Equatorial Solitary Waves. Part I: Rossby Solitons

JOHN P. BOYD

Department of Atmospheric and Oceanic Science, The University of Michigan, Ann Arbor 48109

(Manuscript received 31 October 1979, in final form 20 June 1980)

ABSTRACT

Using the method of multiple scales, I show that long, weakly nonlinear, equatorial Rossby waves are governed by either the Korteweg-deVries (KDV) equation (symmetric modes of odd mode number \(n\)) or the modified Korteweg-deVries (MKDV) equation. From the same localized initial conditions, the nonlinear and corresponding linearized waves evolve very differently. When nonlinear effects are neglected, the whole solution is oscillatory waves which decay algebraically in time so that the asymptotic solution as \(t \to \infty\) is everywhere zero. The nonlinear solution consists of two parts: solitary waves plus an oscillatory tail. The solitary waves are horizontally localized disturbances in which nonlinearity and dispersion balance to create a wave of permanent form.

The solitary waves are important because 1) they have no linear counterpart and 2) they are the sole asymptotic solution as \(t \to \infty\). The oscillatory wavetrain, which lags behind and is well-separated from the solitary waves for large time, dies out algebraically like its linear counterpart, but the leading edge decays faster, rather than slower, than the rest of the wavetrain. Graphs of explicit case studies, chosen to model impulsively excited equatorial Rossby waves propagating along the thermocline in the Pacific, illustrate these large differences between the linearized and nonlinear waves. The case studies suggest that Rossby solitary waves should be clearly identifiable in observations of the western Pacific.

1. Introduction

Equatorial waves have been the subject of dozens of papers in meteorology and oceanography over the past 15 years, but until recently, all treatments have been linear. Domaracki and Loesch (1977) and Loesch and Deininger (1979) have studied resonant triad interactions. Individual waves (as opposed to triads), however, can still be profoundly altered by weak nonlinearity even when triad resonance is unimportant. This present work is the second in a three-part series whose purpose is to pioneer the theory of individual nonlinear equatorial waves.

The first part (Boyd, 1980) treated Kelvin waves, because they are nondispersive, cannot form solitary waves and require different mathematical methods from those used here.

This second part will analyze weakly dispersive equatorial Rossby waves. As noted in the Abstract, such modes give rise both to solitary waves and to oscillatory wavetrains, and both parts are drastically altered by even weak nonlinearity. The solitary waves, however, are of special interest because they are completely missed by linear theory and are much more easily observable in the ocean than the nonlinear modifications to the wavetrain.

Solitary waves, in general, have been known for a century and a half, but it is only in the last 15 years that they have been understood. The reason for this is what one may call the “soliton paradox.” Since the solitary wave requires an exact balance between nonlinearity and dispersion, it was, in the words of Scott et al. (1973) “long considered a rather unimportant curiosity. . . . Since it is clearly a special solution to the partial differential equation, many have assumed that somewhat special initial conditions would be required to launch it and, therefore, that its role in relation to the (general) initial value problem would be a minor one at best.” The paradox is that the solitary wave, in fact, was first discovered in nature by a Scottish engineer in 1834, who observed one generated by a canal boat; by the mid-1960’s, dozens of other examples had been seen in nonlinear optics, plasma physics, and a score of other fields. If solitary waves are mere “curiosities,” why were they so readily observed in nature?

The “soliton paradox” was resolved by Gardner et al. (1967), who showed that a localized but otherwise arbitrary initial value disturbance for the Korteweg-deVries (KDV) equation will generate 1) a conventional oscillatory wavetrain plus 2) a finite number of solitons. Because the oscillatory wavetrain rapidly destroys itself by dispersion, however, the asymptotic solution for large \(t\) consists of solitary waves alone. Further, it was shown that it is easy to create solitary waves; indeed, rather special initial conditions are necessary to not generate at least one soliton.
There is little doubt that when John Scott Russell discovered solitary waves in 1834, the canal boat generated a rich spectrum of waves. Because the wavetrain dispersed within a few feet of the bow, however, he was able to observe the soliton in its pristine state as a “rounded, smooth, and well-defined heap of water” which he was able to follow on horseback for a couple of miles before he lost it in the windings of the canal.

Thus, solitary waves are not special solutions; rather, as \( t \to \infty \) they are the only solutions. In addition, the inverse scattering method shows that when solitary waves collide, they emerge from the collision unchanged in shape or speed. Thus, in the absence of dissipation, solitons are in principle immortal.

This longevity has inspired intensive research: the review by Scott et al. gives 267 references to solitary waves in such diverse fields as water waves, nonlinear optics, elementary particle theory, plasma physics and half a dozen others. Recent reviews by Miles (1980) and Redekopp (1981) summarize a vast number of papers in geophysics alone. Thus, the notion that solitary waves are mere “curiosities” is no longer tenable.

On top of this, there are three reasons that make the study of equatorial solitary waves particularly appealing in comparison to that of other geophysical solitons that have been analyzed in the past. First, mathematical simplicity: Hermite functions give the latitudinal structure here in contrast to the general parabolic cylinder functions of Clarke (1971) and the nonlinear critical-layer analysis of Redekopp (1977) and Brown and Stewartson (1979). Second, latitudinal localization: equatorial solitons are confined to a narrow band about the equator by Coriolis forces, whereas the midlatitude solitary waves of Redekopp (1977) are latitudinally trapped by mean winds, which are much more fickle and hard to analyze. Third, excitation: equatorial solitons would arise naturally in a nonlinear generalization of McCreary’s (1976) oceanic model of El Niño, for example, but most previous work on geophysical solitons has offered no suggestion as to how such solitary waves might actually be generated in the real atmosphere or ocean.

Thus, there is ample motivation for the present investigation even though equatorial solitary waves have not yet been identified in nature.

The plan of the work is as follows: Section 2 is an introduction to the three-part series as a whole, explaining how the analysis of this paper fits into that of the other two. A derivation of the KDV equation for Rossby waves of odd mode number and the MKDV equation for Rossby waves of even mode number is given in Sections 3 and 4, respectively. In Section 5, which is a brief review cobbled together from widely scattered works, a summary is given of those properties of the general inverse scattering solution which are helpful in understanding the numerical case studies presented in Sections 6 and 7 for the KDV and MKDV equations, respectively. In Section 8 a discussion is presented of what happens when nonlinearity initially dominates dispersion so that the Rossby waves may be treated as approximately nondispersive, at least for a time. Resonant triad interactions and their relationship to the present work are presented in Section 9 while Section 10 is a summary and prospectus.

2. An overview of nonlinear equatorial waves

Nonlinear equatorial waves can be divided into three broad classes: nondispersive, weakly dispersive and strongly dispersive. In the strict sense, “nondispersive” means that all waves of a given mode have the same phase speed \( c \) independent of the zonal wavenumber as is true of the Kelvin mode. Here, I will also apply the term to ultralong Rossby waves for which dispersion is very weak in comparison to the nonlinearity. Waves which are “nondispersive” in this broad sense cannot form solitary waves because the dispersion, by assumption, is too small to balance nonlinearity to make a wave of permanent form possible. Instead, an inviscid localized packet of nondispersive waves will steepen and eventually break. A thorough treatment of the nonlinear Kelvin wave is given in the first part of this series (Boyd, 1980).
In the sense that I will use it here, “weak dispersion” means that two waves of a given mode whose zonal wavelengths differ by O(1) will have phase speeds which differ by only O(ε), where ε ≪ 1. The nonlinear evolution of such waves is governed by either the Korteweg-deVries (KDV) equation

\[ u_t + uu_x + u_{xxx} = 0 \]  

(2.1)

or the modified Korteweg-deVries (MKDV) equation

\[ u_t + u^3u_x + u_{xxx} = 0 \]  

(2.2)

depending on the latitudinal mode number as explained in Section 4, where the subscripts denote differentiation with respect to the subscripted variable. Since the dispersion is weak, it can be directly balanced against weak nonlinearity (for both equations) to create a wave of permanent form, i.e., a soliton. Because of this standoff between opposing tendencies, the solitary wave will neither steepen due to nonlinearity nor spread out due to dispersion but will instead propagate without change of shape. The solitons are mathematically described by

\[ u = 3A \text{sech}^2[\frac{1}{2}A^{1/3}(x - c_0t - At)] \]  

(KDV) (2.3)

\[ u = \pm 6^{1/2}A \text{sech}[A(x - c_0t - A^2t)] \]  

(MKDV) (2.4)

where \( c_0 \) is the linear phase speed. Like Scott-Russell’s solitary wave (which is described by (2.3)), each soliton is a single, “rounded, smooth, and well-defined heap of water” as shown schematically in Fig. 1a. To distinguish these one-humped solitary waves from those formed from strongly dispersive waves, which have many crests and troughs, one may use the term “unimontane” from the Latin.

“Strong dispersion” in the present context means that two waves of a given mode whose zonal wave-lengths differ by O(1) will have phase speeds which also differ by O(1). This implies that a wavepacket which initially was singly peaked like a unimontane soliton would break up very rapidly into many crests and troughs, and weak nonlinearity would not be able to prevent dispersion from doing this. Thus, one-humped, unimontane solitons are as impossible when the dispersion is too strong as when there is none at all.

However, there is a way out: a wavepacket whose Fourier transform is strongly peaked about a central wavenumber \( k_0 \) will disperse slowly rather than quickly because the small range of wavenumbers \( k \) implies a correspondingly small range of phase speeds \( c \). The weak tendency of such a wavepacket to spread out with time can be balanced by weak nonlinearity, again allowing solitary waves. However, a wavepacket which consists of the superposition of many waves with all approximately the same zonal wavenumber cannot be a one-humped disturbance like a unimontane soliton but must necessarily have many troughs and crests as shown schematically in Fig. 1b.

The equation that governs the nonlinear evolution of such packets of strongly dispersive waves (whether solitons or not) is the nonlinear Schrödinger equation (NLS)

\[ iu_t + u_{xx} + \lambda |u|^2u = 0, \]  

(2.5)

where \( \lambda \) is a constant and the mathematical form of the solitary waves is

\[ u = Ae^{i(k(x - c_0t)} \text{sech}[A(\lambda/2)^{1/2}(x - c_0t)], \]  

(2.6)

where \( k \) is the central wavenumber of the packet, \( c_0 \) the corresponding phase speed, and \( c_p \) the group velocity. The dotted line in Fig. 1b, which is mathematically described by the hyperbolic secant in (2.6),

<table>
<thead>
<tr>
<th>Table 1. Evolution equations and equatorial waves.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Nondispersive</strong></td>
</tr>
<tr>
<td>Definition</td>
</tr>
<tr>
<td>Soliton type</td>
</tr>
<tr>
<td>Evolution equation</td>
</tr>
<tr>
<td>Kelvin</td>
</tr>
<tr>
<td>Mixed Rossby-gravity</td>
</tr>
<tr>
<td>Gravity</td>
</tr>
<tr>
<td>Rossby</td>
</tr>
<tr>
<td>Reference</td>
</tr>
</tbody>
</table>

N.A. = not applicable.
is the envelope and travels at the group velocity $c_g$. Inside the envelope, the individual crests and troughs propagate with the phase velocity $c_p$. Since the envelope is a hyperbolic secant function like (2.3) and (2.4), the solitary waves of the nonlinear Schrödinger equation are known as "envelope solitons" to emphasize their many-crested nature.

This overall classification scheme for nonlinear equatorial waves is summarized in Table 1. The non-dispersive Kelvin wave has been treated in Boyd (1980). Gravity waves and mixed Rossby-gravity waves, which are always "strongly dispersive" in the present context, will be analyzed along with short Rossby waves in the third paper of the series (Boyd, 1981). In the remainder of this, the second part, I will concentrate strictly on weakly dispersive waves (the second column of the table) except for Section 8.

3. Derivation of the KDV equation

The model adopted here is the same as in Boyd (1978, 1980, 1981)—a single layer of homogeneous fluid of depth $H$. Defining "Lamb's parameter" as

$$ E = \frac{4\Omega^2 a^2}{gH}, $$

(3.1)

where $\Omega$ is the angular frequency of the earth's rotation, $a$ the radius of the earth, and $g$ the gravitational constant, length and time are nondimensionalized in terms of the scales

$$ L = \frac{a}{E^{1/4}} \quad \text{(length scale),} $$

(3.2)

$$ T = \frac{E^{1/4}}{2\Omega} \quad \text{(time scale).} $$

(3.3)

With this nondimensionalization, the nonlinear shallow water equations on the equatorial beta-plane are

$$ u_t + uu_x + vu_y - yv + \phi_x = 0, $$

(3.4)

$$ v_t + uv_x + vv_y + yu + \phi_y = 0, $$

(3.5)

$$ \phi_t + u_x + (u\phi)_x + v_y + (v\phi)_y = 0. $$

(3.6)

It is convenient to transform these equations into a frame of reference moving with the linear wave. Defining

$$ s = x - ct, $$

(3.7)

where $c$ is constant, the equations become

$$ u_t + (u - c)u_x + vu_y - yv + \phi_x = 0, $$

(3.8)

$$ v_t + (u - c)v_x + vv_y + yu + \phi_y = 0, $$

(3.9)

$$ \phi_t - c\phi_x + u_t + (u\phi)_x + v_y + (v\phi)_y = 0. $$

(3.10)

The next step is to apply the method of multiple scales in $s$ and $t$ following Clarke (1971). The twist is that, thanks to the transformation (3.7), the wave is independent of the "fast" scales in $s$ and $t$, which therefore need not even be defined, and is a function only of the "slow" variables

$$ \xi = e^{1/2}s, $$

(3.11)

$$ \tau = e^{1/2}t. $$

(3.12)

Assume

$$ u = e^{r\xi}(\xi,\eta,\tau) + e^{r\eta}(\xi,\eta,\tau) + \cdots, $$

(3.13)

$$ v = e^{r\xi}(\xi,\eta,\tau) + e^{r\eta}(\xi,\eta,\tau) + \cdots, $$

(3.14)

$$ \phi = e^{r\xi}(\xi,\eta,\tau) + e^{r\eta}(\xi,\eta,\tau) + \cdots. $$

(3.15)

Mathematically, the factor of $e^{1/2}$ instead of $e$ in the expansion for $v$ [Eq. (3.14)] is demanded by the requirement that each term of (3.13)-(3.15) be expandable in half-integral powers of $e$. Physically, it expresses the fact that the north-south current is very small in comparison to the zonal current for an ultralong (small zonal wavenumber) Rossby wave, equatorial or not.

Substituting the expansions (3.11)-(3.15) into (3.8)-(3.10) gives the lowest order set

$$ -c u_t^0 - y v^0 + \phi_x^0 = 0, $$

(3.16)

$$ y u^0 + \phi_y^0 = 0, $$

(3.17)

$$ -c \phi_x^0 + u_x^0 + v_y^0 = 0, $$

(3.18)

plus the first-order set (3.24)-(3.26) below. These lowest order equations (3.16)-(3.18) can be reduced to

$$ v_{yy}^0 + \left( \frac{1}{c} - y^2 \right) v^0 = 0. $$

(3.19)

This is merely a simplified form of the usual latitudinal structural equation for equatorial waves. The eigensolutions of (3.19) are

$$ v^0(y,\xi,\tau) = \eta(\xi,\tau)e^{-(1/2)y^2}H_n(y), $$

(3.20)

$$ u^0(y,\xi,\tau) = \eta(\xi,\tau) \left[ \frac{H_{n+1}(y)}{2(1 - c)} + \frac{nH_{n-1}(y)}{(1 + c)} \right] e^{-(1/2)y^2}, $$

(3.21)

$$ \phi^0(y,\xi,\tau) = \eta(\xi,\tau) \left[ \frac{H_{n+1}(y)}{2(1 - c)} + \frac{nH_{n-1}(y)}{(1 + c)} \right] e^{-(1/2)y^2}, $$

(3.22)

with

$$ c = \frac{-1}{2n + 1}, \quad n \geq 1, $$

(3.23)

where $n$ is the latitudinal mode number and $H_n(y)$ is the $n$th Hermite polynomial. Note that (3.19) is simply a one-dimensional eigenvalue problem in $y$ alone. To lowest order, $\eta(\xi,\tau)$ is arbitrary; it will be determined at the next order.
Note also that (3.20)–(3.23) are valid only for $n \gg 1$. It is well known (Holton, 1975) that for $n = 0$ there is only one easterly wave—the so-called “mixed Rossby-gravity” wave—instead of the two (Rossby plus gravity) that exist for higher $n$. The mixed Rossby-gravity wave behaves like a pure gravity wave when it is equatorially trapped, and so it is always strongly dispersive in the sense used here. Therefore, it must be dealt with in the third paper of this series (Boyd, 1981). Holton (1975) gives a full discussion of how the reduction from three equations to (3.19) introduces the spurious root given by (3.23) for $n = 0$, but one can see from the solutions themselves that something is amiss: for $c = -1$, the denominators of (3.21) and (3.22) are infinite.

The first-order equations are

$$-cu^1_y - vy^1 + \phi_x^1 = F_1,$$  
$$yu^1 + \phi_y^1 = F_2,$$  
$$-c\phi_x^1 + u_x^1 + v_y^1 = F_3,$$  
where

$$F_1 = -u^0_x - u^0u^0_y - v^0u_y^0,$$  
$$F_2 = cy^0,$$  
$$F_3 = -\phi_x^0 - (u^0\phi^y)_x - (v^0\phi^y)_y.$$  

It is straightforward to reduce this set down to

$$v^1_{yy} + \left(\frac{-1}{c} - y^2\right)v^1 = F_4,$$  
where

$$F_4 = \left(1 - \frac{1}{c^2}\right)(yF_1 + cF_2c)$$  
$$+ \frac{yF_1}{c^2} + \frac{yF_3}{c} + \frac{F_1y}{c} + F_3y.$$  

In order for (3.30) to have a solution, $F_4$ must be orthogonal to the homogeneous solution of (3.30). This gives the condition

$$\int_{-\infty}^{\infty} e^{-(1/2)y^2}H_n(y)F_4(y, \xi, \tau) = 0.$$  

Explicitly performing the integrals via Gauss-Hermite quadrature (Abramowitz and Stegun, 1965) yields

$$\eta_\epsilon + \alpha_n \eta_\epsilon + \beta_n \eta_{\epsilon\epsilon\epsilon} = 0,$$  
which is the Korteweg-deVries equation with $\alpha_n$ and $\beta_n$ as mode-dependent constants.

It can be shown that when the mode number $n$ is even, the nonlinear terms in $F_4$ are all antisymmetric in $y$ with respect to the equator, whereas $H_n(y)$ itself is symmetric. This implies that

$$\alpha_n = 0, \quad n \text{ even},$$

so that (3.33) collapses into a linear equation. This linear equation provides a perfectly good description of the even $n$ modes in the parameter range assumed by our scaling; the mere fact that it is linear implies the important truth that the even $n$ modes will be much less weakly affected by nonlinearity than those of odd $n$ for comparable amplitudes and length scales. Nonetheless, (3.33) is merely a long-wavelength approximation to the full linear dispersion relation for even $n$, so I will confine myself to the odd $n$ modes in discussing the KDV equation (3.33) in the rest of the paper. Nonetheless it is still possible to balance nonlinearities against dispersion for even $n$ if one makes different scaling assumptions and works to higher order. This leads to the modified Korteweg-deVries (MKDV) equation and is the subject of the next section.

The full lowest order solutions for the lowest three odd $n$ modes are explicitly

$$n = 1$$  
$$c = -\frac{1}{3},$$  
$$v^0 = \eta_\epsilon 2y e^{-(1/2)y^2},$$  
$$u^0 = \eta_\epsilon \left(-9 + 6y^2\right) e^{-(1/2)y^2},$$  
$$\phi^0 = \eta_\epsilon \left(3 + 6y^2\right) e^{-(1/2)y^2},$$  
$$\eta_\epsilon - 1.53666\eta_\epsilon - 0.098765\eta_{\epsilon\epsilon\epsilon} = 0,$$

$$n = 3$$  
$$c = -\frac{1}{7},$$  
$$v^0 = \eta_\epsilon (-12y + 8y^3) e^{-(1/2)y^2},$$  
$$u^0 = \eta_\epsilon (-5y^4 - 35y^2 + 7y^4) e^{-(1/2)y^2},$$  
$$\phi^0 = \eta_\epsilon (-7y^4 + 7y^2 + 7y^4) e^{-(1/2)y^2},$$  
$$\eta_\epsilon - 2.90735\eta_\epsilon - 0.0199916\eta_{\epsilon\epsilon\epsilon} = 0,$$

$$n = 5$$  
$$c = -\frac{1}{9},$$  
$$v^0 = \eta_\epsilon (120y - 160y^3 + 32y^5) e^{-(1/2)y^2},$$  
$$u^0 = 11\eta_\epsilon (-11 + 54y^2 - 28y^4 + 7y^6) e^{-(1/2)y^2},$$  
$$\phi^0 = 11\eta_\epsilon (1 + 6y^2 - 12y^4 + 9y^6) e^{-(1/2)y^2},$$  
$$\eta_\epsilon - 16.6792\eta_\epsilon - 0.008196\eta_{\epsilon\epsilon\epsilon} = 0.$$  

A couple of final comments are in order. First, the three KDV equations given above are meaningless without a specific value for the perturbation parameter $\epsilon$. The normal choice is to pick $\epsilon \ll 1$ so that $\eta \sim O(1)$, but now that $\epsilon$ has fulfilled its role as a bookkeeping device, it is easier to set $\epsilon = 1$ and allow the amplitude of $\eta$ and the scales of $\tau$ and $\xi$ to revert back to their usual values in the
nondimensionalization given by (3.2) and (3.3). Thus,
\[ \tau = \xi, \]  \hspace{1cm} (3.38)
\[ \xi = x - ct. \]  \hspace{1cm} (3.39)

A similar policy will be followed in the rest of this three-part series of papers.

Second, the most striking feature of the three KDV equations is that they contain no parameters. The coefficients of the KDV equation for a given mode are simply pure numbers. Most geophysical problems contain many parameters and this makes it very difficult to give a complete description of the phenomenon. When many parameters are involved, one must therefore retreat to (hopefully) representative case studies.

For equatorial solitons, no such retreat is necessary. The one-soliton solutions given for each of these modes in Section 5 are completely explicit, contain no parameters, and involve nothing more arcane than a hyperbolic secant function. Few theories in geophysics can boast such simplicity. The price for this simplicity is discussed in Section 10.

4. Derivation of the modified Korteweg-deVries equation

In the previous section, I showed that for Rossby modes of even \( n \), the first-order solubility condition (3.23) reduces to a linear equation. The nonlinear terms are not identically zero at first order, however, but merely force a first-order solution which is of opposite symmetry with respect to the equator than the lowest order solution. At second order, the interaction between the nonlinear first-order terms and the lowest order will generate a nonlinear expression which will indeed enter the second-order solubility condition. Because of the higher order, however, the nonlinearity will be cubic instead of quadratic, and the resulting evolution equation will therefore be the modified Korteweg-deVries equation instead of the KDV equation itself.

To arrange that dispersion and nonlinearity will still be of comparable magnitudes when the nonlinearity is a perturbation order weaker than in the previous section implies that dispersion must be weakened also. Since the dispersion tends to zero as the zonal scale increases, this can be done simply by changing the scaling assumptions on \( \tau \) and \( \xi \).

For Rossby modes of even \( n \), let
\[ \tau = \epsilon \tau, \]  \hspace{1cm} (4.1)
\[ \xi = \epsilon \xi, \]  \hspace{1cm} (4.2)
\[ v = \epsilon^2 (v^0 + \epsilon v^1 + \cdots), \]  \hspace{1cm} (4.3)
in place of (3.11), (3.12) and (3.14). As before, let
\[ u = \epsilon (u^0 + \epsilon u^1 + \cdots), \]  \hspace{1cm} (4.4)
\[ \phi = \epsilon (\phi^0 + \epsilon \phi^1 + \cdots). \]  \hspace{1cm} (4.5)

Substituting (4.1)–(4.5) into (3.8)–(3.10) then gives the order-by-order perturbation equations.

Since both here and in the previous section, the lowest order represents linear, nondispersive waves, it is clear that the lowest order equations are still (3.16)–(3.18). At first order, however, the inhomogeneous terms \( F_1, F_2 \) and \( F_3 \) will be different because, via rescaling \( \xi \) and \( \tau \), the linear dispersive terms have been pushed back to second order. One has
\[ -c u_\xi - y v^1 + \phi_\xi = F_1, \]  \hspace{1cm} (4.6)
\[ y u^1 + \phi_y = F_2, \]  \hspace{1cm} (4.7)
\[ -c \phi_\xi + u_\xi + v_\xi = F_3, \]  \hspace{1cm} (4.8)

where now
\[ F_1 = -u^0 u_\xi^0 - v^0 u_\xi^0, \]  \hspace{1cm} (4.9)
\[ F_2 = 0, \]  \hspace{1cm} (4.10)
\[ F_3 = -(u^0 \phi)_\xi - (v^0 \phi)_\xi. \]  \hspace{1cm} (4.11)

As before, the set of three first-order equations can be reduced to a single equation for \( v \) alone, i.e.,
\[ v_{yy} + \left( -\frac{1}{c} - y^2 \right) v^1 = F_4, \]  \hspace{1cm} (4.12)

where \( F_4 \) is again given in terms of \( F_1, F_2 \) and \( F_3 \) by (3.31). Because \( F_4 \) and the homogeneous solution of (4.12) have opposite symmetry with respect to the equator, the solubility condition (3.32) is automatically satisfied at first order. The difficulty is that in order to derive the second-order solubility condition that will yield the MKDV equation, one must first explicitly solve the first-order equations.

If \( F_4 \) is expanded as a series of Hermite functions, that is,
\[ F_4 = \sum_{m=0}^{\infty} \kappa_m H_m(y) e^{-(1/2)y^2}, \]  \hspace{1cm} (4.13)
then solving (4.12) is easy. The Hermite functions satisfy the eigenvalue problem
\[ \left[ \frac{\partial^2}{\partial y^2} + (2m + 1 - y^2) \right] e^{-(1/2)y^2} H_m(y) = 0. \]  \hspace{1cm} (4.14)

Recalling from (3.23) that
\[ -\frac{1}{c} = 2n + 1, \]  \hspace{1cm} (4.15)
it follows from (4.14) that
\[ \left[ \frac{\partial^2}{\partial y^2} + \left( -\frac{1}{c} - y^2 \right) \right] e^{-(1/2)y^2} H_m(y) = 2(n - m) e^{-(1/2)y^2} H_m(y). \]  \hspace{1cm} (4.16)
Substituting the expansion for $F_4$ [Eq. (4.13)] into (4.12) and use of (4.16) shows that
\[
v^1 = \sum_{m=0}^{\infty} \frac{\kappa_m}{2(n-m)} e^{-(1/2)\nu^2} H_m(y). \tag{4.17}
\]
It is obvious that (4.17) is meaningless unless $\kappa_n = 0$, which is equivalent to the solubility condition (3.32). When $n$ is even, however, $F_4$ is antisymmetric with respect to the equator and because of the symmetry of the Hermite functions themselves, this implies that all the even $\kappa_m$ are automatically zero.

Computing the nonzero odd $m$ expansion coefficients requires a modest trick. Each lowest order field $(\nu^0, \nu^0, \phi^0)$ is the product of a polynomial with a Gaussian $(e^{-(1/2)\nu^2})$. $F_4$ in turn involves products of these fields. It is no great feat to multiply two polynomials together and then convert the product into a sum of Hermite polynomials; any polynomial can be written as an exact, finite sum of Hermite polynomials and no approximations are necessary. The sole difficulty is that the nonlinear terms are all the product of one Gaussian with another. Noting that only a single factor of $e^{-(1/2)\nu^2}$ appears in (4.13), one sees that there is one Gaussian too many.

The remedy is to use the known expansion
\[
e^{-\nu^2} = e^{-(1/2)\nu^2} \sum_{m=0}^{\infty} a_{2m} H_{2m}(y), \tag{4.18}
\]
where
\[
a_{2m} = \frac{(-1)^m (2)^{1/2}}{12^m m!}. \tag{4.19}
\]
If this expansion is truncated at $m = M$, then the rest of the computation of the $\kappa_m$'s merely involves multiplying three polynomials together and rearranging the result into a finite series of Hermite polynomials. The tricky question is: how rapidly does (4.18) converge?

Table 2 gives the coefficients $\{a_{2m}\}$, but their rapid decrease is somewhat misleading because the maxima of the Hermite functions also increase without bound. A better measure is to convert (4.18) into a series of normalized Hermite functions
\[
e^{-\nu^2} = e^{-(1/2)\nu^2} \sum_{m=0}^{\infty} \tilde{a}_{2m} \tilde{H}_{2m}(y), \tag{4.20}
\]
where "normalized" means that
\[
\int_{-\infty}^{\infty} e^{-\nu^2} H_m \tilde{H}_m dy = 1 \text{ for all } m, \tag{4.21}
\]
since the normalized functions satisfy the simple bound
\[
|e^{-(1/2)\nu^2} \tilde{H}_m(y)| \leq 0.785 \text{ for all } y \text{ and } m. \tag{4.22}
\]
This implies the absolute value of the largest error for $y$ on the interval $[-\infty, \infty]$—the error in the "$L_\infty$ norm"—can be bounded by the sum of the neglected normalized coefficients, i.e.,
\[
|e^{-\nu^2} - e^{-(1/2)\nu^2} \sum_{m=0}^{M} \tilde{a}_{2m} \tilde{H}_{2m}(y)| \leq \sum_{m=M+1}^{\infty} |\tilde{a}_{2m}|
\]
for all $y$. \tag{4.23}

The normalized coefficients are explicitly
\[
\tilde{a}_{2m} = \frac{\pi^{1/2} (2)^{1/2} (-1)^m (2m)!}{6^n(m!)}, \tag{4.24}
\]
\[
-0.9709835(-1)^m \frac{3^m (2m)!^{1/4}}{6^n(m!)} \quad \text{as } m \to \infty. \tag{4.25}
\]

From the rapid decrease of the normalized coefficients given in Table 2, it is evident that the series converges very rapidly. The maximum error for $M = 5$ is less than 0.001, and the asymptotic form (4.25) shows that each increase of $M$ by 2 will add roughly another decimal place of accuracy. Thus, this "Gaussian expansion" trick is extremely effective, and its usefulness has been explained in some detail because it will be equally crucial in Boyd (1981) for deriving the nonlinear Schrödinger equation for strongly dispersive waves.

Once $M$ has been chosen and (4.18) substituted into $F_4$, solving for $v^1$ is merely a matter of polynomial algebra. Unfortunately, even for modest $M$, one finds oneself working with rather large polynomials—degree 25 even for $n = 2$ and $M = 5$. In consequence, this algebra was performed using the algebraic manipulation language REDUCE2 as explained in the Appendix.

With $v^1$ known, $u^1$ and $\phi^1$ follow from
\[
\phi^1 = \left(\frac{\eta}{2\eta_\epsilon}\right) \left(\frac{1}{1-c^2}\right) (F_1 + c F_3 - c v^1 + y v^1), \tag{4.26}
\]
where the first factor comes from integrating $\eta \eta_\epsilon$.
with respect to $\xi$ to obtain $\eta^2/2$ (recall that $v^0$ is proportional to $\eta$, whereas $u^0$ and $\phi^0$ are proportional to $\eta$ itself) and

$$u^1 = -\frac{\phi_\xi}{y}.$$  \hfill (4.27)

With the first-order solution known, it is now possible to proceed to attack the second-order equations, which are

$$-cu_\xi^2 - y v^2 + \phi_\xi^2 = F_\xi^2,$$  \hfill (4.28)

$$y u_\xi^2 + \phi_\xi^2 = F_\xi^2,$$  \hfill (4.29)

$$-c\phi_\xi^2 + u_\xi^2 + v_\xi^2 = F_\xi^2,$$  \hfill (4.30)

where

$$F_\xi^2 = -u_\xi^2 - u'u_\xi - u'u_\xi - v'u_\xi - v'u_\xi - v'u_\xi,$$  \hfill (4.31)

$$F_\xi^2 = c v_\xi^2,$$  \hfill (4.32)

$$F_\xi^2 = -\phi_\xi - (u^2 \phi_\xi + u^2 \phi_\xi - v^2 \phi_\xi + v^2 \phi_\xi).$$  \hfill (4.33)

The reduction of (4.28)–(4.30) to

$$v_\xi^2 + \left(1 - \frac{1}{c} - y^2\right)v_\xi^2 = F_\xi^2$$  \hfill (4.34)

is identical to that leading to (4.12), and the relationship of $F_\xi^2$ to $F_\xi^2$, $F_\xi^2$ and $F_\xi^2$ is the same as (3.31) except for the superscripts of 2 on the $F$'s. The solubility condition is

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2(2\pi)^2}}H_n(y)F_\xi^2(y)dy = 0.$$  \hfill (4.35)

After performing the integrals via Gauss-Hermite quadrature as before, one finds that the only way this can be satisfied is if $\eta(\xi, \tau)$ obeys the modified Korteweg-de-Vries equation

$$\eta_\tau + \alpha_\eta \eta_\xi \eta_\xi + \beta_\eta \eta_\xi \eta_\xi \eta_\xi = 0.$$  \hfill (4.36)

Since $\beta_\eta$ comes from the linear solution $(u^0, v^0, \phi^0)$ only, it is independent of the truncation $M$ in (4.18), and thus is accurate to at least five decimal places. For the $n = 2$ mode given explicitly below, $M = 5$ and $M = 7$ yielded values for $\beta_\eta$ that differed by less than $1\%$, so the quoted value is accurate to at least $1\%$—considerably more than the assumptions that underlie my model.

As noted in the previous section, one obtains weakly dispersive, pure Rossby waves only for $n \geq 1$—the mixed Rossby-gravity $(n = 0)$ is strongly dispersive—so the lowest mode which evolves according to the MKDV equation is the $n = 2$ mode. Its solution is explicitly

$$n = 2,$$  \hfill (4.37a)

$$c = -\frac{1}{4},$$  \hfill (4.37b)

$$u^0 = 10\eta(\xi, \tau)e^{-\frac{1}{2}(2\pi)^2}(y^2 - 2),$$  \hfill (4.37c)

$$\phi^0 = 10\eta(\xi, \tau)e^{-\frac{1}{2}(2\pi)^2}y^3,$$  \hfill (4.37d)

where $\eta(\xi, \tau)$ satisfies

$$\eta_\tau + 10.6\eta^2 \eta_\xi - 0.0384\eta_\xi \eta_\xi \eta_\xi = 0.$$  \hfill (4.37e)

It will be shown in Section 9 that the signs of (4.37e) are such that it does not allow solitons although it does permit permanent wave disturbances in the form of shock waves. For this reason and also because of computational difficulties for large $n$ that are explained in the Appendix, the corresponding solutions for $n = 4$ will be omitted and the rest of the paper will concentrate mostly on the KDV equation. In Section 7 and part of Section 5, however, the MKDV equation (4.37e) and its solutions will be discussed in some detail.

5. The theory of the KDV and MKDV equations

The pioneering work of Gardner et al. (1967), later extended to other equations by Lax (1968), showed that the KDV equation, although nonlinear, could be solved through a sequence of purely linear steps via the so-called inverse scattering method. This technique, although very powerful from a theoretical viewpoint—it established, for example, that the general solution consists of the wave train plus a finite number of solitons as noted in the Introduction—is useless for numerical purposes. In consequence, the case studies of the next two sections were computed by the finite-difference method described in the Appendix. In addition, the inverse scattering algorithm can only be justified when the initial disturbance is localized in space; it fails for periodic initial conditions, for example.

Nonetheless, some useful information can be extracted from the inverse scattering procedure. First, however, it is necessary to transform the KDV equation from the general form

$$\eta_\tau + E \eta \phi_\xi \phi_\xi + F \eta_\xi \phi_\xi \phi_\xi \phi_\xi = 0$$  \hfill (5.1)

to the canonical form

$$u_t - 6uu_x + u_{xxx} = 0.$$  \hfill (5.2)

Since one has four transformation parameters at one's disposal—one can independently rescale $t, x$ and $\eta$ and multiply (5.1) by an overall constant—and one has only three coefficients of the KDV equation to match to those of (5.2), one is free to choose one of the transformation parameters arbitrarily. Letting

$$t = \tau,$$  \hfill (5.3)

i.e., the time scale is unaffected, one can show
\[ x = \frac{\xi}{F^{1/3}}, \quad (5.4) \]
\[ u = -\frac{E\eta}{6F^{1/3}}. \quad (5.5) \]

For the canonical KDV equation (5.2), the first step in inverse scattering is to solve the one-dimensional Schrödinger equation of quantum mechanics with the initial condition \( u(x,0) \) serving as the potential function. Each bound state then corresponds to a solitary wave, and the associated energy level gives the shape, amplitude and speed of the soliton. Analytic formulas for the Schrödinger energy levels, however, are known only for two simple cases: a \( \text{sech}^2(x) \) potential and a so-called “square-well” or “top-hat” potential (Whitham, 1974).

In consequence, these initial conditions have been used for the case studies of the next two sections so that the numerical results for a finite time can be compared with the calculated solitons. The phases of the solitary waves, however, depend on the normalization integrals of the Schrödinger equation eigenfunctions, so the location of a given soliton is most easily found by direct numerical integration of the KDV equation itself.

Elementary quantum theory does provide two additional useful bits of information. First, if
\[ u(x,0) \geq 0, \quad (5.6) \]
there can be no solitons (a potential hill has no bound states—only a potential valley can). Note that because of the sign change implied by (5.5), this means that an initial condition of depression will generate no solitons for the original KDV equation (5.1) and also for the case studies of the next section. Second, if
\[ \int_{-\infty}^{\infty} u(x,0) dx < 0, \quad (5.7) \]
there will always be at least one soliton (Zabuksky, 1968). Thus, as emphasized in the Introduction, special initial conditions are not needed to generate solitary waves.

The solitons of the MKDV equation are determined by finding the discrete eigenvalues of a more complicated equation which is not Schrödinger’s, but the general principle is the same. As noted in the previous section, however, the signs of the coefficients of the MKDV equation for the \( n = 2 \) Rossby wave are such that it does not admit solitons anyway.

Existing results for the nonlinear KDV and MKDV wavetrains are much more limited than those for the solitons because they are very complicated and give only the form of the solution—there remain some undetermined phase factors. For the linearized KDV/MKDV equation
\[ u_t + u_{xxx} = 0, \quad (5.8) \]
however, it is possible to give a fully determined asymptotic solution, uniformly valid for all \( x \) for sufficiently large time, as
\[ u(x,t) \sim \frac{U([-Z]^{1/2}) \text{Ai}(z)}{(3t)^{1/3}}, \quad (5.9) \]
where \( U(k) \) is the Fourier transform of the initial condition
\[ U(k) = \int_{-\infty}^{\infty} u(x,0) e^{-ikx} dx, \quad (5.10) \]
and where
\[ z = \frac{x}{3t}, \quad (5.11) \]
\[ Z = \frac{x}{3t}, \quad (5.12) \]
and where \( \text{Ai}(z) \) is the usual Airy function. The asymptotics of \( \text{Ai}(z) \) show that \( u(x,t) \) is exponentially small for \( x > 0 \) and sinusoidal for negative \( x \). The wavefront is stationary in the canonical coordinate system of (5.8), but in most physical applications, this is a moving coordinate system—in this case, one traveling westward at the phase speed of the nondispersive linear Rossby wave.

For the MKDV equation in the canonical form
\[ v_t - 6v^2v_x + v_{xxx} = 0. \quad (5.13) \]
Ablowitz et al. (1979) show that the corresponding solution is
\[ v(x,t) \sim \frac{w(z; r([-Z]^{1/2}))}{(3t)^{1/3}}, \quad (5.14) \]
where \( r(k) \) is the “scattering data” for the inverse scattering problem corresponding to the MKDV equation and where \( w(z; r) \) is a nonlinear generalization of the Airy function known as the second Painlevé transcendent (Rosales, 1978). For small values of
\[ r_0 = r(0) = \tanh U(0), \quad (5.15) \]
where “small” means \( r_0 < 0.5, r(k) = U(2k) \) and \( w(z; r) \) is indistinguishable from \( \text{Ai}(z) \) for all \( z \). The solution is essentially linear for all time and space. For \( r_0 \) near 1, however, the wavetrain is strongly nonlinear, but the qualitative form remains similar to (5.9)—an exponentially decaying tail in front and a damped oscillation for negative \( x \).

For the KDV wavetrain, the results are extremely complicated. Ablowitz and Segur (1977) and Miles (1979) have independently attacked this problem, but their descriptions of the asymptotic solution differ even in some qualitative details although Miles expresses the hope that these differences are merely semantic. The most striking difference from the MKDV wavetrain can be found, however, by using elementary scale analysis.
Table 3. Parameters for numerical cases. For sech² initial conditions, \( \eta(x,0) = A \text{sech}^2(Bx) \). For the top hat, \( \eta(x,0) = A \) for \(-B < x < 0\) and is zero elsewhere.

<table>
<thead>
<tr>
<th>No.</th>
<th>Graph</th>
<th>Equation</th>
<th>Initial</th>
<th>A</th>
<th>B</th>
<th>( r )</th>
<th>Number solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Fig. 3</td>
<td>KDV</td>
<td>Sech²</td>
<td>0.12</td>
<td>0.394</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>Fig. 4</td>
<td>KDV</td>
<td>Sech²</td>
<td>-0.12</td>
<td>0.394</td>
<td>100</td>
<td>Note</td>
</tr>
<tr>
<td>III</td>
<td>Fig. 5</td>
<td>KDV</td>
<td>Sech²</td>
<td>0.12</td>
<td>0.279</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>IV</td>
<td>Fig. 6</td>
<td>KDV</td>
<td>Sech²</td>
<td>0.06</td>
<td>0.394</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>V</td>
<td>Fig. 7</td>
<td>KDV</td>
<td>Top hat</td>
<td>0.12</td>
<td>0.394</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>VI</td>
<td>Fig. 8</td>
<td>MKDV</td>
<td>Sech²</td>
<td>0.12</td>
<td>0.394</td>
<td>100</td>
<td>Note</td>
</tr>
</tbody>
</table>

Let \( L(t) \) be the horizontal length scale, \( A(t) \) the amplitude scale, and \( \epsilon \) the initial amplitude \([\approx A(0)]\). The linear solution (5.9)–(5.12) shows that

\[
L \sim O(t^{1/2}),
\]

\[
A \sim O\left(\frac{\epsilon}{t^{1/3}}\right).
\]

For the MKDV equation, it follows that

\[
\frac{u^2u_x}{u_{xxx}} \sim O(A^2L^2),
\]

\[
\sim O(\epsilon^2),
\]

which is independent of time. Thus, if the amplitude of the solution of the MKDV equation is initially small, then it will remain so for all time and the MKDV wavetrain will differ little from the linear solution.

For the KDV equation, however, which is quadratically instead of cubically nonlinear, the same analysis gives

\[
\frac{uu_x}{u_{xxx}} \sim O(AL^2),
\]

\[
\sim O(t^{1/3}).
\]

If \( \epsilon \ll 1 \), then obviously nonlinear effects will initially be small. But (5.21) shows that no matter how small \( \epsilon \) is, nonlinearity will always be as important as dispersion for sufficiently large time. In other words, the effects of dispersion are decreasing as \( t^{-2/3} \) while the amplitude is decreasing more slowly as \( t^{-1/3} \). Thus, the asymptotic wavetrain of the KDV equation is always nonlinear.

Miles (1979) and Ablowitz and Segur (1977) show that the most striking consequence of this nonlinearity for the wavetrain is that the first trough decreases as \( t^{-2/3} \) instead of the \( t^{-1/3} \) of the linear theory. Outside the neighborhood of the first few troughs, however, the KDV wavetrain is similar to (5.9) and (5.14). The numerical case studies of the next section will permit direct comparisons between the linear and nonlinear wavetrains for representative conditions.

6. Case studies of the KDV equation

Without exception, all the case studies presented here and in the following section are unforced initial value problems in an unbounded ocean. Physically, waves can be either generated at coastal boundaries or forced locally by fluctuating surface winds, but neglect of the explicit forcing is not necessarily a good approximation in either case. Stochastically forced waves, however, are not likely to be of much relevance to the present work in any event because it is unlikely that a given mode would receive enough energy in comparison to other modes to appear as an observable solitary wave. Rather, it is large-amplitude waves excited at boundaries by "sudden events" like the onset of the southwest Monsoon in the Indian Ocean or the relaxation of Pacific trade winds that triggers El Niño that are most likely to generate observable solitons.

This in turn implies that, as in Boyd (1980), the signaling problem in which a nonzero wave amplitude at the coast of generation is specified for all time is perhaps more relevant to equatorial solitary waves than the conventional initial value problem. However, the KDV equation, which has three spatial derivatives but only a single-time derivative, is inappropriate as a signaling problem. The physical resolution of this dilemma is that a wave-exciting event will generate gravity waves as well as Rossby waves. Since the gravity waves travel at different rates than the Rossby waves (see Section 9), the two classes of waves will rapidly separate, and the packet of Rossby waves can then be described accurately by the analysis of earlier sections. The KDV equation itself, however, is not a good model of what happens in the neighborhood of the coast when the Rossby and gravity packets are still together. Since a full treatment of coastal effects—one really ought to worry about such matters as coastal Kelvin waves and shelf waves, too—would be very complicated and bury the simple soliton ideas under a mass of detail, the idealization of the unbounded initial value problem seems the better route.

Since the gravest mode is likely to be most strongly excited and has the simplest structure, such solitons should be the most important and the most easily observed. Therefore, all the case studies graphed here are for the \( n = 1 \) mode of the first baroclinic mode although a few remarks about higher modes are made at the end.

Table 3 lists the illustrated cases including the \( n = 2 \), MKDV example of the next section. The following parameters will be used: \( H = 40 \text{ cm (equivalent depth of the first baroclinic mode)} \), \( E = 214,000 \), the nondimensional length scale = 295 km, and the time scale \( T = 1.71 \text{ days} \). [The last three parameters are mathematically defined in (3.1)–(3.3).] If one takes the Pacific Ocean to be 10,000 km across, this
translates into a nondimensional width of 34. Since
the $n = 1$ mode has a (linear) nondimensional phase
speed $c = - \frac{1}{3}$, the linear wave requires about 100
time units to cross the ocean. Since, as already men-
tioned above, boundary reactions to changes in the
wind stress are most likely to generate waves large
enough for nonlinear effects to matter before the
wave collides with a coast or dissipates, propaga-
tion from the Peruvian coast to the Far East is the
normal life cycle of an observable Rossby soliton.
For this reason, all the case studies below will be
followed for the transoceanic propagation time of
100 units.

The first example will be that of a single soliton
without a wavetrain. The solitons of the untrans-
formed KDV equation for the $n = 1$ mode
\[ \eta_t - 1.5366 \eta \eta_x - 0.09877 \eta_{xxx} = 0, \]
form the one-parameter family
\[ \eta(\xi, \tau) = A \text{sech}^2[B(\xi - 0.395B^2\tau)], \]
where
\[ A = 0.772B^2. \]

As a somewhat arbitrary prototype example, we set
\[ A = 0.12 \]
which implies
\[ B = 0.394. \]

This has a maximum zonal current of $-54$ cm s$^{-1}$ at
the equator and a maximum height displacement of
7 cm at $y = \pm 400$ km. Translating this everywhere
upward height displacement in the equivalent baro-
tropic model to the actual physical situation of
propagation along a thermocline with an undisturbed
depth of 100 m, the wave crest becomes a down-
ward displacement of the thermocline by 17 m. In
the east-west direction, the wave decays to half its
maximum value at $\xi = \pm 660$ km relative to the
crest. Both the magnitude and sign of this are
comparable to the Rossby waves of McCreary (1976),
for example, which suggests that this prototype is
a reasonable standard for discussion.

A top view of this soliton is given in Fig. 2. Note
the east-west elongation of the height contours; for
a smaller, slower and longitudinally broader soli-
tary wave, this elongation would be much more
pronounced and the contours would be shaped like
sausages. Note also that the north-south currents
are very weak in comparison to the zonal currents; this
is a general property of long Rossby waves, and it,
too, would be exaggerated for a weaker (and longer)
soliton.

Fig. 3 illustrates the importance of nonlinearity by
comparing the prototype soliton of (6.2)–(6.6) with
the corresponding solution of the linearized KDV
equation at $\tau = 100$. In this and later graphs, the
crest is at $\xi = 0$ at $\tau = 0$, a nondimensional co-
ordinate system moving with the linear phase speed
is used, and the convention that west is in the posi-
tive $\xi$ direction is employed. Thus, the positive
location of the solitary wave’s crest shows that it
has traveled $\sim 20\%$ faster than the crest of the linear
solution. The width of the graph (80 units) is equiv-
alent to a dimensional extent of $\sim 25$ 000 km, which is
obviously wider than any present-day ocean; the
reasons for using an unbounded domain have al-
ready been explained, but the backward extent of
the wavetrain will serve to reemphasize the limita-
tions of neglecting coastal boundaries.

The differences between the soliton and the
linear solution in Fig. 3 are impressive: the soliton
is twice as tall and half as wide as the first crest of

![Diagram](http://journals.ametsoc.org/jpo/article-pdf/10/11/1699/4402837/1520-0485(1980)010_1699_eswpir_2_0_co_2.pdf)
the linear wavetrain. The nonlinear solution is everywhere above mean sea level, but the first linear crest is followed by a trough half its size. The soliton would cross the Pacific in 147 days; the linear peak would take ~40 days longer. It would be difficult to convince a casual observer that the two graphs in Fig. 3 were really generated by the same initial conditions and (except for one term) the same equation.

Fig. 4 is identical with Fig. 3 except the sign of the initial conditions was reversed. An initial condition of depression, which in the inverse scattering method is equivalent to a potential barrier, cannot produce solitons because the equivalent Schrödinger equation has no bound states. As a result, the nonlinear solution is also an oscillatory wavetrain, so the qualitative differences between linear and nonlinear are not as striking as when the initial condition is one of elevation. Nonetheless, nonlinearity does produce pronounced changes from the linear solution.

First, nonlinearity makes crests taller and narrower—for the solitons, this is what counteracts the opposing tendency of dispersion to spread and shorten—and does the reverse for troughs, creating a kind of superdispersion for the latter. The result is to create what one might call “crest-and-trough democracy” in the nonlinear wavetrain: the first crest and trough are of approximately the same magnitude, whereas the linear solution is almost totally dominated by the deep and broad leading trough. The first linear crest has less than half the amplitude of the leading trough, and successive maxima and minima fall off so steeply that at \( \xi = -40 \), the linear peaks are only one-fifth the amplitude of their nonlinear counterparts.

Ablowitz and Segur (1977) and Miles (1979) have described this behavior in mathematical form. They showed that, whereas in linear theory, the first trough decays more slowly than the rest of the wavetrain (as \( t^{-1/2} \) rather than \( t^{-1/2} \) because it is a caustic), in nonlinear theory the first trough decays faster than the rest of the wavetrain as \( t^{-2/3} \). The ratio of the first two troughs in fact decreased from 1.67 at \( \tau = 50 \) to 1.27 at \( \tau = 100 \) for the case illustrated in Fig. 4, so that in a wavetrain only slightly larger or an ocean only slightly wider, the first trough would no longer have been the deepest by the time the wave struck the western boundary.

Figure 5 illustrates another empirical principle. With sech\(^2\) initial conditions, one can show via the inverse scattering method that for certain combinations of height and width, the wavetrain is zero, i.e., the initial crest splits into a finite number of solitary waves and nothing else. The initial conditions for the case shown in Fig. 5 were deliberately chosen to be exactly between the pure one- and two-soliton solutions in the hopes of producing as large and conspicuous a wavetrain as possible. The effort failed miserably. Although the two solitons are not completely separated at \( \tau = 100 \), the amplitudes are already within 3% (tall crest) and 25% of their asymptotic values as calculated via the inverse scattering method from the known bound states of Schrödinger’s equation. The amplitude of the first trough of the wavetrain, however, is only a laughable 2% of the amplitude of the larger soliton. This suggests the notion of “soliton dominance”: a smooth, large initial condition of elevation will generate almost no wavetrain—the solitons completely dominate the solution even for rather small times.

Fig. 7 will show that the shape of the disturbance is not terribly important for “soliton dominance” but the amplitude is. The initial condition for Fig. 6 is the same shape and width as for the prototype shown in Fig. 4 but the amplitude is only half (case IV).

The result of this moderate nonlinearity is that the linear and nonlinear solutions both have well-developed wavetrains following a single dominant peak. Although the soliton has not completely

---

**Fig. 4.** Comparison of nonlinear (solid) with linear (dotted) solutions for a run identical to case I except that the sign of the initial condition was reversed so that the initial disturbance is a trough instead of a crest.

**Fig. 5.** As in Fig. 4 except that the initial crest is 40% broader (case III).
separated itself from the wavetrain in the nonlinear case, comparison of the first peak of the nonlinear solution with the calculated soliton shows that the peak is much closer to the solitary wave in shape and size than to the first crest of the linear solution—it is already essentially a soliton. Since the height ratio of the first peak (soliton) to the first trough (wavetrain) is more than 6 to 1, one could well argue that the principle of “soliton dominance” applies here, too, but at least when the nonlinearity is weaker, the oscillatory tail behind the soliton is large enough to catch the eye. In contrast, the wavetrain is essentially invisible in Fig. 5.

This trend toward “soliton dominance,” i.e., of large smooth disturbances to yield a handful of solitons and not much else, has been observed by previous workers such as Johnson (1972) and Greig and Morris (1976), but they have not remarked on it. “Soliton dominance” cases are much easier to numerically simulate than pure wavetrain situations like that shown in Fig. 4, so these earlier investigators missed some minor numerical difficulties which are therefore discussed in the Appendix.

Note also in Fig. 6 that the nonlinear wavetrain is led by a trough. Ablowitz and Segur (1977) and Miles (1979) show that this is always true whereas a linear wavetrain may begin with either a trough or a crest. What happens in the nonlinear counterpart of the latter case is that—as in Fig. 6—the leading crest runs off to the right as a soliton.

Before discussing solutions of the MKDV equation, it is desirable to look at initial conditions of a very different shape from that considered above. Fig. 7 compares the linear and nonlinear solutions for the “top-hat” or “square-well” initial conditions

\[ \eta(x,0) = \begin{cases} 0.12, & -16 \leq x \leq 0 \\ 0, & x < -16 \text{ or } x > 0 \end{cases} \]

(6.9)

First, note that although the height of the disturbance is nowhere greater than the crest of the prototype one soliton case shown in Fig. 3, three solitary waves are produced by (6.9) because of its greater width. In general, for a given shape of initial condition, the number of solitons is a function of the product of the height with the square of the width. Their size, however, is constrained by Segur’s (1973) remark that the amplitude of the largest soliton can never exceed twice the maximum height of the initial disturbance.

Second, the amplitudes of the three nonlinear peaks (0.21, 0.16, 0.08) agree well with those calculated from the bound states of Schroedinger’s equation with a square-well potential (0.220, 0.162, 0.073). Even though the solitons are still weakly overlapping, the solution at \( \tau = 100 \) consists basically of three solitons plus a small tail. Despite the sharp fronts in the initial condition, soliton dominance is still true.

Third, a straight line can be drawn (approximately) through the peaks of all three solitons. In the absence of dispersion, a nonlinear wave will steepen and break as discussed in Boyd (1980): since the slope in the square-well initial condition is already infinite, the wave would break immediately. When breaking is prevented by diffusion, as in Burger’s equation, it is well known that the asymptotic solution is the so-called triangular wave (Whitham, 1974, p. 106): the leading edge is almost vertical and the trailing edge is a flat slope so that the shape of the disturbance is a right triangle. Zabusky and Kruskal (1965) and Jeffrey and Kakutani (1972) discovered that the corresponding result when dispersion prevents breaking is that the disturbance splits up into a set of solitons with a triangular envelope: for all large times, the solitary waves lie within the triangle determined by the line tangent to the peaks of the tallest and shortest solitons and the vertical line through the middle of the tallest soliton. Thus, whenever the initial disturbance evolves to or begins with a steep leading edge, and the product of the height with the square of the width is large,
 TABLE 4. The wavenumber $B$ and nonlinear phase speed correction $v$ for the lowest three symmetric Rossby waves. The solitons for the $n$th mode form the

$$\eta = A_n \text{sech}^2[B(\xi - c \tau)].$$

For purposes of comparison, the amplitudes have been normalized by setting

$$a_n = A_n E_n^{1/2},$$

where

$$E_n = \int_0^\infty u_n^2 + \phi_n^2 dy,$$

where $u_n$ and $\phi_n$ are the lowest order fields defined by (3.21) and (3.22) but with the factor of $\eta(\xi, t)$ omitted. (Since $v \rightarrow 0$ in the nondispersive limit, this is an approximate energy normalization.)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$E_n$</th>
<th>$B$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.96</td>
<td>0.613</td>
<td>x$^{1/2}$</td>
</tr>
<tr>
<td>3</td>
<td>606x</td>
<td>0.699</td>
<td>x$^{1/2}$</td>
</tr>
<tr>
<td>5</td>
<td>75,400x</td>
<td>0.785</td>
<td>x$^{1/2}$</td>
</tr>
</tbody>
</table>

a triangular envelope of solitons will result as shown in Fig. 7.

All the explicit case studies were done for the gravest $n = 1$ mode because 1) it is likely to receive more energy than the higher order modes and therefore will be the most nonlinear, and 2) for a given amplitude, solitons of higher modes have a zonal scale and nonlinear phase speed similar to the $n = 1$ soliton. In Section 3 the KDV equation was derived in terms of unnormalized modes. This simplifies the algebra but also magnifies the apparent differences between the solitons of different modes because the maximum geopotential of the unnormalized modes are 8.06 ($n = 3$) and 91.8 ($n = 5$) times greater than that for the lowest mode. To compare solitons of different $n$, it is therefore necessary to normalize them in some fashion. Table 4 presents such a normalized comparison between the “wave-numbers” and nonlinear corrections to the phase speed. One can see that for a given normalized amplitude, the higher order solitons have progressively smaller zonal scales and smaller corrections to the linear phase speed, but the differences are not very large.

In summary, there are two principal conclusions. First, the oceans are wide enough so that nonlinearity will significantly modify long equatorial Rossby waves. Second, these modifications can be heuristically described by the principles of “soliton dominance” and “crest-and-trough democracy” described above. Both are consequences of the fact that nonlinearity narrows and raises crests and broadens and lowers troughs. Thus, instead of evolving into a weakly dispersed disturbance, rather dominated by the leading crest or trough, as in linear theory, a long Rossby wave packet will fission into one or more narrow, fast-moving solitons which will carry off much or most of the energy plus a trailing wavetrain—always beginning with a trough—which is strongly dispersed and much more spread out than its linear counterpart.

7. The modified Korteweg-de Vries equation

The first step in solving the MKDV equation is to transform it from the general form

$$\eta_t + E \eta^2 \eta + F \eta_{xxt} = 0,$$

(7.1)

to the canonical form

$$v_t - 6v^2 v_x + v_{xxx} = 0.$$  (7.2)

As for the KDV equation, let

$$t = \tau$$

(7.3)

and then the transformation can be accomplished via

$$x = \frac{\text{sgn}(F)}{|F|^{1/2}} \xi,$$  (7.4)

$$v = \frac{|E|^{1/2}}{6|F|^{1/6}} \eta,$$  (7.5)

provided that, as for the $n = 2$ Rossby mode, $F$ and $E$ are of opposite signs.¹ For the $n = 2$ equatorial Rossby wave, this implies

$$x = -2.964\xi,$$  (7.6)

$$v = 2.294\eta.$$  (7.7)

The parameter $r(0)$, which as shown in the previous section is the crucial measure of the nonlinearity of the wavetrain, is [see Eq. (5.15a)]

$$r_0 = \tanh \left[ \int_{-\infty}^{\infty} v(x,0) dx \right]$$  (7.8)

$$= -\tanh \left[ 6.80 \int_{-\infty}^{\infty} \eta(\xi,0) d\xi \right].$$  (7.9)

If one chooses $\epsilon = 0.12$ and $b = 0.394$, the same values (and shape) as for the “prototype” $n = 1$ KDV soliton discussed in the previous section, one finds

$$|r_0| = 0.9995.$$  (7.10)

This is sufficiently close to 1 to suggest that for physically plausible initial conditions, nonlinearity can play a major role in the asymptotic solution for the $n = 2$ Rossby wave. Fig. 8, which compares the linear and nonlinear wavetrains for this case at $t = 160$ (roughly the length of time for the $n = 2$ mode to cross the Pacific) bears this out.

Since the nonlinear wavetrain solution has been already given in (5.14) and since (7.2) admits no

¹ If $E$ and $F$ have the same signs, then (7.1) cannot be transformed into (7.2), so one actually needs two canonical forms for the MKDV. The other, which is identical to (9.2) except for the sign of the nonlinear term, allows soliton solutions while (9.2) does not.
solitons, there is little point in discussing the MKDV solution further except to note three final points. First, the MKDV equation may be solved numerically by the same algorithm used for the KDV case studies of the previous section. Second, although it admits no isolated solitons, Eq. (7.2) does have permanent solutions in the form of nonlinear shocks given by (Jeffrey and Kakutani, 1972)

$$v(x,t) = \pm A \tanh[A(x - 2At)]$$

(7.11)

Because such disturbances are unbounded, they cannot be fitted into the inverse scattering framework and it is not known under what circumstances they could be generated. Third, Ablowitz et al. (1979) give a full description of the MKDV wavetrain.

8. Nondispersive (ultralong) Rossby waves

When the zonal length scale is very long or, in general, whenever

$$\eta_t \gg \eta_{tec}$$

(8.1)

initially, the Rossby wave will evolve for a time as if it were nondispersive. The ultimate development has already been indicated in discussing the top-hat/square-well example in Section 8: as the wave crest steepens, dispersion eventually will become important, and the end result will be one or more solitons. If more than one, the peaks will all be approximately tangent to a single straight line for all large time.

For small times, however, the dispersive third derivative term of the KDV equation can be neglected to reduce it to

$$\eta_t + E\eta_{tec} = 0$$

(8.2)

in the moving coordinate system. This equation has been thoroughly discussed in Boyd (1980). To apply the strained coordinates solution given there for the Kelvin wave to (approximately) nondispersive Rossby waves, one merely needs to replace $E$ and the linear phase speed $c$ for the Kelvin wave (1.22 and 1.00, respectively) in the solution by the corresponding values for the given Rossby mode (e.g., $-1.54$ and $-1.5$ for $n = 1$). Thus, no separate treatment of nondispersive Rossby waves is necessary.

9. Triad interactions and long Rossby waves

Because of the complexity and variety of the resonant triad interactions between equatorial waves of different modes and classes, no definitive treatment of them has yet been given, so it is obviously not possible to say the last word on the relationship between such triads and equatorial solitary waves. Nonetheless, it is possible to at least make a beginning because Domaracki and Loesch (1977) and Loesch and Deininger (1979) have made a similar beginning in classifying the equatorial triads themselves.

In principle, a long Rossby wave, such as are superposed to form the solitons discussed earlier, can participate in an enormous variety of triads. For present purposes, however, the most important question is stability: under what circumstances does the long Rossby wave transfer the bulk of its energy to the two secondary waves of a given triad?

Domaracki and Loesch (1977) provide a powerful criterion for answering this question: only one of the three waves in a given triad is unstable, and that one is always the mode with the largest absolute value of frequency. By systematically applying this criterion, one can show that a long Rossby wave is only unstable with respect to a very limited class of triads.

For simplicity, the exact dispersion relations will be replaced by approximations accurate for small $k$. Since inclusion of the neglected terms only makes the high gravity wave frequencies higher and the low Rossby wave frequencies lower, for purposes of applying the Domaracki-Loesch comparison criterion, one in fact sacrifices nothing by making these approximations.

The dispersion relations are

$$w_{\text{Rossby}} = \frac{-k}{2n + 1},$$

(9.1)

$$w_{\text{gravity}} = \pm(2n + 1)^{1/2},$$

(9.2)

$$w_{\text{mixed Rossby-gravity}} = -1,$$  

(9.3)

$$w_{n=0 \text{ gravity}} = 1,$$

(9.4)

$$w_{\text{Kelvin}} = k.$$  

(9.5)

Comparing (9.1) with (9.2)–(9.4) immediately gives one striking conclusion: a long Rossby wave with $k < 0$ can never be unstable with respect to any triads involving gravity waves or mixed Rossby-gravity waves. The reason is simply that the frequencies of the gravity waves are so high, and the frequency of the Rossby wave so low, that the Rossby wave cannot be that member of the triad with the lowest absolute frequency. (For short Rossby waves with $k > 1$, this may not be true—but the multiple scales
analysis of earlier sections specifically assumed that
$k \ll 1$.

To examine the stability of Rossby waves in triads
involving only Rossby and Kelvin waves, however,
it is necessary to look at the resonance conditions,
which are

\begin{equation}
 w_p = w_1 \pm w_2, \tag{9.6}
\end{equation}

\begin{equation}
 k_p = k_1 \pm k_2, \tag{9.7}
\end{equation}

where the signs in (9.6) and (9.7) must be the same
and where the subscript $p$ denotes the primary wave
and the subscripts 1 and 2 denote the two secondary
waves. The convention implicit in (9.6) and (9.7)
is that

\begin{equation}
 k_p, k_1, k_2 > 0. \tag{9.8}
\end{equation}

This differs from that used by Longuet-Higgins
and Gill (1967), who allowed wavenumbers of either
sign, but this is strictly a bookkeeping difference.

The Domaracki-Loesch criterion for instability of
the primary wave demands that

\begin{equation}
 |w_p| > |w_1|, \ |w_2|. \tag{9.9}
\end{equation}

In addition, they prove that the mode numbers must
satisfy the symmetry condition that their sum is
odd, i.e.,

\begin{equation}
 n_p + n_1 + n_2 = \text{odd integer}. \tag{9.10}
\end{equation}

If the triad is composed purely of Rossby waves,
then (9.9) implies that only the plus sign in (9.6) and
(9.7) is allowed. The resonance conditions are

\begin{equation}
 \frac{-k_p}{2n_p + 1} = \frac{-k_1}{2n_1 + 1} - \frac{k_2}{2n_2 + 1}, \tag{9.11}
\end{equation}

\begin{equation}
 k_p = k_2 + k_3. \tag{9.12}
\end{equation}

Solving them for $k_1$ and $k_2$ in terms of $k_p$, $n_p$, $n_1$
and $n_2$ gives

\begin{equation}
 k_1 = k_p \left( \frac{2n_1 + 1(n_p - n_1)}{2n_1 + 1(n_p + n_1)} \right), \tag{9.13}
\end{equation}

\begin{equation}
 k_2 = k_p \left( \frac{2n_2 + 1(n_p - n_2)}{2n_2 + 1(n_p + n_2)} \right). \tag{9.14}
\end{equation}

It is easy to see, however, that unless $n_p$
is intermediate meridional mode number, i.e., unless

\begin{equation}
 n_1 < n_p < n_2, \tag{9.15}
\end{equation}

where the subscript 2 has been assigned to the
secondary wave of higher $n$, then one of the secondary
wavenumbers $k_1$, $k_2$ is negative in violation of (9.8).
This would imply the absurdity that the corresponding
secondary Rossby wave was propagating eastward.
The condition (9.15) is the natural analogue of the
well-known condition (Longuet-Higgins and Gill, 1967)
for midlatitude Rossby triads that only the wave of
intermediate total wavenumber is unstable.

When $n_p = n_1$ or $n_2$, one of the secondary waves
has wavenumber 0, i.e., the “wave” is a component
of the zonally averaged flow. This does not violate
(9.8), but consistency with wave-mean flow
interaction theorems (Boyd, 1976) implies that the explicit
interaction coefficients for such a triad are zero as
has been verified for mid-latitude Rossby waves by
Longuet-Higgins and Gill (1967). Thus, both
inequalities in (9.15) are necessary for instability.

This mode number restriction is a severe one for
the low-order modes. The $n = 1$ Rossby wave is not
unstable with respect to any triads composed solely
of Rossby waves. The $n = 3$ mode is unstable only to
triads including (i) $n = 1$ or $n = 2$ and (ii) $n = 4$ or
higher, subject to the additional restriction (9.10).

When the triad includes one or more secondary
Kelvin waves, the minus sign in (9.6) and (9.7) is
appropriate; there are valid triads with plus signs,
too, but then the primary Rossby wave would not
satisfy the frequency criterion for instability (9.9).
When both secondary modes are Kelvin waves, the
resonance conditions are

\begin{equation}
 \frac{-k_p}{2n_p + 1} = \frac{k_2 - k_1}{2n_2 + 1}, \tag{9.16}
\end{equation}

\begin{equation}
 k_p = k_2 - k_1, \tag{9.17}
\end{equation}

and these are incompatible and have no solution.

The one remaining possibility is a triad composed
of a primary Rossby wave, a secondary Rossby
wave and a Kelvin wave. The resonance conditions are

\begin{equation}
 \frac{-k_p}{2n_p + 1} = \frac{-k_R}{2n_R + 1} - \frac{k_K}{2n_K + 1}, \tag{9.18}
\end{equation}

\begin{equation}
 k_p = k_R - k_K, \tag{9.19}
\end{equation}

where the subscripts $R$ and $K$ have been used to
denote the Rossby and Kelvin waves. These can be solved to give

\begin{equation}
 k_R = k_p \left( \frac{2n_R + 1(n_p + 1)}{2n_p + 1(n_p + 1)} \right), \tag{9.20}
\end{equation}

\begin{equation}
 k_K = k_p \left( \frac{n_p(n_p + 1)}{2n_p + 1(n_p + 1)} \right). \tag{9.21}
\end{equation}

Unless

\begin{equation}
 n_R > n_p, \tag{9.22}
\end{equation}

$k_K < 0$ in violation of (9.8), so the Rossby/Kelvin
triad must satisfy this additional constraint.

Thus, there is no particular difficulty in finding
resonant triads long Rossby waves unstable. However,
there is an escape clause: the analysis above
assumed a continuous spectrum in zonal wavenumber.
In reality, however, the finiteness of the earth (or of the ocean basin) will limit the allowable
wavenumbers. For a Pacific Ocean with a width
of 10,000 km and the value of $H$ chosen in Section
7, the allowed nondimensional wavenumbers are
the integral multiples of 0.185. Since (9.13), (9.14),
(9.20) and (9.21) predict explicit values for $(k_1, k_2)$,
given $(k_p, n_p, n_1, n_2)$, there is only an infinitesimally small probability that the allowed discrete wave-
numbers will exactly satisfy these resonance conditions. Under some circumstances, the permitted wavenumbers may not come even close to resonance.

For the $n = 1$ mode, for example, only the Kelvin/Rossby triads are unstable and (9.20) and (9.21) predict that $k_K$ will be small in comparison with $k_p$. For $n_R = 3$, for example, $k_R$ is only $\frac{1}{12}$ of $k_p$, and no permitted Kelvin wave will be anywhere near this broad when $k_p \ll 1$ itself as implicitly assumed in the theory of ultralong Rossby solitons. When $k_p$ is not small, of course, this argument loses its force, but then the multiple scales analysis of Sections 3 and 4 must fail anyway: a Rossby wavetrain composed of such short waves is no longer “weakly” dispersive.

Even so, if a triad is not in exact resonance, it may still be unstable if the amplitude of the primary wave is sufficiently large. However, unless the triad is close to resonance, “sufficiently large” will be too large for the amplitude of the long Rossby wave to be considered “small,” and again, the multiple scales will fail.

Thus, the overall conclusion is that it is consistent with the assumptions of earlier sections—small amplitude and weak dispersion—to neglect resonant triad instability. There is only a very limited class of triads, out of the huge variety in which waves can participate, which would be unstable for an infinitesimal long Rossby wave even in an unbounded ocean, and none at all in a confined ocean until the amplitude of the Rossby wave, i.e., the soliton, is sufficiently large. To say more than this and to determine the critical amplitude for instability for various nonresonant triplets of waves is not possible without a treatment of equatorial triads sufficiently detailed to warrant a separate article.\(^2\)

---

\(^2\) The alert reader will note that I have used the finiteness of the planet to slay the dragon of triad instability while blithely glossing over its effects on the soliton theory, which also assumed an unbounded domain in $x$. In fact, this is justified by the physics: a finite domain has a profound effect on triad instability but only a mild effect on solitary waves.

The reason is that the KDV equation has permanent, finite-amplitude solutions even on a finite periodic domain. They are called “coidal waves” because their shape is described by the square of the elliptic cosine function. The elliptic functions depend on a parameter $m$ which ranges from 0 to 1. When $m = 0$, the elliptic cosine is the ordinary cosine. As $m \to 1$, however, the elliptic cosine becomes the hyperbolic secant function and the coidal wave passes smoothly into a solitary wave. [Dorvanyak and Loesch (1977) give some illustrations of elliptic functions for $m$ very close to 1—see their Fig. 5.]

Thus, on a large but finite periodic domain, one can think of the solitary wave as a very extreme coidal wave. In point of fact, many numerical models of solitons in an unbounded domain have employed a finite periodic domain, and these have worked splendidly in reproducing the results of the unbounded theory. Thus, the reverse procedure of approximating the physics of a finite ocean by a theory for an unbounded one is not unreasonable, and the discreteness of the zonal wavenumber spectrum will not be a serious problem.

10. Summary and prospectus

The most striking conclusion of this present work is that sudden events which displace the thermocline down off the eastern coast of an ocean will generate Rossby solitons which should be detectable, particularly near the western coast. To put it another way, the numbers fall in the right ballpark: the nonlinear and linear solutions in the Pacific at least will differ significantly by the time the disturbance reaches the Far East even without the assumption of unreasonable amplitudes or of zonal scales or time scales incompatible with an ocean of finite width. Such solitons should be observable not only because they are permanent and will always outrun the dying wavetrain, but also because the solitons suppress the wavetrain itself in some circumstances. For a strong initial condition of thermocline depression, the solitons are not merely the leading and long-lasting things, they are approximately the only things as expressed by the notion of “soliton dominance” illustrated in Section 6.

The theory itself has the virtue of great mathematical simplicity. The coefficients of the Korteweg-deVries and modified Korteweg-deVries equations for each mode are simply numbers, explicit numbers. The solitons for each mode form an analytic one-parameter family involving nothing more arcane than a hyperbolic function. Earlier workers have developed both numerical and approximate analytic methods for calculating the solitons and oscillatory tail that evolve from a given initial condition.

The price for this back-of-the-envelope simplicity, however, is that the physics is too simple. The mean fluid is not at rest, but rather flows as an Equatorial Undercurrent which is both fast and strongly sheared in latitude and depth. The oceans really do have coasts and the depth of the ocean really does vary: the former can lead to complicated wave reflections analyzed (in a linear model) by Moore and Philander (1977) and the latter can cause soliton fission when a solitary wave runs onto a continental shelf (Tappert and Zabusky, 1971; Johnson, 1972). There is viscosity and baroclinicity and . . .

Nonetheless, it is possible to extend the present theory to incorporate both baroclinicity and mean shear at least; none of the physical mechanisms omitted from the simple model used here preclude or prevent disturbances that are or closely resemble solitons. The price of such extensions is to sacrifice the simplicity of the theory given above, and thus the shallow water, equivalent depth model adopted here is the proper and sensible choice for a first attack on the subject. It should be clear, however, that much remains to be done; the work presented here is the first word on the subject rather than the last.

For this reason, one is hesitant to spell out a
precise recipe for observing solitary waves. Nonetheless, it is worth reiterating that it is what Sir Walter Scott called the “big bow-wow stuff” that is relevant to solitary waves: large-amplitude impulsive events such as the relaxation of the trade winds which triggers El Niño or the onset of the southwest monsoon in the Indian Ocean. How sudden is sudden depends on the ocean; because the Pacific is so much wider than the Atlantic and Indian, its time scales—including those for impulsiveness and soliton development—are much longer than those for the other two oceans. Because of the simplicity of the theory, it is easy to determine the prospects of measurable solitary waves in a given case. Besides the criterion listed above, Whitham (1974) gives simple but effective ways of estimating the size and number of solitons generated by a given disturbance. Because of the physical omissions of the theory given here, such simple-minded, back-of-the-envelope estimates are likely to be the most useful in wave-hunting. Small-amplitude randomly forced waves or seasonal changes in the winds that are so slow that the ocean merely passes through a series of quasi-equilibrium states in response—one might call them Jane Austen oceanography—will most definitely not produce detectable solitons, however, even though nonlinear effects, in the form of triad interactions, may be important in determining the spectrum of randomly forced waves.

In summary, the present work has presented a first attack on the problem of equatorial solitary Rossby waves. The results are simple enough and the case studies are sufficiently encouraging to suggest that one can now begin to look for solitary waves in the real ocean, provided that one keeps the limitations of the theory in mind. In particular, in “soliton dominant” situations such as those discussed in Section 6, the solitary waves may be the only detectable Rossby wave response to a sudden event generating such waves on the eastern coast of an ocean. However, the present theory needs to be extended in several directions by incorporation of more physics before it can hope to be a faithful model of events in the real sea.

Acknowledgments. My special thanks go to Stanley Jacobs, who provided me with invaluable references and shared his own vast knowledge of nonlinear waves. Computations were performed on the University of Michigan Amdahl 470. This work was supported by the National Science Foundation under Grant OCE-79-09191.

APPENDIX

Computational Method

The derivation of the coefficients of the KDV and MKDV equations as given in Sections 3 and 4 required prodigious amounts of algebra involving polynomials of degree as high as 30. For this reason, I employed the algebraic manipulation language REDUCE2 to perform the calculations. A thorough discussion of REDUCE2 and algebraic manipulation languages and their uses is given in the review by Barton and Fitch (1972), who list about 200 references for applications to quantum electrodynamics, celestial mechanics and general relativity. Besides saving one from bondage to an adding machine and a ream of foolscap, the REDUCE2 calculations are much more reliable. Not only is the computer virtually error-free in simple addition and multiplication—of course, one can still blunder by giving the machine an incorrect expansion for evaluation—but REDUCE2 also performs differentiation so that one can check computed answers by direct substitution into the original differential equations. The programs for the odd mode (KDV) and even mode (MKDV) required only 80 statements and 192 statements, respectively, including all comments and checks, plus a 129-line library of subroutines common to both.

The numerical calculations of Sections 6 and 7 were performed using the hopscotch algorithm of Greig and Morris (1976). Although designed for the KDV equation, it can be applied to the MKDV equation also merely by altering their definition of $f(u)$ from $u^{1/2}$ to $f(u) = u^{3/2}$.

Following Greig and Morris, I approximated the infinite domain by imposing the conditions $u(x_0) = u_2(x_0) = u(x_1) = u_2(x_1)$ at the finite boundaries $[x_0, x_1]$. Although the number of boundary conditions exceeds the order of the equation, this approach was satisfactory for smooth initial conditions provided the wavepacket did not reach either boundary. When it did, spurious oscillations appeared ahead of the first crest or tallest soliton where the exact solution is exponentially small, and small-scale ripples with wavelengths equal to 2–4 times the grid spacing were superimposed on the main part of the wavetrain. The width of the largest allowable boundary is proportional to the time and inversely proportional to the grid spacing. For the case graphed in Fig. 7, for example, $[-80, 80]$ was completely satisfactory for a spatial

---

3 d’Inverno (1975) gives an amusing illustration of these twin advantages of labor-saving and accuracy with the so-called Bondi metric problem, which is one of the few that have been done both by hand and machine. He notes “originally, the calculation had taken something like six months by hand”, but REDUCE2 needed only 90 seconds on a CDC-7600, and the programming is so simple that this problem has been widely used to compare the efficiency of different languages (Cohen et al., 1976). In addition, d’Inverno corrects no fewer than six errors in the published hand calculation.

4 For the top-hat/square-well profile, which is discontinuous, small ripples were observed both ahead of and on the wavetrain even for times small enough so that the numerical solution was $<10^{-16}$ within 10 units of either boundary. The ripples could easily be removed by smoothing, however, since the general shape of the wavetrain was smooth and well-defined even when the raw graph was ugly and jagged looking.
grid spacing \( h = 0.4 \) right up to \( \tau = 100 \), but for \( h = 0.2 \), only up to \( \tau = 50 \). For the fine grid, spurious ripples with an amplitude of \( \sim 1\% \) the depth of the initial trough appeared by \( \tau = 100 \).

For “soliton dominant” cases, these small-scale oscillations were not much of a problem; when there was a large wavetrain, the ripples seemed to be much more noticeable. Since Greig and Morris ran only cases with smooth initial conditions and little or no wavetrain, this probably explains why they did not remark on this problem. It is in any event more a cosmetic difficulty than a physical one except in extreme cases. Since one knows the ripples in front of the wavetrain are spurious, one can simply ignore them. The two-grid-length waves can be removed by applying a low-pass smoother (Shapiro, 1970) to the computed results.

To avoid the need for smoothing and yet give accurate results, the case illustrated in Fig. 7 required a domain of \([-160, 160]\) with 1600 grid intervals \( (h = 0.2) \) and 3200 time steps (grid spacing in time \( t = 1/32 \)) and took 51 s on the University of Michigan Amdahl 470 at a cost of \$5.10 (deferred priority). If one were willing to accept small-amplitude (O(2%)) two-grid-length ripples, however, one could double \( h \), increase the time step by a factor of 8, and reduce the domain by a factor of 2 to decrease the total computational expense by a factor of 32. Although it is only one-dimensional and therefore not a great challenge, the KDV equation, from a numerical analysis viewpoint, is nonetheless one of the thornier members of the tribe.

REFERENCES


