Equatorial Solitary Waves. Part 3: Westward-Traveling Modons

JOHN P. BOYD

Department of Atmospheric and Oceanic Science, The University of Michigan, Space Research Building, 2455 Hayward, Ann Arbor, MI 48109

(Manuscript received 20 January 1984, in final form 26 October 1984)

ABSTRACT

Boyd's previous work on equatorial Rossby solitary waves, which derived the Korteweg-deVries equation using the method of multiple scales, is here extended in several ways. First, the perturbation theory is carried to the next highest order to (i) assess the accuracy and limitations of the zeroth-order theory and (ii) analytically explore solitons of moderate amplitude. Second, using the refined theory, it is shown that Rossby solitary waves will carry a region of closed recirculating fluid along with the wave as it propagates provided that the amplitude of the wave is greater than some (moderate) threshold. The presence of such closed "streaklines", i.e., closed streamlines in a coordinate system *moving* with the wave, is an important property of modons in the theory of Flierl, McWilliams and others. The "closed-streakline" Rossby waves have many other properties in common with modons including (i) phase speed outside the linear range, (ii) two vortex centers of equal magnitude and opposite sign, (iii) vortex centers aligned due north-south, (iv) propagation east-west only and (v) a roughly circular shape for the outermost closed streakline, which bounds the region of recirculating fluid. Because of these similarities, it seems reasonable to use "equatorial modon" as a shorthand for "closed-streakline, moderate amplitude equatorial Rossby soliton," but it should not be inferred that the relationship between midlatitude modons and equatorial solitary waves is fully understood or that all aspects of their behavior are qualitatively the same. Kindle's numerical experiments, which showed that small amplitude Rossby solitons readily appear in El Niño simulations, suggest—but do not prove—that the very large El Niño of 1982 could have generated equatorial modons.

1. Introduction

Boyd (1980) used the method of multiple scales to show that long, small amplitude Rossby waves evolved in longitude and time according to the Korteweg-deVries equation, which implies the existence of Rossby solitary waves. Kindle (1982) supported this theory through direct numerical solution of the full nonlinear shallow water wave equations with no approximations whatsoever. In one experiment without boundaries, he initialized the flow with an exact single soliton[1] (according to the theory). To the extent that the Korteweg-deVries model is accurate, the soliton should have propagated across the Pacific without change in shape or form. Kindle found that this was indeed the case: the little upstream ripples that represented the error in the theory were so small as to be almost invisible on his perspective plots. The differences between the linear and nonlinear solutions were very large, however.

In a second set of experiments, Kindle added rigid boundaries and simulated model El Niños of various intensities. He found—as predicted by the inverse scattering transform theory of the Korteweg-deVries equation—that *general* wind stresses readily excite solitons and that very strong El Niños normally generate a train of two or three solitary waves.

A soliton wavetrain may seem almost a contradiction in terms—the very word "solitary" seems to convey the image of a single peak isolated by vast distances from any other dynamical phenomenon, "alone on a wide, wide sea" as Coleridge would put it—but in fact it is the only reasonable description of Kindle's results. First, the individual peaks, although still overlapping a bit by the time the disturbance reaches the western Pacific, closely resemble in shape, height and speed the exact solitons predicted by the Korteweg-deVries theory for very large times when the peaks are completely separated. This is not surprising: the solitons decay *exponentially* fast in x away from the center of the soliton, so the solitons can be quite close together and yet have almost no overlap at all. Figures 5, 6, and 7 of Boyd (1980) stress this very point. Second, the nonlinear solution in Kindle's simulations is quite different from the corresponding linear results because the nonlinearity has pinched the height field into tall, narrow peaks

[1] "Soliton" will be used as a synonym for "solitary wave" throughout this paper. The purist's view that "soliton" should be reserved for solutions obtainable by the "inverse scattering" method is falling out of favor because it is now known that the special properties implied by "inverse scattering" apply only approximately to solitary waves in nature.

instead of the low, rapid ripples created by linear dispersion. "Soliton formation" is the most reasonable description for this nonlinear steepening/compression since the peaks would separate into discrete solitary waves—with almost no further change in shape—if the ocean were only wider.

Thus, Kindle's results add credibility to the purely analytic theory of Boyd (1980) and fill in some of its gaps. Recent developments, however—one theoretical and one observational—have raised some additional questions which this new work will partially resolve.

The observational development is the 1982 El Niño, which is perhaps the biggest on record. Its magnitude suggests that it may well generate very strong Rossby solitons whose size raises two questions, one mathematical and one physical. The mathematical issue is the rather obvious uncertainty as to whether the analytical theory of Boyd (1980), which was derived through a *small* amplitude perturbation expansion, can be legitimately applied to the moderate to large amplitude solitary waves created by such a powerful El Niño. Extending the perturbation theory to the next highest order, as done here, is the simplest analytical way of resolving this uncertainty.

The physical question is one posed to the author by Dennis Moore (private communication, 1982): can a Rossby soliton isolate a closed region of recirculating fluid and drag it bodily along with the wave? Such a closed recirculation region would be of great interest in particle tracer studies and the like. For small amplitude, the answer is no: the Rossby wave has a finite linear speed, and closed "streaklines" (that is, streamlines in a coordinate system moving with the wave) can occur only if the eddy zonal velocity is greater than (and in the same direction as) the phase velocity. For a sufficiently large solitary wave, however, the direct calculations presented below show that closed streaklines are inevitable.

The theoretical development that also motivated this work is the creation of the theory of modons by Stern (1975) and Flierl *et al.* (1980). Modons come in several varieties, but the simplest type, "riderless" dipoles in the terminology of Flierl *et al.* (1980), have many properties in common with the equatorial Korteweg-deVries type solitons as summarized in the abstract. On the other hand, the property of closed "streaklines", which is essential to midlatitude modons, is almost accidental for equatorial solitons, which propagate as stable, unchanging solitary waves regardless of whether the amplitude is or is not large enough for "closed streaklines". In view of the interest in modon theory in connection with Gulf Stream rings, it seems timely (J. McWilliams, private communication, 1983) to explore these similarities and differences between equatorial solitary waves and modons.

Section 2 extends the perturbation series of Boyd (1980) to the next highest order while Section 3 gives the end results and the ensuing error estimates for the zeroth-order theory. Section 4 discusses the relationship between equatorial solitary waves and modons while the final section is a summary and prospectus. The Appendix discusses an issue which is not part of a direct comparison between modons and solitons, but which is nonetheless important in the real ocean: the role of boundaries. The Appendix shows that it is consistent, at least to zeroth order, to neglect boundaries in applying the Korteweg-deVries theory.

2. Derivation

Boyd (1980) applied the method of multiple scales to the usual nondimensional shallow water waves on the equatorial beta-plane. The wave amplitude was assumed to be $O(\epsilon)$ where $\epsilon \ll 1$ so that nonlinear effects could be computed perturbatively. In order to balance dispersion against nonlinearlity, it is necessary to compute dispersive effects perturbatively also, which can be done by assuming that the zonal length scale is $O(\epsilon^{1/2})$. The zeroth-order equations then describe *linear,* nondispersive Rossby waves. For the lowest symmetric mode, $n = 1$, the solution is (Eqs. (3.35a–e) of Boyd, 1980)

$$v^0 = A_\xi(\xi, \tau)2ye^{-(1/2)y^2}, \tag{2.1a}$$

$$u^0 = A(\xi, \tau)\frac{(-9 + 6y^2)}{4}e^{-(1/2)y^2}, \tag{2.1b}$$

$$\phi^0 = A(\xi, \tau)\frac{(3 + 6y^2)}{4}e^{-(1/2)y^2}. \tag{2.1c}$$

The function $A(\xi, \tau)$, where ξ and τ are the "slow" longitude and time variables, is undetermined at this order. The zeroth-order solution is extremely simple—it contains no free parameters—because the dispersion is postponed to next order; the latitudinal structure functions are the zero wavenumber limit of those for linear Rossby waves of general wavenumber.

The solution of the first-order perturbation equations requires two steps which each yield different information about the wave. The first step is to impose a "nonsecularity" condition on the zeroth-order solution so that the first-order solution is bounded. This demands that $A(\xi, \tau)$ satisfy the Korteweg-deVries equation as shown in Boyd (1980). The second step, which is accomplished here for the first time, is to actually *solve* the first-order perturbation equations to obtain the first-order corrections to what is given in (2.1).

The general first-order equations can be written [(3.24)–(3.26) in Boyd, 1980] as

$$-cu_\xi^1 - yv^1 + \phi_\xi^1 = -u_\tau^0 - u^0u_\xi^0 - v^0u_y^0, \tag{2.2a}$$

$$yu^1 + \phi_y^1 = cv_\xi^0, \tag{2.2b}$$

$$-c\phi_\xi^1 + u_\xi^1 + v_y^1 = -\phi_\tau^0 - (u^0\phi^0)_\xi - (v^0\phi^0)_y, \tag{2.2c}$$

where c is the linear, nondispersive phase speed.

Equations (2.2) are completely general, but in the rest of this work, we will specialize $A(\xi, \tau)$ to a single, isolated solitary wave. Furthermore, since the $n = 1$ mode is dominant in Kindle's (1982) simulations and leads the wavefront across the ocean (being the fastest Rossby waves), explicit calculations will be limited to this one mode. For the $n = 1$ wave,

$$c = -\frac{1}{3}, \tag{2.3}$$

$$A(\xi, \tau) = 0.771B^2 \operatorname{sech}^2\{B[x - (c + c^1)t]\}, \tag{2.4}$$

where

$$c^1 = -0.395B^2 \tag{2.5}$$

is the nonlinear correction to the linear phase speed.

To simplify notation and also to shift into the reference frame moving with the wave, let

$$X \equiv x - \left(-\frac{1}{3} - 0.395B^2\right)t. \tag{2.6}$$

Since (2.2a) and (2.2c) involve u_ξ^1 and ϕ_ξ^1 whereas (2.2b) involves these same quantities in undifferentiated form, it is useful to differentiate the y-momentum equation so that three unknowns are consistently taken as $(u_\xi^1, v^1, \phi_\xi^1)$. Then

$$-cu_X^1 - yv^1 + \phi_X^1 = F_1, \tag{2.7a}$$

$$yu_X^1 + \phi_{yX}^1 = F_2, \tag{2.7b}$$

$$-c\phi_X^1 + u_X^1 + v_y^1 = F_3, \tag{2.7c}$$

where for the $n = 1$ mode

$$F_1 = c^1 A_X \frac{(-9 + 6y^2)}{4} e^{-(1/2)y^2}, \quad -AA_X \frac{(81 + 60y^2 - 12y^4)}{16} e^{-y^2}, \tag{2.8a}$$

$$F_2 = cA_{XXX} 2y e^{-(1/2)y^2}, \tag{2.8b}$$

$$F_3 = c^1 A_X \frac{(3 + 6y^2)}{4} e^{-(1/2)y^2}, \quad -AA_X \frac{(-30 + 24y^2 - 24y^4)}{16} e^{-y^2}. \tag{2.8c}$$

[Note that the zeroth-order phase speed appears in (2.8b) while only c^1 appears in (2.8a) and (2.8c).]

As in Boyd (1983), the most efficient way to solve (2.7) is to expand both the unknowns and the forcing functions F_i as series of Hermite functions, i.e.,

$$v^1 = e^{-(1/2)y^2} \sum_{n=0}^{\infty} v_n(X) H_n(y) \tag{2.9}$$

and similarly for the other variables, to introduce sum and difference variables defined by

$$S^1 = \phi^1 + u^1, \tag{2.10a}$$

$$D^1 = \phi^1 - u^1 \tag{2.10b}$$

to replace ϕ^1 and u^1 as unknowns, and finally to use the "raising" and "lowering" operators to reduce the problem down to a set of *algebraic* equations for the X-dependent Hermite coefficients of S_X^1, v^1 and D_X^1. One equation is the singlet (forced Kelvin wave)

$$S_{0X} = \frac{3}{4}(F_{10} + F_{30}) \tag{2.11}$$

plus the triplets, one for each *odd* integer m,

$$\begin{vmatrix} \frac{4}{3} & -1 & 0 \\ (m+1) & 0 & -\frac{1}{2} \\ 0 & 2m & -\frac{2}{3} \end{vmatrix} \begin{vmatrix} S_{X,m+1} \\ v_m \\ D_{X,m-1} \end{vmatrix} = \begin{vmatrix} F_{1,m+1} + F_{3,m+1} \\ F_{2,m} \\ F_{3,m-1} - F_{1,m-1} \end{vmatrix}, \tag{2.12}$$

where $F_{i,m}$ is the mth Hermite coefficient of F_i, $i = 1, 2, 3$. Details of the derivation of (2.11) and (2.12) are omitted since, apart from minor changes in notation, they are identical with the "forced long wave" equations (3.27) and (3.28) of Boyd (1983) except for the replacement of the group velocity $c_g(k)$ by $c = -1/3$, the zeroth-order linear phase (and group) speed of the wave. The solution of (2.11) is trivial; defining

$$\sigma_m \equiv F_{1,m+1} + F_{3,m+1}, \tag{2.13}$$

$$\delta_m \equiv F_{3,m-1} - F_{1,m-1}, \tag{2.14}$$

the solution of (2.12) is

$$S_{X,m+1} = \frac{3}{2(m-1)}\left(m\sigma_m + \frac{1}{2}\delta_m - \frac{2}{3}F_{2,m}\right), \tag{2.15a}$$

$$v = \frac{(m+1)\sigma_m + \delta_m - \frac{4}{3}F_{2,m}}{(m-1)}, \tag{2.15b}$$

$$D_{X,m-1} = \frac{3}{2(m-1)}\left[2m(m+1)\sigma_m + (m+1)\delta_m - \frac{8}{3}mF_{2,m}\right]. \tag{2.15c}$$

Since the X and y dependence of forcing function F_i is separable, computing the coefficients needed for (2.15) via Gauss-Hermite quadrature is trivial. The "linear" terms in (2.8), i.e., those proportional to A_X rather than AA_X, can be represented as exact finite sums of Hermite functions; the corresponding solutions are given in the next section. The nonlinear terms generate an infinite series because of the extra factor of $\exp(-\frac{1}{2}y^2)$, but the series converges rapidly; the coefficients are listed in Table 1.

The one snag is that the denominators in (2.15) all vanish for $m = 1$, which in turn demands that the

TABLE 1. The Hermite series coefficients for the first-order latitudinal structure functions $V^1(y)$, $U^1(y)$, and $\phi^1(y)$. The first part of the table lists the unnormalized coefficients while the second gives the corresponding normalized coefficients. The latter are useful for error analysis: the normalized Hermite functions are rather tightly bounded by 0.785, so the error in truncating these series is less than the sum of the absolute values of the neglected coefficients.

	v_n	u_n	ϕ_n
		Unnormalized	
0	0	1.789276	−3.071430
1	0	0	0
2	0	0.1164146	−0.3508384E-1
3	−0.6697824E-1	0	0
4	0	−0.3266961E-3	−0.1861060E-1
5	−0.2266569E-2	0	0
6	0	−0.1274022E-2	−0.2496364E-3
7	0.9228703E-4	0	0
8	0	0.4762876E-4	0.1639537E-4
9	−0.1954691E-5	0	0
10	0	−0.1120652E-5	−0.4410177E-6
11	0.2925271E-7	0	0
12	0	0.1996333E-7	0.8354759E-9
13	−0.3332983E-9	0	0
14	0	−0.2891698E-9	−0.1254222E-9
15	0.2916586E-11	0	0
16	0	0.3543594E-11	0.1573519E-11
17	−.1824357E-13	0	0
18	0	−0.377013E-13	−0.17023E-13
19	0.4920951E-16	0	0
20	0	0.35476E-15	0.1621976E-15
21	0.630264E-18	0	0
22	0	−0.2994113E-17	−0.1382304E-17
23	−0.1289167E-19	0	0
24	0	0.2291658E-19	0.1066277E-19
25	0.1471189E-21	0	0
26	0	−0.1178252E-21	−0.1178252E-21
		Normalized	
0	0	2.382126	−4.089104
1	0	0	0
2	0	0.4383689	−0.1321112
3	−0.6177913	0	0
4	0	−0.8523089E-2	−0.4855271
5	−0.1869916	0	0
6	0	−0.3641001	−0.7134305E-1
7	0.9868437E-1	0	0
8	0	0.2037216	0.7012758E-1
9	−0.3547170E-1	0	0
10	0	−0.9094727E-1	−0.3579110E-1
11	0.1113516E-1	0	0
12	0	0.3722793E-1	0.1558009E-1
13	−0.3169244E-2	0	0
14	0	−0.1454972E-1	−0.6310672E-2
15	0.8037800E-3	0	0
16	0	0.5524352E-2	0.2453068E-2
17	−0.1658389E-3	0	0
18	0	−0.2056290E-2	−0.9284615E-3
19	0.1654515E-4	0	0
20	0	0.7543701E-3	0.3449009E-3
21	0.8685529E-5	0	0
22	0	−0.2736964E-3	−0.1263585E-3
23	−0.7992579E-5	0	0
24	0	0.9843520E-4	0.4580056E-4
25	0.4468420E-5	0	0
26	0	−0.3515194E-4	−0.1646057E-4

numerators in (2.15) all vanish, too. Although there would seem to be three conditions, in fact the numerators of (2.15b) and (2.15c) are two and four times the numerator of (2.15a), respectively, so the single nonresonance condition is

$$\frac{3}{2}(F_{1,2} + F_{3,2}) + \frac{3}{4}(F_{3,0} - F_{1,0}) - F_{2,1} = 0, \quad (2.16)$$

which explicitly is (using $\sigma_2 = 0.21688$ and $\delta_0 = 6.48094$)

$$-c^1A_X - (8/81)A_{XXX} - 1.5366AA_X = 0. \quad (2.17)$$

This is the usual Korteweg-deVries equation for the $n = 1$ mode [(3.35e) in Boyd (1980)], specialized to the case of a single soliton.

The mathematical reason for the singular denominators in (2.15) is that the equations in (2.12) are not linearly independent for $m = 1$. Consequently, one can arbitrarily choose one of the unknowns, say v_1, and then solve two of the three equations in (2.12) [the top and bottom rows, for example] to obtain $S_{2,X}$ and $D_{0,X}$. Physically, this arbitrariness arises because the free $n = 1$ Rossby wave is an exact solution of (2.12). To exclude the free wave from the first order solution so that it consists of forced waves only, it suffices to set

$$v_1(X) = 0. \quad (2.18)$$

The remaining steps are then simple: solving for $S_X{}^1$ and $D_X{}^1$, taking their sum and difference to recover $u_X{}^1$ and $\phi_X{}^1$, and finally integrating in X to obtain u^1 and ϕ^1. The results are presented and discussed in the next section.

3. First order solution: Discussion

The first order solution for $n = 1$ mode is

$$v^1 = AA_XV^1(y), \quad (3.1)$$

$$u^1 = c^1A\tilde{U}^1(y) + A^2U^1(y), \quad (3.2)$$

$$\phi^1 = c^1A\tilde{\phi}^1(y) + A^2\phi^1(y), \quad (3.3)$$

where

$$A(X) = 0.7713\,\mathrm{sech}^2(BX), \quad (3.4)$$

$$c^1 = -0.395B^2, \quad (3.5)$$

$$X = x - (c + c^1)t = x - \left(-\frac{1}{3} - 0.395B^2\right)t, \quad (3.6)$$

$$\tilde{U}^1(y) = \frac{9}{16}(3 + 2y^2)e^{-(1/2)y^2}, \quad (3.7)$$

$$\tilde{\phi}^1(y) = \frac{9}{16}(-5 + 2y^2)e^{-(1/2)y^2}, \quad (3.8)$$

and where $V^1(y)$, $U^1(y)$, and $\phi^1(y)$ are given by infinite Hermite series whose coefficients are listed in Table 1.

Figures 1, 2, and 3 graph the latitudinal structure functions at both zeroth- and first-order. Each of the first-order functions must be multiplied by either $A(X)$ or c^1—both of which are proportional to B^2—in comparison to its zeroth-order counterpart to reconstruct the total v, u and ϕ. One finds

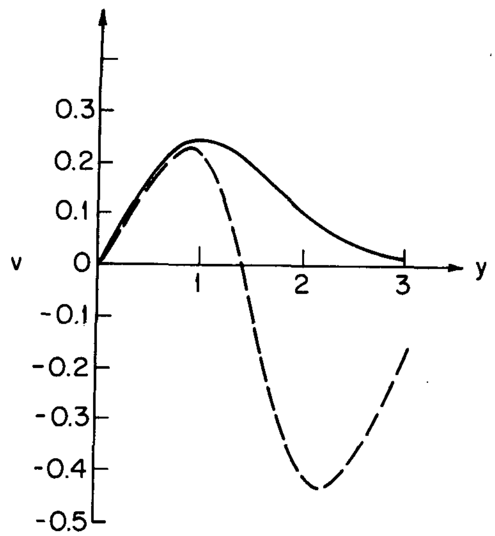

FIG. 1. A comparison of the first- and zeroth-order latitudinal structure functions for the north–south current. Solid line is $v^0/(5A_X)$; dashed line is $V^1(y)$ (note scaling factor).

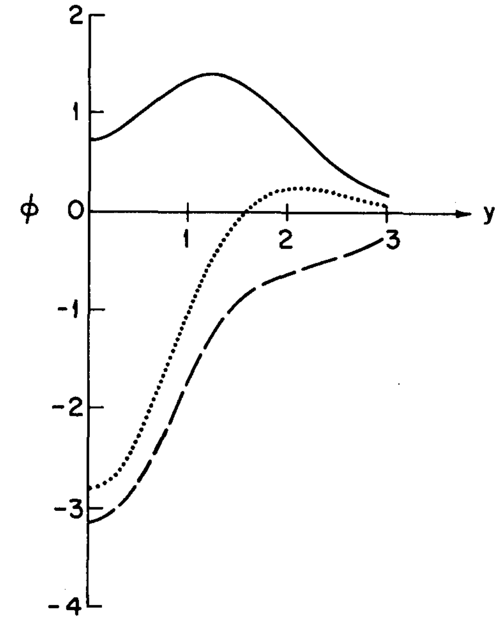

FIG. 3. A comparison of the first-order and zeroth-order latitudinal structure functions for the height. Solid line is ϕ^0/A; dashed line is $\phi^1(y)$; dotted line is $\tilde{\phi}^1(y)$.

$$\frac{\max|v^1/A_X|}{\max|v^0/A_X|} = 0.277B^2, \qquad (3.9)$$

$$\frac{\max|u^1/A|}{\max|u^0/A|} = 0.334B^2, \qquad (3.10)$$

$$\frac{\max|\phi^1/A|}{\max|\phi^0/A|} = 0.924B^2. \qquad (3.11)$$

These ratios are probably the best measures of the relative importance of the first- and zeroth-order terms since the pointwise ratios of u^1/u^0, etc., fluctuate wildly because u^0 and ϕ^0 have zeros in y. There is a partial cancellation between the terms proportional to c^1A and those proportional to A^2 in u^1 and ϕ^1 that keeps the ratios comparatively small.

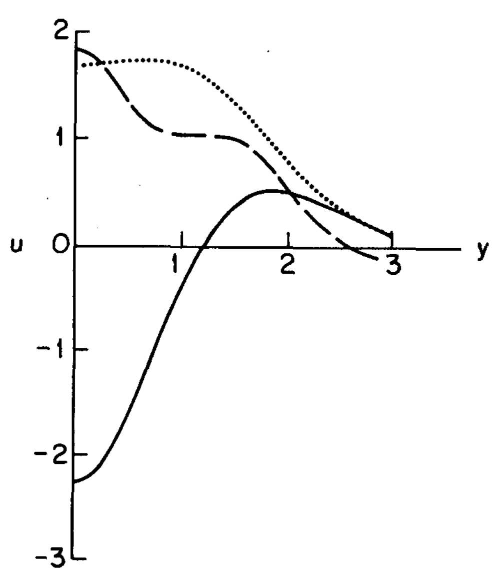

FIG. 2. A comparison of the first-order and zeroth-order latitudinal structure functions for the zonal current. Solid line is u^0/A; dashed line is $U^1(y)$, dotted line is $\tilde{U}^1(y)$ (multiples c^1A).

The first-order height field is far larger than either of the velocity corrections. The reason is that the strong advection tends to fill in the (relative) trough at the equator, but the currents are only weakly affected by ϕ^1, which decays rapidly away from the equator so that it is small at the ridges of ϕ^0 at $y = 1.22$. For comparison purposes,

$$\max|\phi|^0 = 1.093B^2, \qquad (3.12)$$

which is only slightly larger than the ratio in (3.11). Thus, the relative error in ϕ^0 (if we assume the absolute error approximately equals ϕ^1) is roughly equal to the nondimensional magnitude of ϕ^0 itself.

For the "prototype" soliton of Boyd (1980), $B = 0.394$, which implies that the error ratios (3.9) through (3.11) are only 4.3, 5.2 and 14.3%, respectively. Since these ratios are rather conservative in the sense of being defined by maximums, one is not surprised that Kindle (1982) found only very small differences between the zeroth-order analytic solution and the exact numerical circulation.

For the "modon" graphed in Fig. 4 and discussed in the next section, $B = 0.6$ and the error ratios are still only 10.0, 12.0 and 33.3%. Only the first two ratios—both small—are relevant to Fig. 4.

The conclusion is that the zeroth-order solutions of Boyd (1980) remain accurate even in the "modon-like" range of moderate amplitude. One can therefore

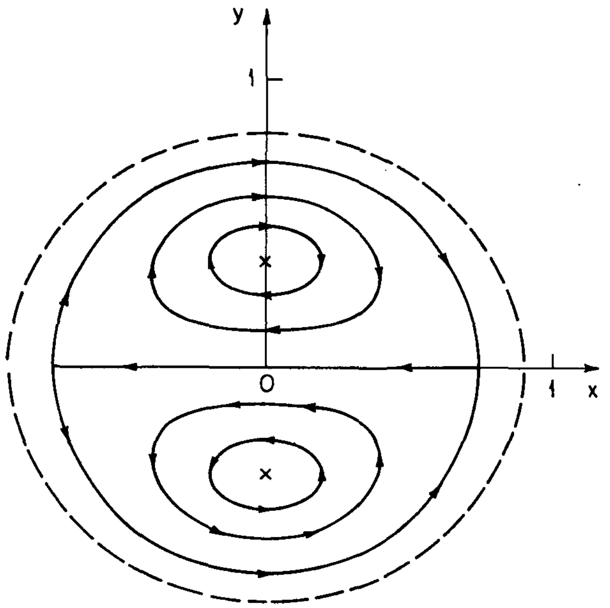

FIG. 4. Streamlines in the coordinate system moving with the wave ("streaklines") for the $n = 1$ Rossby soliton for $B = 0.6$. The dashed line is the outermost closed streakline as computed using the zeroth-order solution alone. The solid curves are the streaklines calculated using the sum of the zeroth- and first-order terms. The solid near-circle of largest radius is the outermost closed streakline. The crosses mark the interior stagnation points.

use the analytic theory to discuss "equatorial modons" as done in the next section.

4. Solitons and modons

The "streaklines", which, following Flierl *et al.* (1980), are defined to be the streamlines in a coordinate system moving at the nonlinear phase velocity of the wave, can be calculated by numerically integrating the ordinary differential equation

$$\frac{dy}{dx} = \frac{v_s}{u_s}, \tag{4.1}$$

where $v_s(x, y)$ and $u_s(x, y)$, the velocities in the moving system, are given by

$$u_s = u(x, y) - c - c^1, \tag{4.2a}$$

$$v_s = v(x, y) \tag{4.2b}$$

with $u(x, y)$ and $v(x, y)$ evaluated from the Korteweg-deVries soliton solutions given in earlier sections. By choosing several different arbitrary starting points, it is trivial to map the general pattern of the circulation. The outermost closed streakline, which is the boundary between the recirculating fluid that always travels with the wave and the other external fluid particles that are eventually left behind as the wave propagates westward, must intersect both equatorial stagnation points. These can easily be found by simply printing out $u_s(x, y = 0)$; the appearance of these stagnation points as the soliton amplitude is increased marks the threshold of the transition from nonrecirculating waves to those with closed streaklines.

Figure 4 shows the results of such an analysis for $B = 0.6$. The solid streaklines were computed using the first-order solution including the one bounding the recirculation region; for purposes of comparison, the corresponding outermost closed streakline as calculated from u^0 and v^0 alone is shown as the dashed line. One sees that the first-order corrections shrink the recirculation region only a little. The threshold for recirculation, which is $B = 0.532$ at first order, is only 7% smaller at zeroth order. Thus, as found through different criteria in Section 3, the formulas of Boyd (1980) are accurate even well into the moderate amplitude, recirculation regime. "Moderate" is the appropriate description for the wave illustrated in Fig. 4 since, taking the undisturbed thermocline to be 100 meters, the solitary wave deepens the thermocline to about 140 m at the two off-equatorial peaks of the wave height ϕ. What is most striking about Fig. 4, however, is its close resemblance to the modon graphs of Flierl *et al.* (1980). In each case (assuming the modon is what Flierl *et al.*, 1980, call "riderless"), the solitary wave is a dipole vortex: the vorticity pattern north of a line drawn parallel to the x axis and through the two stagnation points is the exact mirror image (with reversed sign) of the vorticity pattern to the south. In each case, the mutual interaction of the two vortices accelerates the total nonlinear phase speed outside of the linear range. The phase speed for linear equatorial Rossby waves for example, lies in the range $-1/3 \leq c \leq 0$ for all wavenumbers and mode numbers while the total phase speed of the nonlinear $n = 1$ soliton satisfies $c + c^1 < -1/3$. In each case, the north–south phase velocity is 0: because of the beta effect, the soliton/modon can only propagate along a latitude circle even though it is possible for (dispersing) packets of linear Rossby waves to also move north–south. Finally, the streakline bounding the recirculating flow is roughly circular in shape for both solitons and modons. Strictly speaking, the outermost streakline is a circle in the theory of Flierl *et al.* (1980), but this is probably an artifact of the analysis: McWilliams and Zabusky (1982) give examples of numerically generated modons for which the analytic theory with its perfectly circular boundary is not applicable. The outermost closed streakline in Fig. 4 is not exactly circular, but a circle can approximate the curve to within about 2%.

These similarities are very striking, but there is also at least one noteworthy difference. Closed streaklines are absolutely essential for modons; numerical experiments by McWilliams *et al.* (1981) have shown that when a modon is weakened by dissipation so that there is no longer a recirculation region, the modon rapidly disperses as a packet of linear Rossby waves. In contrast, experiments with the Korteweg-deVries equation plus dissipation show that the soliton gradually becomes shorter and broader but never enters a second, more rapid decay phase dominated by dispersion. No experiments have yet been done with dissipative equatorial solitons *per se*, but the fact that there is a smooth and uneventful transition for undamped solitons from closed streaklines to no closed streaklines suggests that they will behave like any other Korteweg-deVries solitons when weak damping is applied. In summary, the existence of a region of recirculation is life-and-death to the modon, but is irrelevant to the structure and longevity of an equatorial solitary wave. The equatorial trapping, which prevents the north–south dispersion that mid-latitude linear waves experience, is probably responsible for this difference.

Another difference (perhaps) is that modon theory is richer than equatorial soliton theory in the sense of including (i) "riders" and (ii) eastward-traveling waves. Riders, to use the term coined by Flierl *et al.* (1980), are radially symmetric terms that may be added to the basic dipole modon. These riders have been important in attempts to apply modon theory to Gulf Stream rings since the rings appear to be monopoles, i.e., vortices of a single sign, rather than dipoles. The equatorial analogue would be solitons that lack the symmetry with respect to the equator of the $n = 1$ solitons discussed in Sections 2 and 3.

The eastward-traveling modons have the positive and negative vortices interchanged so that the mutual vortex pair interaction drives the wave east instead of west. Far from the center of the vortex, a solitary wave or modon has negligible amplitude and the dynamics is linear, so it is easy to show—for either the midlatitude beta-plane or the equatorial beta-plane—that a necessary condition for a disturbance to be isolated by decaying exponentially in the zonal direction is that c lie outside the range allowed to linear waves. Thus, solitary waves whose nonlinear interaction drives them eastward must have $c > 0$ on either beta-plane. This in turn implies that if an equatorial analogue to an eastward-traveling modon exists—and there seems no reason why not—a small amplitude perturbation theory like that of Sections 2 and 3 will not work. Eastward-traveling modons must have large amplitude so that the vortex pair interaction is strong enough to drive the wave towards the east in spite of the strong tendency of all long, linear Rossby waves to propagate in the opposite direction.

It is possible, however, to create a simple heuristic model that is highly suggestive that eastward-traveling equatorial modons exist. The Korteweg-deVries equation is always derived as a long-wave approximation. Benjamin *et al.* (1972) and Peregrine (1966) independently pointed that it is therefore legitimate to replace the Korteweg-deVries by any other model equation which has the same solutions in the long-wave limit. Specifically, they suggested the so-called "Regularized Long Wave" (RLW) equation, which for the $n = 1$ equatorial Rossby wave is

$$A_t - \frac{8}{27}A_{xxt} - \frac{1}{3}A_x - 1.5366AA_x = 0. \quad (4.3)$$

(Note that this is not written in a moving coordinate system.) For linear sine waves, the dispersion relation for (4.3) is

$$c = \frac{-1}{3 + \frac{8}{9}k^2} \text{ [RLW]}, \quad (4.4)$$

where k is the zonal wavenumber. When Taylor expanded in powers of k, (4.4) agrees to within $O(k^4)$ with the corresponding dispersion relation for the *KdV* equation

$$c = -\frac{1}{3} + \frac{8}{81}k^2 \text{ [}KdV\text{]}. \quad (4.5)$$

More important, however, the RLW formula (4.4) closely agrees with the approximation

$$c = \frac{-1}{3 + k^2} \quad (4.6)$$

which was shown in Boyd (1983) to have a relative error of no worse than 1.5% for all k, not just $k \ll 1$. The factor of $8/9$ in (4.4) gives a better approximation than (4.6) for small k at the expense of a larger error for large k, but the linear RLW dispersion relation (4.4) is nevertheless uniformly accurate to within no worse than 12% for all k.

The reader may allow himself a chuckle: ironically, the "Regularized Long Wave" equation is really more appropriate for Rossby waves than for the long water waves to which it was first applied. The RLW model captures the linear dispersion very accurately, so the only approximation embodied in it is the nonlinear term.

Kasahara (1977), Cane and Sarachik (1976) and Ripa (1982) have independently shown that the linear eigenfunctions of the shallow water wave equations are a complete set. The RLW equation can then be interpreted as a straightforward application of Galerkin's method in which the expansion is truncated to but a single Rossby mode. The multiple scales perturbation theory provides a rigorous justification for such a drastic truncation in the long wave/small amplitude limit: the $n = 1$ mode is in resonance with itself, but not with any of the modes which are discarded by the truncation. For solitary waves which do not have long zonal scales or small amplitudes, however, the only justification we can offer for the RLW model is to point to its success for westward-traveling modons as shown above.

The reason for pursuing this alternative RLW model is that it predicts eastward- as well as westward-traveling solitary waves. The soliton solutions of (4.3) are

$$A(x-ct) = -2.31B^2c\,\mathrm{sech}^2[B(x-ct)], \quad (4.7)$$

where

$$c = \frac{-1/3}{[1 - 1.184B^2]}. \quad (4.8)$$

When $B > 0.92$, c is positive and $A(x - ct)$ is everywhere negative, just as expected from the analogy with eastward-traveling modons. As noted earlier, $|A(0)|$ must be greater than some minimum so that the nonlinear interaction of the two vortices can push the disturbance eastward in spite of the westward phase velocity of all linear Rossby waves: (4.6) and (4.8) show that

$$-A(0) > 0.65 \quad (4.9)$$

for all RLW solitary waves with $c > 0$.

Unfortunately, even this limiting amplitude is more than double that of the westward-traveling modon graphed in Fig. 4, so the RLW model is likely to be only qualitatively accurate. Still, it is highly suggestive. Numerical work to confirm the existence of eastward-traveling equatorial modons is now in progress, using the RLW solution (4.7)–(4.8) to provide a first guess for the Newton's iteration.[2]

Finally, one can mention "monopoles" (McWilliams, 1982). Because these are only quasi-solitons, *slowly* radiating energy away, they are not part of modon theory except in the generalized sense. Nonetheless, one would like to know whether equatorial analogues exist or not, and the analysis of Sections 2 and 3 gives no answer. Since the monopoles—at least in the middle latitudes—must propagate north–south as well as east–west, the strong equatorial trapping that confines equatorial waves to a low latitude channel would seem to preclude equatorial monopoles unless the monopoles are somehow reflected off the turning latitudes to pursue a zig-zag course within a limited latitudinal range.

5. Summary

The extension of the perturbation theory to first order shows that the zeroth-order theory of Boyd (1980) is accurate even for solitons of such large amplitude that closed regions of recirculation appear in a coordinate system moving with the wave. The similarities between these $n = 1$ equatorial Korteweg-deVries solitons and the midlatitude dipole modons of Flierl *et al.* (1980) are so striking that it seemed reasonable to entitle this paper "Equatorial Modons".

However, the presence or absence of closed "streaklines" and recirculation does not have the life-or-death significance for equatorial solitary waves that it does for modons, and the perturbative analysis given here is silent on the subject of equatorial counterparts to "riders," eastward-traveling modons, and "monopoles." There seems no reason why large amplitude eastward-traveling equatorial solitons could not exist, so this work has raised as many questions for the future as it has resolved.

Acknowledgments. I thank Dennis Moore and Jim McWilliams for asking the questions that inspired this study, and also Joseph Tribbia for helpful comments. This work was supported by the National Science Foundation through Grants OCE8108530 and OCE8305648.

APPENDIX

Solitons and Continental Boundaries

The author's earlier work on Rossby solitons, Boyd (1980), completely ignored the effects of boundaries. The practical reasons for this are that the Korteweg-deVries equation can be solved by the inverse scattering method only when the spatial domain is infinite and that a full nonlinear treatment of boundary effects would be extremely complicated. What Boyd (1980) should have done—and did not—was to point out that this neglect of boundaries is in fact consistent with the perturbation theory to lowest order.

To recognize this, one must examine the key assumptions of the perturbation scheme, which are the following: (i) the wave amplitude is $O(\epsilon)$ where $\epsilon \ll 1$ and (ii) the longitudinal scale of the wave is $O(\epsilon^{-1/2})$. Together, these assumptions imply that the effects of both nonlinearity and dispersion are so weak that neither can effect the wave except upon the long time scale $O(1/\epsilon)$. However, the soliton travels away from a continental boundary such as Peru with a phase speed which is $O(1)$. Since its width is $O(\epsilon^{-1/2})$, the soliton will be clear of the coast (except for its exponentially small tail) after a time of only $O(\epsilon^{1/2})$—before nonlinear effects and dispersion can significantly alter it. Thus, to within an error of $O(\epsilon^{1/2})$, which is the magnitude of the nonlinear and dispersive changes in the wave during the time it is close to the coast, one can calculate the boundary generation of solitary waves using a *linear, nondispersive* theory. In this context, nondispersive means that in the linear computation, one may make the so-called "long wave" approximation of perfect geostrophic balance in latitude. One can then take this solution at $t = O(\epsilon^{1/2})$ and use it as the initial condition for the Korteweg-deVries model, which will then describe how nonlinearity and dispersion modify the wave packet as it propagates further towards the west.

---

[2] The RLW model is equally appropriate as a substitute for the Korteweg-deVries equation in other solutions such as Rossby waves in a midlatitude shear flow (Redekopp, 1977) or a flow over bottom topography (Malanotte-Rizzoli, 1982). It is a little surprising that there have been no attempts to investigate eastward-traveling Rossby solitons except within the framework of Flierl *et al.* (1980) modon models.

This computational scheme is not recommended for practical calculations; since $c = -1/3$ for the $n = 1$ mode, the errors in applying it are more like $O(3\epsilon^{1/2})$. The physically interesting solitons and modons, i.e., those that are well-formed by the time they arrive in the western Pacific, must have ϵ large enough, i.e., sufficiently large amplitude, so that the errors in this linear boundary generation/Korteweg-deVries model would be substantial. Nonetheless, this model is of considerable conceptual value. It shows us that the neglect of boundary effects or wave generation at the boundaries as in Boyd (1980), is consistent with the lowest-order perturbation theory. It also reminds us of the physical reason for this consistency: for long, small amplitude Rossby waves, dispersion and nonlinearity can alter the wave packet only at a slow rate. However, since these changes in the wave packet are secular, i.e., they steadily accumulate as the packet propagates towards the west, the wave can be altered by O(1) by weak dispersion and nonlinearity—but only after the wave has long since left the eastern continental boundary behind.

For equatorial Rossby waves of large enough amplitude to behave like modons, the arguments given above may not be quantitatively accurate; one would really like to calculate how nonlinearity and dispersion alter the generation of the waves at the boundaries instead of being content to consider the effects only far downstream. Unfortunately, such a calculation would be very messy and complicated, and therefore is not attempted here.

Nonetheless, one can qualitatively predict what will happen: just as for smaller amplitude, any wind stress that tends to deepen the thermocline will generate one or more solitons at the eastern boundary of the ocean, and if the amplitude of these solitons is large enough, they will behave like modons as they propagate towards the west.

REFERENCES

Benjamin, T. B., J. L. Bona and J. J. Mahony, 1972: Model equations for long waves in nonlinear dispersive systems. *Phil. Trans. Roy. Soc. London,* **A272,** 47.

Boyd, J. P., 1980: Equatorial solitary waves. Part 1: Rossby solitons. *J. Phys. Oceangr.,* **10,** 1699–1717.

——, 1983: Equatorial solitary waves. Part 2: Envelope solitons. *J. Phys. Oceangr.,* **13,** 428–449.

Cane, M., and E. S. Sarachik, 1976: Forced baroclinic ocean motions. I. The linear equatorial unbounded case. *J. Mar. Res.,* **34,** 629–655.

Flierl, G., V. Larichev, J. McWilliams and G. Reznik, 1980: The dynamics of baroclinic and barotropic solitary eddies. *Dyn. Atmos. Oceans,* **5,** 1–41.

Kasahara, A., 1977: Numerical integration of the global barotropic primitive equations with Hough harmonic expansions. *J. Atmos. Sci.* **34,** 687–701.

Kindle, J. C., 1982: On the generation of Rossby solitons during El Niño. *Hydrodynamics of the Equatorial Ocean,* J. C. J. Nihoul, Ed., Elsevier, 353–368.

Malanotte-Rizzoli, P., 1982: Planetary solitary waves in geophysical flows. *Advances in Geophysics,* Vol. 24, Academic Press, 147–224.

McWilliams, J. C., 1982: Isolated vortices in planetary fluid motions. *Proc. Ninth U.S. Congress on Applied Mechanics,* 283–289.

——, and N. J. Zabusky, 1982: Interactions of isolated vortices. I. Modons colliding with modons. *Geophys. Astrophys. Fluid Dyn.,* **19,** 107–227.

——, G. Flierl, V. Larichev, and G. Reznik, 1981: Numerical studies of barotropic motions. *Dyn. Atmos. and Oceans,* **5,** 219–238.

Peregrine, D. H., 1966: Calculations of the development of an undular bore. *J. Fluid Mech.,* **25,** 321.

Redekopp, L. G., 1977: On the theory of solitary Rossby waves. *J. Fluid Mech.,* **82,** 725–745.

Ripa, P., 1982: Nonlinear wave–wave interactions in a one-layer reduced gravity model on the equatorial beta-plane. *J. Phys. Oceangr.,* **12,** 97–111.

Stern, M. J., 1975: Minimal properties of planetary eddies. *J. Mar. Res.,* **33,** 1–13.