J. Fluid Mech. (2005), *vol.* 529, *pp.* 195–219.
doi:10.1017/S0022112005003411 Printed in the United Kingdom

Near-inertial waves in the ocean: beyond the 'traditional approximation'

By THEO GERKEMA[1]† AND VICTOR I. SHRIRA[2]

[1]Laboratoire des Ecoulements Géophysiques et Industriels, BP 53, F-38041 Grenoble, France

[2]Department of Mathematics, Keele University, Keele, ST5 5BG UK
v.i.shrira@keele.ac.uk

(Received 12 May 2004 and in revised form 16 November 2004)

The dynamics of linear internal waves in the ocean is analysed without adopting the 'traditional approximation', i.e. the horizontal component of the Earth's rotation is taken into account. It is shown that non-traditional effects profoundly change the dynamics of near-inertial waves in a vertically confined ocean. The partial differential equation describing linear internal-wave propagation can no longer be solved by separation of spatial variables; it was however pointed out earlier in the literature that a reduction to a Sturm–Liouville problem is still possible, a line that is pursued here. In its formal structure the Sturm–Liouville problem is the same as under the traditional approximation, but its eigenfunctions are no longer normal vertical modes of the full problem. The question is addressed of whether the solution found through this reduction is the general one: a set of eigenfunctions to the full problem is constructed, which depend in a non-separable way on the two spatial variables; these functions are orthogonal and form, under mild assumptions, a complete basis.

In the near-inertial range, non-traditional effects act as a singular perturbation; this is seen from the sub-inertial short-wave limit, which is present whenever the 'non-traditional' terms are there, but disappears under the traditional approximation. In the dispersion relation the sub-inertial modes represent a smooth continuation of the super-inertial ones. The combined effect of the horizontal component of rotation and a vertical inhomogeneity in the stratification is found to play a crucial role in the dynamics of sub-inertial waves. They are trapped in waveguides localized around minima of the buoyancy frequency. The presence of horizontal inhomogeneities in the effective Coriolis parameter (such as shear currents or beta effect) are shown to enable a transition from super-inertial to sub-inertial waves (and thus effectively an irreversible transformation of large-scale into small-scale motions). It is suggested that this transformation provides a mechanism for mixing in the deep ocean.

The notion of critical reflection of internal waves at a sloping bottom is also modified by non-traditional effects, and they strongly increase the probability of critical reflection in the near-inertial to tidal range.

1. Introduction

Near-inertial waves represent the most energetic and, probably, the dynamically most significant part of the internal wave spectrum in the ocean (e.g. Fu 1981). Internal waves of all scales (as well as all other atmospheric and oceanic motions of scales small compared to the Earth's radius) are commonly described as if the Earth

† Present address: NIOZ, PO Box 59, 1790 AB Den Burg, The Netherlands; gerk@nioz.nl

were locally flat, i.e. the motions are considered in a plane tangent to the Earth's surface, attached at the location under consideration. This plane is co-rotating with the Earth's angular velocity $\boldsymbol{\Omega}$. In the equations of motion, written in a Cartesian frame fixed to this plane, the Coriolis vector has two components: a horizontal (strictly meridional) one $\tilde{f} = 2\Omega \cos\phi$, and a vertical one $f = 2\Omega \sin\phi$ ($\Omega = |\boldsymbol{\Omega}|$, ϕ latitude). The neglect of the terms involving the horizontal component $\tilde{f}$ represents the so-called *traditional approximation* (Eckart 1960). The additional assumption of constant f gives the so-called 'f-plane'. We shall instead refer to it as the 'traditional' f-plane, as opposed to the 'non-traditional' one in which $\tilde{f}$ and f are both assumed to be non-zero constants.

Scepticism with regard to the traditional approximation was first expressed by Bjerknes *et al.* (1933), in the context of the Laplace Tidal Equations. The debate that followed was summarized in the overview by Hendershott (1981). It would seem that his conclusion concerning the traditional approximation still holds today: "its domain of validity has not yet been entirely delineated". In discussing the validity, it is important to recognize that two distinct questions are at stake. One relates to phenomena that exist under the traditional approximation, and to whether these phenomena would be significantly modified by non-traditional effects. A key finding with respect to this question was that non-traditional effects generally become negligible for strong stratification, i.e. for small Ω/N, N being the buoyancy frequency. For example, as Hendershott (1981) notes, the assumption of strong stratification 'saves' the Laplace Tidal Equations in the sense that their solutions are, to a good approximation, also solutions of the fuller equations that include the non-traditional terms. Similarly, one finds that strong stratification tends to suppress the non-traditional terms in the dispersion relation for internal waves (Phillips 1968); but, as Phillips notes, non-traditional effects may remain present "in the finer details", even for strong stratification. This brings us to the other question which relates to phenomena that owe their existence to non-traditional effects (and that, therefore, disappear under the traditional approximation).

It has long been known that non-traditional effects enlarge the traditional frequency window for internal waves ($|f| \leqslant \sigma \leqslant N_{\max}$), thus creating a range $\sigma_{\min} \leqslant \sigma \leqslant |f|$ (e.g. Saint-Guily 1970; Kamenkovich & Kulakov 1977; LeBlond & Mysak 1978; Brekhovskikh & Goncharov 1994, Miropol'sky 2001).† We call inertio-gravity waves in this range *sub*-inertial (since $\sigma < |f|$), as opposed to the more familiar super-inertial waves ($\sigma > |f|$). The width of this sub-inertial range becomes smaller for stronger stratification, i.e. $\sigma_{\min} \to |f|$ as $\Omega/N \to 0$, as was first noted by Saint-Guily (1970). (We should remark at the outset that the assumption $\Omega/N \ll 1$ is satisfied in the seasonal thermocline, but not in the bulk of the ocean.)

The question now is whether this class merely forms a minor extension of the class of super-inertial waves, or whether it adds something qualitatively new and dynamically important. In the 1970s it was found (at least for constant N) that in a vertically bounded ocean the dispersion curves of super-inertial waves are relatively weakly modified, while those of sub-inertial waves form a family of branches qualitatively different from the super-inertial ones (Saint-Guily 1970; Kamenkovich & Kulakov 1977; see also Miropol'sky 2001). However, to our knowledge the class of sub-inertial waves has never been properly investigated, perhaps because they were thought to be of little significance, given the relatively small width of their frequency range. Along with

† Due to $\tilde{f}$ there is also an extension of the window above $N_{\max}$ which is of little significance and is not analysed here.

the general aim of clarifying the role of $\tilde{f}$ in internal-wave propagation, a primary and specific aim of our paper is to study these sub-inertial waves and elucidate their unique role in ocean dynamics.

These waves, for any model of stratification, become infinitely short, both horizontally and vertically, as the wave frequency approaches the minimum frequency $\sigma_{\min}$ (§ 3). This sub-inertial short-wave limit is a radical departure from the traditional description of near-inertial wave dynamics, and below we argue that the traditional approximation can be regarded as a singular limit (§ 3.1.1). Another key result of the present paper is that the sub-inertial dispersion curves form a direct smooth continuation of the super-inertial ones (§ 3). This opens the possibility for an effective transformation of one class into another. Moreover, it makes the sub-inertial waves dynamically important irrespective of how small their range of frequencies might be, since in the ocean the bulk of internal wave energy is in the inertial peak, i.e. in the very range where the transformation occurs. In particular, background horizontal variations in the effective Coriolis parameter f transform super-inertial into sub-inertial waves, or vice versa. Such variations are always present in the ocean, due to, for example, shear currents. So, even within the linear framework non-traditional effects bring in an inherent possibility of wave evolution towards smaller and smaller spatial scales (§ 5).

Although we focus on near-inertial waves, most results are applicable to the whole internal-wave range. Upon closer examination it appears that the literature contains more results on non-traditional aspects of internal waves than is generally assumed; we discuss the pertinent references, if not yet cited, in later sections as appropriate. We should note at the outset that we exclude the equatorial region, for which the non-traditional dynamics of internal waves requires a separate treatment (see Maas 2001, and references therein).

The paper is organized as follows. In § 2 we formulate the basic elements of the description of linear internal-wave propagation on the non-traditional f-plane. In § 3 we consider a vertically bounded domain, for which modal solutions are derived and analysed. Mathematically, abandoning the traditional approximation means that one has to deal with a non-separable partial differential equation. We show that this non-separability causes no difficulties so far as linear problems on the f-plane are concerned. Although normal *vertical* modes do not exist, there is a complete set of normal (non-vertical) modes. This takes away the principal reason for making the traditional approximation in the first place – an important point that has been generally overlooked. In § 3.3 we analyse the behaviour of sub-inertial waves trapped near the bottom, or in other regions where the stratification has a local minimum. In § 4, we review briefly the main results on characteristics in a vertically unbounded domain, following Badulin, Vasilenko & Yaremchuk (1991). Furthermore, we show how the notion of critical reflection at sloping bottoms (with constant slope) is modified by non-traditional effects, and briefly outline the implications for near-bottom mixing. Having thus discussed the three classes of near-inertial waves on the non-traditional f-plane (super-inertial, inertial, sub-inertial), we consider in § 5 the transition from one class into the other, due to variations in the effective f. The new picture that emerges is summarized in § 6.

2. Basic equations and properties

We begin with the linearized equations of motion of a density-stratified fluid on the non-traditional f-plane, under the Boussinesq approximation (e.g. Phillips 1966,

§ 5.7):

$$u_t - fv + \tilde{f}w = -p_x, \tag{2.1}$$

$$v_t + fu = -p_y, \tag{2.2}$$

$$w_t - \tilde{f}u = -p_z + b, \tag{2.3}$$

$$u_x + v_y + w_z = 0, \tag{2.4}$$

$$b_t + N^2 w = 0, \tag{2.5}$$

where p is the departure of pressure from its hydrostatic value (divided by a constant reference density); b is the buoyancy. The Cartesian frame with the coordinates x (west–east), y (south–north)-z (vertical, positive upward, with the origin at the unperturbed ocean surface) is used; u, v and w are the corresponding velocity components. We allow the buoyancy frequency N to depend on z (unless stated otherwise). In the traditional approximation one would take $\tilde{f} = 0$.

This set can be reduced to an equation for w, the vertical velocity component:

$$\nabla^2 w_{tt} + (\boldsymbol{f} \cdot \nabla)^2 w + N^2 \nabla_h^2 w = 0, \tag{2.6}$$

where $\boldsymbol{f} = (0, \tilde{f}, f)$, and ∇_h^2 denotes the horizontal Laplacian. Substituting $w = W \exp(\mathrm{i}\sigma t)$ into (2.6) gives

$$(N^2 - \sigma^2)W_{xx} + (N^2 - \sigma^2 + \tilde{f}^2)W_{yy} + 2f\tilde{f}W_{yz} - (\sigma^2 - f^2)W_{zz} = 0. \tag{2.7}$$

For plane waves travelling in the direction $(\cos\alpha, \sin\alpha)$, we find from (2.7) by introducing $\chi = x\cos\alpha + y\sin\alpha$ and substituting $\partial_x = \cos\alpha\,\partial_\chi$ and $\partial_y = \sin\alpha\,\partial_\chi$:

$$\left(N^2 - \sigma^2 + f_s^2\right)W_{\chi\chi} + 2ff_sW_{\chi z} - (\sigma^2 - f^2)W_{zz} = 0, \tag{2.8}$$

where $f_s = \tilde{f}\sin\alpha$. From this expression it is clear that non-traditional effects play no role for plane waves travelling in the east–west direction (i.e. $\alpha = 0$ or π). Equation (2.8) is the starting point of further study in this paper; it is convenient to write (2.8) as

$$AW_{\chi\chi} + 2BW_{\chi z} + CW_{zz} = 0, \tag{2.9}$$

where B and C are constants, while A may depend on z, via N. Without loss of generality, we shall assume that $B \geqslant 0$; this implies that the direction of positive χ is always poleward (except, of course, when propagation is purely zonal). For later reference, we note that $C < 0$ for super-inertial waves ($\sigma > |f|$), while $C > 0$ for sub-inertial waves ($\sigma < |f|$). Under the traditional approximation the mixed derivative would be absent ($B = 0$).

For a vertically bounded domain, (2.9) has to be solved subject to standard boundary conditions. At the ocean surface we use the 'rigid lid' approximation

$$W(0) = 0. \tag{2.10}$$

The boundary condition at the bottom $z = -H(\chi)$ is that of a vanishing normal component of the velocity, which for a flat bottom at depth H reduces to

$$W(-H) = 0. \tag{2.11}$$

For wave propagation to be possible, hyperbolicity of (2.9) is required, which implies $B^2 - AC > 0$. This condition delineates the range of allowable frequencies σ:

$$\sigma_{\min}^2 = \frac{1}{2}\left(\lambda - [\lambda^2 - (2fN)^2]^{1/2}\right) \leqslant \sigma^2 \leqslant \sigma_{\max}^2 = \frac{1}{2}\left(\lambda + [\lambda^2 - (2fN)^2]^{1/2}\right), \tag{2.12}$$

where $\lambda = N^2 + f^2 + f_s^2$. These upper and lower bounds, or equivalent expressions, have been derived before (e.g. Brekhovskikh & Goncharov 1994; the expressions derived by LeBlond & Mysak (1978) apply to the special case $\alpha = \pm\pi/2$). (Note that A vanishes at $\sigma^2 = N^2 + f_s^2$, i.e. within the allowed range of σ.)

Almost everywhere in the ocean one finds $N > |f|$. In the rare situation when $N < |f|$ (which we shall exclude from further consideration below), the frequency window on the traditional f-plane becomes $N < \sigma < |f|$. In this case too the window is enlarged by non-traditional effects; in particular, one finds from (2.12) that $0 < \sigma < (f^2 + f_s^2)^{1/2}$ as $N \to 0$. Interestingly, this implies that for a certain range of frequencies internal waves can continue to propagate vertically if $N \approx 0 \ll |f|$ in part of the column while $N > |f|$ in an adjacent part (a possibility that does not exist under the traditional approximation). This phenomenon has recently been observed in the Mediterranean (Van Haren & Millot 2004).

3. Waves in a vertically bounded and horizontally uniform domain

Equation (2.9) cannot be solved by separation of variables; it is only under the traditional approximation ($B = 0$) that this becomes possible. Yet, the non-separability does not in this case really pose a hurdle, for one can still reduce (2.9), (2.10), (2.11) to a boundary-value problem. This felicitous fact has received little attention in the literature, with the exception of Saint-Guily (1970) and Brekhovskikh & Goncharov (1994). They noticed that substitution of

$$W = \psi(z) \exp \mathrm{i}(k\chi + \delta z) \tag{3.1}$$

into (2.9), with $\delta = -kB/C$, yields

$$\psi'' + k^2 \left[\frac{B^2 - AC}{C^2} \right] \psi = 0. \tag{3.2}$$

Together with the boundary conditions (2.10), (2.11), (3.2) constitutes a Sturm–Liouville problem if we regard σ as given, and k^2 as the eigenvalue to be found (Ince 1956). In explicit form it reads

$$\psi'' + k^2 \left[\frac{N^2(z) - \sigma^2}{\sigma^2 - f^2} + \left(\frac{\sigma f_s}{\sigma^2 - f^2} \right)^2 \right] \psi = 0; \qquad \psi(0) = \psi(-H) = 0. \tag{3.3}$$

Before looking into solutions, we notice an obvious but important point that seems to have passed unnoticed in the literature: mathematically, the non-traditional term – i.e. the term in (3.3) involving f_s – produces merely an additional constant in the expression in square brackets, and hence poses no extra difficulty at all. (As we discuss below, however, the nature of the singularity at $\sigma = |f|$ is changed by this additional constant.)

At the same time, the role of ψ_n is subtly different from that under the traditional approximation, and this raises a number of fundamental questions. What do the eigenfunctions ψ_n actually describe? They are eigenfunctions of the Sturm–Liouville problem (3.3), but are not normal vertical modes of the original problem, since by virtue of (3.1) ψ_n forms only part of the vertical dependence of W_n. Are W_n normal modes then? Is the expansion in normal vertical modes possible at all? Below we show that although the full solutions W_n cannot be regarded as normal vertical modes (Appendix A), the set of W_n understood as unseparable functions of two independent variables (χ and z) is both orthogonal and, under certain assumptions, complete (Appendix B).

In § 3.1 the general properties of the dispersion relation following from (3.3) are studied in detail. The structure of the wave field is investigated in § 3.2, and is shown to be fundamentally different from that under the traditional approximation. This is because the wavenumber k now also affects the vertical structure, via (3.1). Finally, a new type of waveguide for sub-inertial waves, around minima of $N(z)$, is considered in § 3.3.

3.1. *The dispersion relation*

The dispersion relation for constant N, discussed below, was derived earlier by Kamenkovich & Kulakov (1977) without explicitly using the transformation (3.1); the plots were reproduced in Miropol'sky (2001). The main point of their analysis was to verify that for small Ω/N, the corrections to the super-inertial modes are small and the sub-inertial interval shrinks. These results were interpreted as an indication that the traditional approximation is reasonably adequate. Below we put forward a different view, by pointing out that the traditional approximation acts as a singular limit.

3.1.1. *Constant stratification*

It is instructive to consider first the simplest possible model: an ocean with constant stratification $N = N_c$ confined between two horizontal boundaries separated by a depth H. The solution to the boundary-value problem (3.3) then becomes $\psi_n = \sin q_n z$, with $q_n = n\pi/H$, n being an arbitrary integer, and we find the dispersion relation

$$k_n = \frac{\mp q_n C}{(B^2 - AC)^{1/2}} = \frac{\pm q_n(\sigma^2 - f^2)}{\left[\left(\sigma^2 - \sigma_{\min}^2\right)\left(\sigma_{\max}^2 - \sigma^2\right)\right]^{1/2}}, \tag{3.4}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the upper and lower bounds specified in (2.12). The additional vertical dependence of W_n due to the exponent we introduced in (3.1) is specified by $\delta_n (= -k_n B/C)$:

$$\delta_n = \frac{\pm q_n B}{(B^2 - AC)^{1/2}} = \frac{\pm q_n f f_s}{\left[\left(\sigma^2 - \sigma_{\min}^2\right)\left(\sigma_{\max}^2 - \sigma^2\right)\right]^{1/2}}. \tag{3.5}$$

From (3.4) we can derive the group velocity

$$\frac{\mathrm{d}\sigma}{\mathrm{d}k} = \frac{\pm\left[\left(\sigma^2 - \sigma_{\min}^2\right)\left(\sigma_{\max}^2 - \sigma^2\right)\right]^{3/2}}{q_n \sigma((B^2 - AC) + B^2 + C^2)}. \tag{3.6}$$

In figure 1(*a–d*) we show the two branches of the dispersion relation (3.4) for an arbitrary fixed n. The solid thick line corresponds to the plus-branch (i.e. taking the plus-sign in the last expression in (3.4)); the dashed thick line corresponds to the minus-branch. Thin solid and dashed lines show the corresponding curves under the traditional approximation. Notice that the plus-branches are mirror-images (with respect to the vertical axis) of the minus-branches, as is also clear from (3.4). Under the traditional approximation, the waves become infinitely long at their minimum frequency ($|f|$). On the non-traditional f-plane (thick lines), by contrast, the waves become infinitely *short* at their minimum frequency $\sigma_{\min} < |f|$. This is also clear from (3.4), because the denominator vanishes as $\sigma \to \sigma_{\min}$, while the numerator does not. The vertical wavenumber δ_n, given by (3.5), tends to infinity as well, implying that vertical scales too become infinitely short at the minimum frequency (for any mode-number n). We note that the sub-inertial short-wave limit was also evident in figure 4 of Saint-Guily (1970), but was not commented on, and seems to have passed unnoticed since.

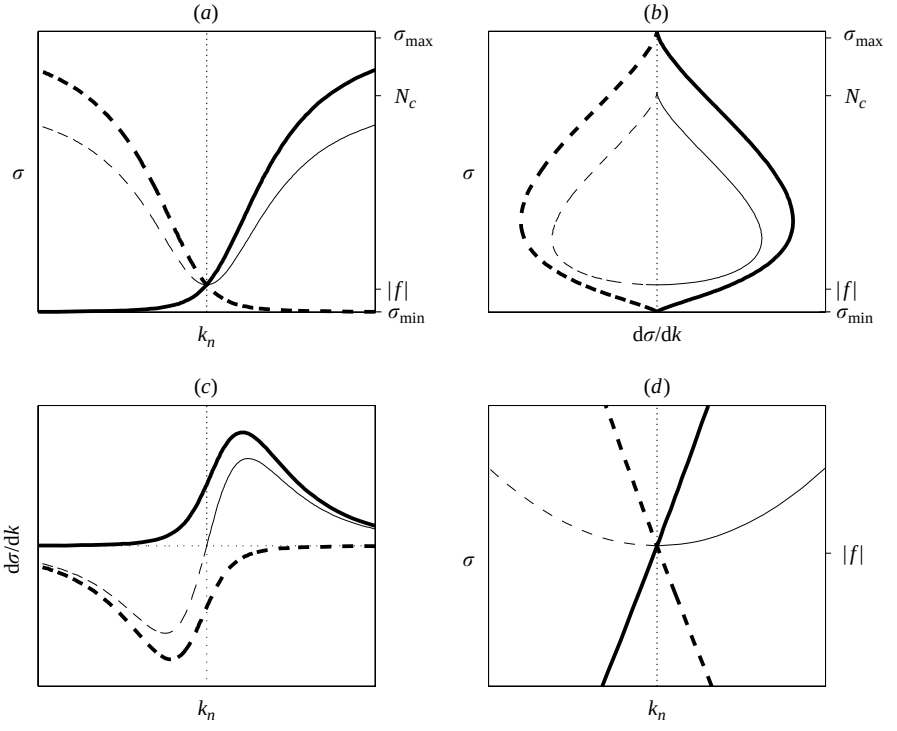

FIGURE 1. Dispersion relation for mode n (n being arbitrary but fixed). (a) Frequency σ versus wavenumber k_n; (b) frequency σ versus horizontal group velocity $d\sigma/dk$; (c) horizontal group velocity $d\sigma/dk$ versus k_n; (d) enlargement of (a) in the vicinity of $|f|$. Thick lines represent the curves for the non-traditional f-plane; thin lines, for the traditional f-plane. Solid lines correspond to the plus-branches in (3.4), (3.6), dashed lines to the minus-branches. In all panels, dots denote a zero axis.

The traditional approximation thus has the character of a singular limit: in the presence of non-traditional effects, no matter how weak, waves become infinitely short at the minimum frequency $\sigma_{\min}$; under the traditional approximation, this short-wave limit is absent. Mathematically, the singular character is expressed by the fact that the nature of (2.9) changes fundamentally under the traditional approximation, which removes the mixed derivative; likewise (3.3) changes as the higher-order singularity (at $\sigma = |f|$) is removed.

As is evident from figure 1(a), in the presence of non-traditional effects the familiar monotonic correspondence between time and length scales no longer exists. This implies that the notion of the Rossby radius of deformation, as a measure to assess the importance of Coriolis effects, becomes invalid. Indeed, on the non-traditional f-plane one encounters the paradoxical fact that Coriolis effects remain of first-order importance in the sub-inertial short-wave limit.

In figure 1(b, c) the horizontal group velocity (3.6) is shown. Another crucial difference between traditional and non-traditional dynamics now becomes apparent. Under the traditional approximation the solid and dashed branches, both, of course, super-inertial, join smoothly at $\sigma = |f|$. By contrast, on the non-traditional f-plane each of the two sub-inertial branches represents the direct continuation of the corresponding super-inertial branch, and a smooth continuation of one super-inertial branch into the other becomes impossible.

3.1.2. *Universal behaviour near the inertial frequency*

The behaviour for frequencies close to f, discussed for constant N by Brekhovskikh & Goncharov (1994) and Miropol'sky (2001), remains, in fact, the same for any $N(z)$. This can be seen as follows. In the expression in square brackets in (3.3), the non-traditional term becomes dominant when $\sigma \approx |f|$, that is, in the vicinity of the inertial frequency. As a consequence, at the leading order the solution is independent of $N(z)$, and the dispersion relation (for the nth mode) becomes

$$\pm k_n \frac{\sigma f_s}{\sigma^2 - f^2} = \frac{n\pi}{H}. \tag{3.7}$$

By exploiting further the proximity of σ to $|f|$ we find

$$\sigma = |f| \pm k_n \frac{f_s H}{2n\pi}. \tag{3.8}$$

Thus, in contrast to the traditional boundary-value problem, where for all eigenmodes $\sigma \geqslant |f|$, we now have, in addition, a family of *sub-inertial* modes with $\sigma < |f|$. As it follows from (3.8), frequencies of the modes of both families tend to the inertial frequency in the long-wave limit. For super-inertial waves this looks similar to the modal behaviour under the traditional approximation. However, there is a qualitative difference: the group velocity, which follows from (3.7), now does not vanish at the inertial frequency:

$$\left.\frac{\mathrm{d}\sigma}{\mathrm{d}k}\right|_{k_n=0} = \pm \frac{f_s H}{2n\pi}, \tag{3.9}$$

but equals the group velocity of the corresponding sub-inertial wave (note that this value is not small; the typical value for the deep ocean is of order $10\,\mathrm{cm\,s^{-1}}$.) So like in the case of constant N (figure 1*d*), for every n each of the two sub-inertial branches represents a direct and smooth continuation of the corresponding super-inertial branch.

This result is 'universal' in the sense that it does not depend on the stratification $N(z)$. Away from the inertial frequency the above asymptotics becomes, of course, invalid. In particular, the behaviour further into the sub-inertial regime depends on $N(z)$, as we discuss in § 3.3.

3.2. *Modification of the vertical structure*

A general solution to the original boundary-value problem (2.9), (2.10), (2.11) can be constructed by superposing eigenmodes $\psi_n(z)$ of (3.3), with the substitution (3.1) taken into account:

$$W = \sum_{n=1}^{\infty} \psi_n(z)\{a_n \sin(k_n\chi + \delta_n z) + b_n \cos(k_n\chi + \delta_n z)\}. \tag{3.10}$$

Here a_n, b_n are arbitrary coefficients. The structure of the solution confirms that the components of the series are not of a separable form unless the traditional approximation were made, in which case $\delta_n = 0$.

To get a better understanding of how non-traditional effects affect the vertical structure of the wave, we select just one mode with an arbitrary n and with time-dependence included:

$$w = \psi_n(z)\cos(k_n\chi + \delta_n z - \sigma t). \tag{3.11}$$

Hence the phase isolines are given by

$$\varphi(\chi, z) = k_n\chi + \delta_n z = \text{const}. \tag{3.12}$$

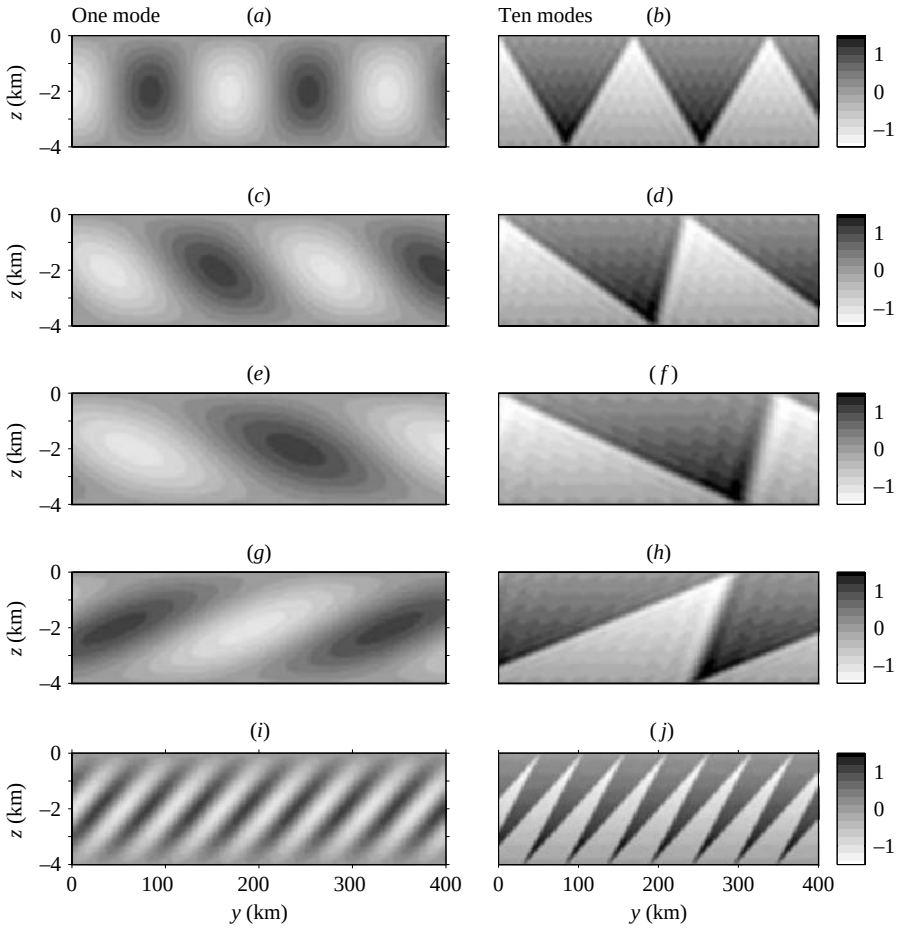

FIGURE 2. Modal solutions (3.10) for constant N, showing northward propagating near-inertial waves at mid-latitude. (a, b) The solution under the traditional approximation for the first mode (a), and of the first ten modes together (b), both for frequency $(\sigma = 1.025f)$. (c, d) The corresponding super-inertial solution without the traditional approximation. The same for (e, f), but at the lower frequency $(\sigma = 1.015f)$. $(g–j)$ Examples of the (non-traditional) sub-inertial solution, in (g, h) for $\sigma = 0.99f$, and in (i, j) for $\sigma = 0.98f$. Parameter values are: $N_c = 5 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$, $\alpha = \pi/2$, $\phi = 45°\mathrm{N}$; hence $\chi = y$ and $f_s = \tilde{f} = f$, except of course in (a) and (b), where $f_s = 0$.

Recall that δ_n, in explicit form, is given by

$$\delta_n = k_n \frac{f f_s}{\sigma^2 - f^2}. \tag{3.13}$$

From (3.12) we see that the phase φ has a vertical dependence, which would be absent under the traditional approximation ($f_s\,, \delta_n \equiv 0$), and which introduces very small vertical scales, even for the first mode. (This point was overlooked by Munk & Phillips (1968) when they attempted to justify the traditional approximation.) Notice that the expression (3.12) holds for arbitrary $N(z)$, since it does not involve the eigenfunction ψ_n; of course, the specific values of k_n do depend on the stratification. In the left-hand panels of figure 2, we illustrate the solution (3.11) for the simple case of constant stratification, where we take the first mode at $t = 0$. Figure 2(a) shows an example

under the traditional approximation (necessarily super-inertial), while figure 2(c, e) shows non-traditional super-inertial solutions, and figure 2(g, i) sub-inertial solutions (frequency decreases downward in the panels). The most obvious change with respect to figure 2(a) is the asymmetry in steepness created by non-traditional effects, with even a retrograde type of propagation at sub-inertial frequencies. The right-hand panels of figure 2 follow the same pattern but now with ten modes superposed in the solution (3.10), where we took $a_n = 0$ and $b_n = n^{-1}$. The superposition of modes creates a confined beam where $\mathrm{d}W/\mathrm{d}z$ is large. Such beams can be constructed in an equivalent but more direct way by using characteristics (see § 4).

3.3. *Trapped sub-inertial modes*

In this section we consider the case of non-constant stratification, which gives rise to a new type of wave motion that does not exist under the traditional approximation: sub-inertial inertio-gravity waves trapped in regions of the weakest stratification. For this purpose, it is convenient to rewrite (3.3) as

$$\psi'' + \tilde{k}^2[-N^2(z) + r^2]\,\psi = 0; \qquad \psi(0) = \psi(-H) = 0, \tag{3.14}$$

where

$$\tilde{k}^2 = \frac{k^2}{f^2 - \sigma^2} \qquad \text{and} \qquad r^2 = \frac{\sigma^2 f_s^2 + \sigma^2(f^2 - \sigma^2)}{f^2 - \sigma^2}.$$

Hereafter we shall restrict ourselves to sub-inertial waves ($\sigma < |f|$), so that we have $\tilde{k}^2, r^2 > 0$. For a sub-inertial wave to exist the expression in square brackets in (3.14) must be positive, i.e. $N^2(z) < r^2$. This confirms the idea, mentioned above, that sub-inertial waves will be trapped in regions around the minima in $N^2(z)$. These regions rapidly widen as σ approaches the inertial frequency, since r^2 is inversely proportional to $f^2 - \sigma^2$.

3.3.1. *Trapped modes for linear stratification*

First we consider the simplest possible case of a vertically varying N: $N^2(z) = N_0^2 + \gamma_1 z$, where N_0 and γ_1 are constants. (We have shifted the origin of the coordinate frame to the bottom, for convenience.) Equation (3.14) then becomes

$$\psi'' - l^2\left[z - \frac{\Delta}{\gamma_1 C}\right]\psi = 0. \tag{3.15}$$

where $l^2 = k^2\gamma_1/C$ and $\Delta = B^2 - A_0 C > 0$, with $A_0 = N_0^2 - \sigma^2 + f^2$. As before, $B = f f_s$ and $C = f^2 - \sigma^2$; the latter is here positive because we consider sub-inertial waves. We now derive the structure and properties of the corresponding bottom-trapped modes. The transformation $\hat{z} = l^{2/3}[z - \Delta/(\gamma_1 C)]$ reduces the equation to the standard Airy-equation:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\hat{z}^2} - \hat{z}\psi = 0.$$

We shall consider a domain that extends indefinitely upwards; then the solution in terms of the original z reads

$$\psi = \mathrm{Ai}\left(l^{2/3}\left[z - \frac{\Delta}{\gamma_1 C}\right]\right). \tag{3.16}$$

We require that ψ be zero at the bottom, here defined at $z = 0$. Let us denote the zeros of Ai by $-s_n$; they are known positive constants (e.g. Abramowitz & Stegun

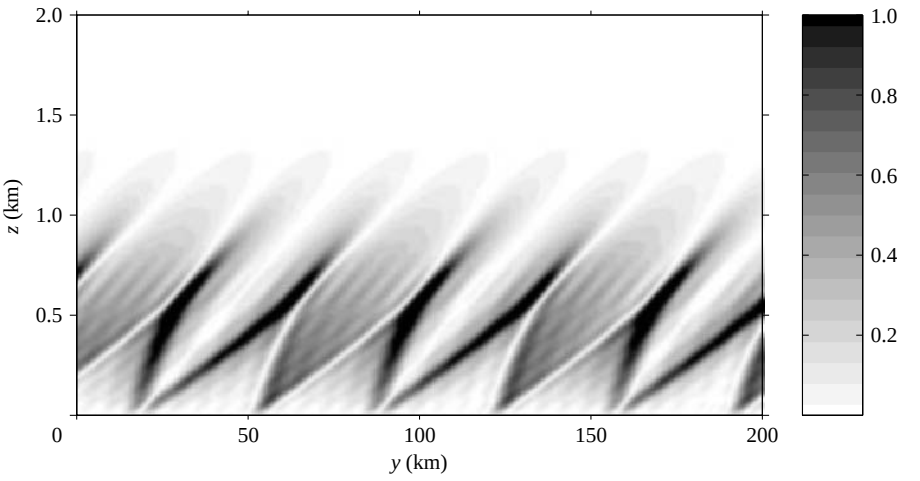

FIGURE 3. An example of a bottom-trapped sub-inertial wave ($\sigma = 0.99f$), given by a superposition of the first five modes in (3.10), with $a_n = b_n = n^{-1}$. The stratification is given by the realistic deep-ocean values $N_0 = 5 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$ and $\gamma_1 = 4 \times 10^{-10}\,\mathrm{rad^2\,m^{-1}\,s^{-2}}$; latitude $\phi = 45^\circ$ and propagation is in the meridional direction: $\alpha = \pi/2$ (poleward to the right).

1965, § 10.4). Then one finds for the wavenumbers l_n: $l_n^{2/3} = s_n \gamma_1 C/\Delta$, so that

$$k_n = \pm\gamma_1 C^2 \left(\frac{s_n}{\Delta}\right)^{3/2}; \qquad \delta_n = \mp\gamma_1 BC \left(\frac{s_n}{\Delta}\right)^{3/2}. \tag{3.17}$$

In explicit form the dispersion relation $k_n(\sigma)$ reads

$$k_n = \pm\gamma_1 C^2 \left(\frac{s_n}{(\sigma^2 - \sigma_{\min}^2)(\sigma_{\max}^2 - \sigma^2)}\right)^{3/2}, \tag{3.18}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are calculated from (2.12), using the minimum value of $N(z)$ (i.e. N_0), which is reached at the bottom. Like in the case of constant N, we find that the horizontal and vertical scales, k_n^{-1} and δ_n^{-1}, become infinitely short as $\sigma \to \sigma_{\min}$.

The group velocity is given by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}k} = \frac{\mp\Delta^{5/2}}{\gamma_1 \sigma C s_n^{3/2}[4\Delta + 3C(A_0 + C)]}. \tag{3.19}$$

The general solution for W can be constructed from (3.10), where ψ_n is now represented by (3.16). An example of $|W|$ made up of the first five modes is shown in figure 3. Of course, no comparison can now be made with traditional results, because this type of wave does not exist under the traditional approximation (for $B = 0$ we get $\Delta < 0$).

3.3.2. *Sub-inertial waveguides for the oceanic stratification*

A typical oceanic stratification profile is shown in figure 4, in which are also indicated the regions of trapping for a certain range of sub-inertial frequencies. In this case, there are three such regions, separated by the upper peak (the seasonal thermocline) and the smaller lower one (the main pycnocline). We shall assume below that the trapped sub-inertial motions decay sufficiently quickly outside their waveguides, so that the waveguides are well separated. If this is not the case, more

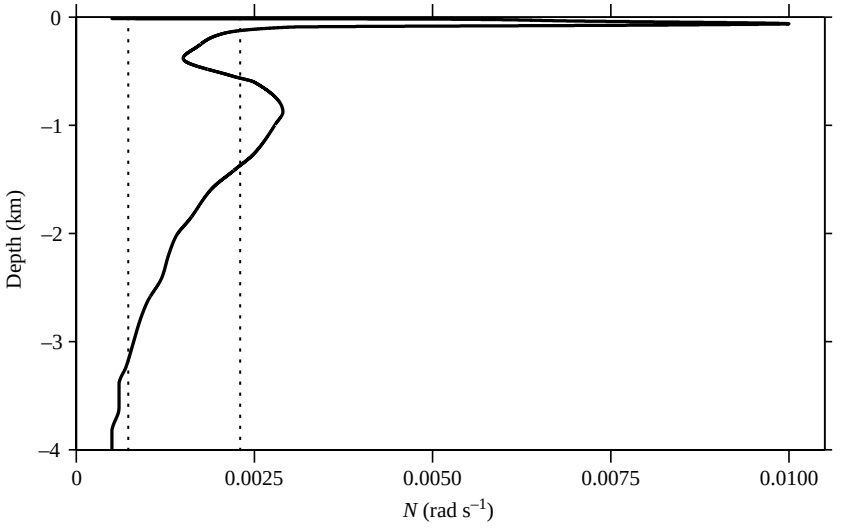

FIGURE 4. An empirical profile of the buoyancy frequency N, from the Bay of Biscay (summer). Sub-inertial waves up to $0.990f$ ($0.999f$) can exist where N lies to the left of the first (second) dotted line.

complicated asymptotic analysis of the type given in the book by Fedoryuk (1993) has to be applied.

The thin upper trapping region coincides with the mixed layer, in which (potential) temperature and salinity are nearly constant. Since N is virtually constant in this layer, the properties of the sub-inertial trapped modes are as those for the case of constant N discussed in § 3.1.1, with the water depth being replaced by the thickness of the mixed layer. Specifically, we have for $\sigma \to \sigma_{\min}$:

$$k_n \sim (\sigma - \sigma_{\min})^{-1/2}; \qquad \frac{\mathrm{d}\sigma}{\mathrm{d}k} \sim (\sigma - \sigma_{\min})^{3/2}. \tag{3.20}$$

The second trapping region lies in between the seasonal thermocline and the permanent pycnocline. At the leading order, the variation in N^2 can now be captured by a parabolic profile:

$$N^2(z) = N_0^2 + \gamma_2(z - z_m)^2 + \cdots.$$

If we place the origin at the point z_m, the boundary-value problem (3.14) takes the form

$$\psi'' + \tilde{k}^2\left[r^2 - N_0^2 - \gamma_2 z^2\right]\psi = 0, \qquad \psi \to 0 \text{ as } |z| \to \infty. \tag{3.21}$$

Here we have supposed (as before) that the solution decays sufficiently quickly outside the waveguide so as not to be affected by the presence of the other waveguides; this allows us to give the boundary conditions the present form. The transformation $\hat{z} = (\gamma_2\tilde{k}^2)^{1/4}z$ reduces (3.21) to the form

$$\psi'' - \hat{z}^2\psi = -\frac{|\tilde{k}|\left(r^2 - N_0^2\right)}{\gamma_2^{1/2}}\psi, \qquad \psi \to 0 \text{ as } |\hat{z}| \to \infty. \tag{3.22}$$

Notice that we can write

$$r^2 - N_0^2 = \frac{\Delta}{C} = \frac{\left(\sigma^2 - \sigma_{\min}^2\right)\left(\sigma_{\max}^2 - \sigma^2\right)}{C}$$

where Δ and C are defined as in § 3.3.1, and $\sigma_{\min}$ and $\sigma_{\max}$ are obtained from (2.12) with N replaced by N_0. Equation (3.22) can be solved in terms of Hermite polynomials (e.g. Abramowitz & Stegun 1965, § 19.13, § 22). In particular, the dispersion relation is found to be $\tilde{k}\Delta/(\gamma_2^{1/2}C) = \pm(2n+1)$, where n is a non-negative integer; hence

$$k_n = \pm(2n+1)\gamma_2^{1/2}\frac{C^{3/2}}{\Delta}$$

and so the group velocity becomes

$$\frac{\mathrm{d}\sigma}{\mathrm{d}k} = \frac{\mp\Delta^2}{\sigma(\gamma_2 C)^{1/2}(2n+1)[3\Delta + 2C(A_0 + C)]}.$$

For $\sigma \to \sigma_{\min}$, we now find

$$k_n \sim (\sigma - \sigma_{\min})^{-1}; \qquad \frac{\mathrm{d}\sigma}{\mathrm{d}k} \sim (\sigma - \sigma_{\min})^2. \tag{3.23}$$

Finally, the bottom-trapped waves in the lowest waveguide can at leading order be described by a linear dependence of N^2 on z. This leads us back to the previous section; specifically, we find for the asymptotic behaviour as $\sigma \to \sigma_{\min}$:

$$k_n \sim (\sigma - \sigma_{\min})^{-3/2}; \qquad \frac{\mathrm{d}\sigma}{\mathrm{d}k} \sim (\sigma - \sigma_{\min})^{5/2}. \tag{3.24}$$

We thus see from (3.20), (3.23) and (3.24) that as $\sigma \to \sigma_{\min}$ the wavenumber always tends to infinity, although the specific asymptotic behaviour differs for each of the types of waveguide.

Although the oceanic stratification with three distinct sub-inertial waveguides, considered above, is indeed the most common, there exist situations with many minima of $N(z)$ due to double-diffusion, which might result in a much more complicated waveguide structure. Nevertheless, the asymptotic behaviour we described is generic and is expected to persist in more complicated circumstances as well.

We finally note that the near-inertial wave dynamics will be very sensitive to variations in the effective f, caused, for example, by a passing Rossby wave or vortex. The reported magnitudes of such variations of the effective f in the ocean (e.g. Kunze 1985) are significant in the present context. If, for example, such a variation transforms the near-inertial wave from super- into sub-inertial, the wave will suddenly find itself trapped in a waveguide. This provides a new and potentially important mechanism of interaction between the near-inertial waves and large-scale motions.

4. Characteristics and waves over sloping bottom

In this section we review the basic mathematical facts concerning wave propagation in vertically unbounded and semi-infinite domains. The main non-traditional effects were discussed by Badulin *et al.* (1991); their discussion is recapitulated and extended below as the starting point for further analysis. The practical implications, such as for critical reflection at a sloping bottom, are dealt with in § 4.3.

4.1. *Characteristics*

First we analyse (2.9) by using the method of characteristics; they are the curves $\xi_{\pm}(\chi, z) = \text{const}$ along which

$$\frac{\mathrm{d}z}{\mathrm{d}\chi} = \frac{B \pm (B^2 - AC)^{1/2}}{A} \equiv \mu_{\pm}, \tag{4.1}$$

see e.g. Zauderer (1989). Recall that

$$A(z) = N(z)^2 - \sigma^2 + f_s^2; \qquad B = ff_s; \qquad C = -(\sigma^2 - f^2).$$

The characteristic coordinates can now be introduced as

$$\xi_\pm = \chi - \int_{z_0}^{z} \frac{\mathrm{d}\tilde{z}}{\mu_\pm(\tilde{z})}, \tag{4.2}$$

where z_0 is an arbitrary constant. Using these characteristic coordinates, we can transform (2.9) into its canonical form:

$$\frac{\partial^2 W}{\partial \xi_+ \partial \xi_-} = \frac{C^2}{4\Delta}\left[(\mu_+^{-1})' \frac{\partial W}{\partial \xi_+} + (\mu_-^{-1})' \frac{\partial W}{\partial \xi_-}\right], \tag{4.3}$$

where primes denote derivatives with respect to z (once a specific form of $N(z)$ has been chosen, the terms involving these derivatives can be cast in terms of the characteristic coordinates). Here $\Delta = B^2 - AC$; recall that the condition for hyperbolicity is $\Delta > 0$. The physical meaning of the terms involving the first derivatives, on the right-hand side of (4.3), is as follows: a wave propagating along a particular characteristic creates at each point it passes reflected waves which will propagate along the other (crossing) characteristic. This is why there is no solution describing a wave propagating only along one family characteristics (unless, of course, we take N constant, in which case the right-hand side becomes zero, see below). It is worth noting that the right-hand side might tend to zero not only when the stratification becomes uniform ($N \to \text{const}$), but also when $\sigma \to |f|$ since then the factor C, and hence the right-hand side, vanish for arbitrary stratification. (We note that the solution, in terms of the original coordinates χ and z, however still depends on N.) For arbitrary $N(z)$ and σ, one cannot in general solve (4.3) analytically. Nevertheless, the characteristics provide important insight into the qualitative behaviour of the solutions; for example, in any point M the solution at that point depends only on the values of W in the area confined by the two characteristics that intersect at M.

Importantly, (4.3) has the same formal structure as under the traditional approximation; in other words, the traditional approximation does not simplify the mathematical problem, once reduced to its core, and can, therefore, be dispensed with. This is in line with what we found in § 3 for the problem in a bounded domain. Specifically, for constant stratification, (4.3) reduces to the simplest possible form:

$$\frac{\partial^2 W}{\partial \xi_+ \partial \xi_-} = 0, \tag{4.4}$$

which is the starting point for further analysis in the remainder of this section. Even though the mathematical problem is, in essence, the same as under the traditional approximation, the physical behaviour differs fundamentally, as becomes apparent when we transform the solution back to the original coordinates, via (4.2).

4.2. *Basic properties*

For constant N (denoted by N_c), the characteristics (4.2) become

$$\xi_\pm = \mu_\pm \chi - z + z_0, \tag{4.5}$$

where $\mu_\pm$ are now constants (for later convenience, we have multiplied the right-hand side by the constant factor $\mu_\pm$). As was discussed by Badulin *et al.* (1991), the orientation of $\xi_- = \text{const}$ depends on the sign of C. Thus, are there three cases to be distinguished, as is illustrated in figure 5. Here propagation is in the meridional

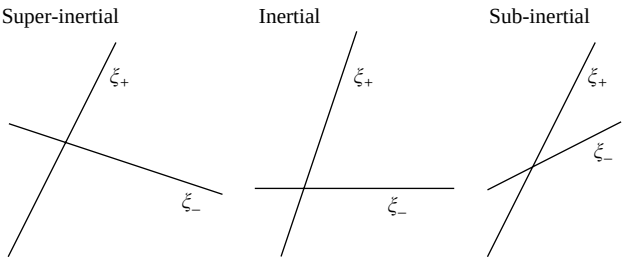

FIGURE 5. Schematic view of the characteristics $\xi_\pm = \text{const}$, for each of the three regimes.

direction (poleward to the right), and we have also assumed that $A > 0$ (a condition normally satisfied in the ocean). For super-inertial waves ($\sigma > |f|$), the two characteristics $\xi_\pm = \text{const}$ always have a different orientation (like on the traditional f-plane), and a different steepness (unlike on the traditional f-plane). For strictly inertial waves ($\sigma = |f|$), $\xi_- = \text{const}$ becomes horizontal. The characteristic $\xi_+ = \text{const}$ (which is non-existent on the traditional f-plane) retains its orientation; notice that its steepness, $2B/A$, depends on N_c. Finally, for sub-inertial waves ($\sigma < |f|$), the two have the same orientation; their steepness tends to the same limit as $\sigma \to \sigma_{\min}$. The class of sub-inertial waves would not exist under the traditional approximation.

The direction of the characteristics coincides with that of energy propagation; this can be verified by deriving the dispersion relation. We substitute $W = \exp \mathrm{i}[k\chi + mz]$ into (2.9) to obtain, with $(k, m) = \kappa(\cos\theta, \sin\theta)$,

$$A\cos^2\theta + 2B\cos\theta\sin\theta + C\sin^2\theta = 0, \tag{4.6}$$

so that

$$\sigma^2 = N_c^2\cos^2\theta + (f_s\cos\theta + f\sin\theta)^2, \tag{4.7}$$

which is a known dispersion relation (e.g. Phillips 1966). The group-velocity vector is given by

$$\left(\frac{\partial\sigma}{\partial k}, \frac{\partial\sigma}{\partial m}\right) = (2\kappa\sigma)^{-1}\left[\left(N_c^2 - f^2 + f_s^2\right)\sin 2\theta - 2ff_s\cos 2\theta\right](\sin\theta, -\cos\theta). \tag{4.8}$$

Its direction in the (χ, z)-plane is $-\cot\theta$. Now, from (4.1) and (4.6) we find that $-\cot\theta = \mu_\pm$, which confirms that energy propagates in the direction of the characteristics.

Moreover, the group-velocity becomes zero (in both components) only if the frequency takes one of its extreme values, $\sigma_{\min}$ or $\sigma_{\max}$. Since $|f|$ is not a lower bound on the non-traditional f-plane, it follows that the group velocity is non-zero; in other words, waves having the inertial frequency do propagate (in the direction indicated in figure 5), unlike on the traditional f-plane.

4.3. *Reflection at constant slopes*

A slope is called critical if the direction of the slope coincides with that of one of the characteristics. The 'traditional' notion of criticality requires modification when we consider the non-traditional f-plane, for two reasons: first, because in the class of super-inertial waves (which exists also under the traditional approximation), there is now an asymmetry in steepness, as illustrated in figure 5; second, because there

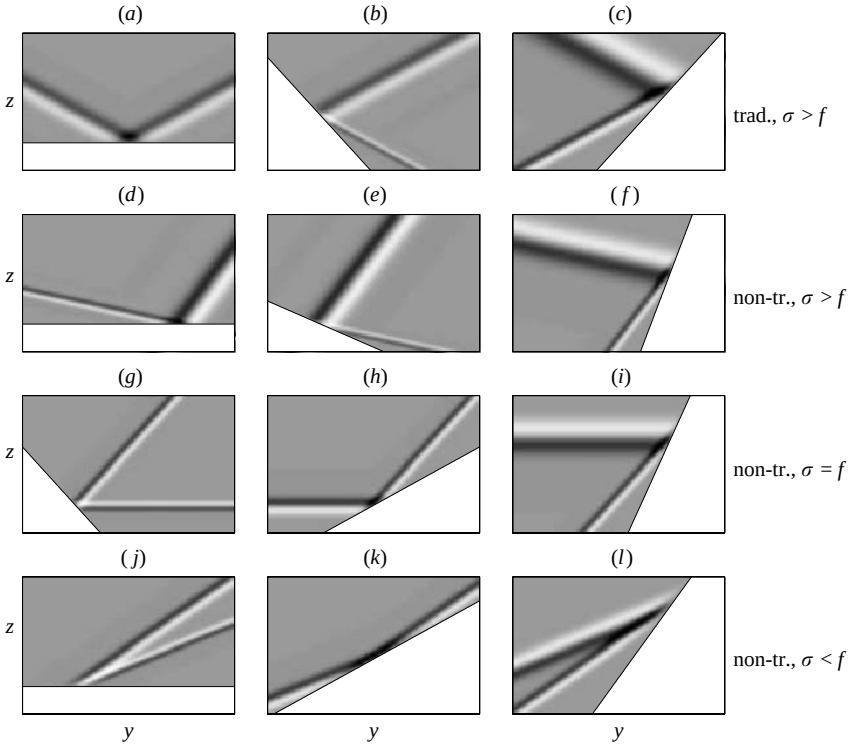

FIGURE 6. Generic cases of non-critical reflection at constant slopes. (*a–c*) For a super-inertial frequency ($\sigma = 1.02f$), under the traditional approximation. All other panels (*d–l*) are without the traditional approximation: (*d–f*) super-inertial ($\sigma = 1.02f$); (*g–i*) inertial ($\sigma = f$); (*j–l*) sub-inertial ($\sigma = 0.98f$). In all panels, parameter values are: $N_c = 5 \times 10^{-4}\,\mathrm{rad\,s^{-1}}$, $\alpha = \pi/2$ (propagation in the meridional direction, poleward to the right), $\phi = 45°\mathrm{N}$ (mid-latitude), hence $f_s = f$ (except of course in *a–c*, where $f_s = 0$).

are now two additional classes of wave-motion, namely the inertial and sub-inertial ones.

Here we consider the simplest case in which the gradient of the slope and both characteristics lie in the same plane. We assume that the bottom has a constant slope: $z = \gamma\chi$. Our starting point is (4.4), but now it is more convenient to use the streamfunction Ψ ($W = -\Psi_\chi$) instead of the vertical velocity W. The streamfunction satisfies the same equation, but vanishes at the slope. The general solution of (4.4) is $\Psi = F(\xi_+) + G(\xi_-)$, where F and G are arbitrary functions. The requirement that $\Psi = 0$ at the slope implies

$$F([\mu_+ - \gamma]\chi) + G([\mu_- - \gamma]\chi) = 0$$

for all χ. Hence, Ψ is given by

$$\Psi = F(\xi_+) - F\left(\frac{\mu_+ - \gamma}{\mu_- - \gamma}\,\xi_-\right). \tag{4.9}$$

We choose $F(x) = \exp(-ax^2)\sin(x)$; this creates the confined beams shown in figure 6. The nine generic non-traditional types of non-critical reflection are illustrated in figure 6(*d–l*). We distinguish between super-inertial, inertial, and sub-inertial frequencies.

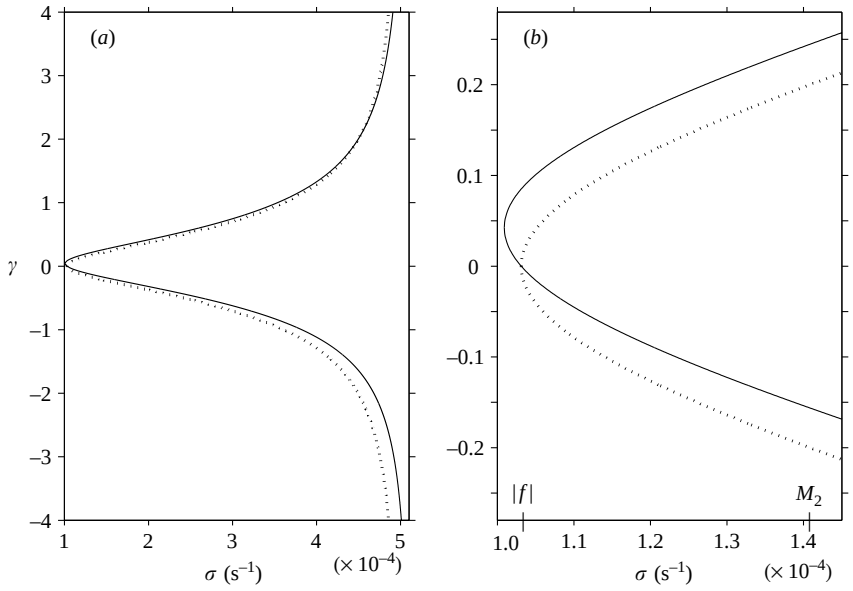

FIGURE 7. (*a*) Critical values of the slope as a function of wave frequency (solid line); (*b*) an enlargement for the near-inertial range. The lower branch represents $\gamma = \mu_-$; the upper branch, $\gamma = \mu_+$. The critical slopes under the traditional approximation are also shown (dotted line); the curve is then symmetric with respect to the horizontal axis; in this case $|f|$ is the lower bound of the frequency domain.

Notice that at inertial frequencies, a flat bottom is critical. Under the traditional approximation this is no longer the case since then there is no energy propagation at the inertial frequency. To describe critical reflection, one needs to include viscous and nonlinear effects (e.g. Dauxois & Young 1999).

In figure 7 we plot the critical γ as a function of wave frequency, for the same f, f_s and N_c as in figure 6. Statistically, near-inertial waves are more likely to encounter a critical slope when the depth decreases in the poleward direction ($\gamma > 0$), than when it increases ($\gamma < 0$); for sub-inertial waves critical slopes do not occur at all in the latter case. This contrasts with the behaviour under the traditional approximation, where no such asymmetry is found (dotted line in figure 7). More importantly, the shift with respect to the dotted line means that critical reflections are much more likely to occur than under the traditional approximation. This is because mildly sloping bottoms are more abundant than those with steeper slopes; hence the shift of the lower branch towards smaller $|\gamma|$ by far outweighs the shift of the upper branch towards higher $|\gamma|$. The shift is still noticeable at the semi-diurnal tidal frequency (M_2). The increased likelihood of critical reflection is expected to account better for intense mixing at boundaries (compared to estimates obtained under the traditional approximation).

For slopes of a more general shape, new phenomena might arise. For example, a concave slope can support slope-trapped sub-inertial solutions. Following a WKB approach, we see that a sub-inertial beam reflected from a supercritical upper segment of the slope will bounce back from the subcritical lower segment, and hit the supercritical segment again, and so on. This issue merits further consideration, but is not pursued here.

5. Transformation towards small scales due to a horizontal inhomogeneity

In § 3 we considered a horizontally uniform ocean confined between two horizontal planes and found that the behaviour of near-inertial waves in the sub-inertial regime differs radically from that in the super-inertial regime, but that at the same time the sub-inertial dispersion branches represent a smooth continuation of the corresponding super-inertial branches. This implies that in the real ocean the character of near-inertial waves will be extremely sensitive to any variation in the 'effective' f, caused either by large-scale inhomogeneities (vortices, Rossby waves, currents) or, in their absence, by the meridional variation of f itself. Such variations can cause a transformation of waves from the super-inertial range to the sub-inertial range, or vice versa. The possibility of a transformation of long super-inertial waves into ever shortening sub-inertial ones is of special interest from a general oceanographic point of view (notably the subject of deep-ocean mixing). Here we briefly outline how and when such a transformation can occur.

To fix ideas consider the following simple example: we trace the kinematics of an initially super-inertial narrow-band wave packet propagating across a vertically uniform zonal current $U(y)$. Let the initial packet be characterized by a central frequency $\omega_{(0)}$ and wavevector $(k_{x(0)}, k_{y(0)})$, while the vertical structure of the packet is described by one of the modes of (3.3); the specific number n of the mode is of little importance for our further consideration as long as it is fixed (the group velocity, of course, depends on n, but the qualitative conclusions that follow do not depend on n). We assume a separation of scales wide enough for the WKB description to be valid; by virtue of the standard conservation laws (e.g. Whitham 1974), the packet's zonal wavenumber $k_{x(0)}$ and frequency $\omega_{(0)}$ will be preserved (in the reference frame moving with $U(0)$ where $y=0$ corresponds to the initial position of the packet), while the meridional wavenumber k_y and intrinsic frequency $\sigma=\omega_{(0)}-k_xU(y)$ will be evolving. At each position y they can be found from the dispersion relation $\sigma(k_x, k_y)$, presumed to be known. Since the group velocity can also be found from $\sigma(k_x, k_y)$ one can write and solve in quadratures the equations for the packet trajectory in the (x, y)-plane:

$$\dot{x}=\frac{\partial\sigma(k_x, k_y(y))}{\partial k_x}+U(y), \qquad \dot{y}=\frac{\partial\sigma(k_x, k_y(y))}{\partial k_y}. \tag{5.1}$$

It is illuminating to follow the packet's progress on the dispersion curve. The curve itself is the result of a cross-section of the still-water dispersion surface $\sigma=\sigma(k_x, k_y)$ with the plane $k_x=k_{x(0)}$, while the position of the packet on the curve is specified by its intrinsic frequency $\sigma=\omega_{(0)}-k_{x(0)}U(y)$ (see figure 8).

If $k_{x(0)}U(y)$ increases, the intrinsic frequency decreases and passes the inertial frequency.† Further into the sub-inertial domain the dispersion curve for a typical density stratification splits into three branches corresponding to sub-inertial waves localized in the surface, bottom and seasonal waveguides discussed in § 3.3.2. We plot just one sub-inertial branch. As the packet approaches the minimal possible frequency $\sigma_{\min}$ specified by (2.12), k_y grows and tends to infinity. Although each of the three guides is characterized by its own minimal frequency $\sigma_{\min}$, and its own dispersion curve, the conclusion regarding the tendency towards infinitely short scales is robust. This tendency holds for any stratification and for any shear flow $U(y)$.

† At the inertial frequency the WKB approximation becomes invalid and scattering of the energy into other modes occurs, which we do not consider here.

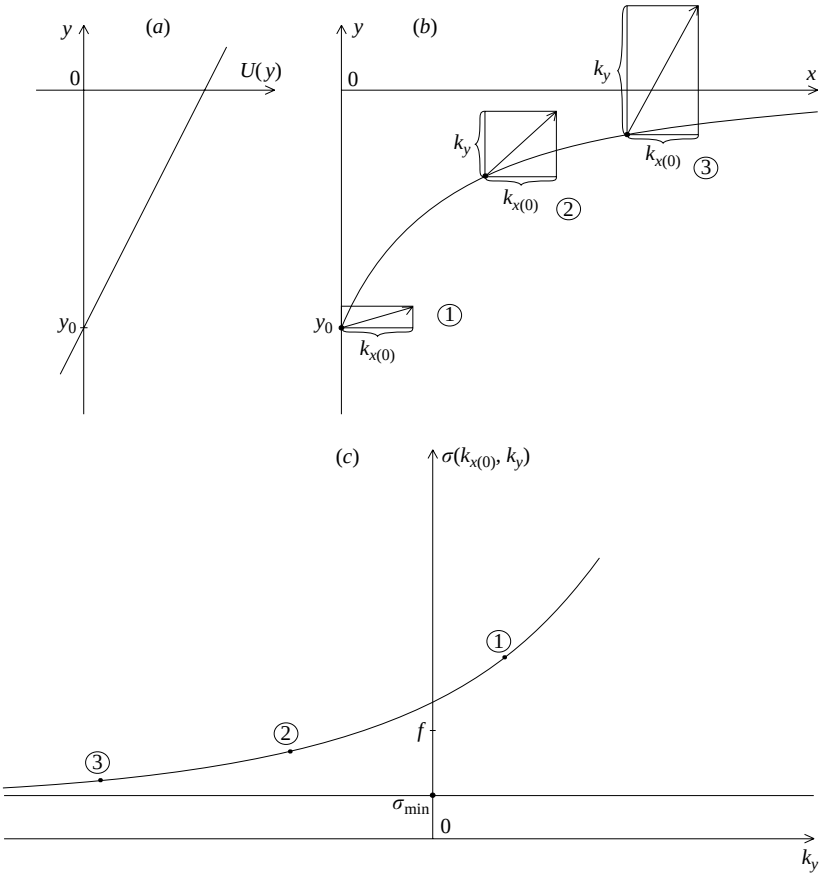

FIGURE 8. Transformation of a wave packet into small scales due to a sheared zonal current. (*a*) Current profile; (*b*) trajectory of a wave packet and evolution of its wave vector shown at selected consecutive moments in time: ①t_0, ②t_1, ③t_2; (*c*) growth of the meridional wavenumber along a cross-section of the dispersion surface as the intrinsic frequency tends from ω to $\sigma_{\min}$.

The remaining questions to be investigated are as follows: (i) Will the packet reach the point $y = y_{cr}$ where $\sigma = \sigma_{\min}$ in finite or infinite time? (ii) What will happen to the packet underway and, in particular, could it get reflected at $y = y_{cr}$?

The answer to the first question is straightforward since the explicit asymptotics of $\sigma = \sigma(k_x, k_y)$ for large k_y have been established in § 3. Indeed, consider the packet in the position far in the sub-inertial domain where k_y is already large and the short-wave asymptotics applies. If we expand $U(y) = U(y_{cr}) + U'(y_{cr})(y - y_{cr})$, we find (after moving the frame origin to $y = y_{cr}$)

$$\frac{\sigma(y)}{\sigma_{\min}} = 1 + \frac{k_{x(0)} U'(0) y}{\sigma_{\min}}, \tag{5.2}$$

which, in particular, implies $\sigma(y) - \sigma_{\min} = k_x U'(0) y$. On the other hand, we know that

$$\frac{\sigma(y)}{\sigma_{\min}} = 1 + g(k_x, k_y), \tag{5.3}$$

where $g(k_x, k_y)$ is known from the short-wave asymptotics of the dispersion relation. This gives an explicit formula for k_y as function of y. For example, for the bottom

waveguide employing (3.24), we find

$$k_y \sim y^{-3/2}. \tag{5.4}$$

The time τ it takes for a packet to advance from y_* to y is given by

$$\tau = \int_{y_*}^{y} \frac{\mathrm{d}y}{c_{gr}(k(y))}. \tag{5.5}$$

Inserting the expressions for the group velocity for the bottom guide from (3.24) into (5.5), we find

$$\tau = \int_{y_*}^{y} y^{-5/2}\mathrm{d}y \sim y^{-3/2}, \tag{5.6}$$

which becomes infinite as $y \to 0$, i.e. as the packet approaches the critical point. The above integrals for τ diverge not only for the example considered here, but for all plausible models of stratification.

There are two important points to be mentioned: first, the decrease of the group velocity by virtue of conservation of wave action leads to an infinite growth of the wave amplitude (Whitham 1974); second, the fact that τ is infinite implies that in practice any reflection from the critical point $y = y_{cr}$ will be very unlikely since the ever decreasing wavelengths are subject to ever increasing dissipation over an infinite period of time and it is likely that the wave will dissipate when the WKB approximation employed is still valid (see the analysis in Badulin, Shrira & Tsimring (1985) of a mathematically similar problem in the context of short internal waves). In the absence of dissipation however, to prove rigorously the absence of reflection is impossible when remaining within the WKB framework since the WKB approximation loses its validity in the vicinity of the critical point and a more detailed investigation of the singularity neighbourhood is required.

Naturally, for the singularity to occur at $y = y_{cr}$ the intensity of the current should exceed a certain threshold specified by the condition

$$U_{\max} \geqslant (\omega - \sigma_{\min})/k_x. \tag{5.7}$$

Such currents are common in the ocean and, therefore, one might expect significant pumping of energy contained in the inertial peak into small scales. Of course, the conclusion is not confined to the type of current chosen for the above illustrative example, but holds for the most general type of ocean inhomogeneity.

Moreover, even in the absence of any currents a similar mechanism will be still at work, since a horizontal inhomogeneity is always present due to sphericity of the Earth, via the corresponding meridional variation of the Coriolis parameter f; this is further discussed elsewhere (Gerkema & Shrira 2005).

6. Discussion

The analysis presented above shows that one finds a qualitatively and fundamentally new picture of near-inertial wave dynamics if three elements are combined: non-traditional effects, a vertical confinement, and background variations in the effective f. In the ocean, it will be noticed, all three are always there. Specifically, the vertical confinement is provided not only by the bottom and the surface of the ocean but also by vertical density variations which lead to the waveguides described in § 3.3. Furthermore, the effective f varies due to the presence of shear currents, or the β-effect.

The analysis of the non-traditional f-plane in a vertically bounded domain reveals that the class of sub-inertial waves, which owes its existence to non-traditional effects (as has long been known at least for constant N), differs fundamentally from the class of super-inertial waves. The sub-inertial waves become infinitely short as the lower bound of the frequency domain is approached (§§ 3.1.1, 3.3). Here the finite vertical extent is essential, since in the absence of horizontal boundaries no such tendency is found (§ 4). If we now add the third element, variation in effective f, the tendency towards ever smaller scales becomes an intrinsic property of the linear propagation of near-inertial waves. The irreversible evolution into short scales is of special importance for deep-ocean mixing (see e.g. the recent review by Garrett & St. Laurent 2002). The sensitivity of inertial-wave dynamics to large-scale motions (like Rossby waves, eddies, and shear currents), as outlined in § 5, might open a new avenue of research of mechanisms of their coupling.

Direct observational evidence of the sub-inertial range is not available at present because of the difficulties in distinguishing these motions within the inertial peak and, probably, because they were not specifically sought. A promising route could be to look at the vertical distribution of the near-inertial wave field. For example, persistent abyssally intensified near-inertial motions have recently been found in the year-long observations by Van Haren, Maas & Van Aken (2002), which is consistent with the idea of a deep sub-inertial waveguide (§ 3.3). Also promising would be a study of the polarization of the velocity records, and the ratio of potential to kinetic energies, see Gerkema & Shrira (2004).

Apart from the overall picture and its constituent elements, a number of results obtained here merit special mention. Non-traditional effects not only significantly alter the conditions for critical reflection for near-inertial waves, but also the importance of criticality. Depending on the orientation of the bottom slope, the critical value of steepness can decrease quite considerably (compared to the r.m.s. bottom slope) for a broad frequency range, which includes the M_2 tide (figure 7). Since, statistically, milder slopes are much more abundant, this implies a large increase in the occurrence of critical reflection of M_2 beams, and hence a substantially increased contribution to mixing. (The notion of critical reflection invites the practical question of what value of N, i.e. at what height and over what vertical scale, actually determines criticality.)

From a mathematical perspective, we have clarified a number of questions concerning the traditional approximation. The usual *raison d'être* for the traditional approximation is to render the equations of motion separable. As we have argued above, for the standard linear internal-wave problem on the f-plane this advantage is spurious: in a bounded domain, the 'non-traditional' Sturm–Liouville problem (3.2) has the same mathematical structure as the one found under the traditional approximation. The resulting two-dimensional eigenfunctions are orthogonal and, under some assumptions, form a complete set. It thus appears that in this kind of problem there is no need to make the traditional approximation.

We also notice that even for waves of frequencies far outside the inertial range, where the traditional approximation seems to work well, it can be justified only in a linear setting, since the higher frequency waves inevitably 'feel' the presence of the sub-inertial branches through nonlinear interactions (whatever the value of Ω/N).

The authors are grateful to Hans van Haren, Leo Maas and Uwe Harlander for comments and discussions. The work was supported by INTAS 01-0025 and INTAS 01-234.

Appendix A. Non-existence of normal vertical modes

The general solution (3.10) merits a special discussion, since the interpretation of its structure is less straightforward than on the traditional f-plane, especially with regard to the existence of normal modes. Even though the functions ψ_n are normal modes within the framework of the reduced eigenvalue problem (3.3), they cannot be treated as the vertical modes in the general solution (3.10), because they carry only part of the vertical dependence. Nevertheless, we can interpret each n-component in the series (3.10) as a mode (an independent way of being), as is clear from the fact that the constants a_n and b_n can be chosen arbitrarily. We now show that vertical 'normality' can be introduced only at the expense of the modal character. For simplicity the consideration is confined to the case of constant $N = N_c$.

We can isolate the full vertical dependence as follows. First we develop the 'non-traditional' z-dependent parts in Fourier cosine series, which are applicable here because we consider the half-range $-H \leqslant z \leqslant 0$ (see e.g. Butkov 1968):

$$\sin(\delta_n z) = \sum_{m=0}^{\infty} s_{n,m} \cos(q_m z) \qquad \cos(\delta_n z) = \sum_{m=0}^{\infty} c_{n,m} \cos(q_m z). \tag{A 1}$$

Expressions for the coefficients $s_{n,m}$ and $c_{n,m}$ can be easily obtained, but are not needed here. The series converge point-wise. We substitute the series in (3.10), and use $\psi_n = \sin q_n z$ with $q_n = n\pi/H$ (since $N = N_c$); applying trigonometric rules now yields

$$W = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} A_{n,m}(\chi)[\sin(q_n - q_m)z + \sin(q_n + q_m)z], \tag{A 2}$$

where

$$A_{n,m}(\chi) = \frac{1}{2}[a_n s_{n,m} + b_n c_{n,m}] \cos(k_n \chi) + \frac{1}{2}[a_n c_{n,m} - b_n s_{n,m}] \sin(k_n \chi).$$

One can re-arrange the double series (A 2) to obtain

$$\begin{aligned} W &= \sum_{k=1}^{\infty} \left\{ \sum_{l=1}^{k} A_{l,k-l} + \sum_{l=k}^{\infty} A_{l,l-k} - \sum_{l=k+1}^{\infty} A_{l-k,l} \right\} \sin(q_k z) \\ &\equiv \sum_{k=1}^{\infty} F_k(\chi) \sin(q_k z). \end{aligned} \tag{A 3}$$

Thus we have split off the vertical dependence in each component. Superficially, (A 3) looks like the modal decomposition one would find on the traditional f-plane. Yet, (A 3) is nothing more than a Fourier sine expansion, which here should not be taken for an expansion in terms of normal vertical modes. The reason is that the constants a_n and b_n are not associated with individual components $\sin(q_k z)$, but with all of them at once; the components are intrinsically coupled. Unlike on the traditional f-plane, we cannot select just one component – and this renders the very usage of the word 'mode' inappropriate.

We can summarize the situation by stating that we have either modes and no normality, as in (3.10), or normality and no modes, as in (A 3). Hence the conclusion that normal vertical modes do not exist. (Likewise, the atmospheric 'normal modes' in Thuburn, Wood & Staniforth 2002 are normal only with respect to a reduced eigenvalue problem, analogous to our (3.3).)

Finally, we note that one may not differentiate the Fourier sine series (A 3) term-by-term, as can be verified by noticing that the second vertical derivative of the actual solution (3.10) is non-zero at the boundaries, whereas the second derivative of (A 3) obviously is zero. This means that one cannot solve (2.9) by substituting (A 3) as an Ansatz (substitution would involve differentiation).

Appendix B. Modal decomposition in the full problem: orthogonality and completeness

Here we show that $W_n(\chi, z)$ are non-separable orthogonal eigenmodes of the 'full' boundary-value problem (2.9), (2.10), (2.11) and that the sequence with $n = 1, 2, \ldots, \infty$, forms a complete set.

By construction, $W_n(\chi, z)$ are eigenmodes of the full problem, which can be checked by direct substitution of $W_n(\chi, z) = \psi_n(z) \exp \mathrm{i}(k_n \chi + \delta_n + \phi_n)$ into (2.9), (2.10), (2.11). It is also straightforward to check that the functions $W_n(\chi, z)$ are orthogonal, with the weight function being proportional to $\Delta \equiv B^2 - AC$, that is

$$\int_{-\infty}^{\infty} \mathrm{d}\chi \int_0^H \mathrm{d}z [N^2(z) - r^2] W_n(\chi, z) W_m^*(\chi, z) = 0 \qquad \text{if} \qquad k_n \neq k_m \,, \tag{B 1}$$

where r^2 is a function of σ only and is specified in (3.14). Indeed, direct substitution yields

$$\int_{-\infty}^{\infty} \mathrm{d}\chi \mathrm{e}^{\mathrm{i}(k_n - k_m)\chi} \int_0^H \mathrm{d}z [(N^2(z) - r^2)\psi_n(z)\psi_m(z)] \mathrm{e}^{-\mathrm{i}(B/C)(k_n - k_m)z} = \tilde{c}_n \delta_{nm}, \tag{B 2}$$

where δ_{nm} is the Kronecker symbol and

$$\tilde{c}_n \equiv \left[\int_0^H (N^2(z) - r^2)\psi_n^2(z) \mathrm{d}z \right]. \tag{B 3}$$

Although the set of eigenfunctions $\psi_n(z)$ of the reduced boundary-value problem is complete by virtue of the known properties of the Sturm–Liouville problems, it is far from obvious whether the set $W_n(\chi, z)$ with $n = 1, 2, \ldots, \infty$ is complete or not. Moreover, it is not even obvious what the notion of the completeness means in this context. Apparently, we cannot expect to be able to expand an arbitrary function of χ and z, since the set of $W_n(\chi, z)$ implies that a particular real frequency σ has been chosen and the spatial manifold corresponding to such motions would be very peculiar and not of evident practical importance. In this context it would be more appropriate to consider either the extended problem which includes time explicitly, or the original boundary-value problem (2.8), (2.10), (2.11) as a problem of finding σ for a given k, rather than vice versa as we have done so far. However, along both routes there certain difficulties requiring special consideration which goes beyond the scope of the present paper.

Remaining within the framework of the original (χ, z) setting, we can state that if the boundary-value problem were Hermitian, then the completeness of the set would follow (e.g. Korn & Korn 2000, § 15.4.6). By definition, to assess this, one has to prove that

$$\int_{-\infty}^{\infty} \mathrm{d}\chi \int_0^H \mathrm{d}z [V^*(\hat{L}U) - U(\hat{L}V)^*] = 0, \tag{B 4}$$

where

$$\hat{L}W = AW_{\chi\chi} + 2BW_{\chi z} + CW_{zz}, \tag{B 5}$$

and U and V are any pair of functions satisfying the same homogeneous boundary conditions (2.10), (2.11), as well as yet unspecified conditions at the lateral boundaries. If we choose these conditions to be those of periodicity, say, set at $\chi = \pm l$, where l is arbitrary, or the condition of vanishing at plus and minus infinity, the criterion (B 4) holds and the boundary-value problem under consideration is indeed Hermitian. These conditions are those most commonly used, and at least for this class of functions the completeness has thus been proved. However, to be sure that this class indeed includes all functions of interest, solving the initial-value problem is probably the most sensible route.

Since we have shown that the eigenfunctions $W_n(\chi, z)$ are orthogonal and the set is complete, the principal incentive for making the traditional approximation has been removed.

REFERENCES

Abramowitz, M. & Stegun, I. A. 1965 *Handbook of Mathematical Functions*. Dover.

Badulin, S. I., Shrira, V. I. & Tsimring, L. Sh. 1985 The trapping and vertical focusing of internal waves in a pycnocline due to horizontal inhomogeneities of density and currents. *J. Fluid Mech.* **158**, 199–218.

Badulin, S. I., Vasilenko, V. M. & Yaremchuk, M. I. 1991 Interpretation of quasi-inertial motions using Megapoligon data as an example. *Izve. Atmos. Ocean. Phys.* **27**, 446–452.

Bjerknes, V., Bjerknes, J., Solberg, H. & Bergeron, T. 1933 *Physikalische Hydrodynamik*. Springer.

Brekhovskikh, L. M. & Goncharov, V. 1994 *Mechanics of Continua and Wave Dynamics*. Springer.

Butkov, E. 1968 *Mathematical Physics*. Addison-Wesley.

Dauxois, T. & Young, W. R. 1999 Near-critical reflection of internal waves. *J. Fluid Mech.* **390**, 271–295.

Eckart, C. 1960 *Hydrodynamics of Oceans and Dtmospheres*. Pergamon.

Fedoryuk, M. V. 1993 *Asymptotic Analysis*. Springer.

Fu, L. L. 1981 Observations and models of inertial waves in the deep ocean. *Rev. Geophys. Space Phys.* **19**, 141–170.

Garrett, C. & St. Laurent, L. 2002 Aspects of deep ocean mixing. *J. Phys. Oceanogr.* **58**, 11–24.

Gerkema, T. & Shrira, V. I. 2005 Near-inertial waves on the "non-traditional" β-plane. *J. Geophys. Res.*, **110**, C01003, doi:10.1029/2004JC002519.

Hendershott, M. C. 1981 Long waves and ocean tides. In *Evolution of Physical Oceanography* (ed. B. A. Warren & C. Wunsch), pp. 292–341. MIT Press.

Ince, E. L. 1956 *Ordinary Differential Equations*. Dover.

Kamenkovich, V. M. & Kulakov, A. V. 1977 On effect of rotation on waves in the stratified ocean. *Oceanologia*, **17** (3), 400–410 (in Russian). English Translation: *Oceanologia*, **17** (3), 260–266.

Korn, G. A. & Korn, T. M. 2000 *Mathematical Handbook for Scientists and Engineers*. Dover.

Kunze, E. 1985 Near-inertial wave-propagation in geostrophic shear. *J. Phys. Oceanogr.* **15**, 544–565.

LeBlond, P. H. & Mysak, L. A. 1978 *Waves in the Ocean*. Elsevier.

Maas, L. R. M. 2001 Wave focusing and ensuing mean flow due to symmetry breaking in rotating fluids. *J. Fluid Mech.* **437**, 13–28.

Miropol'sky, Yu. Z. 2001 *Dynamics of Internal Gravity Waves in the Ocean*. Kluwer.

Munk, W. & Phillips, N. A. 1968 Coherence and band structure of inertial motion in the sea. *Rev. Geophys.* **6**, 447–472.

Phillips, N. A. 1968 Reply *J. Atmos. Sci.* **25**, 1155–1157.

Phillips, O. M. 1966 *The Dynamics of the Upper Ocean*. Cambridge University Press.

Saint-Guily, B. 1970 On internal waves. Effects of the horizontal component of the earth's rotation and of a uniform current. *Deutsche Hydrograph. Z.* **23**, 16–23.

Thuburn, J., Wood, N. & Staniforth, A. 2002 Normal modes of deep atmospheres. II: f-F plane geometry. *Q. J. R. Soc.* **128**, 1793–1806.

Van Haren, H., Maas, L. & Van Aken, H. 2002 On the nature of internal wave spectra near a continental slope. *Geophys. Res. Lett.* **29**, 10.1029/2001GL014341.

Van Haren, H. & Millot, C. 2004 Rectilinear and circular inertial motions in the Western Mediterranean Sea. *Deep-Sea Res.* **51**, 1441–1455. 10.1016/j.dsr.2004.07.009.

Whitham, G. B. 1974 *Linear and Nonlinear Waves.* John Wiley & Sons.

Zauderer, E. 1989 *Partial Differential Equations of Applied Mathematics.* Wiley.