The Saturation of Gravity Waves in the Middle Atmosphere.
Part IV: Cutoff of the Incident Wave Spectrum

COLIN O. HINES
Arecebo Observatory, Arecebo, Puerto Rico

(Manuscript received 4 May 1992, in final form 3 December 1992)

ABSTRACT

A Doppler-spread theory for the saturation of middle atmosphere gravity waves was presented in an earlier member of this sequence of papers. It employed a model in which a broad spectrum of waves subject to linear theory is incident from below. The spectral distribution (in vertical wavenumber \( m \)) is deformed, as it propagates upward, in response to the growing importance of the Eulerian advective nonlinearity imposed on each wave by the total wave-induced wind. The deformation is such as to statistically spread the spectrum towards larger \( m \), with the largest-\( m \) waves being progressively obliterated in quasi-critical-layer interactions. The model invoked a cutoff of the incident spectrum at a vertical wavenumber specified as lying in the range 0.5–1.0 times the local buoyancy frequency divided by the rms wind speed, with the choice 0.5 being adopted tentatively. A qualitative argument for the chosen cutoff wavenumber was presented but was not supported by any more certain quantitative analysis at the time. The present paper derives an analytic form for the cutoff function, illustrates it in application, and provides quantitative support for a value possibly as low as 0.5 in the stratosphere and a value possibly as high as 1.0 in the mesosphere. In addition, it slightly recasts the heuristic approach to the Doppler-spread analysis, and it admits to certain difficulties, associated with the largest-\( m \) waves, whose circumvention appears to require a far more detailed analysis of wave-wave interaction through the advective nonlinearity.

1. Introduction

The irregular horizontal winds of the middle atmosphere commonly exhibit what is termed a saturated spectrum (in vertical wavenumber \( m \)), whereby the spectral form approximates to \( AN^2 m^{-3} \) at sufficiently large \( m \), \( N^2 \) being the stability parameter (buoyancy frequency squared) and \( A \) being an almost universal constant lying in the range \( 1/6 < A < 1/2 \), almost irrespective of meteorological conditions, location, time, and height [e.g., Dewan et al. (1984), Wu and Widdel (1991)], and comparison between them]. This spectrum of winds is most often attributed to a spectrum of irregular gravity waves, though an alternate (or overlapping) interpretation in terms of turbulence has been suggested by some.

In the gravity wave interpretation, the quasi-universal spectral amplitude and form have been attributed to linear instability (Dewan and Good 1986; Smith et al. 1987), to diffusive dissipation imposed by nonlinear wave–wave interactions (Weinstock 1990), and to Doppler spreading imposed by nonlinear wave–wave interactions (Hines 1991a,b; henceforth HI and HI respectively). The last of these echoes a corresponding attribution in the case of corresponding observations in oceanic wave spectra (Flatté et al. 1985; Allen and Joseph 1989), and is the one to be pursued here.

In HI and HI, I established the need, and developed an approximate theory, for the inclusion of the advective nonlinearity of the Eulerian fluid dynamic equations in the description of gravity waves propagating upward through the middle atmosphere. The theory exhibited features that could be identified with the features of observed saturated spectra. It was subject, however, to three distinct criticisms: 1) it treated the wave-induced winds as if they were horizontal background winds, uniform in time and in the horizontal, albeit with retention of the actual \( m \) spectrum; 2) it invoked an incident \( m \) spectrum that extended only up to some cutoff wavenumber \( m_C \), whose existence was inferred and whose value was set only empirically, albeit with subsequent qualitative justification; and 3) it failed to yield the asymptotic form \( m^{-3} \) at large wavenumber, though I argued that it came close enough to that form over a sufficiently broad range of \( m \), near and above \( m_C \), to account for the observations.

The first of these criticisms is intrinsic to the approximate method adopted for dealing with the nonlinear wave–wave interactions, and I have come to conclude that the third follows almost inevitably from it—a point to be elaborated on briefly in later discussion. These criticisms can be overcome, if at all, only.
by the development of an accurate wave–wave interaction theory, and their nemesis may already lie in the analyses of Allen and Joseph (1989) and/or Weinstock (1990)—another point to be elaborated on briefly in later discussion.

The present paper is designed to remove the second criticism. The previously qualitative argument leading to the existence and location of the spectral cutoff is here converted into a fully developed, quantitative analysis that yields a well-defined cutoff function. This is achieved in section 3, after the development of underlying relations in section 2. Evaluation of the cutoff function requires the evaluation of a certain decorrelation distance that depends on the wind spectrum, and that is done in a brief section 4 for a representative spectrum. I illustrate the form of the cutoff function in section 5, and then, in section 6, the form of the observable spectrum into which this cutoff function, in conjunction with the Doppler-spreading effects, converts any initially incident spectrum—a form that is remarkably similar to actually observed spectra in the vicinity of the spectral knee. A summary of the Doppler-spread theory is then given in section 7. The body of the paper virtually ignores the effects of the earth’s rotation, but they are introduced and discussed in the Appendix.

2. Doppler-spread spectra

I adopt the physical model, the conventions, the approximations, and to a large extent the notation, of HI and HII. In the model, the atmosphere is taken to be gravitationally stratified, nonrotating (until the Appendix), windless, and isothermal or nearly so, in the absence of waves. A spectrum of gravity waves is incident from below, having wavenumbers \( \{ k, l, m \} \) and apparent—that is, groundbased—frequencies \( \omega \), with \( m \) and \( \omega \) positive. (Phase variations in the waves are taken to be of the form \( kx + ly - mz - \omega t \) in order that this convenient sign convention may be adopted for the assumed ongoing waves; \( z \) is the vertically upward coordinate.) Since there is no background wind, the apparent frequency of a wave is equal to its intrinsic frequency \( \Omega \). Vertical wavelengths are taken to be no greater than the atmospheric scale height \( H \); hence, \( mH \gg 1 \) and the dispersion relation approximates to

\[
\Omega m = Nh, \tag{2.1}
\]

where \( h = \sqrt{[k^2 + l^2]^{1/2}} \) is the horizontal wavenumber and \( N \) is the buoyancy frequency of the unperturbed atmosphere (denoted \( N_0 \) in HI and HII and taken to be constant); \( \Omega \ll N \) has been assumed, as for a hydrostatic approximation.

It is further assumed that a multiplicity of randomly phased waves or wave packets is present in the observed region of the atmosphere at any given time. The central limit theorem then establishes that, at least in the absence of overriding nonlinearities, the distribution of horizontal wind components induced by the waves shall be Gaussian, in general with different rms values in different azimuths. Though the initial development in HII retains this azimuthal generality, here, for convenience, I assume the more specialized case of an azimuthally isotropic spectrum of waves. This spectrum yields for any one horizontal wind component the rms value \( 2^{-1/2} \sigma_T \), where \( \sigma_T \) is the square root of the total horizontal wind variance.

The Doppler-spread theory makes the basic assumption that the important effects of nonlinear wave–wave interaction are introduced through the advective nonlinearity of the Eulerian equations. As developed to date and again here, I make the further assumption that these effects can be represented adequately, at least as an initial approximation, by treating the statistically described horizontal wind field produced by the ensemble of waves in a rather cavalier fashion: as if it were a background wind field, unchanging in time and in the horizontal, through which all individual waves must propagate. As they propagate, their horizontal wavenumbers \( k \) and \( l \) (and, hence, \( h \)) remain unchanged, as do their apparent frequencies \( \omega \). Their intrinsic frequencies \( \Omega \) suffer Doppler shifting as they move upward from one into another realization of the random, pseudobackground wind field, and by (2.1) their vertical wavenumbers \( m \) then suffer Doppler shifting as well. Statistically, which is all one can hope to determine for general application, the vertical wavenumbers are, in fact, Doppler spread. The objective of the present analysis is to determine the consequences of this spreading on the vertical-wavenumber spectrum of horizontal winds, and to compare these consequences with observation.

As in HII, I shall conduct the analysis in terms of the power spectral density for fractional potential temperature fluctuations, \( \Theta^2(k, l, m) \), at least initially. This is best thought of as a spectral density weighted by the atmospheric density, which permits me to omit from explicit appearance (for simplicity of notation) the standard exponential growth with height in inverse proportion to the decreasing atmospheric density. That growth would have to be reintroduced if ever the wind variance \( \sigma^2_T \) were being calculated from \( \Theta^2 \) explicitly. However, in the present theory the effects of the growth enter the spectral shape—which is all that is of concern in this paper—only via \( \sigma_T \), which is retained explicitly. Consequently, despite the apparent omission of growth, its effects are not at all lost from the analysis.

Consider first a small element of wavenumber space centered on a fixed pair of horizontal wavenumbers, \( \{ k, 0 \} \), and further defined by their apparent horizontal trace speed in the \( x \) direction, \( \omega/k \). (For simplicity of discussion, \( k \) is to be thought of as positive for the time being.) These parameters are unchanging with height, but the vertical wavenumber \( m \) to which they correspond varies with height in accordance with the height-varying pseudobackground wind and the correspond-
ing intrinsic frequency. At some chosen height \( z_1 \), the vertical wavenumber \( m_1 \) will depend on the \( x \) component of pseudobackground wind \( u_1 \), but that can be known only statistically. One must therefore make provision for all possible values of \( u_1 \) and perform a probability-weighted integration in order to obtain statistically observable results. Thus, the spectrum \( \Theta_1^\perp(k, 0, m_1) \), which will be abbreviated to \( \Theta_1^\perp(m_1) \), must be obtained as

\[
\Theta_1^\perp(m_1) = \int \psi_1^\perp(m_1, u_1) P_1(u_1) du_1, \quad (2.2)
\]

where \( \psi_1^\perp(m_1, u_1) \) is a new spectral form introduced for present purposes and

\[
P_1(u_1) du_1 = \frac{1}{\sqrt{\pi \sigma_1}} \exp\left(-\frac{u_1^2}{\sigma_1^2}\right) du_1 \quad (2.3)
\]

is the probability of finding the \( x \) component of wind in the small range \( du_1 \) about \( u_1 \); the integral is to be taken over all relevant \( u_1 \), in a manner to be described, and \( \sigma_1 \) is the value of \( \sigma_T \) at height \( z_1 \).

Standard Doppler shifting, from frequency \( \omega \) in the earth-based frame to \( \Omega = \omega - ku_1 \) in the local windborne frame, combines with the dispersion relation (2.1) to yield

\[
m_1^{-1} + u_1/N = \omega/kN = m_0^{-1}, \quad (2.4)
\]

where \( m_0 \), defined by the identity, is the \( m_1 \) that would be found at \( z_1 \) under the realization \( u_1 = 0 \). Figure 1 illustrates, in the \( \{m_1, u_1\} \) plane, the curve along which the portion of the spectrum at a given \( \omega/k \) (and hence at a given \( m_0 \)) will be distributed under variations of \( u_1 \) into other realizations. Curves in the negative \( m_1 \) half-space correspond to waves that must have passed through critical conditions (of vanishing \( \Omega \), infinite \( m \)); I shall take all such waves to have been obliterated in, or on the approach to, these conditions.

Figure 1 also illustrates the line along which the integration required in (2.2) must be effected: there must be a value for the spectral density \( \psi_1^\perp(m_1, u_1) \) available at each point along the line and, indeed, over the whole of the \( m_1 > 0 \) half-space above the curve \( u_1 = -Nm_1^{-1} \). This curve corresponds to the limiting case \( m_0 \to \infty \); points above it on the line of integration correspond to all lesser but still positive values of \( m_0 \), and so to all relevant waves. [Note: the corresponding lower limit was specified in HII following its (2.13) as \(-Nm_0^{-1}\), but should have appeared as \(-Nm_0^{-1}\).]

The \( \psi_1^\perp \) spectral density cannot be assigned arbitrarily over the \( \{m_1, u_1\} \) plane. Instead, continuity of the pressure perturbation and vertical displacement from one height to another—or as some would prefer, continuity of wave action flux or of momentum flux—requires of the spectrum that \( \psi_1^\perp dm_1 \) at \( m_0 \to \infty \), and \( \psi_1^\perp \) at any other value of \( u_1 \) (except if \( u_1 > Nm_0^{-1} \), when the wave will have been obliterated and \( \psi_1^\perp = 0 \)). This statement is true, at least, if partial reflections are ignored as in a WKB approximation. The derivative here has been found from (2.4) with \( u_1 \) held fixed at its other value. If I now rewrite \( \psi_1^\perp(m_0, 0) \) as \( \Theta_0^\perp(m_0) \), then

\[
\Theta_0^\perp(m_1) = \int_{-Nm_1^{-1}}^{\infty} \Theta_0^\perp(m_0) m_1^{-1} m_0 P_1(u_1) du_1
\]

\[
= \int_{-Nm_1^{-1}}^{\infty} \Theta_0^\perp(m_0) \frac{m_0}{m_1} e^{-x^2/\sigma_1^2} du_1, \quad (2.6)
\]

with the range of integration having been defined as by the vertical line segment in Fig. 1 along which the integration is to be performed. Along that line, the variable of integration can be changed to \( m_0 \) by use of (2.4) once again, now with fixed \( m_1 \).
\[ \Theta_i^2(m_1) = \int_0^\infty \Theta_i^2(m_0) \frac{N}{m_0 \sigma_1} \frac{1}{\sqrt{v \sigma_1}} \times e^{-N(m_0 - m_1)^2/v^2} dm_0. \]  

(2.7)

The succession of steps leading from (2.2) to (2.7) has resulted in the product of derivatives \( \frac{dm_0}{dm_1} \frac{dm_0}{dm_0} \frac{du_1}{dm_1} \) which is the same as \( \frac{du_1}{dm_1} \) obtained from (2.4) with \( m_0 \) held constant.

The form obtained in (2.7) is identical to that found in HII but with subscripts \( i \) replacing the former subscripts \( j \). The subscripted \( i \) referred to some hypothetical "initial" or "incident" wave spectrum \( \Theta_i^2(m_i) \) rising from possibly mythical lower elevations at which the wave system was taken to be perfectly linear, noninteracting. The subscripted \( 0 \) now refers to what is seen to be an equivalent spectrum, local to the height of interest, but one that would be found there only in the "zero wind" realization \( u_1 = 0 \). [In principle, this spectrum could be inferred from a statistically observed spectrum by inversion of the integral equation (2.7), but I have been unable to achieve the inversion.] For some conceptual purposes the new identification is to be preferred, since it does not invoke any hypothetical underlying region, and the new notation will therefore be retained.

The next task is to extrapolate \( \Theta_i^2(m_1) \) upward to overlying heights. This task involves a requirement to leave out of account, at greater heights, all waves that are obliterated via critical-layer processes at height \( z_1 \). This requirement must be imposed, step by step, for each successive step that is taken upward in height. Its implementation successively reduces the intensity of the propagating spectrum, more so at large \( m_0 \) than at small \( m_0 \) because the large-\( m_0 \) waves are more susceptible to Doppler shifting into critical conditions. In HII, I described only the general, qualitative consequence of the obliteration process: that of imposing, in effect, a sharp cutoff (approximated in HII by a step function) of the large-\( m_0 \) waves near a characteristic vertical wavenumber \( N/\sigma_T \). Here, I develop the argument more thoroughly and derive an analytic form for the cutoff function itself.

3. The cutoff function

Consider, then, the spectrum to be found at some greater height \( z_2 \). It will be found, as at height \( z_1 \), from the form

\[ \Theta_i^2(m_2) = \int_{-N^{-1/2}}^{N^{-1/2}} \psi_i^2(m_2, u_2) P_2(u_2) du_2. \]  

(3.1)

with

\[ m_2^{-1} + u_2/N = m_0^{-1} + u_1/N \]  

(3.2)

giving the relationship between points in the \( m_2, u_2 \) plane, each of which defines a corresponding \( m_0 \), and the corresponding curves in the \( m_1, u_1 \) plane from which waves having that \( m_0 \) must have come. The mapping of \( \psi^2 \) occurs, again with conservation of wave action (in the absence of dissipation), via

\[ \psi_i^2(m_2, u_2) = \int_{-N^{-1/2}}^{N^{-1/2}} \psi_i^2(m_1, u_1) m_2^{-1} m_1 P_2(u_1) du_1 \]

\[ = \int_{-N^{-1/2}}^{N^{-1/2}} \Theta_i^2(m_0) m_2^{-1} m_0 P_2(u_1) du_1, \]  

(3.3)

where \( P_2(u_1) \) must represent a joint probability, that of finding \( u_1 \) at height \( z_1 \), given that a specified \( u_2 \) is under consideration at height \( z_2 \). This probability can be calculated in principle, given the wave spectrum, and the integration can be performed in principle, but I prefer to avoid the complexities of joint probabilities. This can be done, approximately, by means of a simple expedient: I take the vertical separation of \( z_2 \) from \( z_1 \) to equal or exceed the "decorrelation distance" \( D \), whereupon \( P_2(u_1) \) becomes independent of \( u_2 \) and simply equal to \( P_1(u_1) \). (The decorrelation distance is not in itself a uniquely defined quantity; it may be different for different purposes. In effect, it is the distance that \( z_2 \) must be separated from \( z_1 \) in order to make the approximate result match closely the true result. It will be discussed again later.) With that done, insertion of (3.3) in (3.1) gives

\[ \Theta_i^2(m_2) = \int_{-N^{-1/2}}^{N^{-1/2}} \int_{-N^{-1/2}}^{N^{-1/2}} P_1(u_1) du_1 \Theta_i^2(m_0) \]

\[ \times m_2^{-1} m_0 P_2(u_2) du_2. \]  

(3.4)

The upper limit on the integration over \( u_1 \), both in (3.3) and in (3.4), is designed to ensure that the portion of the spectrum that was obliterated at height \( z_1 \) is not resurrected and allowed to proceed on to greater heights. This is a crucial point, for it gives rise to the all-important cutoff function in the Doppler-spread theory.

Since \( u_1 \) now appears only in the known (Gaussian) probability factor \( P_1(u_1) \), the integration over \( u_1 \) can be carried out explicitly to give

\[ \Theta_i^2(m_2) = \int_{-N^{-1/2}}^{N^{-1/2}} \left\{ 0.5 + 0.5 \text{erf}\left[ N/m_0 \sigma_1 \right] \right\} \]

\[ \times \Theta_i^2(m_0) m_2^{-1} m_0 P_2(u_2) du_2, \]  

(3.5)

where \( \text{erf}(x) \) is the error function,

\[ 2\pi^{-1/2} \int_0^\infty \exp(-q^2) dq. \]

Had obliteration not been taken into account, the factor within braces in (3.5) would not have appeared but all else would have remained unchanged. As things are, this factor—which is \( 1 \) at \( m_0 = 0 \) and \( 0.5 \) at \( m_0 = \infty \)—reduces the contribution of large-\( m_0 \) waves to the overlying spectrum, as previously predicted. It will lead in due course to the cutoff function.

The necessity for taking all wave obliteration into account requires that \( z_2 \) not be separated from \( z_1 \) by more than \( D \); otherwise, some further, unrecognized
and unaccounted obliteration would take place in between. Indeed, since \( \sigma \) in practice tends to grow with height, it could be argued that the \( \sigma_1 \) in (3.5) should be increased toward \( \sigma_2 \) to account adequately for obliteration between the two levels. This step will be taken shortly, in effect, by the introduction of height integrations.

Were we to continue to a third height \( z_3 \), a distance \( D \) above \( z_2 \), a form similar to (3.5) would be found, with subscript 3 replacing subscript 2, but it would include a further cutoff factor given by \( \{ 0.5 + 0.5 \times \text{erf}(N/m_0\sigma_2) \} \). Successive slabs of thickness \( D \) will introduce successive factors of the same type.

The multiplying together of successive cutoff factors is cumbersome. One would prefer to have a continuously varying function with which to express the statistically continuously occurring effects of critical-level obliteration. In the case of \( \sigma_2 \) and \( D \) constant with height, this could be achieved by raising the single-layer cutoff factor to an appropriate power:

\[
\{ 0.5 + 0.5 \text{erf}(N/m_0\sigma_2) \}^{D^{-1}} \ln \{ \sigma \} \exp, \quad (3.6)
\]

which is obviously the same thing as the multiple product in all cases where the range of integration of \( z \) is an integer number of times \( D \); that is, where \( n \) is an integer. It must be approximately, if not exactly, valid also within layers: \( n \) may be taken to be continuous, with \( z_a \) then being the height of current interest. More generally, with height-varying \( \sigma \) and \( D \), the multiple cutoff factor can be given by the functional form

\[
\chi_1(m_0, z_n) = \exp \int_{z_1}^{z_n} \ln \{ 0.5 + 0.5 \text{erf}(N/m_0\sigma_2) \} D^{-1} d\sigma, \quad (3.7)
\]

wherein the \( \ln \) is necessarily negative. The multiplication of many cutoff factors has, thus, been accomplished via the addition—and hence integration—of their logarithms. The subscripted 1 on \( \chi_1 \) is a reminder of the lower limit of integration, \( z_1 \), which is soon to be changed.

In Hill, the Doppler-spread theory in conjunction with a representative model spectrum was shown to produce a \( \sigma_2 \) that increases exponentially with height, with scale height \( 4H \), so long as the large-\( m \) portion of the spectrum is truly saturated. In that case, \( \chi_1 \) can be rewritten as

\[
\chi_1(m_0, z_n) = \exp \int_{z_1}^{z_n} \ln \{ 0.5 + 0.5 \text{erf}(N/m_0\sigma_2) \} \times 4HD^{-1}\sigma^{-1} d\sigma. \quad (3.8)
\]

[The theory also permits “pseudosaturation,” described by Hines (1993), in which a different behavior of \( \sigma_2 \) would occur at this point.] The Doppler-spread theory also introduces a characteristic vertical wavelength \( \lambda_C \)

\[
= 2\pi/m_C \approx 4\pi\sigma_2/N. \quad \text{If the decorrelation distance} \quad D
\]

is taken to be some constant multiple \( \alpha \) of this wavelength, as it will be shortly, the function becomes

\[
\chi_1(m_0, z_n) = \exp \left[ \frac{HN}{4\pi} \int_{z_1}^{z_n} \ln \{ 0.5 + 0.5 \text{erf}(N/m_0\sigma_2) \} \sigma^{-2} d\sigma \right]
\]

\[
= \exp \left[ 4HDn^{-1}M_0 \right]
\]

\[
\times \int_{M_0^{-1}}^{M_0^{-1}} \ln \{ 0.5 + 0.5 \text{erf}(M_0^{-1}) \} dM_0. \quad (3.9)
\]

where \( M_0 = m_0\sigma_2/N \) is a nondimensionalized form of the vertical wavenumber \( m_0 \), \( M_0 \) being \( m_0\sigma_2/N \) and \( D_n = 4\pi\alpha\sigma_2/N \), the decorrelation distance at height \( z_n \). (The transition here from \( \sigma_2 \) to \( M_0^{-1} \) as the variable of integration constitutes simply a mathematical change having no physical content.) The corresponding spectrum at \( z_n \) is then given by

\[
\Theta_n^2(m_n) = \int_{-\infty}^{\infty} \chi_1(m_0, z_n)\Theta_0^2(m_0)
\]

\[
\times m_n^{-1} m_0 P_n(u_n)du_n. \quad (3.10)
\]

In principle, application of the cutoff theory to observations would require the observational determination of the statistical spectrum \( \Theta_0^2(m_1) \) at some low height \( z_1 \), its insertion in (2.7), inversion of that equation to determine \( \Theta_0^2(m_0) \), and the insertion of this spectrum in (3.10) to provide a theoretical prediction of the statistical spectrum at the overlying height \( z_n \). This sequence is unlikely to be followed in practice. Instead, since the Doppler-spread theory purports to yield a standardized spectral form at all heights, matching a quasi-universal form observed, one may conjecture that the input spectrum \( \Theta_0^2(m_0) \) at height \( z_1 \) would itself take the form of some smoothly and slowly varying (with \( m \)) model spectrum \( \Theta_0^2(m_0) \) modified by the cutoff function \( \chi \) that would have been applicable at \( z_1 \) had the lower limit of integration in (3.7) been taken at a height (possibly mythical) where \( \sigma_2 = 0 \). In that case, the net effect would be to introduce the model spectrum in place of \( \Theta_0^2(m_0) \) in (3.10) and simultaneously extend the range of \( \sigma_2 \) integration in (3.8) down to 0, or the range of \( M_0^{-1} \) integration in (3.9) up to \( \infty \). The statistically observable spectrum at height \( z_n \) then becomes

\[
\Theta_n^2(m_n) = \int_{-\infty}^{\infty} \chi(m_0, z_n)\Theta_0^2(m_0)
\]

\[
\times m_n^{-1} m_0 P_n(u_n)du_n, \quad (3.11)
\]
where

\[ x(m_0, z_n) = \exp \left[ 4HD_n^{-1} M_{0n} \times \int_{M_{0l}}^{\infty} \ln(0.5 + 0.5 \text{erf} M_{0l}^{-1}) dM_{0l}^{-1} \right] = \left\{ \exp \left[ M_{0n} \times \int_{M_{0l}}^{\infty} \ln(0.5 + 0.5 \text{erf} M_{0l}^{-1}) dM_{0l}^{-1} \right] \right\}^{4HD_n^{-1}}. \]

(3.12)

in which \( x_\infty \), defined by the identity as the function within braces in the middle line, has the advantage of being a function of \( m_0 \) and \( z_n \) only via the local normalized wavenumber \( M_{0n} \). It may be noted that (3.11) is equivalent to what would have been obtained from the use of a mythical underlying region extending down to \( z = -\infty \), where the wave system would have been free from nonlinearity, had there been a spectrum given by \( \Theta^2_{\infty}(m_0) \) there. But there is certainly no requirement for such a region to exist in order that the formulation be valid. Comparison of (3.11) with (2.6) will show that we could again assign to \( \Theta^2_{\infty}(m_1) \) any form we chose—presumably a form that matched observations—whereupon (3.11) would define a corresponding \( \Theta^2_{\infty}(m_0) \) purely mathematically without any need to conjure up an underlying region. What the form (3.11) does, however, is suggest that the \( \Theta^2_{\infty}(m_0) \) spectrum at the lowest height of observation might be expected to consist of a slowly varying function of \( m_0 \), given by \( \Theta^2_{\infty}(m_0) \), multiplied by a possibly rapidly varying cutoff function \( x_\infty(m_0, z_1) \) that severely limits its extension to large \( m_0 \). In point of fact, we are free to choose any form of cutoff function at height \( z_1 \) that will produce a match to the observations there; the proposed choice of \( x_\infty(m_0, z_1) \) as the cutoff function at height \( z_1 \), if it is made, imposes on the Doppler-spread theory a constraint that is not intrinsic to it. Nevertheless, I consider it appropriate that this choice be made and its consequences be examined, simply because it will tend to produce a universality of form such as the relevant observations appear to favor. In practice, once we have ascended to a height \( z_n \) above \( z_1 \), such that \( \sigma_n \gg \sigma_1 \), the actual choice of initial cutoff function at \( z_1 \) will be of no further consequence: the cutoff will have moved to much smaller values of \( m_0 \) and will have removed from relevance the region of \( m_0 \) in which that initial cutoff occurred.

The subscripted \( \infty \) on \( x_\infty \) is intended to indicate that, in the derivation of \( x_\infty \), obliteration was imposed only when \( m \) was Doppler shifted to \( \infty \) in the fashion of a true critical layer. In fact, obliteration is likely to occur in the course of an approach toward a critical layer. The effect is often taken, and was taken in HII, to be representable by obliteration if \( m \) is Doppler shifted to some maximum permissible value, \( m_M \). In HII, this maximum was taken (for the middle atmosphere) to be imposed by the onset of instability in the wave-induced wind system, such as will occur if the large-\( m \) tail of the wave spectrum extends to sufficiently large \( m \), namely, \( m_M \). (It may be, of course, that significant dissipation occurs at \( m < m_M \), as was already acknowledged in HII. My purpose in ignoring it here is to determine how far the Doppler-spread theory can go "on its own" at \( m < m_M \), and how much, if any, is left over, requiring further analysis and, perhaps, further physics.)

Given this more realistic behavior, one can repeat the foregoing analysis with appropriately altered limits of integration at various points along the way. [The lower limit of integration in (2.6) becomes \( N(m_0^l - m_1^l) \), the upper limit in (2.7) becomes \( m_M \), the upper limit in (3.3) becomes \( N(m_0^l - m_1^l) \), and likewise in similar later forms. The limits in (3.7)–(3.9) remain unchanged.] The end result is to replace (3.11) and (3.12) by

\[ \Theta^2_{\infty}(m_0) = \int_{N(m_0^l - m_1^l)}^{\infty} (m_0, z_n) \Theta^2_{\infty}(m_0) m_0^{-1} m_0 P_n(u_n) du_n \]

\[ = \int_{N(m_0^l - m_1^l)}^{\infty} (m_0, z_n) \Theta^2_{\infty}(m_0) \times \frac{N}{m_0 m_n \sqrt{\pi} \sigma_n} e^{-N(m_0^l - m_1^l)^2/\sigma_n^2} dm_0 \]

(3.13)

and

\[ \Theta^2_{\infty}(m_0) = \{ \Theta^2_{M_M}(m_0) \}^{4HD_n^{-1}}, \]

(3.14)

respectively, where

\[ \Theta^2_{M_M} = \exp \left[ M_{0n} \int_{M_{0l}}^{\infty} \ln \left\{ 0.5 \right. \right. \]

\[ + 0.5 \text{erf}(M_{0l}^{-1} - M_{0l}^{-1}) \]}

\[ dm_0 \right\}^{4HD_n^{-1}}, \]

(3.15)

and \( \Theta^2_{M_M} = \Theta^2_{\infty} \) as previously defined when \( M_M = \infty \). These functions will be illustrated shortly.

The argument of the error function in (3.15), namely \( M_{0l}^{-1} - M_{0l}^{-1} = N/m_0 \sigma_T - N/M_M \sigma_T \), is a generalization of the corresponding argument in (3.9), where the integration is initially to be taken over \( \sigma_T \) and only subsequently over \( M_{0l}^{-1} \). On the assumption that \( m_M \) is determined by instability of the wave system, it was found in HII that this \( m_M \) would be inversely proportional to \( \sigma_T \) and so \( M_M \) would be a constant. (This constant was taken to be 11.5 in one example and 2.4 in another, the choice being dependent on the critical wind-shear variance—\( N^2 \) or \( 0.5N^2 \), respec-
tively—that was thought to be required for the maintenance of marginal instability.) Accordingly, when the integration is transformed to an integration over $M_0^{-1}$, as in (3.15), $M_M$ is to be treated as a constant.

The foregoing results apply initially only to the horizontal wavenumber pair \( \{k, 0\} \) for which they were derived. With azimuthal isotropy having been assumed, they apply equally to the general pair \( \{k, l\} \) and then, after integration over all \( \{k, l\} \), to the one-dimensional \( m_n \) projections of the full three-dimensional spectra, since the participating functions other than $\Theta_\theta^0$ are independent of \( k \) and \( l \).

With \( z_n \) replaced by the general \( z \), subscripted \( n \)'s may be dropped from the notation entirely, and \( \sigma_n \) may be replaced by its original form, \( \sigma_T \).

4. The decorrelation distance \( D \)

The problem of defining the decorrelation distance \( D \), both in principle and in practice, remains. As in HII, I shall take as a representative model a “modified Desaubies” (MD) spectrum in which the horizontal wind-power spectral density is proportional to \( m / [1 + (m / m_c)^{p}] \). This spectrum, introduced by VanZandt and Fritts (1989) as being closely representative of middle atmosphere spectra, increases almost linearly with \( m \) until a rounded peak is reached at \( 3^{-1/4} m_c \). Thereafter it decreases with further increase of \( m \) and very rapidly adopts the asymptotic form \( m^{-3} \) that is commonly taken to be characteristic of saturation. (This spectrum is illustrated in the vicinity of its peak by the dotted-line curve in Fig. 8.)

The autocorrelation function of the winds produced by such a spectrum is shown in Fig. 2. It exhibits a decrease to and through zero as the lag distance is increased from zero and then a gradual increase asymptoting to zero as the lag is increased indefinitely. Similar spectra will produce similar behavior.

As the decorrelation distance \( D \) I shall take the lag distance that is required to produce the first zero of the autocorrelation function. In the illustrated case, this is approximately \( \lambda_c / 4 \), where \( \lambda_c \) is the vertical wavelength corresponding to \( m_c \), namely, \( 2\pi / m_c \). (A pure sine wave of wavenumber \( m_c \) has the first zero of its autocorrelation function at \( \lambda_c / 4 \) exactly, but thereafter the autocorrelation function oscillates between \( +1 \) and \( -1 \).) This behavior is likely to be typical of spectra similar to the chosen spectrum, so I adopt

\[
D = \lambda_c / 4
\]  

(4.1)

(hence, \( \alpha = \frac{1}{4} \)) for use in relations such as (3.9). The choice \( m_c = N / 2\sigma_T \), already made for use in the derivation of (3.9), has been adopted from the analysis in HII and will be subject to testing shortly.

5. Illustrative cutoff functions

The basic cutoff function \( x_\infty \) is illustrated in Fig. 3 by the rightmost curve, drawn with a continuous line. Accompanying it, as broken-line and dotted-line curves, are the cutoff functions \( x_{11.5} \) and \( x_{2.4} \), corresponding to \( M_M = 11.5 \) and 2.4, respectively. (In the case \( M_M = 2.4 \), there would be a further cutoff to be effected at \( M_0 = 2.4 \) (not shown), dropping the curve to zero there; likewise for \( M_M = 11.5 \).)

The two additional trios of curves (continuous, broken, and dotted) lying successively farther to the left
of these \( x \) curves represent the three cutoff functions raised to the power \( p = 16 \) and 112, respectively. These values may be considered to be representative of the mesosphere and the stratosphere in turn, being derivable from (for example) \( H = 6 \) km and \( \lambda_C = 6 \) km for the mesosphere, \( H = 7 \) km and \( \lambda_C = 1 \) km for the stratosphere. (The fact that the relevant power is greater for the stratosphere than for the mesosphere may seem counterintuitive, since the mesospheric waves will have been subject to obliteration over a greater height range. It results, however, from the decorrelation distance being larger, in the mesosphere, relative to the scale height of rms wind growth: there are fewer slabs of atmosphere having statistically independent winds underlying a mesospheric height of observation, and yet having rms winds comparable to those at the mesospheric height, than there are underlying a stratospheric height of observation, and yet having rms winds comparable to those at the stratospheric height.)

The qualitative effect of these cutoff functions warrants discussion. As may be seen from (3.13), and as was noted already in HII, the spectral density \( \Theta^2(m) \) asymptotes to the form \( m^{-1} \) as \( m \to \infty \), not to the form \( m^{-3} \) that observations seem to require. Nevertheless, Doppler spreading was shown in HII to produce something close to an \( m^{-3} \) form for about an order-of-magnitude range of \( m \) on the large-\( m \) (tail) side of the spectral peak, which is about all that was called for observationally, provided that the cutoff function could be approximated by a step function situated at \( M_0 = 0.5 \).

A semiquantitative discussion of the nature of the cutoff led to the conclusion that the real cutoff function might just about meet this requirement, though perhaps the approximating step function might better be situated at a somewhat larger value of \( M_0 \), such as \( M_0 = 1 \). In that case, the Doppler-spread tail spectrum would not be quite as steep as \( m^{-3} \) even locally, and additional processes, such as dissipation within the tail, would have to be invoked to steepen it.

The \( p = 112 \) and \( p = 16 \) curves in Fig. 3 now give confirmation that a step-function cutoff is probably not a bad approximation, but at the same time they indicate that the siting of the step might properly be at some \( M_0 \) greater than 0.5, even though the value 0.5 is not a bad approximation for the stratosphere. If the step were situated at the \( M_0 \) where the true cutoff function has the value 0.5, for example, then \( M_C \), would lie between 0.59 and 0.75 in the stratosphere and between 0.83 and 1.14 in the mesosphere, for the range of \( M_M \) discussed and illustrated here (2.4 \( \leq M \leq \infty \)). The fact that these figures are generally larger than 0.5 tends to imply that the corresponding tail spectra will not be as steep as \( m^{-3} \), with consequences that will be discussed in section 7. Before that discussion begins, though, the implications of the cutoff curves in application to model spectra, producing theoretical "observable" spectra, will be illustrated in section 6.

The effects of the earth's rotation are discussed in the Appendix but are illustrated briefly here in Fig. 3. A dot--dash curve in the \( p = 112 \) group shows the modification imposed on the neighboring dotted curve by inclusion of the Coriolis force in application to a horizontal wavelength of about 60 km. A portion of a double dot--dash curve at the extreme left does the same thing for a horizontal wavelength of 600 km.

Finally, the basic cutoff function \( x_\infty \) is illustrated once again in Fig. 4, now with a logarithmic vertical scale, for purposes of a later discussion.

6. Illustrative "observable" spectra

One of the characteristics demanded of any proposed mechanism of saturation is that it shall produce a quasi-universal form for the tail spectrum: no matter what the initially input spectrum of waves may be, or what its variety of sources may be, or how far above those sources the observations may be made, the mechanism must be capable of producing a tail spectrum that is virtually unchanging in form and intensity. The Doppler-spread theory as developed in HII sought to meet this requirement with the explanation that initially incident waves having \( M_i > M_C \) (read \( M_0 > M_C \), now) are essentially all obliterated below the height of observation, and the waves observed in the tail (having \( M \geq M_C \)) are positioned there simply by Doppler spreading from the initially incident waves that have \( M_i \) (or now \( M_0 \)) just less than \( M_C \). This spreading was of a universal form, given the Gaussian distribution of horizontal winds; hence universality of form was assured. (The universal form itself was uncertain, however, being dependent on the "correct" choice for \( M_C \), as just discussed.) The validity of that argument can now be put to a test by use of the true cutoff functions that replace the former step-function cutoff.

The effects of the \( x_\infty \) cutoff functions for the me-

![Fig. 4. The curve explicitly labeled $\infty$ in Fig. 3, replotted on a logarithmic scale.](image-url)
sosphere \((4H/D = 16)\) and stratosphere \((4H/D = 112)\) are shown in Figs. 5a and 5b, respectively. Each displays the “observable” spectrum produced, via convolution as in (3.11), by each of two distinct model input spectra. The input spectra are shown by dotted lines, given by \(\Theta_3^S(m_0) \propto m_0 \) and \(\Theta_2^S(m_0) = \text{constant} \), and the corresponding “observable” spectra are shown by a continuous line and a broken line, respectively. The curves have been positioned vertically so as to provide equal values of \(\sigma_f\) for the two “observable” spectra, and the model spectra adjust themselves to match the “observable” spectra at small \(M\). As will be seen, the “observable” tail spectra are virtually unchanged in form and intensity as they pass from one input spectrum to the other, despite the great and ever-growing disparity between the two inputs at these wavenumbers. These tail spectra are seen to retain virtually no information as to the content of the initially input spectra at the same wavenumbers. They derive instead, much as was argued in HIL, almost exclusively from the Doppler spreading of input waves in the range \(0.5 < M_0 < 2\), and the input spectra could have been cut off by whim at \(M_0 = 2\), for example, without producing any significant change in the “observable” spectra at any \(M\) whatsoever.

This conclusion may be reconfirmed and even strengthened with \(M_M\) taken to be finite rather than infinite and, furthermore, with Coriolis effects taken into account: the range of \(M_0\) that contributes to the tail narrows and moves to somewhat smaller \(M_0\) values as these effects are added. It reveals the tremendous power of the cutoff function in obliterating almost all memory of the initially input, or zero-wind, model spectrum at wavenumbers now lying well into the tail.

That power is revealed further by the cutoff curve \(\chi_\infty\) in Fig. 4. If the cutoff function is thought of as being locally of the form \(M_0^{p1}\), it will be seen (having regard to the difference of vertical and horizontal scales) that \(p = 1\) at about \(M_0 = 3\) and \(p\) increases rapidly at greater \(M_0\). But this value of \(p\) must be multiplied by 16 for the mesosphere and by 112 for the stratosphere, yielding the forms \(M_0^{16}\) and \(M_0^{112}\), respectively, at \(M_0 = 3\). These forms must surely be sufficiently steep to obliterate any detectable memory of any natural input spectrum at \(M_0\) as great as or greater than 3.

The value of \(M\) at the peak of the spectrum, \(M_p\), is indicated for each spectrum in Fig. 5 et seq. (The peaks in the case of the \(\Theta_3\) = const spectra are shallow but real. They result from the downward-shift portion of the Doppler-spread waves in the range \(0.5 < M_0 < 2\).) For comparison, the MD spectrum, which corresponds most closely to the continuous-line curves of Figs. 5a and 5b, would have had \(M_p = 0.38\) if \(M_C\) were chosen arbitrarily to be 0.5, as in HIL.

The steepest log–log slopes in the tail spectra of Fig. 5, also indicated, are found to range from \(-1.5\) to \(-2.2\), in contrast to the value \(-3\) that is conventionally sought. Even worse, they ultimately tend at very large \(M\) to the value \(-1\), as previously noted. Readers are invited to defer judgment on this apparently departure from observation until it is discussed in section 7.

Because of the near invariance of the tail spectra under changes of model spectrum, further illustrations need be presented only for one model spectrum. The choice \(\Theta_3^S(m_0) \propto m_0\) is adopted here—this being the one that produces “observable” spectra corresponding at small \(m\) to the MD spectral form \(m/[1 + (m/m_C)^4]\).

Observationally, the form of the tail spectrum is said to remain unchanging with height. Figure 6 provides a comparison between the “observable” \(\Theta^3(m)\) spectra calculated for \(4H/D = 16\) (dotted line) and 112 (continuous line), representative of the mesosphere and
stratosphere, respectively. The two curves have been positioned vertically so as to produce identical values of \( \sigma_T \). The two tail spectra are seen to be similar in form, though the mesospheric tail does not fall as steeply as the stratospheric tail. Observationally once again, the absolute (as distinct from the density weighted) spectra are said to be unchanging in intensity with height at all \( m \) lying in the tail. To examine the corresponding theoretical conclusion we first note that, with the curves having been drawn for identical \( \sigma_T \), the horizontal scale may be read as an absolute wavenumber scale, the same for both curves. To account for the actual increase of \( \sigma_T \) with height, the dotted (mesospheric) curve should be displaced upward and, in order to keep the horizontal scale as a common measure of absolute \( m \), to the left. The mathematical scaling is such as to produce a shift of one decade to the left for every three decades upward, a scaling that permits the usual \( m^{-3} \) tail spectrum to slide along itself, remaining unchanged in intensity at a given \( m \). Here, however, because of the way the two tail slopes differ from \(-3\), the shifting would result in an increase of tail intensity in the mesosphere relative to that in the stratosphere. As before, readers are invited to defer judgment of this apparent discrepancy until it is discussed in section 7.

The remaining illustrative spectra are drawn for the stratospheric case \( 4H/D = 112 \) only, but similar spectra result for the mesospheric case. Figure 7 exhibits the effects of wave obliteration at finite \( M = 11.5 \) (broken line) and 2.4 (dotted line), in comparison with, and in contrast to, the case \( M_M = \infty \) (continuous line) that has been illustrated up to this point. Not unexpectedly, the obliteration of waves before they actually reach a critical layer is seen to have an effect similar to that of increasing \( 4H/D \)—an effect already witnessed in Fig. 3—in addition to imposing a sharp cutoff (indicated by the vertical portion of each curve) at \( M_M \). The steepest log-log slope illustrated is now \(-2.7\), which closely approaches the standard \(-3\) value, in contrast to \(-2.2\) for the case \( M_M = \infty \). This steep slope is available over slightly less than a decade in \( M \), however, a point that will call for discussion in section 7.

The value \( M_M = 11.5 \) was chosen for illustration on the basis of its appearance in HII, where it was found to be the value required to produce a wind-shear variance just equal to \( N^2 \), which is a representative criterion for marginal instability. It had that characteristic, however, only for the specific model spectrum that was employed in the calculation of the variance: the MD spectrum, with characteristic wavenumber \( M_C = 0.5 \). Since the present spectra are not quite of the MD form, the obliteration at \( M_M = 11.5 \) now implies a somewhat different wind-shear variance. Computation reveals this variance to be \( 4.1N^2 \). Evidently a somewhat smaller value of \( M_M \) would be needed if the variance were required to be \( N^2 \) exactly.

The value \( M_M = 2.4 \) was chosen to produce a shear variance of \( 0.5N^2 \) in HII, the variance chosen by Smith et al. (1987) as the critical value for marginal instability. Computation reveals that it produces \( 0.48N^2 \) here, essentially the value sought. This case provides a further, quite remarkable, comparison. Its spectrum—the dotted line in Fig. 7—is reproduced in Fig. 8, now as a continuous line (and with its natural extrapolation beyond \( M_M \) shown by a broken line). Superimposed on that spectrum is another, given by the dotted line, which is the MD spectrum \( M/M_C^2 \) beyond \( M_M \) shown by a broken line).
mosphere data, with the choice $M_C = 0.5$ being taken from HII, where the same form of spectrum was adopted for model purposes. The comparison of these two spectra will be taken up in section 7, but their close correspondence in the region of the peak, and even over almost an order of magnitude in $M$ to the right of the peak, may be noted already.

The effects of the earth’s rotation have been virtually ignored up to this point. As was noted in appendix A of HII, introduction of the Coriolis force results in a steepening of the slope in the tail region of the $\Theta^2(m)$ spectrum. That conclusion was to some extent misleading, however, for it properly applies in the first instance only to the $\Theta^2(m)$ spectrum, whereas it is the spectrum $H^2(m)$ of the total horizontal wind that is normally reported observationally. The two have identical forms in the absence of the Coriolis force, and the preceding graphs of $\Theta^2(m)$ can equally well be read as graphs of $H^2(m)$; but with the Coriolis force included, the $H^2(m)$ spectrum actually asymptotes to a constant value as $m \to \infty$. The relevant analysis is presented in the Appendix. Its consequences are illustrated in Fig. 9, where the Coriolis force is first ignored (continuous line), then included for horizontal wavelengths of 6 km (broken line), 60 km (dotted line), and 600 km (dot–dash line) with $M_M = \infty$ for all. Figure 10 illustrates the corresponding curves when $M_M = 11.5$ for the first pair and $M_M = 2.4$ for the second pair. These combinations have no particular significance except qualitatively; they are shown primarily in further illustration of the mathematical results.

It is clear that the addition of the Coriolis force has done little if anything to “improve” the spectral shape in the first decade of $m$ above the spectral peak, and goes totally awry if attention is extended to still larger values of $m$. The latter result, if taken to be observationally significant, adds further weight to the need for amendments such as those now to be discussed.

7. Discussion

This analysis has shown that a cutoff function similar to that invoked in HII should indeed be operating in the atmosphere, and that its application to a wide range of “initially incident” or “zero wind” spectra would lead to an almost universal form of Doppler-spread...
spectral tail not unlike the form usually attributed to the observed spectral tail. The chief defect of the computed results seems to be that the tail does not drop sharply enough with increase of vertical wavenumber: if the form of the tail is taken to be $M^{-\rho}$, then $\rho$ typically lies in the range 1.5–2.7 over the ranges of $M$ illustrated, rather than equalling the reported observational value $\rho = 3$. The computed spectra go further awry at large $M$ with Coriolis effects included, as noted in the preceding paragraph.

One possible explanation of the defect is that Doppler spreading really has no role to play in the formation of the tail: that the analysis fails to incorporate the appropriate basic physical process and therefore should not be expected to match observations. Because of the calculated close approach to the observations near and just above the spectral peak, however, and because of the physical necessity for waves to accommodate in some fashion when they find themselves in horizontal winds comparable to and exceeding their own horizontal travel speeds, I consider it to be virtually established that Doppler spreading—or, more properly, the physical process of wave–wave interaction via the Eulerian advective nonlinearity—is operative in initiating the tail, and one should look to modifications of, or superimposed on, that process when seeking an explanation of the defect of the analysis at higher wavenumbers.

A very likely explanation of the defect lies in the manner by which some of the underlying physics were approximated, specifically 1) in my “nonlinear” treatment of the wave-induced winds as if they were background winds, strictly horizontal, and unchanging in time and in horizontal planes; perhaps 2) in my adoption of an action-conserving, WKB type of approximation for the vertical propagation of waves through that wind system; and perhaps 3) in my imposition of dissipation only suddenly, in step-function fashion, at $M_M$. Indeed, these “approximations” are so far from the truth that one might have been inclined to dismiss them out of hand, without analysis. The degree to which they seem to have been successful here then becomes the most remarkable feature of the present analysis, and one is tempted to try to understand the nature of the failure in the hope that it might be remedied by some better approximation.

I consider 1) to be the most fundamental defective approximation of the analysis to date. I base this belief on the results of Allen and Joseph (1989), who were able to obtain the $m^{-3}$ asymptotic form for the tail spectrum when this approximation was avoided (but who did not illustrate the transitional portion of the spectrum that has been illustrated here or identify in any analytic form the location of that transition). Relative to this defective approximation, one might expect the general tendency of true wave–wave interactions to be such as to bias the actual spectrum from any initially incident spectrum toward higher frequencies and horizontal wavenumbers. Such a bias would tend to minimize the Coriolis effects that appear, at least in Fig. 9 if not in Fig. 10, to be unacceptable.

Quite apart from the temporal and horizontal variations of the random, pseudobackground, wave-induced winds, there remains the fact that those winds in actuality include a vertical component $w$. The Doppler-shift relation $\Omega = \omega - ku$ that led to (2.4) should really be written $\Omega = \omega - ku + mw$ (with the present sign convention), and so (2.4) should in principle be replaced by

$$m_0^{-1} = \frac{\omega}{kN} = \frac{u_1}{N} + m_1^{-1} - w_1m_1/Nk,$$  \hspace{1cm} (7.1)

whose relevant solution is

$$m_1^{-1} = (m_0^{-1} - u_1/N)\{0.5 + 0.5[1 + 4w_1/Nk(m_0^{-1} - u_1/N)^{1/2}]\}^{1/2}.$$  \hspace{1cm} (7.2)

Evidently the form (2.4) is at best an approximation that deteriorates at Doppler-shifted vertical wavenumbers near and exceeding $(Nk/w)^{1/2}$ in individual realizations of the wind field and whose applicability to the statistical state is at least questionable at Doppler-shifted wavenumbers exceeding $(Nk/\sigma_w)^{1/2}$, where $\sigma_w$ is the rms vertical wind. The frequency spectrum of middle-atmosphere irregular winds is often taken to be of the form $\Omega^{-5/3}$ between the inertial frequency $f$ and the buoyancy frequency $N$, and for such a spectrum it is easily shown that $\sigma_w \approx 2^{-1/2}(f/N)^{1/3} \sigma_T \approx 0.1 \sigma_T$ in typical midlatitude conditions, $f$ being the inertial frequency or Coriolis parameter. Thus, the foregoing analysis is already questionable validity at vertical wavenumbers as great as the characteristic value $m_c = N/2\sigma_T$ unless $k$ considerably exceeds $m_c/20$. Typically, this would require that the horizontal wavelengths of waves in the saturated tail should be considerably less than 20 km in the stratosphere, 120 km in the mesosphere, or equivalently, that the intrinsic periods in the tail spectrum should be considerably less than 2 h. Whether or not this is the case in practice is simply not yet known. But unless it is the case, apparent successes such as that illustrated by Fig. 8 would seem to be largely accidental.

The condition for validity at and above $m_c$ is certainly not met by the more extreme of the modeled results in Figs. 9 and 10, including in particular the horizontal asymptotic forms of the spectra, which can therefore be dismissed out of hand as the unfortunate consequences of a defective analysis. The $m^{-1}$ asymptotic form obtained in the absence of Coriolis effects may likewise be dismissed out of hand, as lying at $m$ values for which the original Doppler-shift formula is an inadequate approximation. In that case, one is forced back to the analysis of Allen and Joseph (1989), which incidentally does include Coriolis effects, for an
indication of what the true asymptotic form should be, and that form is \( m^{-3} \) as already noted.

With respect to the defective approximations (2) and (3), it can be said that any additional mechanism that diminishes the spectral densities increasingly with increasing Doppler shift will lead to a tail form that falls more steeply than has been found here. Partial reflection of the upward-propagating waves might serve the purpose. Dissipation that sets in at \( M < M_M \) and increases with \( M \) thereafter certainly would help—whether sufficiently or not, in practice, cannot be told without further analysis. The comparison presented in Fig. 8 suggests, however, that a smoothing of the dissipative process, if properly introduced, might well serve to produce the reported observational form: a gradually increasing diminution from the broken-line curve, replacing the sudden drop of the continuous-line curve, might well produce the dotted-line curve in the same region and simultaneously leave the effects on the smaller-\( M \) portion of the theoretical curve unchanged. In view of the conclusions concerned with the omission of vertical velocities, however, further pursuit of such tuning-up seems unlikely to be worth the attempt.

Despite the difficulties just discussed, the analysis has proven to be remarkably adequate to its task up to a point. In particular, the continuous-line curve in Fig. 8 is almost identical to the illustrated MD spectrum up to the indicated limit at \( M_M = 2.4 \). Beyond that point, the wave system may be taken (as in HII) to have decayed into turbulence, it having already provided sufficient wind shear to produce instability if the critical wind-shear variance is indeed \( 0.5 N^2 \). This turbulence may well have the characteristic \( M^{-3} \) spectral form [see Weinstock (1985), invoking the Lumley–Shur theory of the buoyancy subrange], thereby extending the “observable” spectrum in that form to greater \( M \) values. In the case shown, there would be a slight kink in the curve at \( M_M \), but that kink could be removed by a slight adjustment of parameters or by extending the dissipative effect of the turbulence to slightly smaller \( M \) values. In this case, the Doppler-spread theory would be invoked to account only for the position and form of the spectral peak, and for most of the first decade (in vertical wavenumber) of the tail, leaving wave-induced turbulence to account for the remainder of the tail. The spectral peak would occur at about \( M_p = 3^{-1/4} M_0 = 0.38 \) (i.e., \( m_p = 0.38 N / \sigma_T \)), and then the transition from wave dominance to turbulence dominance would occur at about \( M = 2.4 \) (i.e., \( m = 2.4 N / \sigma_T \)).

The continuous-line curve of Fig. 8 is found to provide, as the integral of \( \Theta^2 dM \), the value 0.192. If the total-wind spectral density in absolute wavenumber \( m \) is taken to be \( A \Theta^2 \), then the total-wind variance \( \sigma_T^2 \) will be given by 0.192 \( (N / \sigma_T) A \), with the \( (N / \sigma_T) \) factor deriving from \( dm / dM \), and so \( \sigma_T^2 = 0.192NA \). Then too, the tail spectral density approximates to \( AM_C^4 / M^3 \), with \( M_C = 0.5 \) as is seen by comparison with the MD spectrum. These facts combine to produce a tail spectrum of the form \( 0.326N^2m^{-3} \), which may be compared with \( 0.318N^2m^{-3} \) (\( = \pi^{-1} N^2m^{-3} \)) in the final model of HII, and with \( 0.17N^2m^{-3} \), as advocated by Smith et al. (1987) and claimed by them to represent the observational data well. The differences between these numerical multiples of \( N^2m^{-3} \) is of no consequence, given the uncertainties and variability of the relevant observations. The conclusion is that the continuous-line curve of Fig. 8, which has been derived by the (approximate) Doppler-shift theory from first principles (plus the assumption that critical shear variance = \( 0.5 N^2 \)), provides an admirable approximation to the observations at least up to \( M = M_M = 2.4 \). This corresponds to vertical scale sizes down to about 200 m in the stratosphere, 1200 m in the mesosphere, with turbulence then being assumed to continue the spectrum to smaller scales.

A further point should be made with respect to the Doppler-spread theory as I have developed it to date: I have assumed a statistical distribution of winds from which to obtain a statistical estimate of spreading. In reality, however, the wave systems will be facing one realization of winds on one occasion, another on another, and the average of the results may not equal the result for the average. I see no quantitative way around this problem except, perhaps, by numerical simulation of many successive cases, as was done for certain (oceanic) circumstances by Flatté et al. (1985), with generally supportive conclusions. But qualitatively, it should be noted that the cutoff function applicable on a given occasion will depend on the actual wind field on that occasion. In particular, an abnormally large wind at some underlying level will act to cut off the \( m_0 \) spectrum at a smaller \( m_0 \) than has been inferred here from probability considerations alone, and the effect of that change would be to steepen the portion of the tail spectrum that is contributed by waves propagating in the direction of the abnormally large wind. This behavior will occur, moreover, whether the abnormally large wind results from statistical accident within the wave-induced wind system or from a superimposed, true background wind—a feature not yet taken into account in the theory. Thus, there is potential within the theory to account for steeper tail spectra than those illustrated here, at least on occasion.

And then, going beyond its own assumptions, the Doppler-spread theory might at some point link onto the nonlinear-diffusion theory of Weinstock (1990) at the larger wavenumbers, to provide a marriage not only of convenience but of physical necessity. (It seems possible—even likely—that Weinstock’s theory provides but a different, somewhat heuristic, representation of the tail region of Allen and Joseph’s more direct analysis. It must derive, after all, from the advective non-
linearity of the Eulerian equations.) The combined theory would then simply unite two aspects of nonlinear wave-wave interaction: Doppler shifting imposed by the large-scale, wind-important waves on waves of middling scale, forcing them into a small-scale tail that then imposes its own limitations on the waves of middling scale via diffusive dissipation of them. If this marriage of the two theories is performed, however, there is no obvious role for instability and turbulence to play in the formation and form of the tail, though both might still occur.

Finally, there is some need to call into question the presumed universality of the intensity and \( m^{-3} \) form of the tail spectrum. Recent lidar observations (Beatty et al. 1992; Senft et al. 1993; Wilson et al. 1991) have revealed what appear to be tail spectra (in that the log-log slope appears to remain constant over as much as a decade in vertical wavenumber) having quite a range of slopes, both greater than and less than \(-3\). Moreover, the lidar observations have often revealed a tendency of the apparent tail spectra to increase in intensity with an increase of height, on passing from the middle stratosphere to the upper mesosphere and lower thermosphere, in contrast to the usually quoted constancy of tail spectra. A possible explanation for this tendency has been given by Hines (1993) in terms of “pseudo-saturation,” a phenomenon available to the Doppler-spread theory. Even without invoking pseudosaturation, the present analysis provides a further possible explanation based on the curves of Fig. 6.

As previously noted, that figure may be converted into one for absolute spectra as functions of absolute wavenumber \( m \) by shifting the dotted curve upward and to the left relative to the continuous curve, one decade to the left for every three decades up—the shifts resulting from an increase of \( \sigma_T \) with height. Had the log-log slopes of the tails been \(-3\), this would have permitted a sliding of the one tail spectrum more or less along the other, leaving those tails virtually superimposed and producing no growth of intensity with height at a fixed \( m \). As things are, however, the shifting would produce an amplification of the mesospheric tail intensity relative to the stratospheric tail intensity, as the lidar observations report. This effect would recur with the curves for finite \( M_M \), even if the stratospheric tail were thereby brought into the \( m^{-3} \) form, simply because the mesospheric cutoff function always leads to a shallower slope than does the stratospheric cutoff function. (Always, that is, in the absence of abnormally large intervening winds.) The reported variation of tail slopes about the usually quoted value \(-3\) could be explained in present terms, such as those employed in the discussion of Fig. 8, by variations in the effective \( M_M \). These variations would presumably result from variations from occasion to occasion of the actual presence or absence of turbulence, since a specified \( M_M \) (or a corresponding wind-shear variance) determines only the probability of finding instability, not the actual occurrence of instability (see HI), much less the occurrence of turbulence itself.

The present analysis has established the form of the cutoff function that should replace the step-function cutoff of HI in the Doppler-spread theory under the current approximations of that theory. It has provided an objective measure for the decorrelation distance appropriate to wind spectra of the general form observed and has thereby permitted a quantitative calculation of the cutoff function, somewhat different at different heights. It has shown that use of these cutoff functions does produce a high degree of universality for the form of the tail spectra, in confirmation of the claims made for the Doppler-spread theory in HI, and, in combination with the criterion that wind-shear variance should equal \( 0.5N^2 \) for marginal instability, it has produced (in Fig. 8) a spectral form almost identical to that commonly attributed to the observations. Finally, it has provided sufficient flexibility to account for variations, such as have been reported from lidar data, from the standardized attributes otherwise claimed observationally for saturated spectra. The theory is still embryonic in that it is based on a crass approximation to the true Doppler-shift effects of nonlinear wave-wave interactions, but it has proven to be remarkably successful in accounting for the position of the onset of the tail region and for the shape in much or most of the first decade of the tail. An improved analysis, less dependent on crass approximation, or else additional physics appears to be required at larger \( m \).

Further development of the Doppler-spread theory will probably require an approach extending that of Allen and Joseph (1989) to atmospheric circumstances (specifically: arbitrary input spectra, predominantly upgoing waves, and a tendency for growth of amplitude with height). Their analysis invokes a cutoff at certain undetermined wavenumbers in their Lagrangian spectrum. That cutoff appears to be the analog of the one investigated here via the cutoff function. The primary virtue of the present analysis may be nothing more than pointing the way to the proper determination of the cutoff when the more accurate analysis is, in due course, undertaken.

Acknowledgments. I am indebted to Prof. D. C. Fritts for prodding me into the analysis presented here; to Dr. M. P. Sulzer for several discussions and for providing Fig. 2; and most particularly to Dr. Q. Zhou for the computations leading to, and the provision of, Figs. 3–10. The Arecibo Observatory is a part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with the National Science Foundation. The National Science Foundation has provided further financial support via Grant ATM-9201040.
APPENDIX

Effects of Coriolis Force

Inclusion of the earth's rotation in the Doppler-spread theory adds complications of mathematical formulation but none of principle. The necessary revisions may be obtained by application of forms found in Gossard and Hooke (1975), for example, under the approximations $mH \gg 1$ and $\Omega \ll N$ already employed. They add terms containing the inertial frequency $f$, equal to twice the vertical component of the earth's angular velocity at the latitude of observation. The dispersion equation (2.1) is modified to

$$\Omega^2 - f^2)^{1/2} m = Nh, \quad (A.1)$$

and the Doppler-shift relation (2.4) to

$$u_n = (N^2 / m_0^2 + f^2 / h^2)^{1/2} - (N^2 / m_n^2 + f^2 / h^2)^{1/2}$$

$$= N \sigma_0^{-1} (1 + Q_0) - N \sigma_0^{-1} (1 + Q_n) \quad (A.2)$$

where

$$Q_n = f m_n / Nh \quad (A.3)$$

is a dimensionless parameter that provides a quantitative measure of the potential importance of Coriolis effects; its definition includes the case $n = 0$. (One would normally employ the ratio $f / \Omega$ as a measure of the importance of Coriolis effects; the use of $Q$, which is not quite the same, is dictated by the present requirement to exhibit operational formulas in terms of $m$ for purposes of the vertical-wavenumber spectrum.) The ratio $f / N$ is typically of the order 0.004 at temperate latitudes in the middle atmosphere, which value will be adopted here for illustrative purposes.

The requirement for vertical continuity of pressure and displacement at wave-displaced horizontal interfaces—or, alternatively, the requirement for conservation of wave action or of vertical flux of horizontal momentum (now angular momentum)—in the absence of dissipation can be shown to require the invariance of $\Theta^2 m^{-1} dm$ as before, leading to the recovery of (2.5) for application in the new circumstances.

The derivatives $dm_0 / dm_0$ and $du_n / dm_0$ are to be obtained from (A.2), although they ultimately combine (as was noted earlier) to provide the only needed derivative,

$$du_n / dm_n = N \sigma_0^{-2} (1 + Q_n^2)^{1/2} \quad (A.4)$$

The spectral transformation equivalent to (2.7) becomes

$$\Theta_1^2(m_1) = m_1^{-1} (1 + Q_1^2)^{-1/2} \int_0^\infty \Theta_1^2(m)$$

$$\times \frac{N}{m_0 \sigma_1} e^{-u_1 / \sigma_1} dm_0, \quad (A.5)$$

with $u_1$ taken as a function of $m_0$ and $m_1$ from (A.2). The upper limit of $u_1$ integration in (3.4), giving the $u_1$ value that corresponds to infinite $m_1$, is now $N \sigma_0^{-1} (1 + Q_0^2)^{1/2} - Q_0$. With obliteration imposed at $m_n$, this limit changes to $N \sigma_0^{-1} (1 + Q_0^2)^{1/2} - N \sigma_0^{-1} (1 + Q_n^2)^{1/2}$, where

$$Q_M = f m_m / Nh. \quad (A.6)$$

This leads in turn to a first-layer cutoff factor

$$0.5 \pm 0.5 \text{ erf} \left( N \sigma_0^{-1} (1 + Q_0^2)^{1/2} \right)$$

$$- N \sigma_0^{-1} (1 + Q_n^2)^{1/2} \right) \quad (A.7)$$

in place of the factor in braces in (3.6). Ultimately, (3.13) is replaced by

$$\Theta_1^2(m_n) = m_n^{-1} (1 + Q_n^2)^{-1/2} \int_0^{m_n} \chi(m_0, z_n)$$

$$\times \Theta_1^2(m_0) \frac{N}{m_0 \sigma_1} e^{-u_1 / \sigma_1} dm_0, \quad (A.8)$$

with $u_n$ taken to be a function of $m_0$ and $m_n$ as given by (A.2), and $\chi(m_0, z_n)$ being given by

$$\chi(m_0, z_n) = \left\{ \chi_{M_M} \right\} ^{4HD_m} \quad (A.9)$$

as before, but now with

$$\chi_{M_M} = \exp \left[ N^{-1} \sigma_0 \int_{N \sigma_0^{-1} z_n}^\infty \ln \left\{ (A.7) \right\} dM_0 \right] \quad (A.9)$$

in place of (3.15). For computational purposes in (A.9), (A.7) should be rewritten as

$$0.5 \pm 0.5 \text{ erf} \left[ (M_0^2 + M_0^2 f^2 / h^2)^{1/2} \right.$$}

$$- (M_n^2 + M_n^2 f^2 / h^2)^{1/2} \right], \quad (A.10)$$

with $m_0$ and $M_M$ being held constant as the integration over $M_0$ proceeds. Unfortunately, $\chi_{M_M}$ is no longer a function of $m_0$ and $z_n$ only via the local normalized wavenumber $M_0$, because of the explicit appearance of $m_0$ in (A.10) without a corresponding $\sigma_0$.

This completes the analysis as it is required for $\Theta^2(m)$. The factor $(1 + Q_n^2)^{-1/2}$ in (A.8) asymptotes to $N \sigma_0 / f m_n$ at large $m_n$, making $\Theta_1^2(m_n)$ asymptote to the form $m_n^{-2}$ (rather than $m_n^{-1}$ as previously, or $m_n^{-3}$ as might be wanted to match the observations), as was pointed out in appendix A of H11. But observers typically report the total horizontal wind-power spectrum $H^2(m)$, rather than $\Theta^2(m)$. Except for a constant of proportionality, the (density weighted) $\Theta^2(m)$ spectrum is given by the corresponding (density weighted) vertical displacement spectrum, which in turn is given by $\Omega^{-2}$ times the corresponding (density weighted) vertical velocity spectrum, which itself is given by $h^2 m_n^{-2} (1 + f^2 / \Omega^2)^{-1}$ times the (density weighted) hor-
horizontal-wind spectrum. [See, for example, the relevant polarization relations in Gossard and Hooke (1975) under approximations as before.] In combination with (A.1), this sequence yields

$$H_n^2(m_n) \propto (1 + 2Q_n^2)\Theta_n^2(m_n). \quad (A.11)$$

Insertion of this in (A.8) yields, finally,

$$H_n^2(m_n) = \frac{1 + 2Q_n^2}{m_n(1 + Q_n^2)^{1/2}} \int_0^{m_n} \chi(m_0, z_n)H_M^2(m_0) \times \frac{N}{m_0\sqrt{\pi} \sigma_n(1 + 2Q_0^2)} e^{-u_n^2/\sigma_n^2} dm_0, \quad (A.12)$$

with \(u_n\) taken to be a function of \(m_0\) and \(m_n\) from (A.2), as before. This is the relation employed in section 6 to produce Figs. 9 and 10, with \(H_M^2(m_0) \propto m_0\). The additional factor \((1 + 2Q_n^2)\) asymptotes to the form \(m_n^2\) at large \(m_n\), thereby offsetting the form \(m_n^{-2}\) that was found at large \(m_n\) for \(\Theta_n^2(m_n)\), and leading to the horizontal portions of the curves in Fig. 9. As is discussed in section 7, neglect of the vertical wind component in the determination of the Doppler-shifted frequency renders these portions irrelevant in any event.

Equation (A.12) is written for the density-weighted horizontal-wind spectrum. For the absolute horizontal-wind spectrum, the right-hand side should contain the further factor \(\exp(z/H)\).

REFERENCES


