ATMOSPHERIC GRAVITY WAVES: A NEW TOY FOR THE WAVE THEORIST

by

C. O. Hines,

Department of the Geophysical Sciences, University of Chicago, Chicago, Ill.

Abstract

This article surveys briefly some features of atmospheric gravity waves, with the objective of bringing their characteristics and their problems to the attention of wave theorists from other disciplines. The inherent anisotropy of their propagation, and further anisotropies and reflection processes that arise in practice, find analogues in ionospheric and other plasma propagation. They provide a novel field for the application of old techniques, and they pose challenging problems that will give birth to new.
1. Introduction

Waves provide a common mathematical framework which unites a wide variety of geophysical phenomena, and indeed unites many of the broader range of physical disciplines as well. Early studies of elastic, hydrodynamic, acoustic, seismic, optical and electron waves helped to establish a familiarity with wave behavior that sped the development of a vast literature on ionospheric radio propagation, and, more recently, on plasma waves in general. These topics in turn have had their influence on the older studies, by providing a deeper physical understanding and an expanded range of mathematical techniques, that were born in response to the challenging new situations encountered.

The objective of the present paper is to draw the attention of ionospheric and plasma wave physicists to yet another type of wave process, one whose practical relevance and difficult problems have come to be appreciated only in recent years. Its objective is, moreover, to solicit the active interest of some who read it, in the expectation that an understanding of the relevant phenomena will be furthered thereby.
The waves in question are to be found in an atmosphere that is stratified by a gravitational field. They may be termed 'atmospheric gravity waves' for convenience; but they differ somewhat from the gravity waves that form on the surface of the seas, and in the atmosphere they merge to some extent with acoustic and tidal waves. Their oscillation and propagation are anisotropic, as a consequence of gravity, and they often occur in media whose properties change substantially within a wavelength. They then give rise to problems analogous to those which are familiar in ionospheric and other plasma propagation, and it is for this reason that the foregoing solicitation of active interest seems appropriate.

Though the subsequent discussion will develop primarily along more abstract lines, some indication of the practical relevance of atmospheric gravity waves is in order at this point. Substantial interest in them dates from the late 1940s, when Queney [1947, 1948] and Scorer [1949] first applied them to the 'lee waves' that are set up by winds flowing over mountains. (See Queney [1948] for references to earlier work.) Martyn [1950] employed them, in a superpositioning of modes that led to what he termed 'cellular waves', in order to explain microbarometric oscillations in the troposphere. This explanation was soon confirmed in substance by the observations and analysis of Gossard and Munk [1954]. Eckart [1960, p. 120] goes so far as to suggest that gravity waves, again in a cellular pattern, may constitute the commonly observed fluctuating component of normal winds (aside from that portion which is generated by flow around local obstacles).
Martyn [1950] applied the cellular-wave concept to a class of moving ionospheric deformations that are commonly termed 'traveling ionospheric disturbances'. It later appeared [Hines, 1960] that the observed waves, while indeed gravity waves, were probably not cellular at ionospheric heights. This later analysis also showed that atmospheric gravity waves could account for a wide range of winds and irregularities nearer the base of the ionosphere, that had come under intensive study by various techniques during the intervening years.

It has recently been inferred [Hines, 1964] that the gravity waves may heat the ionosphere at least as strongly as does solar radiation. Similarly, the high temperature of the solar corona may be maintained by gravity waves that are launched upward from the photospheric surface [Whittaker, 1963].

The subsequent discussion will be concerned, not with these various applications and their consequences, but rather with the more formal aspects of the wave analysis. Even in that area it will be brief. Readers wishing to pursue the mathematical treatment further are referred to an extensive analysis of atmospheric (and ocean) waves by Eckhart [1960] or a more compact review by Tolstoy [1963], both of which provide a general perspective, while some topics not treated by them are developed in research papers that are cited in context below.
2. Wave Perturbations of an Idealized Atmosphere

We shall be concerned exclusively with an atmosphere whose properties, in the unperturbed state, vary only in the (vertical) direction specified by the $z$ coordinate of a cartesian system. The gravitational acceleration $\mathbf{g}$ is constant, with components $(0, 0, -g)$; the ratio of specific heats ($\gamma$) and the mean molecular mass ($\mathcal{M}$) are likewise constant; and the atmosphere is free from rotation.

In the present section, the unperturbed atmosphere is taken to be at rest relative to the cartesian system; it has constant temperature ($T$) and correspondingly constant 'speed of sound' ($C \equiv [\gamma k T/\mathcal{M}]^{1/2}$, where $k$ is Boltzmann's constant). It supports the propagation of waves that are only of perturbation magnitude, and these are free from dissipation. The various restrictions that are imposed by these idealizations will be relaxed, each in turn, in succeeding sections.

Under the influence of gravity, the unperturbed gas density ($\rho$) decreases exponentially with increasing height:

$$\rho(z) = \rho_0 \exp \left( -\frac{z}{H} \right)$$

where $\rho_0 = \rho(0)$, and $H$ is the 'scale height' of the atmosphere:

$$H \equiv C^2/\gamma g \equiv k T/\mathcal{M} g.$$  

The unperturbed gas pressure ($p$) follows the same exponential variation as $\rho$, in order that the force density $\rho \mathbf{g}$ may be balanced by $-\mathbf{\nabla} P$. 
Perturbations from rest, with velocity $\mathbf{U}$, are accompanied by fluctuations $\rho'$ and $p'$ in the gas density and pressure. The gas is then subject to the additional forces $\rho' \mathbf{g}$ and $-\nabla p'$, which balance the inertial force $\rho \frac{d \mathbf{U}}{dt}$ and lead to oscillation. The oscillation is subject to the equation of state for the gas, and to the equation of mass conservation.

The present discussion will be confined for simplicity to the two-dimensional case in which $\partial / \partial y = 0$ (say). Then standard wave analysis reveals that $U_y = 0$, and that $\dot{U}_x, \dot{U}_z, \rho'/\rho$ and $p'/p$ all have the same exponential form, being proportional to

$$\Phi \equiv \exp \left( \frac{z}{2H} \right) \cdot \exp \left( \frac{i (\omega x - k_x x - k_z z)}{2} \right)$$

with

$$k_x^2 = (\omega / \alpha)^2 - 1) k_x^2 + (\omega^2 - \omega_0^2) \gamma^2$$

where

$$\omega_g \equiv (\gamma - 1) \frac{k}{g/C} \quad \text{and} \quad \omega_a \equiv \gamma g / 2 C.$$ 

In practice, $\gamma$ is always less than 2 and therefore $\omega_g < \omega_a$.

In application to the earth's atmosphere, the periods corresponding to $\omega_g$ and $\omega_a$ lie typically in the range 5-15 minutes.
The factor \( \exp \left( \frac{z}{2H} \right) \) in (3) leads to an increase of amplitude with height (provided it is not offset by an imaginary component of \( k_x \)), and in practice this helps to account for the importance of the waves at ionospheric heights. Its physical interpretation lies in the fact that it just offsets the decrease of gas density, in maintaining the vertical flux of wave energy (\( \propto \rho \mathbf{U}_x^2 \)) constant. This result is exact, despite the fact that certain properties of the medium, namely \( \rho \) and \( \mathbf{p} \), vary with \( z \); difficulties that are normally encountered in the treatment of inhomogeneous media do not arise.

The anisotropy imposed by gravity is already apparent in (4), since \( k_x \) and \( k_z \) enter that equation differently except when \( g = 0 = \omega_0 \). Its effect is further evident when the 'polarization relations' between \( \mathbf{U}_x, \mathbf{U}_z, \rho'/\rho \) and \( \mathbf{p}'/\mathbf{p} \) are examined.
We shall be concerned in this section with propagating waves for which $k_x$ is pure real. From (4) it is apparent that $k_x$ may be pure real or pure imaginary; the waves may be 'internal' or 'surface' waves, respectively, the latter category being equivalent to 'evanescent' waves in ionospheric propagation. If attention is confined to the internal waves, examination of (4) reveals that the frequency range $\omega < \omega < \omega_a$ is prohibited, and the available spectrum then falls into two distinct categories. In the high-frequency limit, (4) approximates to
\[
(k_x^2 + k_z^2) = \omega^2/c^2,
\]
which corresponds to simple sound waves; the frequency range $\omega > \omega_a$ is conveniently termed the 'acoustic' range. Complementary to this is the low-frequency 'internal gravity wave' range, $\omega < \omega$, toward which interest is directed in the present paper. Many applications require, however, that evanescent waves and waves with complex $k_x$'s be included, and then the two separate categories must be recognized as divisions of a broader class of 'acoustic-gravity waves'.

The pure 'acoustic' and 'internal gravity' sequences are illustrated, by means of their dispersion relation (4), in figure 1. The diagram actually depicts the relation between $k_x$ and $k_z$ for a variety of wave periods, but it employs the normalized coordinates
\[
\frac{n_x}{\lambda} \equiv \frac{k_x c}{\omega}, \quad \frac{n_z}{\lambda} \equiv \frac{k_z c}{\omega},
\]
which correspond to components of a refractive index vector.
Any combination \( \frac{n_x}{n_z} \) implies a particular direction of phase propagation, of course, while the related direction of energy propagation may be inferred by means of the geometrical construction illustrated in the inset diagram. The fact that the vertical components of phase and energy \( \text{velocity} \) are oppositely directed, in the gravity waves, leads to an illusion if surfaces of constant phase are kept under observation: the waves appear to descend from higher levels, when their energy is actually propagated upward from underlying sources.

The shorter-period waves are seen to propagate nearly isotropically, just as we expect of sound waves, whereas the longer-period waves are highly anisotropic. Analogies may be drawn with VHF waves in the ionosphere, for example, which maintain nearly isotropic propagation, and then with the whistler waves that propagate quite anisotropically at much longer periods. Indeed, the form of the gravity-wave curves in figure 1 is not unlike the corresponding curves that would be drawn for whistlers, if the geomagnetic field were taken as the \( x \)-axis. There is this difference, however: with the \( x \)-axis the axis of symmetry, as for whistlers, energy propagation tends to be confined to directions about that axis, whereas with the \( z \)-axis the axis of symmetry, as for gravity waves, energy propagation tends to be confined to directions about the perpendicular (i.e., horizontal) plane.
A further distinction between gravity waves and whistler waves (or magneto-ionic waves in general) bears mention: while the gravity waves are anisotropic, they are not gyrotropic. There is no double refraction -- no splitting into differently polarized components with different phase speeds -- and, in the absence of atmospheric rotation effects, the oscillatory motion in any given mode is confined to the vertical plane of propagation.

3. Effects of Temperature Variation; Ducting

Consider now an atmosphere in which \( T = T(z) \), from which \( C = C(z) \), and begin with the case in which the stratification takes the form of a series of isothermal layers. The earlier equations will apply to each layer in turn, and the matching of boundary conditions will demand the constancy of \( k \) and \( \omega \) through all layers. Evidently \( k \) will vary from layer to layer, and the processes of refraction and partial reflection will occur. If \( k \) becomes imaginary at some level, and remains imaginary in the half-space above, total reflection will result. Total reflection can also occur at ground level in practice, and it is possible then to establish a wave duct in the atmosphere. This duct will be subject to complications of the type met in the ducting of LF radio waves below the ionosphere, and a similar mode analysis will be applicable to it. Computations of some of the modal characteristics have been published by Pfeffer and Zarichny [1962] and Press and Harkrider [1962], based on a multilayer approximation to the actual atmosphere.
Imperfect ducting is also a distinct possibility, as will be recognized readily. The ground-level boundary condition in this case requires that $k_x$ be complex in the modes of interest [e.g. Gossard, 1962], leading to a decay of energy in the direction of horizontal propagation. (This statement contrasts with the assumption of Eckart [1960], that $k_x$ be real, which led him to confine consideration to 'eigenmodes' of a type that have here been termed 'cellular' waves.) Imperfect ducting may be established both by the troposphere [Gossard, 1962] and by the mesosphere [Hines, 1960], while the leakage energy may account for the waves observed at ionospheric heights; waves strongly ducted by the mesosphere, but nevertheless still leaking energy upward, appear to account well for the traveling ionospheric disturbances previously mentioned [Hines, 1960].

The use of multiple isothermal layers as an approximation to a real atmosphere has its analogue, of course, in problems of radio propagation, and there it has been thoroughly justified. No similar justification has yet been presented for the gravity-wave problem, nor indeed does it appear at first sight that one can be established. The reason for this will now be outlined.
In the variable-temperature case, the unperturbed pressure assumes the functional form

$$ p(z) = p_o \exp \left( - \int_0^z \frac{dz}{H} \right), \quad (8) $$

with $H$ given by (2) again, as one might expect. But the atmospheric density assumes the somewhat different form

$$ \rho(z) = \rho_o \frac{H}{H_0} H^{-1} \exp \left( - \int_0^z \frac{dz}{H} \right), \quad (9) $$

where $H_0$ is the value of $H$ at $z = 0$. The normalized derivative $-\frac{dp}{\rho dz}$ enters the governing equations -- both through the equation of state and through the equation of mass conservation -- and appears ultimately as a parameter in any dispersion equation that may be inferred. This derivative was given by $H^{-1}$ in the isothermal case, and entered (4) by way of $\omega$ and $a$. Here, however, it is given by $H^{-1}(1 + \frac{dH}{dz})$, and there is no provision for the derivative $\frac{dH}{dz}$ in the dispersion equation as previously written.

It is worth noting explicitly that the correction term appearing here is of order $\frac{dH}{dz}$, and not, for example, of order $k \frac{H^{-1}}{2} \frac{dH}{dz}$. More specifically, this term is independent of the vertical wavelength, and does not necessarily become small if the medium is taken to be 'slowly varying' in the traditional sense (i.e., variations within a wavelength being small). Accordingly, the problem is not of the type that is normally encountered in ionospheric radio propagation, nor does it appear to have any immediate bearing on approximations of the WKB type.
It might be supposed that a multi-layer technique could be applied despite the foregoing difficulty, if only a suitably revised form of (4) were developed and applied. Martyn [1950], employing a variable \( \phi \) proportional to \( p^{\frac{1}{2}} \text{div} \mathbf{U} \), obtained for it a differential equation that transforms to

\[
\frac{\partial^2 \phi}{\partial z^2} + H^{-1} (\frac{dH}{dz}) (\frac{\partial \phi}{\partial z}) + q^2 \phi = 0 \tag{10}
\]

in present notation, where

\[
q^2 = (p_B^2 - p^2 - 1) \frac{k_a^2}{\gamma} (p^2 - p_a^2) / c^2 \tag{11}
\]

and

\[
\omega_B^2 \equiv \left( \gamma - 1 + \gamma \frac{dH}{dz} \right)^{\frac{1}{2}} g / c \tag{12}
\]

A WKB approximation to the solution of (10) can be found readily:

\[
\phi \propto \left| qH \right|^{-\frac{1}{2}} \exp \left[ -\int q \, dz \right] . \tag{13}
\]

The conditions for validity of this approximation are of the usual type, e.g., \( |dH/dz| \ll |qH| \), etc. It appears, then, that the multi-layer technique might well be applied if (4) were replaced by (11), (3) by (13), \( k_2 \) by \( q \), and so \( \omega \) by \( \omega_B \); in the neighborhood of zeros of \( q \), more sophisticated matching approximations would of course be required. (See Martyn [1950] for one example; see also Yamamoto [1957] for a fuller development of a four-layer model, with constant \( dT/dz \) in each layer.)
This non-isothermal approach to the multi-layer method is probably quite valid, though it has yet to be established rigorously. But even it is fraught with certain dangers of interpretation. For, if a solution were sought for some variable other than \( \varphi \), the relevant dispersion equation would be found to have a form that could differ from (11): \( \frac{dH}{dz} \) would enter the equation for \( q^2 \) in some different way. Tolstoy [1963], for example, derives a highly complicated relation -- given by his equations (4.4.) and (4.10) -- for the \( q^2 \) that applies to the vertical displacement, and similarly complicated relations may be established for other oscillatory parameters. (Tolstoy's apparently simpler equation, (4.18), has no real application to the atmosphere, except in the elementary isothermal case.)

The danger implicit in a conceptual application of the revised multi-layer approach should now be apparent. One normally expects severe reflection to occur when \( q^2 \) becomes negative and remains negative over a substantial height range. But the conditions that give rise to this situation will vary with the parameter for which \( q^2 \) was determined, whereas the question of reflection applies simultaneously to all parameters. Reflection might be judged to be serious, on inspection of the \( q^2 \) of one parameter, and it might be judged to be negligible on inspection of the \( q^2 \) of another. These apparent discrepancies should not survive a thorough mathematical treatment, of course, but until they are understood they will inhibit a proper conceptual evaluation of reflection processes.

A rigorous analysis by Weston [1961, 1962] avoids any discrete layering of the temperature profile, and succeeds in deducing rather general conclusions about the conditions that will produce ducting and about the ducted modes. It represents only a beginning, however, to the task that lies ahead.
The revised dispersion relation, (11), if accepted at face value, can lead to qualitative differences from the isothermal case. In particular, the condition $\omega_B > \omega_a$ can be met in practice, and there is then no sharp division between acoustic and internal gravity waves. A continuous transition now occurs, as illustrated in Fig. 2 for representative conditions. The orientation of the energy velocity relative to the direction of phase propagation can be determined as before, by drawing normals to the contours of constant period, but the sense of the energy velocity requires more careful examination. It may be shown that the vertical components of the energy and phase velocity are similarly directed for some modes (represented by broken contour lines) and are oppositely directed for others (solid contours); the transition from the one behavior to the other, along a given constant-period contour, is accompanied by infinite energy velocities (as measured by $\omega/\partial k_x$, $\omega/\partial k_z$).

4. Effects of Horizontal Background Winds.

The occurrence of horizontal background winds can affect the propagation of gravity waves, and so can lead to further problems of refraction and reflection. The analysis is relatively direct if a multi-layer model is assumed once again, with a constant wind in each of the layers. Conceptually, one could establish co-moving coordinates for each layer in turn, and then the relevant formalism would be that for zero wind. The matching at layer interfaces would demand a constancy of $k_x$, while $\omega$ would suffer a doppler shift from one layer to the next if the increment of wind contained an $x$ component; $k_z$ would have to adjust accordingly.
As in the case of the temperature variation, however, a continuous variation of wind with height introduces new terms into the dispersion relation [e.g., Scorer, 1949; Martyn, 1950], and these terms do not become negligible as $k_z$ is increased. All of the difficulties indicated above now recur, and have yet to be understood in detail.

The problem of winds adds a new facet to the discussion, for they will automatically introduce an anisotropy with respect to the various directions of horizontal propagation. In terrestrial applications, for example, winds of the mesosphere can act as a directional filter on the gravity-wave spectrum that is propagated upward from below, and the waves that actually penetrate to ionospheric heights then carry information about the winds through which they have passed. All but superficial analysis of this filtering action remains for the future.

Winds add yet another phenomenon to the study, for they are capable of exchanging energy with the waves [Eliassen and Palm, 1961]. Though the processes are totally different in character, this exchange finds an analogue in the traveling-wave amplification of plasma physics.
5. Energy Dissipation

As gravity-wave energy rises into the rarer atmosphere, the effects of viscosity and thermal conduction become increasingly severe. They act to dissipate the energy of the wave, and convert it into heat. As already indicated, the heat produced in this manner may well be significant both to the terrestrial ionosphere and to the solar corona, and of course we may speculate about other planetary atmospheres as well.

The problems of energy dissipation are largely problems for a fluid dynamicist, though they contain some interesting and only partly resolved difficulties for wave theorists in general [cf. Pitteway and Hines, 1963]. Perhaps the most fascinating aspect of the dissipation, both from a layman's and from an expert's point of view, is that the attenuation can be virtually seen in progress in the vast laboratory of the upper atmosphere. Shorter-wave modes (≈ 10 km, vertically) are 'quenched' at heights of 110-120 km, in distances of the order of a wavelength, while longer-wave modes proceed upward to be quenched at some greater height [Hines, 1960, 1964b; Kochanski, 1964]. The increasing use of rocket-released vapor trails, for the study of winds at ionospheric heights, will put this quenching process more clearly into view and will render it susceptible to more detailed analysis.
Hydromagnetic dissipation, resulting from the currents that are induced in the ionosphere when a gravity wave propagates through it, becomes important in practice at periods of half an hour or so. A new anisotropy is introduced, and selective damping occurs. Gravity waves that propagate in the magnetic meridian plane, with a suitable inclination of wave front, will suffer little from this process because the neutral gas tends to oscillate along the magnetic field lines. Waves of different tilt, or different azimuth of advance, establish neutral-gas motions that cross magnetic field lines. Such waves suffer varying degrees of hydromagnetic dissipation, depending on the precise parameters that characterize them.

The hydromagnetic problem has been examined only in preliminary fashion, to date, and most of what has been done remains unpublished as yet. Its relevance is worth noting, however, if only because it introduces processes that will be more familiar to some readers.
6. Nonlinear Effects

As previously noted, the amplitude of internal gravity waves tends to increase with altitude. The exponential growth that is indicated in (3) may not obtain in full, because of reflection and dissipation; but the possibility of substantial growth is quite real, it must be catered to in principle, and it is often realized in practice. Along with it goes the possibility of the waves exceeding perturbation amplitude, and hence of nonlinear effects occurring. Representative estimates reveal, in fact, that much of the wind system present in the earth's upper atmosphere, if indeed due to gravity waves, already constitutes an appreciably nonlinear system. Wave interaction, and perhaps a cascading of energy through the wave spectrum, must then be contemplated. This aspect of gravity waves remains virtually virgin territory, and awaits with untapped charms the onslaught of the first venturesome wave theorist.

Acknowledgment

This paper was prepared with the support of the National Aeronautics and Space Administration, under grant NsG-467 Research.
References


Figure Captions

Fig. 1. Contours of constant period in the refractive index domain, for internal waves in an isothermal atmosphere. The elliptical (broken-line curves) sequence to the left represents acoustic waves, and the (solid-line curves) hyperbolic sequence to the right represents internal gravity waves. The periods, measured in minutes, are shown in boxes on each contour in turn. The basic parameters adopted are $\gamma = 1.4$, $g = 9.5 \text{ m/s}^2$, and $H = 6.0 \text{ km}$, which obtain in the terrestrial atmosphere near the 90-km level. The relation between phase and energy progression is indicated by the insert diagram.

Fig. 2. As for Fig. 1, but now for a temperature-varying atmosphere. The vertical refractive index, $n_z$, is given by $aC/\omega$, with $a$ given by equation (11). The basic parameters adopted are $\gamma = 1.4$, $g = 9.5 \text{ m/s}^2$, $H = 8.7 \text{ km}$, $dH/dz = 0.59$, which obtain in the terrestrial atmosphere near the 110-km level. The sense of energy flow, along the normal to the 3.40-minute contour, is indicated by two short arrows. The sense of energy flow for other contours may be obtained by continuous transformation, the broken-line and solid-line segments being treated separately.