PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Motions in the Ionospheric D and E Regions

C. O. Hines

Phil. Trans. R. Soc. Lond. A 1972 **271**, 457-471
doi: 10.1098/rsta.1972.0016

Phil. Trans. R. Soc. Lond. A. 271, 457–471 (1972) [457]
Printed in Great Britain

Motions in the ionospheric D and E regions†

By C. O. Hines
Department of Physics, University of Toronto, Canada

Winds in the D and E regions of the Earth's upper atmosphere are analysed here into contributions derived from a general circulation, planetary waves, tidal oscillations, gravity waves and turbulence. Some characteristics of each contribution are outlined in turn. All contributions occur naturally in a superimposed state, and the interactions between them are in many respects crucial to an understanding of the system as a whole. Many of the anticipated interactions are described, and suspected consequences to other areas of aeronomical investigation are indicated, explicitly and implicitly. Both the mapping of individual contributions, and the study of interactions and consequences, are found to provide much scope for further investigation. Some topics are highlighted, that might benefit particularly from concerted observations over Europe. A critique of 'ionospheric drift' measurements is included, since these measurements contain evidence of motion that must be associated with the winds of the regions to which they refer: but it is stressed that the manner of association is not yet unambiguous.

1. Introduction

This paper seeks to outline what we know already from past research, and what we would now like most to know, about motions in the D and E regions of the Earth's upper atmosphere (at heights of 60 to 140 km, say). These motions vary on horizontal scales ranging from thousands of kilometres to a few metres, and on time scales ranging from several months to a few seconds. They may be analysed for simplicity into several superimposed dynamical systems, but these systems are found to be complex in their interactions. The individual systems have yet to be mapped in anything that would approach adequate detail, and observations that would examine their interactions are almost totally lacking. This outline is theoretically oriented in consequence, though based on real data wherever possible.

The systems of motion will be identified here as 'general circulation', 'planetary waves', 'tidal oscillations', (shorter-period) 'gravity waves', and 'turbulence', in a scheme that is somewhat arbitrary but nevertheless useful. Each will be defined and surveyed in turn, and its presumed interactions with the others will be noted as convenience permits. At the end, the ubiquitous topic of 'ionospheric drifts' will be given separate treatment, with the aid of insight gained in earlier sections.

Before the dynamical systems are discussed at any length, two general attributes of the D and E regions should be brought into mind. First, the atmosphere in these regions is far less dense than it is in the underlying meteorological regions, by factors that may exceed 10^8 if extremes are compared; and the density decreases exponentially with increase of height even within the D and E regions, the 'scale height' of the exponential ranging from 5 to 30 km there. The wave systems that are to be discussed tend to conserve energy flux when they propagate upward, and the winds they produce tend to grow in amplitude with height in order to offset the diminution of gas density. This rather basic characteristic accounts in large measure for the importance of

† This paper constitutes a written version of an invited oral review paper. The circumstances in which it was perforce prepared have precluded any extension of scope beyond that appropriate to an oral presentation. Interested readers will find further discussion and reference to original sources in earlier reviews by the same author (Q. J. R. met. Soc. 89, 1–42 (1963) and Met. Mono. 9, 114–121 (1968) and in a collection of papers from a Symposium on upper atmospheric winds, waves and ionospheric drifts (J. atmos. terr. Phys. 30, no. 5 (May 1968)).

waves in the upper atmosphere, but it has a further consequence to which increasing attention is being directed: when the waves interact with the background flow of the general circulation, they do so with a strength that derives largely from the dense regions beneath, to which the tenuous regions of present concern must give compliance.

Secondly, and complementary to the first, the decrease of gas density is accompanied by an increase of kinematic viscosity (and of thermal conduction, whose consequences in this context are often much the same). The molecular kinematic viscosity increases exponentially from 5×10^{-2} m²/s near the 60 km level to 10^4 m²/s at 140 km. At the lower levels, its effects are negligible in all the dynamical systems except turbulence; but turbulence itself introduces an eddy kinematic viscosity that is estimated at 10^2 to 10^3 m²/s, at least at the upper levels of its occurrence, at meteor heights (i.e. from 85 up to 105 km, or thereabouts). This eddy viscosity, and then the greater molecular viscosity at greater heights, is of some consequence to the other dynamical systems. It may, for example, offset or even reverse the growth of wave amplitude with height that would otherwise occur; and, by dissipating wave energy, it may force a transfer of wave momentum to the background flow. For virtually all of the components of motion to be discussed, the transition from a nearly inviscid behaviour to a viscously controlled behaviour occurs within the span of the D and E regions. It will provide a recurrent theme in the sections that follow.

2. General circulation

'General circulation' will be taken here to mean, in some ill-defined way, just what the term conjures up in one's mind without further precise specification. It is to be thought of as the gross motion of the atmosphere, global in scale, slowly (e.g. seasonally) changing for the most part, but subject to modification in a matter of hours or days if the processes that govern it alter substantially. It does not, however, include the planetary waves that are isolated for separate discussion below, nor for that matter the other dynamical systems that are to be treated. Instead, it provides a background field of motion upon which the others may be thought to be superimposed.

The general circulation is not well mapped even up to the 100 km level, much less above. Most of the relevant data come from a few sites at middle latitudes. They show winds that are directed toward the east in winter, with some hint of a reversal occurring near the 100 km level, and winds that are directed toward the west in summer up to 80 km or so, reversing to an eastward flow at higher levels. Within the uncertainties of the observations, the flow is geostrophically balanced: the Coriolis force associated with the east–west motion on our rotating planet just offsets the north–south pressure-gradient force, the pressure gradients being those that result from north–south variations of temperature when a hydrostatic balance is maintained in the vertical direction under the influence of gravity.

The temperature variations cannot be attributed solely to radiative processes. Instead, there must be dynamical processes operative near the top of the D region, at 90 km, which serve to heat the atmosphere in winter at latitudes of 50 to 60°, and perhaps to cool it in summer at similar latitudes. Possibly the whole of the high-latitude zone is involved, poleward of 50°: existing data are quite inadequate, and must be augmented (as they could be, by suitable measurements made over northern Europe) if the full global pattern is to be known.

A subsidence of the atmosphere through the 90 km level in the winter hemisphere, and a rising in the summer hemisphere, could provide the requisite heating (by adiabatic compression, aided by a release of the energy of atomic oxygen recombination) and the cooling. These vertical motions would have to be accompanied, at greater heights, by a flow of gas from the summer to the winter hemisphere, and so by a meridional component of the general circulation. Such a component is to be expected, in order to provide (through its associated Coriolis force) a balance with the zonal stress that is generated by viscosity in the presence of vertically varying shears in the east–west circulation. Estimates of the required vertical speeds ($\sim$ 1 cm/s) are compatible with estimates of the expected horizontal speeds ($\sim$ 1 to 10 m/s) and the demands of mass conservation, but these speeds lie within the probable errors of any direct means of detection. They do gain indirect confirmation in magnitude and sense from certain interpretations of winter-time oxygen airglow and summer-time noctilucent clouds, however. If the meridional winds were enhanced at levels of 110 to 130 km, as they might well be by virtue of the increased kinematic viscosity there, they could account (via horizontal transport) for an observed excess of helium at still greater heights, at high latitudes in winter. But in that case, too, they should be directly detectable at the higher levels, and identifiable as a part of the general circulation. They should be sought.

The general circulation provides, as has been said, a background field of motion upon which are superimposed the other dynamical systems. It can alter them, particularly the planetary and gravity waves, and does so in a fashion that will be discussed in later sections. But it can also be affected by them, through processes we may now consider.

One effect of turbulence has already been treated, in that it would be responsible (at levels below about 110 km) for the viscous stress that would drive the suspected north–south circulation. Turbulence will also contribute to the heating that is required near the top of the D region at high latitudes in winter: it is expected to be enhanced at the place and time in question, and even at its more typical intensities it is thought to provide a heating (through dissipation of its energy) of 1 to 10 K/day, which is of the general order required.

The wave systems that are to be discussed carry into the E region each day an amount of momentum, scalarly averaged, that probably exceeds the momentum already there. Most of this momentum must be deposited in the general circulation, as the waves themselves are dissipated. The tidal contributions must be strong over much of the globe, though to some extent variable in direction with latitude. The gravity-wave contributions are likely to be irregularly distributed, but equally important in total unless the vectorial averaging of their contributions leads to strong cancellations in each area affected. This question of momentum transfer by waves, with potentially significant consequences to the general circulation, has only recently come into prominence. It will attract much attention in coming years, and should be supported observationally whenever possible.

The tidal oscillations achieve such large amplitudes in the E region that inherent nonlinearities in their cyclic variation become substantial. These introduce 'd.c.' components of motion that might well pass as, and indeed might well be counted as, a part of the general circulation. Tidal interaction with the ionization of the region, which itself is varying diurnally, leads to further 'd.c.' components; but these become of sensible magnitude only as the very top of the E region is approached.

Some consequences of dynamical processes upon other aspects of aeronomy will bear mention. The possible influence of the general circulation upon airglow, noctilucent clouds, and the distribution of helium, will already be evident from the foregoing discussion and should be tested by studies specific in their design. The distribution of atmospheric constituents in the D region, particularly of neutral trace constituents and of associated esoteric ionic species, is likewise subject to influence. Variations in the latter species high in the D region are thought to be a likely cause of the 'winter anomaly' that occurs in ionospheric radio-wave absorption, which itself constitutes an area of major research activity at the present time. The suggested variations might result from a number of processes of dynamical origin, including regional variations of the general circulation brought about by momentum transfer from waves, or brought about more directly by a specific type of perturbation to which we now turn.

3. Planetary waves

Planetary waves are global in scale, with horizontal wavelengths of a few thousand kilometres: the full range of longitudes may encompass a single dominant oscillation ('wavenumber 1') or a few such oscillations, while waves of somewhat smaller scale (with longitudinal wavenumbers to 10 or 15, say) appear in lesser strength as weather systems form and decay. These waves carry a large part of the energy of the troposphere, and produce much of the deformation from purely zonal flow that is depicted on standard weather maps. Their strongest members are often stationary relative to the ground, being generated by continental configurations, or else move slowly with the weather systems to which they are linked. Because of their long time scales, measured in days, they appear (as on weather maps) to be deformations of the general circulation and may indeed be counted as part of it for many purposes. But they have certain properties that make their isolation as a separate dynamical system a useful exercise.

Chief amongst these, for present purposes, is their ability to propagate upward. With this goes the ability to produce wind perturbations that increase with height, and the ability to carry energy and momentum (and potential vorticity, etc.) with them to the upper levels, for transfer ultimately into the background flow. Indeed, if this ability were not offset in large measure by other effects, the atmosphere above the D region would be altered out of all recognition from the state in which we actually find it.

The principal offsetting process results from interaction with the background flow of the general circulation at stratospheric and mesospheric heights. Upward transmission is totally prevented, in fact, if the waves reach a 'critical layer', at which their Doppler-shifted frequency (as measured in a coordinate system moving with the background flow) is reduced to zero. Transmission may be strongly suppressed under other conditions as well, when the wave energy is reflected back to lower levels. Detailed but rather idealized calculations have indicated that, at middle latitudes, there should be little penetration into the D region from the troposphere whenever the winds of the intervening levels are westward as they are in summer, or eastward and strong ($\geqslant$ 40 m/s) as they are in winter, while penetration may be suspected near the equinoxes or in suitably abnormal circumstances. An idealized wave-packet analysis, designed to take latitudinal variations of the general circulation into account, suggests that substantial penetration might be achieved to the D region polewards of 50° latitude in the winter hemisphere. The potential consequences implicit in this suggestion are so great, that major efforts at direct observational verification would seem to be well warranted.

Among these consequences is some role in the production of the 'winter anomaly' of radio absorption. The precise role is not yet clear, but one line of argument evolves from the fact that the winds produced by the waves might carry a parcel of atmosphere some hundreds of kilometres horizontally, and simultaneously several kilometres vertically, in a matter of hours or days. Thus they can bring to a given observational point an atmosphere that retains, aside from relaxation effects, the composition appropriate to quite a different location and height. Important changes might be effected in this way, particularly in the concentration of minor constituents whose variation with height is rapid. Alternatively, adiabatic heating and cooling produced by the wave may have a significant effect if, as is suspected, the ion chemistry is strongly temperature-sensitive. In any event, there is a widespread hope that planetary waves will prove to be the key to the occurrence of anomalous absorption, and this hope is strengthened by a correlation of anomalous days with peculiar 'stratospheric warmings' which are themselves associated with planetary waves.

Planetary waves that do penetrate to the D and E regions must ultimately be dissipated. For 'typical' wave parameters, well away from critical layers or reflection levels, this will occur when the effective kinematic viscosity attains a value of 10^2 m^2/s or so. This value appears to be reached low in the E region, if not below, in the presence of 'typical' intensities of turbulence. But, as has been remarked already, more intense turbulence may well occur in the winter at high latitudes; dissipation well within the D region would then be expected, and would call into question the relevance of planetary waves to the winter anomaly. Clearly, speculation can be piled on speculation in this area. What is needed now is direct observational evidence of planetary-wave winds, or of their absence, in the D region at high latitudes in winter, preferably in correlation with measurements of composition, thermal structure, and other parameters that might be linked to absorption, not omitting absorption itself.

4. Tidal oscillations

The tides of the atmosphere are generated in part gravitationally as in the oceans, both by the Sun and by the Moon, but in greater part thermally by the Sun. They are inherently global in scale, and their dominant components propagate westward to maintain a constant phase at the subsolar or sublunar point as appropriate. Their corresponding speeds of propagation are sufficiently great, that they should be relatively unaffected by the winds of the general circulation. Though an 8 h solar tide can be detected in meteor-radar winds on occasion, and a 12 h lunar tide can be inferred from magnetic and radio-absorption records of sufficient duration, nevertheless the only components that play a major role in the dynamics of the D and E regions are the 24 and 12 h solar oscillations. These may be discussed best by further analysis into individual 'modes'. Each mode is characterized by a specific latitudinal variation, defined with the aid of a corresponding 'Hough function', and by an associated specific form of vertical structure. The Hough functions for the modes of greatest interest here are depicted in figure 1; the wind, pressure, density, and temperature fluctuations that are produced by these modes depend on the Hough functions and on their latitudinal derivatives.

4.1. *The 'propagating' diurnal tide*

For many purposes, greatest interest centres on a certain component of the 24 h tide, one that will be designated here as the '1, 1' mode (though it is termed the '1, 3' mode in some parts of the recent literature, a nomenclature that appears to be disappearing). This mode propagates in the vertical, with a vertical wavelength of 20 to 30 km through the region of interest, and its important effects are confined to latitudes below 50°.

The '1, 1' mode is generated strongly in the ozonosphere by heating there, but over a depth commensurate with its wavelength and hence in a manner that, through interference effects, is inefficient for purposes of propagation to other levels. The bulk of the '1, 1' energy that reaches the E region, where the most dramatic consequences are to be found, comes from insolation absorption by water vapour in the troposphere, whose depth is more comparable to half a vertical wavelength. The energy, once it is injected into the tidal oscillation, propagates up to the E region in the course of some four or five days. The associated winds increase in amplitude with height, and are readily detectable at meteor heights as a common occurrence. Indeed, with

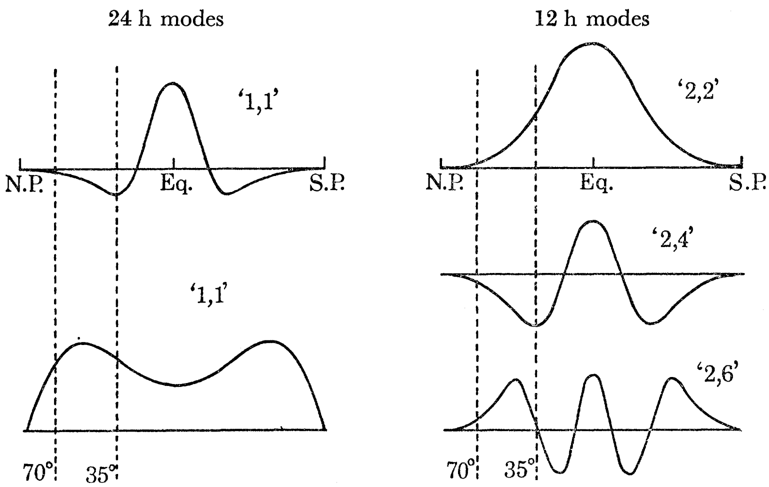

FIGURE 1. Latitudinal structure, from North Pole (N.P.) through the equator (Eq.) to South Pole (S.P.), of the Hough functions for certain solar tidal modes of importance to the D and E regions of the Earth's upper atmosphere. Two diurnal modes (of 24 h period) are illustrated: the '1, 1' mode (also designated the '1, 3' mode in some recent literature) and the '1, −1' mode (also designated, with increasing frequency, the '1, −2' mode). The former propagates vertically while the latter is inherently evanescent in the vertical. Three semidiurnal modes (of 12 h period) are illustrated, none of which is inherently evanescent: the '2, 2', '2, 4' and '2, 6' modes. Higher-order modes no doubt exist, but are relatively unimportant. Modes that are antisymmetric about the equator also exist, and may be excited in some degree under appropriate conditions. The wind, temperature, pressure and density variations produced by a given tidal mode contain terms proportional to the corresponding Hough function itself and to the latitudinal derivative of that function. Vertical dotted lines represent latitudes 70° and 35° N, the range between them being roughly the range spanned by Europe.

wind speeds of 50 m/s and more on occasion, they may come to dominate completely the wind spectrum near the 105 km level; and, with the relatively short vertical wavelengths that characterize the mode, their zig-zag wind profiles contain some of the strongest shears detected at those heights. Associated temperature fluctuations are believed to become so severe, in an equatorial belt at meteor heights, that the atmosphere is rendered convectively unstable and turbulence is presumed to result.

Conversely, the '1, 1' mode is subject to severe energy dissipation when the kinematic viscosity comes to exceed 300 m^2/s, as it does precisely at these levels, in part because of turbulence. Indeed, there must be quite a fine balance attained between normal upward amplification of the tide on the one hand, and dissipative losses to turbulence on the other, for maximum amplitude to be reached at a level as high as 105 km. Certainly it could not be reached much higher, because of the increase of molecular kinematic viscosity with height; but it might be reached much lower, as a consequence of the eddy viscosity that accompanies turbulence. It is possible that the sensitivity of this balance accounts for one observed feature, that the strength of the '1, 1' mode at meteor heights is highly variable over periods of a few days; but correlative observations on such a time scale have yet to be made, and an alternative explanation might be found in the variation of tropospheric water vapour or in complex interference effects.

The '1, 1' mode has been identified observationally in the E region almost exclusively on the basis of its vertical wavelength at a given location. There has been only the grossest form of confirmation on the basis of latitudinal variations of intensity – a comparison of the 24 h components of meteor-level winds at 35° S with those at 53° N – and a much more systematic study, such as might be made over Europe, is highly desirable.

Higher order modes of the 24 h propagating tide, having more latitudinal fine structure than the '1, 1' mode, are more closely confined to low latitudes and even there are generated with lesser intensity. They also have shorter vertical wavelengths, and so are more subject to dissipation by viscous effects. It has been suggested that they might account for vertical fine structure in winds at heights of 90 to 105 km, such as has been revealed by the deformation of individual meteor trails and rocket-released vapour trails; but it is a moot question, whether they could survive the effects of dissipation adequately to do so. (The critical kinematic viscosity for the '1, 3' mode is only one sixth that of the '1, 1' mode, and that of the '1, 5' mode is only one fourteenth, these being the next-order modes symmetric about the equator.)

4.2. *The evanescent diurnal tide*

A distinct category of diurnal oscillations exists, whose individual modes are inherently evanescent (non-propagating) in the vertical, and whose energy therefore tends to reside near its level of insertion. The upward decay of energy density is more rapid than the decline of gas density, so the associated winds decrease in amplitude above as well as below the region of generation. The most important of these modes will here be termed the '1, −1' mode, in conformity with most of the literature that deals with it (though one systematic method of numbering, which may well grow in prominence, leads to the designation '1, −2' for it).

This mode is generated strongly in the troposphere, but with no consequence to the D or E region. The ozonosphere provides a further source, but one that is thought to produce winds no greater than 10 m/s in the D region, and less in the E region. A further significant source is thought to occur within the E region itself, very likely as a consequence of insolation absorption by molecular oxygen. Evidence for its existence, is, however, indirect: it comes from ground-level quiet-day magnetic variations, which result from 'dynamo currents' in the E region, the pattern of these currents being linked in turn by dynamo theory to a wind pattern whose major component is that of the '1, −1' mode. This mode would be highly efficient in the production of ground-level magnetic fluctuations, because of its freedom from phase variations in the vertical. (Contrast the '1, 1' mode, for example, whose oscillatory winds would tend to produce offsetting currents at different levels.) Nevertheless, despite its efficiency, wind speeds of the order 50 m/s are required of the '1, −1' mode to account for the observed magnetic records. Confirmation of such winds in the dynamo region, particularly at high latitudes where they are most needed, is not yet available from direct measurements; it should be sought. Conversely, magnetic records obtained from ground stations provide a convenient means of maintaining a watch on this mode, once their interpretation in these terms is accepted.

Higher order modes of the evanescent type exist, with an emphasis at high latitudes that complements the low-latitude bias of the propagating modes. They are unlikely to occur in significant measure, however, because of their extreme mismatch with the latitudinal distribution of insolation. At present, it seems appropriate to ignore them.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

4.3. *The semidiurnal tide*

Wind variations with a 12 h period are detected routinely at meteor heights, with an amplitude of 20 to 40 m/s. They may result in part from nonlinear effects in the 24 h tide, but are almost certainly dominated by distinctive 12 h modes that are generated directly by solar heating. The relevant modes are not of the inherently evanescent type, although one of their members becomes locally evanescent in the mesosphere as a consequence of the particular temperature profile of that region. This distinctive member is the '2, 2' mode, which is dominant at lower elevations, but which is attenuated relative to the other 12 h modes as its energy leaks upward through the region of evanescence. Chief among its competitors are the '2, 4' and '2, 6' modes, which may be expected to produce winds comparable to those of the '2, 2' mode in the E region, and which may indeed have been correctly identified there on occasion.

The '2, 2' mode originates predominantly in the ozonosphere, at heights of 35 to 55 km, through insolation absorption. Its energy reaches the overlying D and E regions in part by the direct path, and in part indirectly after reflexion from the ground (or indeed, after multiple reflexion between ground and mesosphere, the leakage energy being a superposition of energy inputs from several days). The '2, 4' and '2, 6' modes are likewise generated in the ozonosphere, at least in part; but, as with the '1, 1' mode whose vertical wavelength is comparable, this generation process is inefficient for purposes of propagation, and the bulk of their energy in the D and E regions is likely to have come from insolation absorption by water vapour in the troposphere. This suggests that the '2, 4' and '2, 6' modes might correlate well with the '1, 1' mode, if amplitude variations over a few days are a consequence of variations in the water-vapour content, but no such correlations have yet been sought.

Indeed, separation of the 12 h tide into individual modal contributions has been attempted only on rare occasions. The distinctive vertical structures of these modes provide one method by which separation might be achieved: in the E region, the 'local' vertical wavelength of the '2, 2' mode is of the order 100 km, while that of the '2, 4' mode is variable (with height) over the range 30 to 55 km and that of the '2, 6' mode over the range 20 to 35 km. Though the peak winds of the '2, 4' and '2, 6' modes may not be as strong as those of the '2, 2' mode, they may nevertheless be more apparent in individual wind profiles as a consequence of their more rapid height variations. Their shorter vertical wavelengths are accompanied by greater attenuation under the increasing effects of viscosity and thermal conduction: the '2, 6' mode should attain its maximum amplitude at a height of 110 km or so, while the corresponding elevation for the '2, 4' mode is about 125 km; the '2, 2' mode may continue to increase in amplitude with height throughout the E region, coming to an asymptotic value, which itself is in part electrodynamically influenced, in the overlying F region.

To date, such modal identification as has been made has rested upon the vertical structure almost exclusively. Complementary identification by latitudinal structure is much to be desired, preferably by means of direct wind measurements. It will be seen from figure 1, or better from curves that properly compound the Hough function and its latitudinal derivative, that the latitudinal span of Europe provides an excellent opportunity for the detection of the anticipated latitudinal structure. In this connexion it may be remarked that a statement commonly made about tidal oscillations is incorrect or at best misleading: the winds of individual modes (other than the '1, 1' and '2, 2' modes) do *not* rotate clockwise with the passing of time at all northern latitudes (nor counterclockwise at southern latitudes), even though they may do so over most latitudes; certainly the combined winds of superimposed modes need not do so at all.

The '2, 2' mode in the E region, with its relatively slow height variation of phase, is an efficient generator of magnetic variations at ground level. It appears to be second only to the '1, −1' mode in its absolute contribution, and sufficiently strong that that contribution can be distinguished. Identification in this case involves the latitudinal structure inherently, and may be pursued on a more economical (if less certain) basis than is possible by direct wind measurements at height. It should be exploited more fully in coming years.

5. Gravity waves

The semidiurnal tide, and the non-evanescent diurnal modes, are members of a spectrum of atmospheric waves that carry the general appellation 'gravity waves' (or sometimes, particularly if high frequencies are to be included, 'acoustic-gravity waves'). But they are special members, generated in a precise fashion by systematic processes that recur day after day, and for this reason they have been treated separately here. We turn now to a more irregularly occurring portion of the spectrum, and reserve the term 'gravity waves' for that portion.

Gravity waves may be taken to have a short-period cut-off at a certain resonant (Brunt–Väisälä) period, lying in the range 5 to 10 min in the D and E regions. Their strongest members occur at somewhat longer periods, typically 1 to 3 h, where they begin to overlap high-order tidal harmonics (though the latter are not normally in evidence above the 'noise' imposed by the former).

The only certain natural sources of these waves in the upper atmosphere are provided by the auroral electrojets, and perhaps, on one occasion each, by an eclipse and by an earthquake, while nuclear detonations in the troposphere and at higher elevations have provided unmistakable man-made sources. Tidal oscillations within the E region may act, through nonlinear processes, both to generate gravity waves locally and to transfer energy within a pre-existent gravity-wave spectrum. There is reason to believe that meteorological sources provide the bulk of that spectrum in the D and lower E regions – sources such as weather systems, jet streams, mountain lee waves, etc. – but only the most tentative identification of such sources has yet been possible, and that only on rare occasions. Isolation of sources in individual cases, and identification of their statistical distribution by type, season, and geographical position, constitute outstanding problems in the study of these waves.

Whatever their sources, the superimposed gravity waves of the D and E regions introduce irregular wind components, of the order 1 to 30 m/s, maximizing near a height of 105 km. Their vertical scale size (as measured by a full wavelength – a 'zig' and a 'zag' in the vertical wind profile) is typically 5 to 50 km, and the corresponding horizontal scale size 10 to 300 km. Individual waves propagate with vertical trace speeds of 1 to 20 m/s and horizontal trace speeds of 20 to 100 m/s, more or less. Through the constantly changing interference of superimposed waves, the patterns of irregular wind fields move with similar speeds and superimposed random fluctuations.

The spectrum of gravity waves that reaches the D and E regions from source regions low in the atmosphere is only a filtered fraction of the spectrum that is launched. The filtering is caused in part by internal reflexion, partial or total, associated with the background thermal and wind structure of the intervening levels, and perhaps more seriously by removal at critical layers where individual waves are Doppler-shifted to zero frequency and their energy and momentum are lost to the background. The background flow of the general circulation, and the background temperature distribution that drives it, will account for the gross features of the filtering process. Perturbations caused by planetary waves and tidal oscillations can be equally effective in certain circumstances, however, for their relatively long time-scales make them appear to be part of the 'background', as that background is viewed by the shorter-period gravity waves. It seems likely that planetary waves in the middle atmosphere are effective in varying the filtering of gravity waves at middle and high latitudes, at least in wintertime, and so are effective in varying the dynamical state of the D and E regions indirectly even if they fail to penetrate in a strength adequate to produce a direct change. (Indeed, by enhancing local vertical shears in the general circulation, they may also give rise to instabilities that generate gravity waves.) Likewise, the very strong winds of the '1, 1' tide at latitudes below 45°, near the 105 km level, may be responsible for scavenging much of the gravity-wave energy that reaches this level and so for causing an observed sharp reduction of the energy immediately above. These possibilities are important, but as yet difficult to test, in large part because of the difficulty of isolating individual wave components and of identifying their sources.

Some possibilities for overcoming these difficulties are beginning to emerge. Noctilucent clouds often reveal well-defined families of gravity waves, such that ray-tracing backward in time may be carried out; and a preliminary examination has suggested a linkage to a distorted flow of the tropopause jet stream. One may, with good but uncompelling reason, suggest that the statistical enhancement of sporadic E layers over Japan is associated with gravity waves launched from the Himalayan Mountains; and one would then be led to suspect that a corresponding effect might be found in Europe in association with the Alps (and in the Americas, in association with the Rocky and Andes Mountains, of course).

Gravity waves must ultimately be filtered by dissipative processes, even if they escape substantial reflexion by the background wind and temperature structure. Dissipation occurs automatically at critical layers (unless those layers are themselves unstable), but otherwise it requires an adequate value of the effective kinematic viscosity. High in the E region, the increase of molecular kinematic viscosity (and thermal conduction) with height appears able to account for the decreasing spectral range of the gravity waves that are found, the smaller-scale and longer-period components being removed progressively; in the overlying F region, the spectrum is reduced and the irregular pattern of superimposed waves gives way to quasi-sinusoids whose most frequent manifestation is in the form of travelling ionospheric disturbances (t.i.d.). Low in the E region, on the other hand, and in the D region, dissipation is controlled by turbulence. Indeed, it seems likely that the very existence of turbulence at these heights is a consequence of gravity waves, at least outside the narrow equatorial belt in which the '1, 1' tidal mode is thought to be unstable. This is a matter to which we now turn.

6. Turbulence

Various authors mean various things by the term 'turbulence', not excluding randomly superimposed planetary and gravity waves. Here the term will be confined to a system of motion that is irregular in its nature, describable in any reproducible fashion only by its statistical attributes, inherently diffusive rather than propagating, inherently dissipative rather than conservative of energy, local in that its sources and sinks of energy cannot be far removed, and above all nonlinear in its internal interactions, to the point that a linearized mathematical description (which could be developed automatically in terms of waves) is quite inadequate to provide even a first approximation to its characteristics.

There is observational evidence that leads to a widespread belief, that turbulence as thus defined extends 'typically' through the D region and into the E region, to heights of 100 to 110 km. This evidence derives from the diffusive growth of meteor trails and rocket-released vapour trails, from observed irregular patterns of motion and of ionization on a scale too small in space and too long in time to be attributed to any known type of wave, from a requirement to maintain a well-mixed atmosphere up to heights of the order 100 km in order to account for the concentrations of minor constituents at higher levels, from a requirement to account for the manner of transition from molecular to atomic oxygen near the 100 km level, and from a requirement for a mechanism adequate to remove from the overlying levels the radiative (not to mention the dynamical) energy that is constantly being deposited there in the form of heat.

All of this evidence is compatible with the following as 'typical' parameters describing the turbulence at a height of 95 km, say: spacial scale sizes (of 'eddies') ranging continuously from 500 to 15 m, more or less, and corresponding temporal scales from 300 to 30 s; an eddy diffusion coefficient and kinematic viscosity in the range 10^2 to 10^3 m^2/s; and a dissipation rate of 10^{-2} to 10^{-1} W/kg, which would produce a heating of about 1 to 10 K/day. Very similar values are to be expected at other heights, except that the small-scale end of the spectrum must be expected to increase in scale with height as the molecular kinematic viscosity increases, until, somewhere near the 105 km level, it comes to match the large-scale end of the spectrum and turbulence ceases to exist. This is, of course, an over-simplified picture of the true behaviour, based upon the very crudest of approaches to the study of turbulence. Extensions of observation and theory are currently being effected, but often along discordant lines: the topic is replete with controversy, and the best avenues for further advance are by no means clear.

The energy dissipated by turbulence must be continuously or repeatedly replenished, if the turbulence is to remain in being. It can be derived from convective instabilities or from shear in the background régime. Convective instability arises when the temperature decreases sufficiently rapidly – 'superadiabatically' – with increase of height. As previously implied, the '1, 1' tidal mode is believed to produce superadiabatic gradients of temperature in an equatorial belt, over heights within the range 85 to 110 km. Again, gravity-wave winds at middle latitudes imply gravity-wave temperature gradients that are superadiabatic, over a similar height range and perhaps at lower elevations as well. And direct measurement of temperature gradients at heights of 45 to 95 km, believed to be induced by gravity waves, reveal them to be superadiabatic on frequent occasion, particularly at high latitudes in winter. The regions of superadiabaticity are suspected to be the primary regions of initiation of turbulence. They occur in layers, and so may be expected to produce layers of enhanced turbulence, though the turbulence, once formed, may extract further energy from the shearing winds as the waves propagate by. (Some isolated layers of turbulence have in fact been reported.) In the case of gravity-wave generation, turbulence may also be initiated directly in the regions of shear even in the absence of superadiabatic gradients.

This description of sources leads to the anticipation that the 'typical' parameters listed earlier might be representative only of some particular set of circumstances or of some effective global average. They are in fact taken from fine-structure observations at middle latitudes, and are compatible with calculations of vertical diffusion that employ globally representative background data. One must suspect an enhancement of turbulence intensity in an equatorial belt, another enhancement at high latitudes in winter, and perhaps a suppression at those latitudes in summer. The height of termination of turbulence is likely to vary as well, particularly at high latitudes; heights of 115 km in winter and 85 km in summer need not be surprising, for example. And even regional variations might occur, associated with orographic sources of gravity waves.

Such variations as these have yet to be sought, much less found. The closest approach has been provided by an analysis of meteor winds at a single mid-latitude (35 °S) station. It revealed a seasonal variation of turbulence intensity, with maxima in late summer and late winter. This correlated positively with the intensity of the 24 h tidal winds, in what was an atypical year for those winds. The intensity of gravity waves was not recorded separately, but rather as an extension of the turbulence spectrum to greater scales; a positive correlation was implicit. These limited results should stand as the forerunners of many more extensive studies in coming years.

Some consequences of turbulence have aready been discussed. The damping of planetary waves, of the propagating 24 h tidal oscillations, and of the smaller-scale gravity waves will be recalled readily. A contribution to the temperature distribution of the D and E region, and thence perhaps to the pattern of the general circulation, was likewise mentioned. We may add here the possibility of a contribution to the winter anomaly of ionospheric absorption, which might be brought about either through variations in the vertical mixing of minor atmospheric constituents or through heat deposition if the ionic reactions are sufficiently temperature-sensitive. Turbulence could thus provide a link between anomalous absorption and gravity waves, and through the latter a link to stratospheric warmings. Turbulence below the mesopause is also thought by some to be essential to the formation of noctilucent clouds, by providing a mechanism for raising water vapour to the requisite heights. And turbulence in the D and E regions can affect the thermal budget, the chemical composition, and through it the shape of the ionospheric layer, in the overlying F region. For all these reasons, in addition to the interest it captures on its own, turbulence in the upper atmosphere warrants much fuller study than has been given it to date.

7. Ionospheric drifts

Radio signals reflected from the ionosphere exhibit a rapid fading of intensity of a type that must be attributed to dynamical processes. When signals from a single transmitter are recorded at a number of closely spaced receivers, the fading patterns at the different receivers are usually found to show good correlation with one another, and this correlation can be improved if appropriate time shifts are introduced. The time shifts are taken to be indicative of a motion and distortion of some ground-level pattern of enhanced and diminished signal strength, which in turn is taken to be indicative of some motion (with half the speed) and distortion of an irregular distribution of ionization near the height of radio reflexion. The motions thus deduced are categorized as 'ionospheric drifts'.

The term 'drifts' was chosen at the outset of these studies in order to make it neutral with respect to the geophysical content of the data, as between, for example, a bodily motion of air parcels (or ionization) in what might be termed a wind, or a propagational motion of patterns that might be formed by waves. Attempts to maintain this neutrality until the issue was resolved have not always proven successful. Much of the literature concerned with drifts has adopted

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

a 'wind' terminology, in some cases explicitly and with warning, but in others tacitly and without warning; and it has gone so far as to speak of departures from mean motions as being due to turbulence. Drift data have, with only passing comment, been included with meteor and other observational data in various attempts to map out the general circulation and the winds of superimposed wave systems (as distinct from the *patterns* that might be produced by those systems). Nevertheless, the matter is still inadequately resolved.

Conceptual justification for a 'wind' interpretation is not difficult to find, in regions where turbulence exists. Turbulence might well deform the pattern of ionization distribution sufficiently to produce the irregular fluctuations of amplitude that are observed and that constitute the very basis of the whole technique. Turbulence may be expected to be carried with the winds of the region in which it exists, those winds being a composite of the general circulation and the winds of planetary, tidal, and gravity waves, all superimposed. A long-term mean motion in the drift data might then be expected to yield the general circulation, with successively shorter-term deviations representing the various wave-induced winds. A final short-term spectrum of deviations would remain to be associated with variations in the pattern of turbulence and in the associated process of radio diffraction.

Such a picture, plausible as it appears and may well be, collapses completely in regions where there is no turbulence. Turbulence cannot then produce the irregularities whose existence is fundamental to the technique, and its motion with or without the winds of the region is a meaningless concept. And yet drift observations extend regularly upward into the E region, and even into the F region, well above the levels in which turbulence is believed to occur on any regular basis.

In the absence of turbulence, one must seek an alternative mechanism for the production of the ionospheric irregularities whose existence is so crucial to the technique. And, if such a mechanism is found, one must suspect that it might also be relevant even in regions where turbulence does exist.

Such a mechanism is provided by gravity waves. The winds that these waves produce and the density and temperature variations as well deform the distribution of ionization from its quiescent smooth form. The precise nature of the deformation is a matter for detailed examination, but the *pattern* of deformation must result from point-to-point variations in the wind–density–temperature field, and hence must result from variations in the phase of the wave or, more generally, from the interference effects that are associated with phase variations of superimposed waves. The horizontal wavelengths of gravity waves include a range that is of the right order (10 to 50 km, say) to produce more than one reflexion point (or principal Fresnel zone) for the probing radio waves, and hence, by interference of those waves, to produce fading such as that observed. (Planetary and tidal waves cannot be important in the direct production of the irregularities that produce the fading simply because their horizontal wavelengths are too great.)

Given such a pattern of irregularity, any motion of that pattern such as the drift technique would reveal would then be associated with the phase propagation of the waves, not with the winds that are caused by the waves. If the wave spectrum were broad, interference effects would be severe, and the net motion of the pattern as a whole would bear only some indirect relation to the phase progression of the individual components; it might well be meaningless for geophysical purposes. If the wave spectrum were narrow, at least as measured by the range of horizontal (vector) trace velocities, then the motion of the pattern as a whole (divided by two) would match that of the horizontal trace velocity common to all the waves. The result would be geophysically meaningful and useful for the study of the waves, but it would not bear any *a priori* relation to the winds of the region.

Despite this forewarning, circumstances might arise that would permit the gravity waves to reveal the background wind (including general circulation and the winds induced by the planetary waves and tides). A particular case is illustrated by gravity waves nearing a critical layer, where they have properties that might tend to enhance their observability in drift measurements. The patterns produced by such waves necessarily move with the background wind, as a consequence of the very meaning of a critical layer. There is some possibility that this selection process would be operative some of the time if not all of the time; but advantage could be taken of this possibility only if it were known to be realized in fact, on a given occasion or statistically in given circumstances. We have as yet no firm basis on which to make such a judgement.

The current observational status may be summarized as follows, at least for middle latitudes. The characteristics of drift irregularities high in the E region, and in the F region, blend smoothly with those of t.i.d. in the F region. The latter are caused by gravity waves unambiguously. In so far as there is a transition of characteristics on passing from the former to the latter, it is precisely the transition that is to be expected from increased dissipation in small-scale waves and growth of amplitude with height in larger-scale waves. The stronger t.i.d. are, moreover, occasionally accompanied by the formation of sporadic E layers that propagate with them; if such strong E-region irregularities are induced on some occasions by gravity waves, then weaker ones, still perfectly adequate for purposes of the drift measurements, must be anticipated as a common occurrence at other times. In short, all indications point to gravity waves as being the source of the drift irregularities above the levels of turbulence, which is to say, above a height of about 110 km quite typically. The motions revealed by drifts at such levels are then to be associated with phase propagation, in a fashion that could be meaningful for wave studies in certain circumstances. Whether or not these motions might also be representative of the background winds is quite a different matter, not yet open to resolution. It should be pursued.

Contrasting with this situation, there is a substantial body of literature that establishes a close correlation between drift-deduced motions and real winds, at heights below 110 km. Both the general circulation and the tidal components of wind appear to have been resolved successfully, and a direct correlation with meteor observations has been obtained. These findings lend empirical support to the use of drift observations for purposes of determining the background winds, on a widespread basis even when correlative measurements are lacking. As will be apparent from earlier arguments, however, it does not automatically decide between turbulence-induced irregularities and wave-induced irregularities (or indeed, a mixture of the two), though it would demand of the latter some selection process associated with critical layers.

Until the matter is more clearly resolved than at present, there must remain the nagging doubt that the drift observations even at these lower heights are wave-related, and that the selection process associated with critical layers will not always be operative, hence that the data cannot be guaranteed a valid interpretation. This doubt is if anything enhanced by measurements recently made. Some of these have shown substantial discrepancies between E-region drifts and winds revealed by vapour trails in about half the data. Others, still to be published, have examined the structure of the fading radio signal more carefully and have indicated that the E-region irregularities responsible for the fades have wavelike properties, being composed of a number of long ridges separated horizontally by distances of the order 10 to 50 km. Such shapes and scales are not compatible with a turbulence interpretation, at least if 'turbulence' is taken to be as described here.

It must be clear that much remains to be done in establishing the true significance of drift data, at least in the E region if not in the D.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES