A Consistent Theory for Linear Waves of the Shallow-Water Equations on a Rotating Plane in Midlatitudes

NATHAN PALDOR
Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem, Israel, and Division of Meteorology and Physical Oceanography, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Miami, Florida

SHIRA RUBIN
Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem, Israel

ARTHUR J. MARIANO
Division of Meteorology and Physical Oceanography, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Miami, Florida

(Manuscript received 6 June 2005, in final form 6 April 2006)

ABSTRACT

The present study provides a consistent and unified theory for the three types of linear waves of the shallow-water equations (SWE) in a zonal channel on the β plane: Kelvin, inertia–gravity (Poincaré), and planetary (Rossby). The new theory is formulated from the linearized SWE as an eigenvalue problem that is a variant of the classical Schrödinger equation. The results of the new theory show that Kelvin waves exist on the β plane with vanishing meridional velocity, as is the case on the f plane, without any change in the dispersion relation, while the meridional structure of their height amplitude is trivially modified from exponential on the f plane to a one-sided Gaussian on the β plane. Similarly, inertia–gravity waves are only slightly modified in the new theory in comparison with their characteristics on the f plane. For planetary waves (which exist only on the β plane) the new theory yields a similar dispersion relation to the classical theory only for large gravity wave phase speed, such as those encountered in a barotropic ocean or an equivalent barotropic atmosphere. In contrast, for low gravity wave phase speed, for example, those in an equivalent barotropic ocean where the relative density jump at the interface is $10^{-3}$, the phase speed of planetary waves in the new theory is 2 times those of the classical theory. The ratio between the phase speeds in the two theories increases with channel width. This faster phase propagation is consistent with recent observation of the westward propagation of crests and troughs of sea surface height made by the altimeter aboard the Ocean Topography Experiment (TOPEX)/Poseidon satellite. The new theory also admits inertial waves, that is, waves that oscillate at the local inertial frequency, as a genuine solution of the eigenvalue problem.

1. Introduction

The shallow-water equations (SWE) provide the very fundamental description of the dynamics of an incompressible fluid that occupies a sufficiently thin layer such that the horizontal velocity is uniform across the layer’s height. Mathematically, the SWE are nothing but the 2D Euler equations for a compressible gas in which the pressure is quadratic with density, but with the density of the gas replaced by the fluid height. Linear waves of the SWE in the presence of rotation fall traditionally into two categories: The first is high-frequency waves (Kelvin waves and inertia–gravity, or Poincaré, waves), which represent rotationally modified gravity waves of the nonrotating SWE; the second type is the low-frequency planetary (Rossby) waves that originate from the dependence of Coriolis frequency on latitude, $f(y)$. The derivation of the former type in the context of the SWE is done straightforwardly on the f plane, where $f(y)$ is replaced by a constant $f_0$. In contrast, Rossby waves are derived on the β plane by making additional simplifying assumptions on the flow, for example, near nondivergence or quasigeo-
trophy, both of which are consistent with the smallness of \( f(y) - f_0 = \beta y \) relative to \( f_0 \) (Pedlosky 1979; Gill 1982; Cushman-Roisin 1994).

Although the dispersion relation of planetary (Rossby) waves can be derived directly from the linearized form of the conservation of potential vorticity, the meridional structure of the velocity and height eigenfunctions can only be derived by a perturbation expansion near a simple (steady, geostrophic) state. The small expansion parameter used for evaluating the eigenfunctions is \( \beta \), but both the deviation of the velocity from geostrophy and the velocity field’s divergence are assumed small to the same order. Both the dispersion relation of Rossby waves and the heuristic explanation for their westward propagation are based on vorticity conservation, so changes in the relative vorticity are essential for their existence [see Fig. 3.16.1 in Pedlosky (1979)]. However, the flow divergence in these waves is an essential physical element without which the velocity field is time independent, that is, geostrophic [see section 12.3 and Fig. 12.2 in Gill (1982)]. Thus, while Kelvin and inertia–gravity waves are derived directly from the SWE without making any assumption on the form of the solutions, Rossby (planetary) waves can only be derived by making some assumptions on the solutions. These assumptions limit the generality of the solutions and imply that each type of wave originates from a different physical setup that yields a different set of differential equations.

In the present study we provide a single theory that yields three types of waves—Kelvin, inertia–gravity (Poincaré), and planetary (Rossby)—straightforwardly from the SWE without making any additional assumptions. The new theory yields the following theoretical advances: (i) a derivation of the Kelvin and Poincaré wave solutions on the \( \beta \) plane, (ii) a consistent derivation of the Rossby wave solution that includes the variation of \( f(y) \) everywhere and does not let \( f = f_0 \) in some terms while \( \beta \neq 0 \) in others, and (iii) faster phase speed of Rossby waves in the range of parameters relevant to the ocean. The last result is in accordance with the altimeter observations made aboard the Ocean Topography Experiment (TOPEX)/Poseidon satellite, which show that Rossby waves in the thermocline of the ocean propagate westward faster than predicted by the classical theory (Chelton and Schlax 1996; Osychny and Cornillon 2004).

2. Linear waves of the shallow-water equations

and the eigenvalue equation

In vectorial form the linearized SWE with rotation are given by

\[
\frac{\partial \mathbf{V}}{\partial t} + f\mathbf{k} \times \mathbf{V} = -g \nabla \eta \quad \text{and} \quad \frac{\partial \eta}{\partial t} = -H \nabla \cdot \mathbf{V}. \tag{2.1}
\]

Here \( f \) is the Coriolis frequency, \( H \) is the unperturbed height (thickness) of the shallow layer of fluid, and \( \eta \) is the deviation of height \( h \) from \( H \) (i.e., \( h = H + \eta \)); \( \nabla \) is the two-dimensional nabla operator, \( \mathbf{V} \) is the two-dimensional (horizontal) velocity vector, \( t \) is time, and \( \mathbf{k} \) is a unit vector in the direction perpendicular to the velocity vector \( \mathbf{V} \); and \( g \) is the gravitational constant in barotropic cases and the reduced gravity (i.e., \( g' = g \Delta \rho / \rho_0 \)), where \( \Delta \rho / \rho_0 \) is the relative density difference between the lower and upper layers) in equivalent barotropic cases.

In Cartesian \((x, y)\) coordinates, where \( x \) \((y)\) is directed eastward \((\text{northward})\) for \( \mathbf{V} = (u, v) \), where \( u \) \((v)\) is the velocity component in the eastward \((\text{northward})\) direction, and for the linearly varying Coriolis parameter \( f(y) = f_0 + \beta y = 2\Omega (\sin \phi_0 + \cos \phi_0 y / R) \) (where \( \phi_0 \) is a mean latitude and \( \Omega \) and \( R \) are earth rotation frequency and radius, respectively), the scalar form of these vectorial equations is

\[
\begin{align*}
\frac{\partial u}{\partial t} &= (f_0 + \beta y)u = -g \frac{\partial \eta}{\partial x}, \\
\frac{\partial v}{\partial t} &= (f_0 + \beta y)v = -g \frac{\partial \eta}{\partial y}, \quad \text{and} \\
\frac{\partial \eta}{\partial t} &= -H \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right). \tag{2.2}
\end{align*}
\]

Since earth radius is the only length scale in the equations we take it to be the length scale in nondimensionalizing these equations. The time scale is \( (2\Omega)^{-1} \) and these length and time scales yield the velocity scale: \( 2\Omega R \). If, in addition, we scale the height, \( h \) and \( \eta = h - H \) by the mean height \( H \), then the nondimensional counterpart of system (2.2) is

\[
\begin{align*}
\frac{\partial u}{\partial t} &= \left[ \sin(\phi_0) + \cos(\phi_0) y \right] u = -\alpha \frac{\partial \eta}{\partial x}, \\
\frac{\partial v}{\partial t} &= \left[ \sin(\phi_0) + \cos(\phi_0) y \right] v = -\alpha \frac{\partial \eta}{\partial y}, \quad \text{and} \\
\frac{\partial \eta}{\partial t} &= \left. \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \right|_{\phi_0}. \tag{2.3}
\end{align*}
\]

where \( \alpha = gH(2\Omega R)^2 \) is the only parameter of the nondimensional equations that augments the four dimensional parameters: \( g \), \( H \), \( \Omega \), and \( R \). The reader is reminded that, although the variables in systems (2.2) and (2.3) are designated by the same symbols, they are dimensional in the former and nondimensional in the latter and that the nondimensional \( y \) coordinate measures the latitudinal distance from \( \phi_0 \) in radians. For the
rest of this work nondimensional parameters and variables will be used unless otherwise explicitly stated.

Anticipating linear wave solutions of system (2.3) we let \( u, v \) and \( \eta \) vary with \( x \) and \( t \) as a zonally propagating wave with wavenumber \( k \) and phase speed \( C \), that is, \( e^{ik(x-ct)} \). For this form of \((x,t)\) dependence in system (2.3) the \( x \) momentum equation yields \( u \) as a linear combination of \( V(y) = iv(y)/k \) and \( \eta(y) \),

\[
    u = \frac{\sin(\phi_0) + \cos(\phi_0)y}{C} V + \frac{\alpha}{C} \eta. \tag{2.4}
\]

Substituting this expression for \( u \) in the latter two equations of system (2.3) and rearranging the terms one gets the following two linear first-order ordinary differential equations:

\[
    \frac{dV}{dy} = -\sin(\phi_0) + \cos(\phi_0)y \frac{V}{C} V + \left( \frac{\alpha}{C} - C \right) \eta \tag{2.5a}
\]

and

\[
    \frac{d\eta}{dy} = \frac{k^2 C^2}{\alpha C} \left[ -\sin(\phi_0) + \cos(\phi_0)y \right]^2 V - \sin(\phi_0) + \cos(\phi_0)y \frac{\eta}{C}. \tag{2.5b}
\]

An important difference between the present theory and the classical quasigeostrophic (QG) theory should be pointed out: For sufficiently small \( C \) (e.g., Rossby waves) Eq. (2.5a) implies that \( |\eta| \sim O(f/\alpha) \sim 1 \) \( \{f = \sin(\phi_0) + y \cos(\phi_0)\} \), and Eq. (2.4) then yields \( u/V \sim O(1/C) \gg 1 \). In contrast, in the QG theory \( u/V \) is derived from the assumption of near nondivergence, so \( u \sim dV/dy \sim IV \) (\( I \) is the meridional wavenumber); that is, \( u/V \) is \( O(1) \) instead of \( O(C^{-1}) \gg 1 \).

System (2.5) has unique solutions only when boundary conditions, such as vanishing \( V(y) \) at two zonal walls, are specified. In this case, Eq. (2.5) and its associated boundary conditions constitute an eigenvalue problem for small amplitude waves that develop in a zonal channel on the \( \beta \) plane. The channel is centered at latitude \( \phi_0 \), which defines the mean Coriolis frequency \( \sin(\phi_0) \) (dimensionally \( 2\omega \sin(\phi_0) \)), and the channel walls are located at latitudes \( \phi_{\text{wall}} = \phi_0 \pm \delta \phi \), so the boundary conditions are \( V(y = \pm \delta \phi) = 0 \). The non-dimensional Coriolis frequency, \( \sin(\phi_0) + y \cos(\phi_0) \), varies linearly with \( y \) from a minimal value of \( \sin(\phi_0) - \delta \phi \cos(\phi_0) \) on the south wall to a maximal value of \( \sin(\phi_0) + \delta \phi \cos(\phi_0) \) on the north wall. Though system (2.5) can be formally applied to an infinite \( y \) domain, the neglect of higher-order terms in the expansion of \( \sin(\phi - \phi_0) \), as well as the metric terms of the spherical earth, is justifiable only in a bounded range of \( y \).

One can immediately notice that when \( C^2 = \alpha \) in system (2.5) the coefficient of the \( \eta \) term in Eq. (2.5a) vanishes, so \( V(y) \) can be solved independently of \( \eta(y) \) in Eq. (2.5b). Since the solution of the first-order equation for \( V(y) \) is exponential [see the solution for \( \eta(y) \) below], the vanishing of \( V(y) \) at the channel walls implies that \( V(y) \) has to vanish identically. Letting \( V(y) = 0 \) in Eq. (2.5b) yields a first-order equation for \( \eta(y) \) [that differs from (2.5a) without the \( \eta \) term only by the sign of \( C \)] that solves exactly into

\[
    \eta(y) = \eta_0 \exp \left[ \frac{-\sin(\phi_0) + \cos(\phi_0)y}{2C \cos(\phi_0)} \right]. \tag{2.6}
\]

Thus, for \( C = + (gH)^{1/2} \) the corresponding height amplitude \( \eta(y) \) decreases monotonically with \( y \); that is, the maximal height amplitude of this eastward-propagating wave is located on the channel’s southern wall. In contrast, for \( C = -(gH)^{1/2} \), the height \( \eta(y) \) increases monotonically with \( y \); that is, the maximal height amplitude of this westward-propagating wave is located on the channel’s north wall. This is the well-known pair of Kelvin waves whose phase speed is that of nonrotating gravity waves, while rotation only determines the variation of the height profile with \( y \): \( \eta(y) \). Equation (2.4) with \( V = 0 \) and (2.6) implies that, although the Coriolis frequency is \( y \)-dependent in both positive and negative modes the zonal velocity, \( u(y) \) is in geostrophic balance with the slope of the height amplitude \( d\eta/dy \):

\[
    u(y) = \frac{\alpha}{C} \eta = -\frac{\alpha}{\sin(\phi_0) + \cos(\phi_0)y} \frac{\partial \eta}{\partial y}. \tag{2.7}
\]

The \( C = 0 \) limit is another degenerate case of system (2.5) where both equations yield (after multiplying each of them by \( C \) and setting \( C = 0 \)) the exact same algebraic relation: \( f(y)V = -\alpha \eta \) [with \( f(y) = \sin(\phi_0) + y \cos(\phi_0) \)], which is known the geostrophic relation \( u(y) = \alpha / f(y) \partial \eta / \partial x \) for \( V = iv/k \) and \( \eta = (-ik) \partial \eta / \partial x \). The geostrophic relation for \( u \) is obtained by eliminating \( V \) from Eq. (2.5a) using Eq. (2.4) and setting \( C = 0 \).

For all values of \( C \neq \pm (gH)^{1/2} \) and \( C \neq 0 \) system (2.5) is nondegenerate [i.e., its \( V(y) \) and \( \eta(y) \) solutions are coupled], so the two first-order equations can be transformed to a single second-order equation. Taking the \( y \) derivative of Eq. (2.5a) and employing Eq. (2.5a) to eliminate \( \eta(y) \) and Eq. (2.5b) to eliminate \( d\eta/dy \) from the resulting equation yields

\[
    \frac{d^2 V}{dy^2} \left[ \frac{\cos(\phi_0)}{C} + k^2 \left( 1 - \frac{C^2}{\alpha} \right) + \frac{(\sin(\phi_0) + \cos(\phi_0)y)^2}{\alpha} \right] V = 0, \quad \text{with} \quad V(y = \pm \delta \phi) = 0. \tag{2.7}
\]
A generalization of this equation to continuously stratified oceans for solutions that are not necessarily zonally propagating waves is given in Eq. (5.15) of LeBlond and Mysak (1978). Owing to its complexity the generalized equation is solved for its vertical modes only while the dispersion relation of its horizontal modes is derived only for the case \( f(y) = f_0 \) (see below).

In classical linear wave theory the (nondimensional) Coriolis frequency in the last term on the left-hand side of Eq. (2.7), \( f(y) = \sin \phi_0 + y \cos \phi_0 \), is replaced by its value at the channel center, \( f_0 = \sin \phi_0 \) [i.e., by omitting the \( y \cos \phi_0 \) term from \( f(y) \)], in which case no coefficient in this equation is \( y \) dependent. In this (rather artificial) case, the solutions of the constant-coefficient equation satisfying the boundary conditions at \( y = \pm \delta b \) are given simply by \( V_n(y) = V_0 \sin(n+1)(y + \delta b)/(2\delta b) \) for \( n = 0, 1, 2, \ldots \) and for arbitrary \( V_0 \). Since for this solution \( V_{yy} = -(n+1)\pi/(2\delta b)^2 \) \( V \), the phase speeds \( C \) are given by the roots of the cubic

\[
\cos \frac{\phi_0}{C} + \frac{(n+1)^2 \pi^2}{4(\delta b)^2} + \frac{\sin \frac{\phi_0}{C}}{\alpha} - \frac{k^2}{\alpha} \cos \frac{\phi_0}{C} = 0,
\]

(2.8)

The dispersion relation for the (slow) Rossby waves is obtained from Eq. (2.8) by assuming \( C^{-1} \alpha \gg C^2 \), while the dispersion relation of the fast, inertia–gravity (Poincaré) waves is obtained by assuming \( C^2 \alpha \gg C^{-1} \). The resulting expressions of \( C(k; \alpha, \delta b, \phi_0) \) for these two waves are precisely the nondimensional counterpart of the dimensionless expressions found in textbooks on the subject, for example, Pedlosky (1979), Gill (1982), and Cushman-Roisin (1994). The dimensional form of Eq. (2.8) is given by Eq. (15.18) of LeBlond and Mysak (1978).

The goal of the present study is to extend the aforementioned classical theory to the case where the \( \beta y \) term (= \( y \cos \phi_0 \)) is not neglected in comparison with \( f_0 \) (= \( \sin \phi_0 \)). In this case, the coefficients of Eq. (2.7) are not constant, and the solutions, \( V(y) \), are not harmonic oscillations across the channel. However, a solution of Eq. (2.7) satisfying the boundary conditions at the channel walls will still yield the dispersion relation \( C(k; \alpha, \delta b, \phi_0) \) but as a more complex expression than Eq. (2.8).

Before solving Eq. (2.7) we first transform its independent variable \( y \) to \( z = y/(\delta b) \) so that the parameter \( \delta b \) (one-half the channel width) disappears from the boundary conditions (that are applied at \( z = \pm 1 \)) and appears explicitly in the differential equation. After some trivial rearrangement of Eq. (2.7), the differential equation and the corresponding boundary conditions can be written in terms of \( z \) as

\[
e^{-2z} \frac{d^2V}{dz^2} + \left[ E - (1 + b z)^2 \right] V = 0 \quad \text{and} \quad V(z = \pm 1) = 0,
\]

(2.9)

where the new nondimensional parameters are

\[

\varepsilon = \frac{\sqrt{\alpha}}{\sin(\phi_0) \delta b},
\]

(2.10a)

\[
b = \frac{\cos(\phi_0) \delta b}{\sin(\phi_0)}, \quad \text{and}
\]

(2.10b)

\[
E = -\frac{\alpha}{\sin^2(\phi_0)} \left[ \frac{\cos(\phi_0)}{C} + k^2 \left( 1 - \frac{C^2}{\alpha} \right) \right].
\]

(2.10c)

The parameter \( \varepsilon \) [Eq. (2.10a)] is the ratio between the (nondimensional) radius of deformation \( \alpha^{1/2}/\sin \phi_0 \) and the (nondimensional) channel half-width, \( \delta b \). From a mathematical viewpoint the value of \( \varepsilon \) is unrestricted, but for oceanographic applications (radius of deformation about 50 km and half-channel of over 100 km) it should be less than 0.05. The parameter \( b \) [Eq. (2.10b)] is the maximal relative change in \( f(y) \) across the channel, which should be less than 1 so as to ensure that the neglect of the higher-order terms in the expansion of \( \sin \phi_0 \) is justified. A solution of the Schrödinger equation (2.9) yields the eigensolution, made up of the eigenfunction \( V(z) \) and the associated eigenvalue \( E(\varepsilon, b) \) [defined in Eq. (2.10c)]. The dispersion relation, \( C(k; E) \), is determined from \( E(\varepsilon, b) \) by inverting Eq. (2.10c) to get a cubic \( C(E) \) relation for given \( \phi_0 \) [see Eq. (3.1) below].

The general solution of the differential equation [Eq. (2.9)] can be expressed as a linear combination of parabolic cylinder functions (see chapter 19 in Abramowitz and Stegun 1972). However, the eigenvalues are determined by applying the boundary conditions \( V = 0 \) to a linear combination of these functions, which is as complicated as constructing the solution by numerically integrating the equation. Below, we will solve the linear eigenvalue problem for \( V(z) \) and \( E \), Eq. (2.9), analytically for special values of \( b \) and numerically for general \( b \) and \( \varepsilon \) values, and deduce from these solutions the desired dispersion relation \( C(k, E) \). Before doing so we draw some qualitative consequences of the consistent formulation, Eq. (2.9).

3. Qualitative consequences of the consistent formulation

Since Eq. (2.9) is a Sturm–Liouville problem with \( p(z) = e^{-2z}, r(z) = 1 \), and \( q(z) = -(1 + b z)^2 \) [see section 1.8 in Bender and Orszag (1978) for notation and for details of the following brief discussion], Sturm’s theo-
rem ensures that it has an infinite number of eigen-solutions, \( V_n(z), E_n, n = 0, 1, 2, \ldots \). All of its eigenvalues, \( E_n \), are real positive with \( E_n \to \infty \) when \( n \to \infty \), and the associated eigenfunctions, \( V_n(z) \), have exactly \( n \) zeros between \( z = -1 \) and \( z = +1 \) [so \( V_n(z) \) has one sign throughout \(-1 < z < 1\)]. Each eigenvalue \( E \) yields three \( C(E) \) roots via Eq. (2.10c):

\[
k^2 \frac{C^3}{\alpha} - \left[ \frac{E \sin^2(\phi_0)}{\alpha} + k^2 \right] C - \cos(\phi_0) = 0. \tag{3.1}
\]

This equation determines \( C(k; E, \phi_0, \alpha) \) [where \( E = E[\alpha(\delta \phi, \phi_0, \alpha), b(\delta \phi, \phi_0)] \) is the eigenvalue of Eq. (2.9)] and is the counterpart of Eq. (2.8) of the classical, \( b = 0 \), theory. The difference between Eqs. (2.8) and (3.1) is that the former results from the application of the boundary conditions to the analytic solution of the differential Eq. (2.7) where the \( k \) term is neglected, while the latter results from the general features of solutions of the exact eigenvalue problem (2.9) without solving it (as it can only be solved numerically; see section 4).

The dispersion relation for the slow (i.e., Rossby, planetary) waves is given by the small \( C \) root of Eq. (3.1). An approximate expression for this small-\( C \) root is obtained by neglecting the \( k^2 C^3/\alpha \) term in Eq. (3.1) relative to the \( k^2 C \) term there (recall that \( C^2 \ll \alpha \) for Rossby waves). Solving for \( C \) one then gets

\[
C_n^{\text{Rossby}} = -\cos(\phi_0) \left[ k^2 + \frac{\sin^2(\phi_0)}{\alpha} E_n \right]. \tag{3.2}
\]

From this expression for the phase speed and from the fact that \( E_n \) is an increasing series with \( n \), it is clear that the first, \( n = 0 \), mode has the largest (in absolute value) phase speed.

The dispersion relation for the fast (Poincaré, inertia–gravity) waves, with \( C^2 > \alpha \), obtains easily from Eq. (3.1) by dropping the \( \cos(\phi_0) \) (i.e., \( \beta \)) term and dividing the resulting equation through by \( C \) (\( \neq 0 \)). One then gets the quadratic equation expression

\[
(C_n^{\text{Poincaré}})^2 = \alpha + E_n \frac{\sin^2(\phi_0)}{k^2} \tag{3.3}
\]

Since the phase speed of Poincaré waves is greater than \( \alpha^{1/2} \), while that of Rossby waves is less (in absolute value) than \( \alpha^{1/2} \) (i.e., these speeds are separated by the phase speed of the westward propagating Kelvin wave, \( C = -\alpha^{1/2} \)), Eqs. (3.2) and (3.3) provide fairly accurate approximations to the roots of the cubic equation (3.1).

In addition to these two wave types there are two degenerate cases of the consistent equation that were already highlighted in section 2—the steady, \( C = 0 \), solution and the Kelvin wave solution, \( C^2 = \alpha \). Both solutions appear as special cases of Eq. (3.1) when one lets

\[
Ck^2 \left( \frac{C^2}{\alpha} - 1 \right) = 0. \tag{3.4}
\]

However, this equation also implies, according to Eq. (3.1), that

\[
CE \frac{\sin^2(\phi_0)}{\alpha} = -\cos(\phi_0). \tag{3.5}
\]

which can be satisfied for \( C = 0 \) only on the \( f \) plane (where \( \beta = \cos(\phi_0) = 0 \)) and for \( C^2 = \alpha \) only by the negative Kelvin mode, \( C = -\alpha^{1/2} \) [where Eq. (3.5) is satisfied by \( E = \cos(\phi_0)\alpha^{1/2}/\sin^2(\phi_0) > 0 \)].

A solution of Eq. (2.9) includes, in addition to the eigenvalues \( E_n \), their associated eigenfunctions \( V_n(z) \), so the eigensolution is independent of \( C \) (which is derived from \( E_n \)). Therefore, the eigenfunctions \( V_n(z) \) are identical in the two waves, and for the same meridional wavenumber \( n \) the \( V_n(y) \) of Poincaré waves is precisely that of the Rossby wave. On the other hand, the \( u_n(z) \) and \( \eta_n(z) \) eigenfunctions are different for the two waves since they are related to \( V_n(z) \) by the phase speed \( C \), which is different for the two waves [see Eqs. (2.4) and (2.5a)]. These qualitative consequences are valid even though no assumption was made on the smallness of either the \( \beta \) term or the divergence or the ageostrophic velocity component.

4. Eigenvalues of Eq. (2.9) and the corresponding phase speeds

Although Eq. (2.9) is a Schrödinger equation, which has been studied extensively in theoretical physics, it has no known solutions for arbitrary \( \varepsilon \) and \( b \). The reason is that Eq. (2.9) applies on the finite interval \(-1 < z < 1\), where the term \((1 + bz)^2 \) (referred to as the harmonic oscillator potential) has no symmetry (Fig. 1). But at \( z = \pm 1 \) the boundary condition is \( V = 0 \), which corresponds to an infinite potential well at these points. For \( b = 0 \) and \( b = 1 \), \((1 + bz)^2 \) (the potential) is symmetric (Fig. 1) and analytic solutions can be found, while for \( 0 < b < 1 \) Eq. (2.9) can be easily integrated numerically from \( z = -1 \) to \( z = +1 \) (no singular points exist) so that the eigensolution can be found numerically.

a. Analytic solution for \( b = 0 \)

For \( b = 0 \) the differential Eq. (2.9) has constant coefficients so its eigenfunctions can be solved exactly and the eigenvalues \( E_n \) can be determined from these (purely oscillatory) solutions by applying the boundary conditions. The eigensolutions are then
Likewise, Eq. (3.3) yields the dispersion relation for inertia-gravity (Poincaré) waves,

\[
(C_{n}^{\text{Poincaré}})^2 = \alpha + \frac{E_n \sin^2(\phi_0)}{k^2} = \frac{\sin^2(\phi_0)}{k^2} + \frac{\alpha}{k^2} \left[ k^2 + \frac{(n + 1)^2 \pi^2}{(2\delta \phi)^2} \right].
\]  

(4.3)

The frequency associated with this phase speed, \(k^2 c^2\), is the Pythagorean sum of the inertial frequency \(\sin^2(\phi_0)\) and the gravitational frequency, \(\alpha \kappa^2\), where \(\kappa = [k^2 + (n + 1)^2 \pi^2/(2\delta \phi)^2]^{1/2}\) is the total (zonal and meridional) wavenumber.

b. Power series solutions for \(b > 0\)

For \(b > 0\) the solution of Eq. (2.9) can be written in the form \(V(z) = \theta(z)e^{(1 + bz)/(2\delta b)}\), where \(\theta(z)\) is a solution of the differential boundary value problem:

\[
\frac{d^2 \theta}{dz^2} - 2\left( \frac{1 + bz}{\epsilon} \right) \frac{d\theta}{dz} + \left( \frac{E}{\epsilon^2} - \frac{b}{\epsilon} \right) \theta = 0, \quad \theta(z = \pm 1) = 0.
\]  

(4.4)

Since Eq. (4.4) is regular for all \(\epsilon > 0\), it has a regular series expansion. The boundary condition \(\theta(z = -1) = 0\) suggests the following series expansion (obtained by a change of variables \(x = 1 + z\)):

\[
\theta(z) = \sum_{j=1}^{\infty} a_j (1 + z)^j,
\]  

(4.5)

where the series starts at \(j = 1\) (and not \(j = 0\)) to ensure that \(\theta(z = -1) = 0\). Substituting this series into Eq. (4.4) and equating like powers of \((1 + z)\) yields the recursion relation for \(\{a_j\}\):

\[
a_1 = 1; \quad (\text{this is a trivial normalization condition})
\]

\[
a_2 = \frac{1 - b}{\epsilon};
\]

\[
a_3 = \frac{b}{2 \epsilon} - \frac{E}{6 \epsilon^2} + \frac{2(1 - b)^2}{3 \epsilon^2};
\]

\[
\vdots
\]

\[
a_{j+2} = \frac{2(1 - b)(j + 1) a_{j+1} + \left[ b(2j + 1) - \frac{E}{\epsilon} \right] a_j}{\epsilon(j + 1)(j + 2)}; \quad j \geq 1.
\]  

(4.6)

From the series expansion it is clear that when two of its consecutive coefficients vanish the infinite series terminates at some \(j\) and becomes a polynomial in \((1 + z)\) (which is also a polynomial in \(z\)). The series can be employed for calculating the solution for \(\theta(z)\) and the parameter \(E\) can then be varied to find those values at
which \( \theta(z = +1) = 0 \), which is as efficient computationally as direct integration of the ODE [be it Eq. (4.4) or Eq. (2.9)] with a standard high-accuracy integration algorithm. However, the series expansion is helpful in finding analytic expressions for the eigenvalue problem in the special \( b = 1 \) case.

c. Analytic solutions for \( b = 1 \)

Although from a geometric viewpoint only \( b \) values smaller than \( \pi/4 \approx 0.79 \) (in fact for reasons detailed below \( b \leq 0 \) are acceptable, the \( b = 1 \) case is the only special case (other than \( b = 0 \)) where analytic solution of the eigenvalue problem exists, which confirms our numerical solution for general \( b > 0 \). For \( b = 1 \) the parabolic potential in Eq. (2.9), \((1 + z^2)z^2\), is symmetric about \( z = -1 \) on the \(-3 \leq z \leq 1 \) interval. The change of variables \( x = (1 + z)/2 \), which maps the \(-3 \leq z \leq 1 \) interval to the \(-1 \leq x \leq 1 \) interval, yields a classical symmetric potential about \( x = 0 \) where the eigenvalues are those of the Hermite equation (see Table 22.6 in Abramowitz and Stegun 1972) \( E_n = (2m + 1)e \). Since we are looking for eigenfunctions that vanish at \( x = 0 \) (i.e., \( z = -1 \)), only odd eigenfunctions of the symmetric (Hermite) equation are also solutions of the eigenvalue problem (2.9). Thus, setting \( m = 2n + 1 \) in the eigenvalues of the Hermite equation, \( E_n = (2m + 1)e \), yields the eigenvalues of our problem: \( E_n = (4n + 3)e \).

This simple result can be drawn directly from the power series expansion, Eq. (4.6), by noticing that for \( b = 1 \) all even-indexed coefficients, \( \{a_{2j}\} \), vanish [this reflects the symmetry of Eq. (2.9) about \( z = -1 \)], so the recursion relation for the odd-indexed coefficients is

\[
\begin{align*}
\alpha_0 &= 1; \\
\alpha_3 &= \frac{\alpha_0}{2e} - \frac{E}{6e^2} = \frac{3e - E}{6e^2}; \\
\vdots \\
\alpha_{2j+3} &= \frac{(4j + 3)e - E}{2j(2j + 1)e^2} \alpha_{2j+1}, \quad j \geq 0.
\end{align*}
\]

The eigenvalues \( E_n \) are determined by requiring that the infinite series in (4.7) becomes a polynomial of degree \( 2j + 1 \) (\( j = 0, 1, 2, \ldots \)) with odd powers of \((1 + z)\), which implies for \( E_n \):

\[
E_n = (4n + 3)e, \tag{4.8}
\]

when the mode index \( n \) is identified with the polynomial index \( j \). In particular, for \( n = 0 \),

\[
E_0 = 3e. \tag{4.9}
\]

Substituting the expression for \( E_n \), Eq. (4.8), into the dispersion relations of the two waves, Eqs. (3.2) and (3.3), yields the following approximate dispersion relations for \( b = 1 \):

\[
C_{n,\text{Rossby}} = -\cos(\phi_0) \left[ k^2 + E_n \frac{\sin^2(\phi_0)}{\alpha} \right] \\
= -\cos(\phi_0) \left[ k^2 + \frac{\sin(\phi_0)(4n + 3)}{(b \phi)^2 \sqrt{\alpha}} \right]
\]

and

\[
(C_{n,\text{Poincare}})^2 = \alpha + E_n \frac{\sin^2(\phi_0)}{k^2} = \alpha + \frac{\sin(\phi_0)(4n + 3)\sqrt{\alpha}}{(b \phi)^2}. \tag{4.11}
\]

These dispersion relations for \( b = 1 \) differ markedly from the corresponding relations for \( b = 0 \), Eqs. (4.2) and (4.3): The \( b = 0 \) eigenvalues, \( E_n(e, 0) \), are all larger than 1.0, Eq. (4.1), while those for \( b = 1 \), \( E_n(e, 1) \), are significantly smaller than 1.0 for sufficiently small \( e \) [Eq. (4.8)]. What is still unclear at this point is whether this decrease of the eigenvalues \( E_n(e, b) \) with \( b \) for fixed \( e \) is monotonic with the increase in \( b \) from 0. This question can be answered by solving the eigenvalue problem (2.9) numerically for general \( b \) values.

d. Numerical calculation of the eigenvalues for general \( b \)

A standard (shooting) method for solving the eigenvalue problem consists of integrating the differential equation from \( z = -1 \) (starting with \( V = 0 \) and, say, \( dV/dz = 1 \) there) to \( z = 1 \) (we used a fifth-order Runge–Kutta method with \( 10^{-10} \) tolerance) and varying the values of the parameters so as to satisfy the \( V(z = +1) \) boundary condition. Except for \( e = 0 \) the solution is regular, as can be verified by the expansion in section 4b, so the numerical integration yields a very accurate value of \( V(z = +1; E, e, b) \) [viz., the value of \( V(z = +1) \) for given values of the three parameters]. For fixed values of \( e \) and \( b \) we find (numerically) the roots of the nonlinear equation \( 0 = F(E) = V(z = +1; E, e, b) \). The resulting \( E_0(e, b) \) contours are shown in Fig. 2 from which it is easy to verify that the numerical solutions along the \( b = 0 \) and \( b = 1 \) ordinates are exactly those given analytically in Eqs. (4.1) (with \( n = 0 \) and (4.9), respectively. Two points should be now made with regard to the contours: The first is that for large \( e \) values the eigenvalue \( E_0 \) varies only slightly with \( b \), so \( E(e, b) \approx E(e, 0) \) so that \( C \), too, is close to its value in the classical, \( b = 0 \), theory. We have verified
(results not shown) that for \( \epsilon \) values larger than 0.6
\( (\epsilon = 2, 10, \text{and } 25) \) the \( E \) contours become even more
horizontal [i.e., the values of \( \delta E = E(\epsilon, 1) - E(\epsilon, 0) \)
decrease at larger values of \( E \)], so the near indepen-
dence of \( E \) on \( b \) for large \( \epsilon \) is not limited to the small
range of \( 0.3 < \epsilon < 0.6 \) shown in Fig. 2. The second point
is that at low values of \( \epsilon \) the slopes of the \( E \) contours are
all positive, which implies that for small fixed \( \epsilon \) an in-
crease in \( b \) results in a decrease in \( E \). This decrease of
\( E(b) \) for fixed and small \( \epsilon \) is drastic: at \( \epsilon = 0.05 \), for
example, \( E(b = 0.5)/E(b = 0) = 0.5! \)

A somewhat different view of the \( E_n(\epsilon, b) \) relation-
dship discussed up to this point is obtained by plotting
\( E_n(b; \epsilon) \), namely, by regarding \( \epsilon \) as a parameter in the
\( E_n(b) \) relationship. Figure 3 shows the resulting \( E_n(b) \)
curves for the indicated values of \( \epsilon \) in the interval \( 0 \leq b \leq 1 \).
The three panels clearly demonstrate that for small values of \( \epsilon \approx 0.6 \) (upper panel) the value of \( E_0 \)
decreases monotonically with \( b \) throughout the entire
\( 0 \leq b \leq 1 \) interval, while for large values of \( \epsilon \approx 0.75 \)
the value of \( E_0 \) increases monotonically but only slightly
with \( b \). However, in a narrow region of \( \epsilon \) near 0.7 the
variation of \( E_0(b) \) is not monotonic in the \( 0 \leq b \leq 1 \)
range.

e. Phase speed of Rossby waves

To apply the results obtained in the preceding sub-
section for the eigenvalues of Eq. (2.9) to observations
in the ocean, they have to be translated into estimates
for the phase speed of Rossby waves. This requires that
the three parameters of system (2.9)—\( E, \epsilon, \) and \( b \)—be
transformed to the five parameters of system (2.7): \( C, k, \delta \phi, \phi_0, \) and \( \alpha \). We thus fix \( \phi_0 \) at some (midlatitude)
value, so according to Eqs. (2.10a) and (2.10b) \( \delta \phi = \delta \tan \phi_0 \)
and \( \alpha = (eb \tan \phi_0 \sin \phi_0)^2 \). Any pair of values of
\( \epsilon \) and \( b \) determines \( E_n(\epsilon, b) \) (via Fig. 2), \( \alpha \) [via \( \alpha = (eb \tan \phi_0 \sin \phi_0)^2 \)], and \( \delta \phi \) (via \( \delta \phi = \delta \tan \phi_0 \)) from which the
dispersion relations \( C(k) \) are determined by numeri-
cally finding the three roots of (3.1).

The two panels in Fig. 4 compare the exact dispersion
curves, \( C(k) \), of Rossby waves for indicated values of \( \alpha, \delta \phi, \) and \( \phi_0 \) based on the same \( E_n(\epsilon, b) \) curves of Fig. 2.
The point of these graphs is that \( C \) in the new theory is
2 times that in the classical, \( b = 0 \), theory for a “North
Pacific” channel that lies between 11.5° and 51.5° (i.e.,
\( \phi_0 = 31.5° \) and \( \delta \phi = 20° \); upper panel) for \( (gH)^{1/2} = 3 \)
m s\(^{-1} \). As expected from the results of Fig. 2 for the
same values of \( \phi_0 \) and \( \delta \phi \), the error of the \( b = 0 \) theory
decreases with the gravity wave speed \( \alpha (\sim \epsilon^2) \), so in the
atmosphere where \( (gH)^{1/2} = 30 \) m s\(^{-1} \) (lower panel) \( C \)
of the classical theory is close to that of the new theory.
As can be expected, larger errors are encountered for
larger \( \delta \phi \) (results not shown).

f. Inertial waves

A class of waves that exists in rotating fluids but is of
lesser relevance to the ocean (or the atmosphere) is
inertial (also called gyroscopic) waves. These waves
exist even when the pressure gradient vanishes iden-
tically, so on the \( f \) plane the wave frequency is the
Coriolis frequency; that is, \( k^2 C^2 = \sin^2 \phi_0 \) in the present
way for expressing the eigenfunction $V(z)$ for given values of $E$, the summation of a large number of terms is less efficient numerically than a straightforward integration of Eq. (2.9). The summation suffers from the slow convergence of the series at $1 + z = 2$ ($z = 1$), which is not guaranteed for all values of $E$, $b$, and $\varepsilon$ with a fixed number of terms.

In solving the eigenvalue problem Eq. (2.9) numerically we use the fact that the differential equation and its associated boundary conditions are homogeneous, so the amplitude of the eigenfunctions is undetermined and the normalization of the eigenfunctions is arbitrary. The eigenfunction that corresponds to an eigenvalue $E_n$ is found by integrating the differential equation (2.9) from $z = 1$ with $V = 0$ and $dV/dz \neq 0$ (the value is arbitrary) to $z = +1$, and the choice of $E_n$ for $E$ guarantees that $V(z = +1) = 0$. The accuracy of the fifth-order Runge–Kutta scheme was determined a priori to a relative accuracy of $10^{-10}$. The corresponding height and zonal velocity eigenfunctions [i.e., $\eta(z)$ and $u(z)$] are then easily obtained by substituting the numerically found solution for $V(z)$ and $dV(z)/dz$ into Eqs. (2.5a) and (2.4), respectively. One should only replace $y$ in these expressions by $\varepsilon \phi$ and select the value of $C$ that is relevant to the particular wave, either planetary (Rossby) or inertia–gravity (Poincaré). As was noted above, these two waves share the same $V(z)$ eigenfunctions.

Based on the horizontal shape of the $E$ contours in Fig. 2 for sufficiently large $\varepsilon$ one expects the classical, $b = 0$, theory [Eq. (4.1) and section 4a] to provide an accurate approximation for the eigenfunctions for sufficiently small $b$. For fixed values of $k$ and $\phi_0$ (the results shown below are for $k = 1.0 = \phi_0$), the parameters of the eigenvalue problem—$E_n$, $\varepsilon$, and $b$—determine uniquely the values of $C$, $\alpha$, and $\delta \phi$ via the relations (2.10). For large $\varepsilon$ ($\varepsilon = 2; b = 0.1$), where $E_0 = 10.871$, Eq. (4.1) provides an excellent estimate for the eigenvalue, $E_0 = 1 + \pi^2 = 10.870$, so the relative error in $E_0$ of the $b = 0$ theory is only $10^{-5}$ and $C_{\text{Rossby}} = -0.00478$, which is the same value (to three significant digits) as in the classical theory.

In accordance with this accurate estimate of the eigenvalue, the structure of the associated $V(y)$ eigenfunction shown in the upper-left panel of Fig. 5 clearly confirms the predicted structure of the $b = 0$ theory: a pure sinusoidal variation that vanishes only at the boundaries. However, even in this case the $u(y)$ and $\eta(y)$, which vary across the channel in a nearly sinusoidal manner, are not identical with their $b = 0$ counterparts. For Rossby waves (upper-right panel) $\eta(y)$ does not vanish at the boundaries, in accordance with Eq. (2.5a), while $u(y)$ is not exactly $90^\circ$ out of phase relative
to $V(y)$ and, as anticipated in section 2, its amplitude is $O(1/C) \sim 200$ times the amplitude of $V$ (instead of $1/\delta \phi \sim 7$). Similar changes occur for $u(y)$ and $\eta(y)$ of the Poincaré wave (lower-left panel), where a close inspection shows that they are not exactly antisymmetric with $y$. The Gaussian Kelvin wave (lower-right panel) in the present theory differs only slightly from the exponent of its $b = 0$ theory, and most of the difference occurs in the center of the channel owing to the $\delta \eta$ theory and most of the difference occurs in the center of the channel owing to the $\eta(y = \delta \phi) = 1$ normalization and the exponential decay of $\eta(y)$ with distance from the wall.

In contrast, for small $\epsilon$ ($\epsilon = 0.055, b = 0.15$), where $E_0 = 0.862 (C_{Rosby} = -0.000103)$, the $b = 0$ theory yields [Eq. (4.1)], $E_0 = 1 + 0.0557\pi/4 = 1.0075$; that is, the eigenvalue is in significant error. It is, thus, not surprising that the structure of the associated eigenfunction, $V(y)$, is also far from purely sinusoidal as anticipated by the classical, $b = 0$, theory (upper-left panel in Fig. 6). In fact, it is easy to show that the term $E - (1 + bx^2)$ in Eq. (2.9) is negative for $z > (E^{1/2} - 1)/b$ and positive for $-1 < z < (E^{1/2} - 1)/b$. Therefore, for the present values of $E (=0.862)$ and $b (=0.15)$, $V$ should decay exponentially for $z > -0.48$ [i.e., $y > -0.48 \tan(1) = -0.11$] and oscillate for $-1 < z < -0.48$ [i.e., $-b \tan(1) < y < -0.11$]; thus, even near symmetry about $y = 0$ should not be expected. This asymmetry is expected in light of the expansion of $V(x)(z)$ in section 4b into a symmetric Gaussian about $1 + b^2 = 0$ times a power series in $(1 + z)$, neither of which is symmetric about $z = 0$. The $\eta(y)$ and $u(y)$ functions of the Rossby (upper-right panel) and Poincaré (lower-left panel) waves follow from $V(y)$ with the different values of $C$ in Eqs. (2.4) and (2.5a). The exponent of the Kelvin wave (lower-right panel) undergoes a much more drastic change across the channel than in Fig. 5 because of the small radius of deformation $\alpha = 0.000117$, so $\alpha^{1/2}/\sin(1) = 0.129$ as compared with the relatively wide channel $2\delta \phi = 2b \tan(1) = 0.46$.

In the classical linear wave theory Rossby waves can be derived from the vorticity equation and their existence relies upon the $B$ term, so they have no counterpart in nonrotating fluids. To solve the problem the divergence field is assumed small (but not zero) and this wave is therefore considered quasi nondivergent. To compare our theory (which derives from the SWE without imposing any assumption on the divergence)
with the classical theory we need to calculate the divergence ($\delta$) and vorticity ($\zeta$) associated with the velocity fields in our theory. The definitions of these variables and the relation $\nu = -ikV$ imply

$$\delta = \frac{du}{dx} + \frac{dv}{dy} = ik(u - V_y) \quad \text{and}$$

$$\zeta = \frac{dv}{dx} - \frac{du}{dy} = k^2V - \frac{du}{dy},$$

which upon substituting Eqs. (2.4) and (2.5a) yields

$$\frac{\delta}{ik} = -\frac{fC}{\alpha - C^2}V + \frac{C^2}{\alpha - C^2}V_y \quad \text{and}$$

$$\zeta = -\left(\frac{\beta}{C} + \frac{f^2}{\alpha - C^2}\right)V + \frac{fC}{\alpha - C^2}V_y, \quad (5.1)$$

where $f = \sin \phi_0 + y \cos \phi_0$ and $\beta = \cos \phi_0$. Upon substituting $\delta/(ik)$ into $\zeta$, PV conservation yields

$$ikC\zeta - \beta\nu - f\delta = 0. \quad (5.2)$$

The divergence and vorticity curves shown in Fig. 7 and Fig. 8 correspond to the parameters and eigensolutions of Fig. 5 and Fig. 6, respectively. It is evident that, for both large (Fig. 7) and small (Fig. 8) values of $C$, the divergence field of Rossby waves is nearly negligible everywhere relative to the vorticity field of these waves. This is a result of the low $C$ value of these waves, which according to the continuity equation makes the divergence field small to order $C$, or equivalently Eq. (5.2).
implies that $\delta \zeta \sim C$. As is evident from Eq. (5.1), $\delta(y)$ is approximated by $Cf_b(y)z\xi + O(C^2)$ for Rossby waves, which supports the numerical results in the upper panels of Figs. 7 and 8. Since for $n = 0 \ V(y)$ does not vanish inside the channel, it is clear that for near-internal points where the vorticity vanishes (where its sign changes) the divergence is larger than the vorticity, so $\delta \zeta$ is not uniformly small throughout the channel.

For Poincaré waves (the bottom two panels in Figs. 7 and 8), where $C$ is $O(1)$, $\delta$ is of the same order as $\zeta$, so these waves cannot be considered nonrotational (as are gravity waves in nonrotating systems). In Kelvin waves, where $v(y) = 0 = V(y)$, the vorticity and divergence vary across the channel as $-du/dy$ and $u$, respectively.

6. Concluding remarks and summary

The unified formulation for all linear waves on the midlatitude $\beta$ plane advocated in this study was first suggested for Rossby waves in the $b = 0$ case by Lindzen (1967). A similar unified formulation was developed by Matsuno (1966) for the equatorial $\beta$ plane, which was modified to an eigenvalue problem similar to Eq. (2.9) by Erlick et al. (2007, manuscript submitted to Quart. J. Roy. Meteor. Soc., hereinafter EPZ). The equatorial eigenvalue equation can be obtained directly from Eq. (2.9) by multiplying it through by $\sin \phi_0$ and then substituting $\phi_0 = 0$. When the resulting equation is divided through by $e^y$ the $\sin \phi_0$ factors in $E$ and $b$ disappear, and the potential $(1 + b z)^2$ in Eq. (2.9) becomes $\delta^2 z^2$ with $\delta = (\delta \phi)^2/\alpha^{1/2}$. The entire problem now rests upon solving the one-parameter Schrödinger eigenvalue equation $E(\delta)$ (where $E^\ast$ is the modified eigenvalue). It turns out that in the present formulation the analysis of the equatorial case is much simpler than that for the midlatitude case. More details on the application of the Schrödinger equation approach to the equatorial $\beta$ plane are given in EPZ.

Kelvin waves are traditionally derived on the $f$ plane by letting $V(y) = 0$. The present study shows that they exist on the $\beta$ plane with the same dispersion relation, $C^2 = gh$, which can be anticipated since $f$ does not appear in the dispersion relation. The $\eta(y)$ eigenfunction is trivially modified on the $\beta$ plane by replacing $f_0$ by $f_0 + \beta y$ [i.e., $\sin \phi_0$ by $\sin \phi_0 + y \cos \phi_0$] in the first-order equation (2.5b) with $V = 0$. This yields a Gaussian decay of $\eta(y)$ with distance from the channel walls, Eq. (2.6), while on the $f$ plane $\eta(y)$ decays exponentially: $e^{-\sin(\phi_0)y/C}$.

In the classical theory both Kelvin and Poincaré (inertia–gravity) waves are derived on the $f$ plane without assuming any limitation on the flow (e.g., OG, near nondivergence), while Rossby waves are derived on the $\beta$ plane where even the geostrophic flow is divergent. The dependence of $f$ on $y$ on the $\beta$ plane introduces a nonconstant coefficient into the governing equations, and consequently Rossby waves can only be derived by making some simplifying assumptions, such as that the flow is nearly nondivergent (but the small divergence is essential for its dynamics). Thus, Rossby waves are classified as nearly nondivergent, while the former two waves are classified as nonrotational. In the present theory all three waves are derived in a single consistent $\beta$-plane theory starting from identical equations without making any assumptions on the nature of the solutions to these equations.

Inertia–gravity (Poincaré) waves are also classically developed only on the $f$ plane, and the present study shows both that they also exist on the $\beta$ plane and that they originate from the exact same equations as the Rossby waves. This derivation of the dispersion relations of inertia–gravity waves and planetary (Rossby) waves as different roots of the same solution of the eigenvalue equation, (2.9), attributes their rotational and divergence characteristics to their different phase speeds. According to the continuity equation the velocity divergence is order $C$, since $\nabla \cdot \mathbf{V} = -\partial \eta/\partial t = i k C \eta$, so for the fast inertia–gravity waves [in which the phase speeds are given by the large $C$ roots of Eq. (3.1)] the velocity divergence is also large while the opposite holds for the slow planetary waves where $C$ is the smallest root of Eq. (3.1) and the velocity divergence is also small.

Perhaps the most significant result of the new derivation proposed in this study has to do with Rossby waves. Previously, these waves were only derived by perturbation expansion procedures in which $\beta = df/dy$.
is treated as a small, but nonzero, parameter of the governing equations (either the SWE or the vorticity equation), while at the same time \( f(y) \) is replaced by \( f_0 \) (i.e., \( \beta y \) is neglected relative to \( f_0 \)) everywhere else in these equations. These two conflicting assumptions are made so as to include the \( \beta \) effect in the equations but, at the same time, leave the coefficients in these equations constant to ensure that they can be solved analytically. In the present theory \( f(y) = f_0 + \beta y \) is assumed everywhere, and both Rossby and inertia–gravity waves obtain from different roots of the same eigenvalue equation. The \( V(y) \) eigenfunctions in the classical perturbation theory are all pure trigonometric functions that vanish on the boundaries, while in the new theory these functions have a much more complex form (e.g., Hermite functions) that changes with the values of the model parameters. The analysis presented in this study focuses on the first eigenvalue, \( E_0 \), but applies also to the higher eigenvalues, that is, \( E_n(e, b > 0) > E_0(e, b = 0) \) nearly everywhere. However, since \( C_{Rossby}^{n} \) decreases with \( n \) [see Eq. (3.2)], the effect of letting \( b > 0 \) is strongest for the \( n = 0 \) mode.

Inertial waves, whose frequency of oscillation is \( f_0 \), appear in the present theory as the regular solutions of the eigenvalue problem in the range \( \alpha \sim eb \rightarrow 0 \) and \( E \approx 1 \). This is in contrast to the classical theory where their dispersion relation is that of inertia–gravity waves in the limit \( g = 0 \), but the corresponding eigenfunction cannot satisfy the boundary conditions.

The limitation on the meridional length scale in the present theory is the same as that of the classical \( \beta \)-plane theory, and the higher accuracy of the present theory is obtained only by retaining the \( \beta y \) term in system (2.2). Since \( \beta y \) is neglected in the classical theory, the error there in \( f(y) \) is \( O(\beta y/f_0) \), while in the present theory this error is \( O(\beta y/f_0)^2 \). Thus, to ensure a 10%–20% accuracy in the classical theory one has to require \( \beta y/f_0 < 0.15 \) or \( \text{max} |y|/R = \delta b < 0.15 \tan \phi_0 \), while in the present theory this accuracy is achieved for \( \text{max} |y|/R < 0.4 \tan \phi_0 \); that is, \( \delta b \cot \phi_0 = b < 0.4 \). So, at \( \phi_0 = \pi/4 \) the maximal length scale is increased to 2500 km as compared with 1000 km in the classical theory. These considerations apply to the two planar models where terms associated with the spherical geometry are ignored, which can modify the present planar theory at \( \delta \phi \approx R \). In the future we plan to extend the same method to the spherical earth, taking full account of the latitude dependence of the metric terms [e.g., \( v \tan \phi \) in the continuity equation] and the metric coefficients [e.g., \( d/dx = (1/\cos \phi) d/d\lambda \), where \( \lambda \) is longitude].

The above comparison between the accuracy of the new and classical, \( \beta \)-plane theory focuses on the relative size of similar terms in the governing equations. However, comparing solutions of the two theories (either the contours of Fig. 2, the dispersion curves in Fig. 4, or the eigenfunctions in Figs. 5 and 6), one realizes that the modification to the solutions of the classical theory by the present theory is significantly different from the naïve predictions from the value of \( \delta b \) (or \( b \)) only, and for large \( e \) (i.e., \( gh \) or \( \alpha \)) the modification is negligible even for \( b = 1 \). On the other hand, for \( b = 0.3 \) (\( \delta b = 0.17 \) at \( \phi_0 = \pi/6 \) and \( e = 0.05 \) (deformation radius, \( Ra^{1/2}/\sin \phi_0 = 55 \text{ km} \) and channel of half-width \( \delta b R = 1100 \text{ km} \)) the eigenvalue in the present theory is \( E_0 = 0.7 \), while in the classical theory \( E_0 \) always exceeds 1.0. At \( k < 10^3 \) [so that \( k^2 \ll \sin^2 \phi_0 (E/\alpha) \) in Eq. (4.2)] the phase speed of Rossby waves is well approximated by \( \mathcal{C} \approx -\alpha \cos \phi_0 / (E \sin^2 \phi_0) \), so the relative error in \( \mathcal{C} \) equals that in \( E_0 \), that is, both errors are about 50%, though \( \delta \phi \) is 0.17 only (and \( b = 0.3 \)).

The last point we wish to make regards the possible application of the present theory results to observation of the westward propagation of Rossby waves. The dimensional form of the dispersion relations (3.2) and (3.3) is

\[
\begin{align*}
-kC_{Rossby}^n &= -\frac{\beta k}{k^2 + E_n R_d^2} \\
(kC_{Poincaré}^n)^2 &= ghk^2 + E_n f^2,
\end{align*}
\]

where \( R_d = f/(gh) \) is the radius of deformation. The channel width and meridional wavenumber are implicit in \( E_n \), so prior to comparing the phase speed of Rossby waves with observations (e.g., TOPEX/Poseidon) one has to find an explicit expression for \( E_n \). A theoretical estimate of this expression is currently under way. It is also clear from Eq. (6.1) that the speeding up of \( C_{Rossby} \), which is achieved in the present theory by decreasing \( E_n \) (Fig. 2), is achieved in the classical theory by reducing \( f \) (or \( \phi_0 \)). The reduction of \( \phi_0 \) at small \( e \) (\( \sim gh \)) in the phase speed is commensurate with the southward shift of the latitude where \( \eta(y) \) (the SSH signature) is maximal (so it is detected by the satellite) from \( \phi_0 \) toward the south wall when \( gh \) gets smaller [cf. the \( \eta(y) \) panels for Rossby waves in Figs. 5 and 6].

Acknowledgments. This work was supported by Grant 579/05 of the Israel Science Foundation to HU. We thank two anonymous reviewers for their constructive comments. We are indebted to Dr. A. Sigalov of HU and to Dr. P. Killworth for instructive discussions and for pointing out errors in earlier versions of the manuscript.
REFERENCES