 h GMT. The southernmost latitude circle shown is 20° N; the others are 30° N and 60° N. Map projection is polar stereographic. For units see equation (5) onwards. Contour interval is 2 units. Values greater than 4 units are lightly shaded, and greater than 6 units heavily shaded.

seen in the stratosphere, including layers with small vertical scales. Our neglect of the horizontal component of $\nabla\theta$, an approximation valid for large Richardson number, may not apply on the smallest scales either. Fortunately, there are reasons for supposing that the large-scale dynamics 'sees' mainly a coarse-grain, spatially averaged approximation to Q and θ, varying on vertical and horizontal scales not much smaller than those resolvable by the satellite data, and moreover that many of the smaller-scale fluctuations in $\nabla\theta$ and ζ are sufficiently ill-correlated to be ignored, to a first approximation, in a coarse-grain estimate of the right-hand side of equation (1) exploiting the smoothness of the satellite weighting functions. If some such coarse-grain approximation to Q were not dynamically meaningful, numerical model simulations of the large-scale behaviour of the atmosphere would hardly be practicable. The particular stratospheric wave events under discussion were, in fact, remarkably well simulated by a high-resolution numerical forecasting model[3], as far as we can tell from existing diagnostics.

We further caution that all scales are subject to ill-understood systematic errors from the algorithms used to retrieve vertical temperature profiles from IR radiances, and to errors in the radiances themselves[1]. The difficulties in estimating such errors are well known, being partly due to a paucity of independent temperature measurements by other means such as rocket probes in a sufficient variety of conditions. A preliminary assessment of their effects on Q is given in ref. 2. A full investigation is beyond our present resources as it would require the careful intercomparison of all available data, including newly available satellite data on other quasi-conservative tracers such as ozone. This must be left as a major problem for the future, progress with which could eventually lead to improved data exploitation based on simultaneous estimation of the distribution of tracers and potential vorticity.

One of the main points to be made here, however, is that despite the data errors, the present coarse-grain Q maps exhibit certain qualitative features which are strikingly in accordance with expectations from well-explored lines of theoretical reasoning, supported by rational analytical models, by numerical experiments of various kinds, and by direct estimates of relevant material trajectories based on the satellite-derived wind fields. We feel justified, therefore, in claiming that those qualitative features in the Q fields to which we shall draw attention are almost certainly real, and that to this extent the data are dynamically consistent—more so, perhaps, than hitherto appreciated.

This view has received still further support from unpublished maps of another quasi-conservative quantity, ozone mixing ratio, recently obtained from the LIMS limb-scanning satellite radiometer[4]. Daily hemispheric ozone maps on the 10-mbar isobaric surface for the same period were shown to us by Professor C. B. Leovy after a first version of the present paper was circulated to colleagues and submitted for publication. This time sequence of maps shows qualitative features which, within horizontal resolution, strongly support our interpretation of the Q maps. They also provide more information in the tropics, where our estimates of Q are completely unreliable. All we can safely say about the distribution of Q in the tropics is that it must go through zero near the Equator, for reasons of dynamical stability[5,6].

Linear waves versus breaking waves

Two of the 850 K potential vorticity maps are shown in Fig. 2, for 17 and 27 January 1979, the same dates as in Fig. 1. It should be remembered that they are at best coarse-grain maps, that some of the smaller visible features may not be real, and that the real Q fields are likely to have unresolvable fine-grain structure. Owing to data problems in the tropics, details in the outermost contours are especially to be distrusted wherever they extend south of 20° N, the southernmost latitude circle shown. It is probably reasonable to think of the maps as resembling a blurred view of reality seen through a pane of knobbly glass, the size of the knobbles being of the order of many hundreds of kilometres, and more in the subtropics owing to the larger spacing of the satellite orbits there[1,2]. The quantity plotted is

$$\hat{Q} = g^{-1} H_0^{-1} p_0 Q \qquad (5)$$

in units of 10^{-4} K m^{-1} s^{-1}, p_0 being a standard sea-level pressure taken as 1,000 mbar and H_0 a standard pressure scale height taken as 6.5 km. Values greater than 6 units are picked out by the heavy shading. To obtain a feel for the numerical values, one may note from equations (1), (4) and (5) that, for a given numerical value of $\hat{Q}$, the vertical absolute vorticity component Z_0 which would be realized if $\partial\theta/\partial p$ were brought to a standard value $\partial\theta_0/\partial p$, by means of an adiabatic, frictionless,

rearrangement of mass, satisfying equations (2) and (3), is given by

$$Z_0 = \frac{-Q}{g\partial\theta_0/\partial p} = \frac{-H_0\hat{Q}}{p_0\partial\theta_0/\partial p} \tag{6}$$

Conveniently, this equals $0.2\hat{Q}$ in units of 10^{-4} s^{-1}, for example, 2.0×10^{-4} s^{-1} on the contour marked 10 (which may be compared with the maximum planetary vorticity $2\Omega = 1.458\times10^{-4}$ s^{-1}), when $-\partial\theta_0/\partial p$ is taken as 32.5 K mbar^{-1}. The latter value happens to be close to the area average of $-\partial\theta/\partial p$ on the 850 K isentropic surface north of 45° N on 17 January, according to the satellite data analysis scheme, albeit about 15% too high for the 27 January data.

Comparing Fig. 2*a* with Fig. 1*a*, we see that the main circumpolar vortex appears to be a region of substantial cross-stream potential vorticity gradients, suggesting that it is well able to support Rossby-wave propagation. That is no more than has been shown from previous analyses using eulerian averaging around latitude circles, but Fig. 2*a* gives a better idea of the tightness of the gradients and their instantaneous spatial distribution. A large part of the overall gradient between equatorial and polar regions seems to be concentrated near the edge of the heavily shaded region, while systematic poleward gradients are comparatively weak throughout most of the surrounding, unshaded region. It seems unlikely that this structure is a result of radiative heating or cooling, especially when its observed time evolution is taken into consideration (S. B. Fels and D. L. Hartmann, personal communication).

A more likely explanation will emerge from the following discussion, and will suggest in turn that the gradients near the edge of the heavily shaded region may, in reality, be even tighter than they appear on the maps. Taken at face value, as they appear in Fig. 2*a*, the numerical values of these gradients are still quite large. Expressed as isentropic gradients of Z_0 (locally equivalent to isobaric gradients of 'quasi-geostrophic potential vorticity')[7], they reach values exceeding the local planetary vorticity gradient $\beta = 2\Omega a^{-1}\cos\phi$ by well over an order of magnitude, over Canada and eastern Siberia for instance, a being the Earth's radius and ϕ the latitude. The largest contributor is the gradient of ζ. Gradients in $\partial\theta/\partial \ln p$ contribute in the same sense, as does the planetary vorticity gradient itself, but appear to be numerically less important for the coarse-grain Q distribution in the main vortex. (This suggests that the balance of terms discussed by Simmons[8] should give a good first approximation to the wave structure, insofar as it can be described by linear theory.)

Outside the heavily shaded region, there are clear signs that the dynamics is less wave-like, and indeed highly nonlinear. This is suggested particularly strongly by Fig. 2*b*, and also by the Q maps for a number of other days, omitted for brevity. Especially notable is the gross shape of the '4' contour in Fig. 2*b*, picked out by the lighter shading and emphasized in the coloured version on the front cover of this issue. A long tongue of high-Q air seems to have been pulled out from the main vortex, and to be in the process of being mixed quasi-horizontally and irreversibly into the surrounding region of weaker gradients. As we shall see, this can be accounted for in terms of clockwise advection by the large secondary vortex centred north-east of the Aleutian Islands in Fig. 1*b*. Like the rest of the pattern, this Aleutian vortex was nearly stationary in position, although growing in size, during the previous few days. The Aleutian vortex, it appears, was eating its way into the potential vorticity gradient at the edge of the main vortex, systematically reducing the area best capable of supporting Rossby-wave propagation.

Although this erosion of the main vortex, and irreversible mixing of its material into lower latitudes, is a highly nonlinear process, quite outside the scope of linear wave theory, it is familiar from various theoretical model studies. Essentially the same process is illustrated, for instance, by the time-dependent theory of 'nonlinear critical layers'[9,10]. An example is shown in Fig. 3, in which x corresponds to longitude west and $-y$ to latitude (see Fig. 3 legend for further detail). This solution was obtained analytically, using the method of matched asymptotic expansions, thus avoiding any questions about numerical resolution. It provides a dynamically consistent model example of the effect of an indefinitely persistent 'Aleutian vortex' on a set of material contours, coinciding with Q contours, which initially lie parallel to the x direction. The model Aleutian vortex is twisting up the contours like spaghetti on a fork, destroying the pre-existing, overall gradient of Q. The shading in Fig. 3 emphasizes the fact that long, thin tongues of material, like that suggested by Fig. 2*b*, are produced by this advective process. The same process has been simulated in various kinds of numerical experiment[10–14]. Figure 3 also illustrates the well known fact that such advective processes tend to produce Q fields of increasingly fine spatial scale, exemplifying the so-called 'potential enstrophy cascade' studied in the theory of geostrophic turbulence[15]. This, of course, is one of the reasons that Q is likely to vary on much finer scales than could possibly be resolved by observational data.

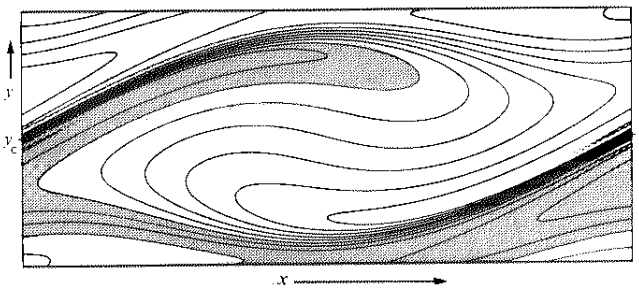


Fig. 3 Analytical solution from the time-dependent theory of nonlinear critical layers[9,10], exhibiting irreversible deformation of material contours due to advection by a two-dimensional (height-independent) 'Aleutian vortex' set up by a stationary Rossby wave on a shear flow. An equation of the form of equation (3) holds exactly, so that the contours are both material contours and contours of Q, or equally its two-dimensional equivalent, the vertical component of absolute vorticity. The shading picks out values of Q intermediate between the higher and lower values (unshaded) at bottom and at top/centre, respectively. The solution is periodic in x. The y scale is exaggerated. Initially the contours lie parallel to the x axis and represent a monotonic gradient of Q. The initial flow is in the x direction before the waves are excited, and its velocity is proportional to $(y - y_c)$ so that there is a 'critical line', where flow speed equals wave phase speed, at $y = y_c$. The time elapsed is 0.57 of the time for an air parcel to make one complete trip around the centre of the vortex or 'cat's eye'.

It might still be questioned whether the tongue of high-Q air appearing in Fig. 2*b* is real, since for reasons already indicated there is little hope of showing this directly from the data alone, especially in view of data limitations in the subtropics. Still less is there any point in trying to make direct estimates of the advection term $\mathbf{u}\cdot\nabla Q$ on the left-hand side of equation (3), because the extra differentiation involved in estimating ∇Q would make the effect of small-scale errors even worse than in Q itself. A better check is to use satellite-derived wind fields to calculate isentropic trajectories, which apart from errors in the wind fields would represent the paths of material parcels if equation (2) were exactly satisfied. A number of such trajectories were calculated, integrating backwards in time from a selection of positions within and near the tongue as it appears in Fig. 2*b*. The trajectories were found to extend back around the growing Aleutian vortex to positions close to where the edge of the main vortex had been on the appropriate date. For instance, an air parcel at 177.5° W, 35° N on 27 January, near the tip of the visible tongue, was estimated to have been over the Arctic Ocean about 4 days earlier, around 80° N and close to the 4 contour at the edge of the shaded region. The trajectory crosses the north coast of mainland Canada near Victoria Island before swinging out over the Pacific via northern California. In computing this trajectory we had to correct for the error in the gradient-wind approximation due to strong deceleration of the

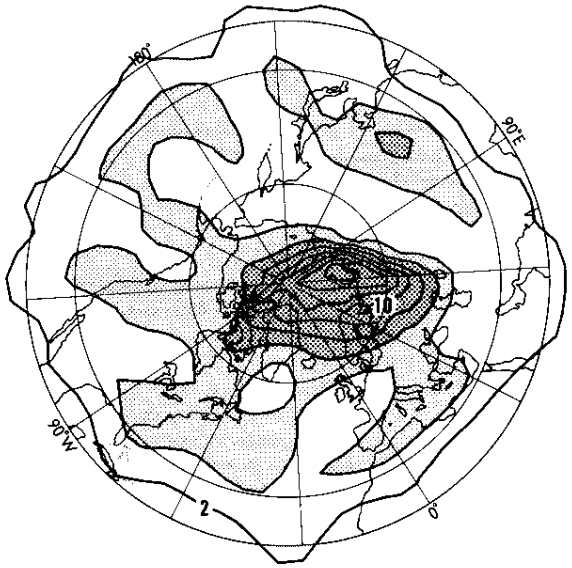


Fig. 4 Ertel's potential vorticity as in Fig. 2, but for 16 February 1979. The shading refers to same contour values as in Fig. 2. The small area of the main vortex suggests that the stratosphere was highly 'preconditioned', that is, susceptible to a major warming.

parcel as it came out of the main vortex[16]. Another air parcel, at 117.5° W, 35° N on 27 January, near Los Angeles, was estimated to have passed close to the North Pole about 2 days earlier, well within the shaded region at the time. Such estimates strongly suggest that advection of high-Q air from the edge of the main vortex could, indeed, have produced a thin tongue having the length, position and gross shape suggested by Fig. 2*b*. They suggest, moreover, that the process would have taken only a few days, so that equations (2) and (3) are likely to have been relevant despite the possible effects of radiative heating or cooling.

The phenomenon seen in Figs 2*b* and 3 is analogous, in a very fundamental sense, to the breaking of ocean waves approaching a beach, and it is appropriate to speak of a breaking Rossby or planetary wave. The basic criterion for saying whether a wave of any kind is breaking is whether material contours and surfaces are being irreversibly deformed, rather than simply undulating back and forth as is assumed in linear wave theory. Wave breaking, in this sense, is undoubtedly a ubiquitous phenomenon and is one of the most effective means whereby waves in naturally occurring flows can cause systematic redistribution not only of potential vorticity, but also of pollutants, angular momentum and other quantities of interest. Other significant examples include (1) the breaking, in the upper troposphere, of packets of Rossby waves radiated upwards and equatorwards by occluding tropospheric depressions[17–19] (vital to the global potential-vorticity and angular momentum balances, but less accessible to direct observation than the present example because the spatial scales are smaller), and (2) the breaking of upward-propagating internal gravity waves in the mesosphere, revealed by noctilucent cloud patterns and by modern radar techniques, and now believed on good evidence to be the key to an old enigma about the global angular momentum balance at altitudes above 50 km (see, for example, ref. 20). Unlike potential vorticity and material tracers, momentum and angular momentum can be transferred between, and not merely within, the sites of wave generation and wave breaking. General theories of wave, mean-flow interaction imply that the distinction between breaking and non-breaking waves—defined, as here, in terms of irreversibility or reversibility of the deformation of material contours—is fundamental to all the aforementioned processes[21]. The distinction seems to be more fundamental, for instance, than questions of detail such as whether local instabilities have any role in deforming the material contours, or whether or not the breaking can be associated with a 'critical layer' as in Fig. 3.

Although the data resolution can hardly be expected to be good enough to reveal wave breaking events of smaller scale than that in Fig. 2*b*, we may note in passing that there is more than a hint of such an event (as seen through 'knobbly glass') in the Z-shaped contour, labelled 2 units, near the bottom of Fig. 2*a*. Despite the severe data problems in this region, the feature is broadly consistent with advection by the observed velocity field during the previous few days. Perhaps more significantly, however, the theoretical considerations mentioned earlier show that the occurrence of such breaking is more or less inevitable in any case, given a few long-accepted observational facts. At most altitudes in the stratosphere, the eulerian-mean flow $\bar{u}$ around latitude circles usually vanishes in at least some part of the tropics or subtropics, as seen in a frame of reference moving longitudinally with the waves of largest amplitude. The waves are often nearly stationary, or moving slowly eastwards (as happens to be the case for the wave-3 disturbance seen in Fig. 1*a*), whereas $\bar{u}$ goes from large eastward values in the winter hemisphere to large westward values in the summer hemisphere. The kinematics of the situation implies the existence of secondary vortices, in the waves' frame of reference, somewhere in the tropics or subtropics, which for realistic wave amplitudes and durations are capable of twisting up material contours irreversibly in the manner illustrated by the analytical solution in Fig. 3.

We therefore conclude (1) that the region of weak potential vorticity gradients surrounding the main vortex is effectively a gigantic 'surf zone', in some part of which planetary waves are breaking most if not all the time, and (2) that the resulting quasi-horizontal mixing is a likely explanation of why the gradients in this zone are observed to be weak.

Encroachment of the surf zone

The analogy between the region surrounding the main stratospheric vortex and the surf zone on an ocean beach appears to be both fundamental and useful, but like any other partial analogy it should not be pushed too far. Two differences may be important. First, theoretical evidence has accumulated suggesting, perhaps surprisingly, that the stratospheric surf zone may be a fairly good Rossby-wave reflector in at least some of the conditions of interest (ref. 11 and refs therein). This is quite unlike an ordinary ocean beach, which is a good wave absorber. Second, whereas ocean beaches are more or less fixed in position, the stratospheric surf zone can encroach polewards whenever erosion of the main vortex is strong enough.

Figure 4, showing Q for 16 February 1979, seems to be an excellent illustration of this last point. The reduction in the heavily shaded area is unmistakable when Fig. 4 is compared with Fig. 2*a*. Indeed, the heavily shaded area in Fig. 4 is perhaps more appropriately compared with the lightly shaded area in Fig. 2*a*, if we define the area of the main vortex by reference to the outer edge of the region of tight gradients. We consider erosion due to wave breaking to be the most plausible explanation. It can simultaneously explain the reduction in the area of the main vortex, the corresponding broadening of the surrounding zone of weak gradients, and the persistently tight gradients marking the interface between them (apparent in all the available Q maps for the period). Gradients as tight as this, or even tighter, are characteristic of the sharp interface between a well mixed and a less mixed region when there is an overall gradient of some quasi-conservable quantity across the two regions. The phenomenon is familiar in several contexts, and has often been demonstrated in small-scale laboratory experiments[22]. The idea that vigorous quasi-horizontal mixing was occurring outside the main vortex seems consistent not only with Figs 1*b* and 2*b* but also with the entire sequence of daily Q maps for late January and early February. Most of these maps show clear signs of large-amplitude wave breaking, on a scale which seems ample to account for the observed broadening and encroachment of what we have called the surf zone.

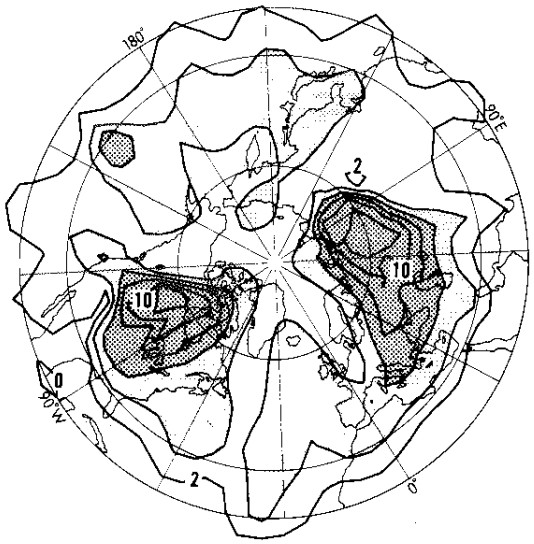


Fig. 5 Ertel's potential vorticity as in Fig. 2, but for 23 February 1979, showing the effect of the subsequent major warming.

Additional supporting evidence comes from the LIMS ozone maps shown to us by Professor Leovy. They show what seems to be the same main-vortex, surf-zone structure, evolving in time in the same way. Erosion by wave breaking can easily explain this structure and evolution because, like Q, the ozone mixing ratio is a quasi-conservative quantity in the middle stratosphere, approximately satisfying an equation of the form of equation (3) over time scales of the order of days. Alternative explanations in terms of diabatic or photochemical effects seem less likely to be successful because, although such processes cannot be expected to be negligible over the whole period in question, there is no obvious reason that they should affect ozone in the same way as Q.

Resonance, preconditioning, sudden warmings

The abovementioned differences between the stratospheric surf zone and an ordinary beach may be significant for various reasons, one of which is the much-discussed possibility[11] that the main vortex may behave like a resonant cavity. If so, changes in its area induced by wave breaking could provide a robust and effective mechanism for the nonlinear self-tuning or detuning of the cavity. In a recent theoretical study, Plumb[23] has pointed out that an approach to resonance via nonlinear self-tuning (that is, 'topographic instability')[24] can explain not only the growth of stationary wave-1 amplitude, but also the simultaneous slowing down of a weaker, westward-travelling wave-1 component in the height field, such as was observed to precede the large-amplitude wave-1 event of Fig. 1*b*[25,26]. Whilst the idea of a self-tuning resonant cavity fits closely with such observations, other explanations of the apparent travelling-wave behaviour are also possible, and the very interesting theoretical questions thus raised have yet to be settled definitively.

Another significant aspect is the 'preconditioning' of the stratosphere to make it susceptible to the occurrence of a major sudden warming. From the present viewpoint, it seems natural to regard major warmings as wave events which break up the main vortex completely, advecting low potential vorticity air over the Pole along with ozone and other tracers. The accompanying descent of air parcels, due to the adjustments which maintain approximate geostrophic and hydrostatic balance, causes adiabatic compression, and hence the large polar temperature rises giving such events their customary name. This notion of a major warming as involving complete breakdown of the main vortex does not quite coincide with the definition currently adopted by the World Meteorological Organization, but seems closer to dynamical fundamentals and is consistent, moreover, with views long held by some synoptic meteorologists[27]. The idea that some kind of preconditioning process precedes a major warming goes back to the remarks by Quiroz, Miller and Nagatani in 1975 (see, for example, ref. 11 and refs therein), and it is now widely believed that such an idea is needed to explain many of the observed facts, including the fact that major warmings do not occur more often.

Preconditioning has often been thought of in terms of changes in the eulerian-mean state of the stratosphere, defined by taking averages around latitude circles. But it now seems evident that a much more fundamental and practically useful measure of preconditioning would be in terms of the reduction in the area of the main polar vortex, or degree of encroachment of the stratospheric surf zone. The advantage of fixing attention on the area of the main vortex is that it eliminates from consideration the large but purely temporary dynamical effects, on the eulerian mean at high latitudes, of transient, reversible disturbances to the main vortex, such as bodily migrations towards or away from the pole accompanied by little or no wave breaking. Instead, it fixes attention on the more permanent, irreversible dynamical effects, in which potential vorticity is being drastically rearranged rather than simply being moved back and forth.

We predict that unusually small values of the area A of the main vortex, as it appears on daily stratospheric potential-vorticity maps, will be found to be characteristic of the conditions observed just before major warmings. The smaller the value of A, the smaller the region best capable of supporting Rossby-wave propagation. Small values of A may thus give rise to a focusing effect, tending to concentrate any subsequent wave disturbance into a smaller and less massive region than usual (a linear model of the disturbance structure again involving Simmons' approximate balance of terms[8]). Evidence suggesting the importance of this focusing effect in the events leading to a major warming can be found both in the observational data and in numerical simulations[11]. An objective definition of A which could be used to investigate its day-by-day behaviour is proposed in ref. 16.

Experiment in stratospheric dynamics

When the area of the main vortex is as small as Fig. 4 suggests it was on 16 February 1979, the preconditioning process is far advanced, according to our hypothesis, and a subsequent major warming very likely. In this case a major warming did occur, several days later. This warming was of exceptional interest for two reasons. The first was that the wave event precipitating it appeared to be unusually simple in form, with wave 2 dominating the eddy fluxes which measure the upward propagation of wave activity from the troposphere. Very large wave-2 amplitudes developed, splitting the main vortex cleanly in half in the course of a few days. The resulting coarse-grain potential vorticity distribution is indicated in Fig. 5, the map for 23 February. The unusual weakness of wave 1 during the splitting process provided direct evidence that large wave-1 amplitudes are not essential to that process[11].

The second reason that this particular warming was especially interesting was the fact that it did not occur sooner. On the basis of past case studies, one might well have expected a major warming to have occurred by mid-February[25]. In this case, however, the preconditioning process was unusually well separated in time from the warming itself. These circumstances, the lateness of the warming and the unusually simple wave structure when it did occur, combined to make it one of the most useful 'controlled experiments' in stratospheric dynamics ever performed by the real atmosphere. This has already been taken advantage of in computer simulations which have successfully 'repeated' the experiment, and also varied its conditions in several different ways[3,28]. The possibilities of such numerical experiments are still far from being exhausted.

Most of the other major warmings observed during the past two decades involved large-amplitude, complicated-looking

mixtures of waves 1, 2 and higher. However, drawing on what has been learned from the case of January–February 1979, we may anticipate that much of this apparent complexity will disappear as soon as attention is focused on the main vortex as the dynamical centre of things, rather than on any analysis tied to latitude circles, such as the conventional Fourier analysis of the height and temperature fields into separate wavenumbers. For instance, in the celebrated major warming of January 1977, the main vortex split in much the same way as in February 1979, the only difference being that it happened to be displaced well away from the Pole at the time, over Baffin Island[29]. Only the Fourier analysis was complicated, not, it seems, the physical phenomenon. Regarded as a wave-2 disturbance relative to the displaced main vortex[30], rather than relative to latitude circles, the observed phenomenon appears remarkably similar to that of February 1979. Indeed, the whole sequence of events leading to the warming, and no doubt those leading to other major warmings, can probably be viewed in essentially the same simple way as above, the main difference being that the preconditioning process (erosion of the main vortex) may not be well separated in time from the subsequent major warming. We await with interest the potential vorticity maps which will help to confirm or disprove these suggestions.

Conclusion

Our understanding of large-scale dynamical and eddy transport processes in the atmosphere would be greatly improved if daily isentropic maps of Ertel's potential vorticity Q were to become available on a routine basis. For example, they would increase our ability to judge the relevance of numerical model experiments and of theoretical concepts such as 'wave propagation', 'instability', 'critical layers' and so on. The above examples from the middle stratosphere seem sufficient to show that the effort would be worthwhile, despite the impossibility of resolving the finest scales in the potential vorticity patterns.

The coarse-grain Q maps presented here used nothing more than standard data-processing techniques applied to routinely-available meteorological data, but besides sharpening our insights into stratospheric dynamics in general, they have already suggested some specific, testable hypotheses. One such hypothesis is that large-scale wave breaking, leading to erosion of the main circumpolar vortex, will be found to be characteristic of the circumstances leading to major stratospheric warmings. The existence of long, thin tongues of material coming off the edge of the main vortex, a process which we believe we are seeing for the first time in Fig. 2*b* albeit in a blurred and distorted form, may soon be independently verifiable in this and other cases by means of data now becoming available from the newest satellite radiometers. An additional check may come from Q maps and air-parcel trajectories derived from high-resolution numerical simulations. For instance, the latest numerical forecasting models have far better horizontal resolution than data from polar-orbiting satellites, and should be able to represent a feature like the long tongue in Fig. 2*b* in considerable detail.

The prediction that A, the area of the main vortex on a given day, will be found to decrease systematically during the buildup to a major warming, should be testable even with data whose horizontal resolution is too coarse to show wave-breaking details. For this purpose the only demand on the data is that an objective measure of A be feasible[16], based on the anticipated tightness of isentropic gradients of Q near the edge of the main vortex. If time series of daily values of A, on some convenient isentropic surface, could be obtained from long, homogeneous runs of data, then other possibilities would be opened up. Such time series should be very effective as indicators of the general state of the stratosphere during a succession of winters. They represent a good combination of simplicity and finesse, since the use of daily estimates of A avoids the loss of information incurred by taking eulerian space or time means. Time series of A might substantially aid current attempts to correlate the interannual variability of the winter stratosphere with other long-term variations such as the quasi-biennial oscillation in the tropical stratosphere, or El Niño and the Southern Oscillation in the troposphere.

Assuming that our findings are confirmed, they also have implications for the next generation of stratospheric tracer transport models, with which the possible effects of pollutants on the ozone layer will be studied. Short of attempting to simulate the three-dimensional motion of the entire troposphere, stratosphere and above[31], some success has already been achieved with 'two-dimensional' models in which the material transports are represented by means of eulerian eddy fluxes relative to latitude circles. A number of significant suggestions for improving models of this kind have recently been made[32–36], and the observational picture now emerging should be a useful input to such modelling efforts. Indeed, it suggests that the most efficient two-dimensional modelling strategy might, in fact, be to abandon altogether the description tied to latitude circles, and instead to partition each isentropic layer in the stratosphere into several regions on the basis of the coarse-grain distribution of Q—say, a main vortex, a surrounding surf zone, and other regions representing the tropics and the summer hemisphere. From a theoretical point of view, this would amount to a low-resolution 'modified lagrangian-mean description'[37] of the stratosphere. The mass of each region would change with time according to observational estimates of erosion rates, in competition with the rate of replenishment of potential vorticity by diabatic processes such as IR cooling in the polar night. The detailed implementation of such a model has yet to be worked out.

For the troposphere, the best hope of obtaining the needed Q maps may be to use dynamically consistent 'data' derived from high-resolution numerical weather forecasting models and analysis schemes, rather than making any attempt to use raw observational data. We may anticipate that Q maps produced as part of the daily numerical forecasting operation will prove valuable in the routine assessment of the forecasts themselves, since they make the model dynamics highly visible. For instance, evidence now becoming available suggests that such maps will lead to immediate insights into the formation and maintenance of the mid-latitude 'blocking' situations which lead to anomalous spells of weather (B. J. Hoskins, G. J. Shutts, personal communication; see also refs 38, 39). They would also make it easier to see what dynamical instability mechanisms are likely to be operating in a given locality. Another promising area concerns the interaction between middle latitudes and the subtropics, one aspect of which is the reflectivity of the subtropics to the packets of Rossby waves radiated upwards and equatorwards by occluding mid-latitude depressions[17–19]. The theory of nonlinear critical layers suggests that this reflectivity, whose effects could be important for medium-range weather forecasting, will be dependent from day to day on the distribution of Q on isentropic surfaces in the high troposphere equatorward of the subtropical jet, where the waves may be expected to break.

A longer version of this article, with more figures and a more complete bibliography, is to appear as ref. 16. An extensive bibliography and background discussion may also be found in a recent review article by one of us[11].

We thank S. A. Clough and N. S. Grahame for generous help with the estimation of Q, P. D. Killworth and J. Venn for producing Fig. 3, J. Austin for valuable assistance with some of the isentropic trajectory calculations, C. B. Leovy for his kindness in showing us the unpublished LIMS ozone maps, also D. G. Andrews, S. A. Clough, S. B. Fels, D. L. Hartmann, P. H. Haynes, E. O. Holopainen, J. R. Holton, B. J. Hoskins, C. P. F. Hsu, K. Labitzke, R. S. Lindzen, J. D. Mahlman, R. McIntyre, A. O'Neill, D. R. Pick, R. S. Quiroz, P. B. Rhines, G. J. Shutts and A. F. Tuck for useful discussions or correspondence. The data were kindly supplied by the Met. O. 20 stratospheric analysis team.

Note added in proof: For related work on large-scale ocean dynamics, see ref. 40.

Received 20 April; accepted 18 August 1983.

1. Pick, D. R. & Brownscombe, J. L. *Adv. Space Res.* **1,** 247–260 (1981).
2. Grahame, N. S. & Clough, S. A. Met. O. 20 Tech. Note, available from U.K. Meteorological Office (1983).
3. Simmons, A. J. & Strüfing, R. *Q. Jl R. met. Soc.* **109,** 81–111 (1982).
4. Gille, J. C. & Russell, J. M. *J. geophys. Res.* (submitted).
5. Dunkerton, T. J. *J. atmos. Sci.* **38,** 2354–2364 (1981).
6. Stevens D. E. *J. atmos. Sci.* **40,** 882–893 (1983).
7. Charney, J. G. & Stern, M. E. *J. atmos. Sci.* **19,** 159–172 (1962) [see Eq. (2.31)].
8. Simmons, A. J. *Q. Jl. R. met. Soc.* **100,** 76–108 (1974).
9. Stewartson, K. *Geophys. Astrophys. Fluid Dyn.* **9,** 185–200 (1978).
10. Warn, T. & Warn, H. *Stud. appl. Math.* **59,** 37–71 (1978).
11. McIntyre, M. E. *Jl met. Soc. Jap.* **60,** 37–65 (1982).
12. Hsu, C.-P. F. *J. atmos. Sci.* **38,** 189–214 (1981).
13. Matsuno, T. *Proc. US–Japan Seminar on the Dynamics of the Middle Atmosphere* (eds Holton, J. R. & Matsuno, T., Terrapub, Tokyo (in the press).
14. Allam, R. J. & Tuck, A. F. *Q. Jl. R. met. Soc.* (submitted).
15. Rhines, P. B. *A. Rev. Fluid Mech.* **11,** 401–441 (1979).
16. McIntyre, M. E. & Palmer, T. N. *J. atmos. terr. Phys.* (in the press).
17. Gill, A. E. *Atmosphere-Ocean Dynamics*, 13.9 (Academic, New York, 1982).
18. Edmon, H. J. *et al. J. atmos. Sci.* **37,** 2600–2616 (1980) (corrigendum **38,** 1115; 1981).
19. Hoskins, B. J. in *Large-scale Dynamical Processes in the Atmosphere* (eds Hoskins, B. J. & Pearce, R. P.) 169–199 (Academic, New York, 1983).
20. Holton, J. R. *J. atmos. Sci.* **39,** 791–799 (1982).
21. Andrews, D. G. & McIntyre, M. E. *J. Fluid Mech.* **89,** 609–646 (1978).
22. Turner, J. S. *Buoyancy Effects in Fluids*, 9.1.1. (Cambridge University Press, 1973).
23. Plumb, R. A. *J. atmos. Sci.* **38,** 2514–2531 (1981).
24. Charney, J. G. & DeVore, J. G. *J. atmos. Sci.* **36,** 1205–1216 (1979).
25. Labitzke, K. *Beilage zur Berliner Wetterkarte*, SO 10/79 (Institut für Meteorologie der Freien Universität Berlin, 1979).
26. Madden, R. A. & Labitzke, K. *J. geophys. Res.* **86C,** 1247–1254 (1981).
27. Labitzke, K. *Mon. Weath. Rev.* **105,** 762–770 (1977).
28. Butchart, N. *et al. Q. Jl. R. met. Soc.* **108,** 475–502 (1982).
29. O'Neill, A. & Taylor, B. F. *Q. Jl. R. met. Soc.* **105,** 71–92 (1979).
30. Palmer, T. N. & Hsu, C.-P. F. *J. atmos. Sci.* **40,** 909–928 (1983).
31. Fels, S. B. *et al. J. atmos. Sci.* **37,** 2265–2297 (1980).
32. Harwood, R. S. *Phil. Trans. R. Soc.* A**296,** 103–127 (1980).
33. Matsuno, T. *Pure appl. Geophys.* **118,** 189–216 (1980).
34. Danielsen, E. F. *J. atmos. Sci.* **38,** 1319–1339 (1981).
35. Holton, J. R. *J. geophys. Res.* **86C,** 11989–11994 (1981).
36. Tung, K. K. *J. atmos. Sci.* **39,** 2330–2355 (1982).
37. McIntyre, M. E. *Phil. Trans. R. Soc.* A**296,** 129–148 (1980).
38. Shutts, G. J. *Q. Jl. R. met. Soc.* (in the press).
39. Illari, L. & Marshall, J. C. *J. atmos. Sci.* (in the press).
40. Holland, W. R., Keffer, T. & Rhines, P. B. *Nature* (submitted).

Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3′ consensus sequence AAUAAA

Craig Montell*, Eric F. Fisher†, Marvin H. Caruthers† & Arnold J. Berk*

* Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, California 90024, USA
† Department of Chemistry, University of Colorado, Campus Box 215, Boulder, Colorado 80309, USA

A single U→G transversion in the 3′ consensus sequence AAUAAA of the adenovirus early region 1A gene was constructed and the effect of this mutation on processing of the 3′ end of the nuclear early region 1A RNAs was analysed. The results demonstrate that the intact AAUAAA is not required for RNA polyadenylation but is required for the cleavage step preceding polyadenylation to occur efficiently.

THE AAUAAA hexanucleotide and closely related sequences are the only primary structures common to the 3′ ends of mRNAs in higher eukaryotes, excluding histone mRNAs[1]. Because most non-histone mRNAs are polyadenylated, this observation led to the suggestion[1,2] that AAUAAA is part of a recognition signal required for proper processing and polyadenylation of eukaryotic mRNAs. This was confirmed by Fitzgerald and Shenk[3], who showed that in a Simian virus 40 (SV40) mutant, deletion of the entire AAUAAA sequence shifted polyadenylation to just downstream from the next most proximal AAUAAA. The mechanism of polyadenylation in higher eukaryotes has not been established. However, it has been shown that in the synthesis of papovavirus[4–7], adenovirus[8–14] and several cellular mRNAs[15–16], transcription proceeds beyond the polyadenylation site of the mature message. An endonucleolytic cleavage probably occurs in these primary transcripts at the site of polyadenylation to generate a substrate for a poly(A) polymerase which polymerizes adenylate residues onto the 3′ hydroxyl group of polynucleotides[13,14,17–21]. To explore the function of the AAUAAA sequence in the RNA cleavage and polyadenylation steps, we constructed a point mutation in this hexanucleotide in an adenovirus 2 (Ad2) transcription unit, E1A and analysed the effects of the mutation on processing of the 3′ ends of E1A nuclear transcripts.

The mRNAs synthesized from early region 1A (E1A) of adenovirus 2 are typical of mRNAs in higher eukaryotes; they are spliced[22–24], and contain a 5′-terminal cap[25], noncoding regions at both the 5′ and 3′ ends and a poly(A) tail[26]. Figure 1 shows the sequence of the distal portion of the 3′-untranslated sequence encompassing the AAUAAA sequence and extending to the previously reported polyadenylation site near nucleotide 1,630 (ref. 26). Although some variants of this hexanucleotide have been reported[27–30], the U in the third position seems to be completely conserved. Therefore, we constructed an adenovirus mutant having a T→G transversion at this position in the DNA sequence.

RNA cleavage and polyadenylation

The T in the third position of the adenovirus E1A AATAAA sequence was changed to a G by oligonucleotide-directed mutagenesis of an M13·E1A clone[31–34]. The mutation was transferred into the adenovirus genome[35] to generate a mutant virus designated Ad2pm1610 (pm1610 = point mutation at nucleotide 1,610 in the Ad2 sequence[36]).

To analyse the effect of the point mutation on the processing of the 3′ ends of the E1A nuclear RNAs (nRNAs), we prepared nRNA from HeLa cells[11] infected with either wild-type Ad2 or Ad2pm1610. Total nRNA was fractionated into poly(A)$^+$ and poly(A)$^-$ RNA (that is, poly(A) tails shorter than 15 residues) by repeated passages over oligo(dT)–cellulose columns[37] and then analysed by the hybridization/S_1 nuclease method[38]. In this technique the RNA is hybridized to ^{32}P-labelled DNA probes in DNA excess, digested with S_1 nuclease to hydrolyse the single-stranded RNA and DNA and the S_1-protected RNA/DNA hybrid fragments are denatured and fractionated on polyacrylamide gels. As the hybridizations are performed in DNA excess and proceed to near completion, the intensities of the S_1-protected bands seen in the autoradiograms are a measure of the concentration of RNA in the analyses.

Figure 2*a* shows the S_1-protected fragments resulting from hybridization of the Ad2 E1A nRNAs to a DNA probe 3′ end labelled with ^{32}P at 1,337 in the Ad2 sequence. This probe (probe A, Fig. 2*d*) is labelled near the 3′ end of E1A and continues past the E1A polyadenylation site to include the 5′ end of the next transcription unit, E1B. The major 290-nucleotide band seen in the analysis of Ad2 nRNA is protected by the

0028–0836/83/410600—06$01.00