PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Dynamic Response of the Indian Ocean to Onset of the Southwest Monsoon

M. J. Lighthill

Phil. Trans. R. Soc. Lond. A 1969 **265**, 45-92
doi: 10.1098/rsta.1969.0040

DYNAMIC RESPONSE OF THE INDIAN OCEAN TO ONSET OF THE SOUTHWEST MONSOON

By M. J. LIGHTHILL, Sec.R.S.

(*Received* 28 *November* 1968)

CONTENTS

The linearized theory of unsteady wind-driven currents in a horizontally stratified ocean is applied to the northern part of the Indian Ocean. This is argued to be a suitable area for detailed application and evaluation of the theory because (i) the theory has certain advantages near the equator (for example, influence of detailed bottom topography is reduced, thermoclines are somewhat less variable in character, and speeds of baroclinic propagation are enhanced relative to current speeds), and (ii) the wind-stress pattern undergoes a well marked change with onset of the Southwest Monsoon, a change to which the pattern of currents shows a more or less identifiable, and rather quick, response which may be compared with theoretical predictions. Response is predicted to be found principally in two modes as far as vertical distribution of current is concerned; to a somewhat lesser extent in the barotropic mode with uniform distribution, and to a somewhat greater extent in the first baroclinic mode with current distribution as in figure 7, concentrated predominantly in the uppermost 200 m (see Appendix for detailed analysis of the modes appropriate to the equatorial Indian Ocean).

Of particular interest is the strong Somali Current, that flows northward along the Somali coast only during the northern hemisphere summer (after monsoon onset) but during that time is comparable in volume flow (about 5×10^7 m^3/s) to other western boundary currents such as the Gulf Stream. Detailed discussion of the application of linearized theory to equatorial oceans with western boundaries leads the author to conclude, both in the barotropic (§ 2) and baroclinic (§ 4) cases, that 'wave packets' of current pattern reaching such a boundary deposit the 'flux' they carry (velocity normal to the boundary integrated along it) in a boundary current which rather rapidly takes a rather concentrated form. Linear theory with horizontal transport neglected indicates that such flux requires of the order of 10 days to become concentrated in a current of 100 km width, but that thereafter it continues to become still thinner; however, with horizontal transport included, a steady-state finite thickness of current is reached. In reality, nonlinear effects would play an important additional part in limiting steady-state current thickness to the observed 100 km or thereabouts, but the time scale required to bring the thickness down to this value is probably given reasonably well by linear theory.

Calculations for a zonal distribution of winds, which rather rapidly make a reversal of direction and increase of strength somewhat north of the Equator characteristic of the onset of the Southwest Monsoon, predict westward propagation of both barotropic and baroclinic wave energy at comparable speeds of the order of 1 m/s; the marked contrast here with other oceans (in the comparability of speeds) is given particularly detailed study. Calculations indicate that the barotropic signal is considerably distorted (figure 3) by the fact that low-wavenumber components reach the western boundary first. Baroclinic propagation takes the form of special planetary-wave modes concentrated near the equator (§3), of which perhaps four, delivering flux patterns depicted in figure 5, and possessing wave velocities of 0.9,

Vol. 265. A. 1159. (Price £1. 5*s*.; U.S. $3.25) 6 [Published 30 October 1969

0.55, 0.4 and 0.3 m/s towards the west, are specially relevant to generation of the Somali Current. Peak surface flows in that current are predicted to be influenced about three times as much by this baroclinic propagation as by the barotropic.

Theory indicates 1 month (of which two-thirds is needed for propagation of current patterns and one-third for their concentration in a boundary current) as characteristic time scale for formation of the Somali Current (see figure 6 in particular for the calculated baroclinic component) in contradistinction to the 'decades' predicted by the same type of theory in mid-latitude oceans (Veronis & Stommel 1956). Observations do, indeed, make clear that the time scale is not significantly more than 1 month, although the possibility that it might be still less cannot yet be decided on the basis of observational evidence. The flow is calculated as reaching 40% of a typical maximum value (observed in August) already within 1 month of monsoon onset (May), even though no effect of wind stress acting within 500 km of the coast has been taken into account. The linearized theory predicts the current as reaching as far north as 6° N or 7° N, but nonlinear terms are generally found in computational studies (Bryan 1963; Veronis 1966) to bring about some 'inertial overshoot' in concentrated boundary currents, which may explain why the current does not in fact separate until about 9° N.

1. Introduction

The ocean over which there is the strongest seasonal fluctuation in the prevailing winds is the northern Indian Ocean, which therefore is particularly suitable for studies of the dynamic response of ocean current patterns to changes in wind-stress patterns. Such dynamic response is studied in the present paper principally with regard to the part of the Indian Ocean north of about the 10° S parallel, although from time to time comparisons are made with the characteristics of other oceans.

The regular variation in surface currents between summer and winter in this part of the Indian Ocean, observed by mariners for over 1000 years (Warren 1966), is depicted in all good atlases: Schott (1935), for example, portrays it excellently. Soon after the onset (normally in June) of the Southwest Monsoon, when winds north of the Equator go into reverse and an extensive region of strong south-west wind comes into being, there is an approximate reversal of most current directions in this area, which oceanographers have normally viewed as related to the changes in prevailing winds. Important knowledge of the corresponding distributions of current in depth was added during the International Indian Ocean Expedition.

Of special interest among currents present only during the northern hemisphere summer is the Somali Current, which flows northward from the neighbourhood of the Equator to about 9° N, where it separates from the coast (Warren, Stommel & Swallow 1966; Swallow & Bruce 1966). Except near its inception this is a strong current with a velocity maximum normally around 2 m/s (and, indeed, Swallow (1967) observed a local surface speed of over 3 m/s near the separation point); a current which like other western boundary currents such as the Gulf Stream and the Kuroshio is regarded as too strong to be a merely local response of the ocean to the local winds. By analogy with those others it tends to be interpreted rather as part of the ocean's dynamic response to the pattern of wind stress over a large part of it.

Even for interpreting patterns of mean current in an ocean such as the Atlantic, Pacific or Antarctic, in terms of mean meteorological input, studies of dynamic response are helpful. They can be used, for example, to find which steady responses to such an input not only are theoretically possible, but could be arrived at by some process in time; while an estimate of such a process's duration may help to give understanding of the mechanics of current generation. In addition, to answer the question of which kinds of variation in meteorological input produce which kinds of variation about the mean current pattern, analysis of dynamic response is essential. However, accurate checks on such analysis are difficult in oceans where the variation in meteorological

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

input is of a highly random character; the northern Indian Ocean, by contrast, performs a 'controlled experiment', every summer, suitable for comparison with theory.

For the important task of estimating time scale of response, a linear theory is very useful. Veronis & Stommel (1956) derived from a linear theory their celebrated estimate that the time scale of the baroclinic part of, say, the North Atlantic Ocean's response to changes in the pattern of wind stress is of the order of decades (see also § 3 below). The Somali Current is markedly baroclinic; that is, there is a strong variation of velocity with depth, associated with horizontal density gradients. Yet it comes into being, not decades after the winds over the northern Indian Ocean change to the Southwest Monsoon pattern, but in a matter of a month at most.

It is important to investigate, therefore, whether a linear theory of dynamic response of the Indian Ocean permits the baroclinic component of the western boundary current to approach full strength in a time scale of order a month, in contrast to the estimate of decades for the North Atlantic. The present paper indicates that proximity to the Equator does indeed cause the time scale to be so greatly reduced. Unfortunately, observations of the time for establishment of the Somali Current have not been made precisely enough in relation to the time of monsoon onset to answer the question whether something like a month is actually needed, as the present theory predicts, or whether by contrast it builds up still more suddenly.

For predicting other aspects of the ocean's pattern of dynamic response, including amplitude and general configuration, a linear theory might be expected not to be so effective, but is found below to give results in quite reasonable accord with observation in the Indian Ocean. Nonlinear effects should certainly be significant in the Western boundary current itself, where the general character of the modification they bring about (a sort of inertial overshoot) is known from comparison of computed steady current patterns with and without nonlinear inertial terms (Bryan 1963; Veronis 1966); such modification would further improve the agreement with observation. Elsewhere, nonlinear effects may be less important because typical current speeds are not large compared with the speeds of propagation of baroclinic response, which are predicted as being much higher near the Equator than elsewhere.

One of the simplest kinds of linearized equation for dynamic response, say to a wind force (F, G) per unit mass of ocean (where the x and y directions are to the east and north respectively), is that for a homogeneous ocean of uniform depth H with turbulent momentum transfer neglected. For this simplified case, on the long-wave approximation, which is closely correct if (as we shall assume) the radian wavenumbers of all significant forcing components are less than about $0.4H^{-1} \simeq 0.1\,\mathrm{km}^{-1}$, the resulting horizontal velocity field (u, v) is uniform with respect to depth and satisfies (in the suffix-derivative notation)

$$u_{tt} - fv_t = F_t + gH(u_{xx} + v_{xy}), \tag{1}$$

$$v_{tt} + fu_t = G_t + gH(u_{xy} + v_{yy}). \tag{2}$$

Equations (1) and (2) are time derivatives of the linearized equations of eastward and northward momentum, respectively, the fluctuations of depth about the undisturbed value H being eliminated through use of the equation of continuity.

The elimination of u from these equations is possible, even though it is essential in ocean-current theory to take into account that the Coriolis parameter f has a positive gradient $\beta = \mathrm{d}f/\mathrm{d}y$ in the northerly direction. It gives

$$v_{ttt} - gH\nabla^2 v_t + f^2 v_t - gH\beta v_x = G_{tt} - fF_t - gH(G_{xx} - F_{xy}). \tag{3}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The terms in (3) without an f or β factor describe forcing of long waves in a non-rotating system; these are the familiar waves with constant free-wave velocity $\sqrt{(gH)}$. In a uniformly rotating system, the terms involving f are added, making the system dispersive, although still isotropic, with a low-frequency cutoff for long waves below the Coriolis frequency f. By contrast, in a system with β-effect ($\beta > 0$), waves of far lower frequency are possible; namely, the dispersive, anisotropic Rossby waves.

When the radian frequencies ω of all significant forcing components, and also the Coriolis parameter f, are small compared with $k\sqrt{(gH)}$, where k is the lowest wavenumber of such a forcing component, only the terms in (3) with the gH factor, that is, the classical Rossby wave terms, are significant. Such a condition can be satisfied in the part of the Indian Ocean near the Equator even if k is as low as $1/(2000\,\mathrm{km})$. For with a typical ocean depth $H = 4\,\mathrm{km}$, this gives $k\sqrt{(gH)} = 10^{-4}\,\mathrm{s}^{-1}$, compared with which any radian frequency ω less than $1/10\,\mathrm{h} = 2.8 \times 10^{-5}\,\mathrm{s}^{-1}$ is already small, while the Coriolis parameter f also is not more than $2.8 \times 10^{-5}\,\mathrm{s}^{-1}$ for latitudes less than 12° (that of, for example, Socotra). Thus the linearized response of a homogeneous ocean near the Equator may rather accurately take the form of Rossby waves, although at higher latitudes the $f^2 v_t$ term in (3) is potentially more important for forcing components of low wavenumber.

Under the conditions given in the previous paragraph, the terms carrying the gH factor dominate not only in (3); equations (1) and (2), similarly dominated, allow us to deduce that the divergence $u_x + v_y$ is small compared with individual velocity gradients. Hence a stream function ψ can be introduced, satisfying approximately

$$\psi_x = v, \quad \psi_y = -u. \tag{4}$$

The terms carrying the gH factor in (3) then take, after integration with respect to x and a change of sign, the standard Rossby form

$$\nabla^2 \psi_t + \beta \psi_x = G_x - F_y, \tag{5}$$

relating rate of change of vertical vorticity to the curl of the applied wind stress. We may note here, also, that Rossby's 'β-plane' approach which analyses equations such as (3) or (5) by ignoring the curvature of the surface described by the (x, y) coordinates, and by taking β constant, is known (Longuet-Higgins 1965) to be particularly accurate near the Equator, where the (x, y) coordinates are locally geodesic and β is stationary.

The above equations for a homogeneous ocean, incorporating the unrealistic feature that currents are distributed uniformly with respect to depth, are of more value than might be expected for real oceans possessing stratification of both density and current velocity. A reasonable linearized model of a real ocean is one based on small perturbations (whose squares are neglected) to a stationary ocean with horizontal stratification of density (that is, with density depending only on the vertical coordinate z), and with turbulent momentum transfer still neglected. In this fairly well-known model (see, for example, Eckart 1960), a simplified analysis of which is given for convenience in the Appendix, any distribution of the horizontal velocity (u, v) with respect to depth can be expanded as a sum of normal modes, each with its own characteristic distribution (proportional to $c_n(z)$; $n = 0, 1, 2, \ldots$) with respect to depth; furthermore, the coefficients of each of these functions $c_n(z)$ in (u, v) satisfy precisely (1) and (2), provided that the depth H is replaced by an 'effective depth' H_n characteristic of the mode in question.

As is well known, the pressure-dependent variations in density have no dynamical effect; only

variations in the pressure-corrected density, here written ρ, that would at atmospheric pressure be associated with the actual stratification of temperature and salinity, are influential dynamically. To a very close approximation (based on the fact that these density variations are at most a few parts per thousand), the normal modes are as follows. The barotropic mode $n = 0$ has $c_0(z) = 1$ (that is, current distribution the same at all depths) and H_0 equal to the actual depth, say H^0, as for the homogeneous ocean. Thus the whole discussion for the homogeneous ocean given above applies without change to the propagation of the barotropic mode. However, the first baroclinic mode already has effective depth H_1 of the order of 1 m at most, and the H_n for higher n are very considerably smaller.

A classical Sturm–Liouville problem determines these baroclinic modes; in mechanical terms, the values of $(1/H_n)$ are the squares of the frequencies of a string stretched to unit tension between $z = 0$ and $z = H^0$, the mass of which in any interval Δz is the decrease $(-\Delta\rho/\rho)$ in relative pressure-corrected density within that interval. There is, accordingly, a general tendency for the $1/H_n$ to increase in rough proportion to the squares of the natural numbers. The slope of the string gives the distribution of current velocity $c_n(z)$ in the nth mode. It is uniform in any well mixed region where ρ is unchanging (corresponding to a massless portion of the string). The integral of $c_n(z)$ between surface and bottom is zero, so that there is zero net volume flow in each baroclinic mode.

This is a special case of the fact that all the eigenfunctions $c_n(z)$ are orthogonal; not only to the barotropic mode $c_0(z) = 1$, but also to each other. Furthermore, any distribution of velocity with depth, provided that it is uniform within any well mixed region, can be expanded as a linear combination of the $c_n(z)$ for $n = 0, 1, 2, \ldots$, whose coefficients can be obtained from this orthogonality property. Also, under conditions of forcing, any distribution of force per unit mass (F,G) with respect to depth can be similarly expanded, and (1) and (2), with H replaced by H_n, then describe the relationship between the coefficients of $c_n(z)$ in (u, v) and (F, G) respectively. For forcing at the surface $z = H^0$ by a wind stress (τ_x, τ_y), this coefficient of $c_n(z)$ in (F, G) is

$$\frac{c_n(H^0)\,(\tau_x, \tau_y)}{\int_0^{H^0} \rho c_n^2(z)\,\mathrm{d}z}. \tag{6}$$

Since, under such forcing at the surface, the coefficient of $c_n(z)$ in (u, v) satisfies equations with (6) as the forcing term, the relative forcing effect of different modes on currents near the surface may be estimated by multiplying (6) by the surface value of $c_n(z)$, namely $c_n(H^0)$. Calculations in the Appendix for oceanic models representing the northern Indian Ocean show this estimate of relative surface-current response in different modes to be greatest for the first baroclinic mode $n = 1$, about 7 % as much for the barotropic mode $n = 0$,† and less than 1 % as much for the higher baroclinic modes. This indicates the irrelevance of the higher baroclinic modes to any explanation of the seasonal nature of the surface currents off the Somali coast; also in the Appendix, their low relevance even to the distribution of current in depth is inferred from the study of response times.

The neglect of turbulent mixing in the model may, as far as vertical mixing is concerned (horizontal mixing is discussed below), be most serious near the free surface, where the sheared Ekman layer generated by wind stress, and penetrating downwards a distance depending on the

† But we shall see (end of §4) that advantages of the barotropic mode over baroclinic modes with respect to type of propagation partly counteract this big disadvantage with respect to forcing term.

vertical exchange coefficient and the Coriolis parameter, is ignored. If, however, this Ekman layer lies within the well-mixed layer near the surface, as is usual, then the velocity independent of depth within the well-mixed layer, which the model predicts, must represent a correct mean value in the layer because the total rate of change of momentum due to vertical exchange integrates to zero. Indeed, Stommel (1958) already showed how divergence within the Ekman layer operates (by the mechanism since christened 'spin-up') to transfer the effect of wind-stress curl to the whole of such a well-mixed region.

We may consider now how the wind-stress pattern associated with the onset of the Southwest Monsoon excites both the barotropic and baroclinic modes of dynamic response in the Indian Ocean. Although the main point of novelty is the predicted short time-scale of baroclinic response, it is necessary to consider also barotropic response, as the Somali Current will be viewed as a combination of the barotropic ($n = 0$) mode and of, in the main, the first baroclinic ($n = 1$) mode. The arguments relating to barotropic response will be given first, since they involve ideas which are relatively more familiar from the existing literature.

It has been pointed out already that the barotropic component of the velocity field (u, v) satisfies the momentum equations (1) and (2), with H equal to the actual depth, provided that wavenumbers of significant forcing components are less than about $0.1\,\mathrm{km}^{-1}$; while, on the other hand, if those wavenumbers exceed about $0.0005\,\mathrm{km}^{-1}$, and associated frequencies are less than about $0.1\,\mathrm{h}^{-1}$, the Rossby equation (5) may be deduced as a good approximation near the Equator. Remembering furthermore that, for a velocity field varying arbitrarily with depth, the barotropic component is obtained by integrating from surface to bottom, we may say in the barotropic case that (1) and (2) relate to momentum integrated from surface to bottom, and that (5) is an equation for rate of change of 'total vorticity'; that is, the vertical component of vorticity integrated from surface to bottom.

Longuet-Higgins (1964) pointed out that Rossby waves, that is, those solutions of (5) with the right-hand side replaced by zero which take a wavelike form, for example, proportional to

$$\exp\{-\mathrm{i}\omega t + \mathrm{i}lx + \mathrm{i}my\}, \tag{7}$$

have the property that their group velocity has magnitude $\beta/(l^2+m^2)$ and makes an angle with the eastward direction twice that which the wavenumber vector (l, m) makes. Pedlosky (1965) pointed out furthermore that, in the common oceanographic case when the wind-stress curl forcing the Rossby waves is distributed approximately zonally, waves with wavenumber vector nearly north or south must in the main be generated, and therefore that their energy must be propagated approximately westward. As far as the barotropic components of ocean currents are concerned, this gives a clearer foundation for understanding the phenomenon of westward intensification than do explanations based on particular approximate nonlinear and/or dissipative steady-flow corrections to the equation of motion.

The onset of the Southwest Monsoon creates a region of strong negative wind-stress curl in a zone centred roughly on the equator; across this zone, the eastward component of wind force F changes from its negative value in the trade winds, somewhat south of the Equator, to a considerably larger positive value in the monsoon itself, somewhat north of the Equator. Since the zone possesses a much greater extent as regards longitude than as regards latitude, Pedlosky's arguments apply, and the ocean's barotropic response, neglecting boundaries (whose effect will be considered presently), consists mainly of waves with l much smaller than m, whose energy moves roughly westward at speed β/m^2. They leave behind a southward motion determined by

the Sverdrup relation (steady form of (5)), balanced by an opposing northward flow which the waves carry progressively farther towards the west.

This response, with boundaries neglected, is calculated in detail from the theory of group velocity in § 2, but we may note here a simple explanation of the general behaviour in terms of the properties of vorticity. When negative wind-stress curl is 'switched on' throughout a long narrow zone extending along the Equator from A to B, at first it causes a build-up of negative vorticity in the zone, generating eastward velocity on the north side of the vortex layer and westward on the south side. The northward return flow near A convects planetary negative vorticity into that region; continuation of this process permits a wave of negative vorticity (with northward flow near its front) to propagate to the west. Meanwhile, a southward return flow near B builds up, and finally becomes steady when it has reached the Sverdrup magnitude, which balances by convection of planetary positive vorticity the creation of negative vorticity by wind stress.

The westward propagation of a pattern of negative vorticity, preceded by northward flow, is slowed down as it approaches a western boundary, essentially by the action of the image vorticity behind the boundary. When a pattern of northward flux approaches the boundary, it tends to pile up there in a stream that becomes gradually more and more concentrated near the boundary, although at a gradually decreasing rate. The way in which a quantity of northward flux ψ_0, which would have reached a boundary $x = 0$ (say) at time $t = t_0$ if the image vorticity had not slowed down its propagation, thereafter proceeds to concentrate itself, is more and more accurately represented by

$$\psi = \psi_0[1 - J_0\{2\sqrt{[\beta x\,(t - t_0)]}\}], \tag{8}$$

a well-known exact solution, to (5) with zero on the right-hand side, whose graph is as in figure 1.

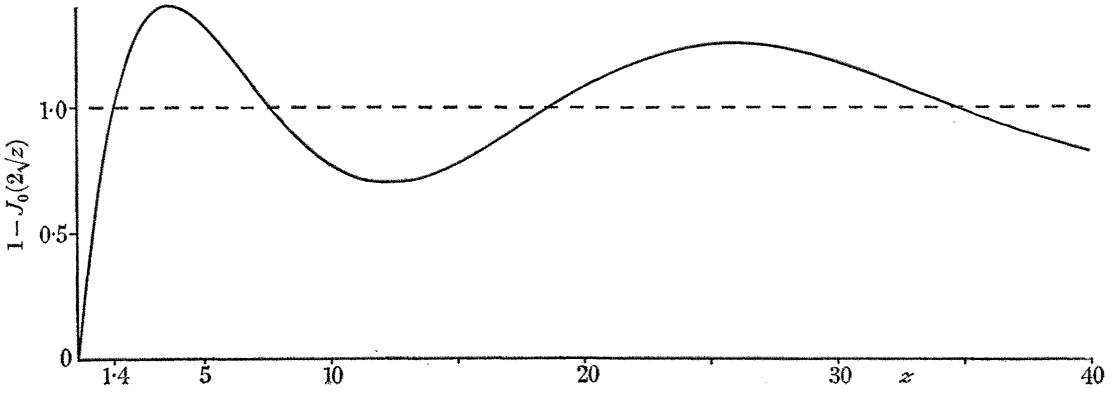

FIGURE 1. Plain line: the function $1 - J_0(2\sqrt{z})$ appearing in (8). This indicates the modification, due to a western boundary, of a flux in the form of a step function (broken line) incident upon it.

The way in which the concentration proceeds is that each element of negative vorticity, together with its image in the boundary, produces northward flow between it and the boundary, convecting planetary negative vorticity into that region, and so intensifying the negative vorticity there. At the same time, southward flow to the east of the element convects planetary positive vorticity into that region and so reduces any negative vorticity there. Figure 1 shows that, at time t, this mechanism for moving the main pattern of negative vorticity nearer the boundary has concentrated the whole flux ψ_0 into a region of width $1.4/\beta(t - t_0)$. There is a time lag of

about 1 week ($t-t_0 = 6\times10^5$ s) before this width is reduced to a typical value 100 km characteristic of the Somali Current or other western boundary currents;† this time lag has to be added to the propagation time lag needed for the main northward-flow signal to arrive at the western boundary.

Any of the mechanisms, used in the steady-flow theory to fix the width of western boundary currents, can come into play after this time lag of about a week to prevent further constriction. For example, it is shown in detail in § 2 how the horizontal-transport mechanism would do this. Other mechanisms are nonlinear and cannot properly be applied to the barotropic component alone, but may be none the less effective for limiting the further narrowing of the combined barotropic–baroclinic Somali Current.

The Rossby-wave theory of propagation of the barotropic component of ocean currents has been criticized for ignoring major perturbing effects of bottom topography, but such effects are very much reduced near the Equator. The equations for the barotropic component have been identified already as integrations from the surface to the bottom of conservation equations, either for momentum, or in the case of (5) for vorticity. The modification that depth variation makes to the left-hand side of (5), if ψ is taken as a stream function for the velocity field integrated from surface to bottom, is an extra term

$$(f/H)\,(H_x\psi_y - H_y\psi_x), \tag{9}$$

due to stretching of vertical vorticity as the barotropic component of flow, independent of z, moves fluid into regions of greater depth. The importance of this term (9), relative to the beta term in (5), is greatly reduced near the Equator, where $|f/\beta|$ is least.

Actually, it will be shown in § 2 that attenuation of the westward propagating energy of waves with $|l| \ll |m|$ is produced, through the interaction term (9), principally if the bottom topography includes Fourier components of relatively high radian wavenumber, around 1/(25 km). Only these make resonant interactions in which energy is transferred to wave modes propagating it in substantially different directions. The relative smallness of the interaction term (9), coupled with the fact that bathymetric data in the northern Indian Ocean (though admittedly incomplete on a small scale) do not indicate notable fluctuations with wavenumbers of this order, suggests that much of the barotropic northward flow pattern, which Rossby waves might, after onset of the Southwest Monsoon, propagate to the west and cause to become concentrated near the boundary, may actually be found there.

Our estimate of the time scale of barotropic response is, in any case, unaffected by these uncertainties regarding its amplitude. A typical propagation speed β/m^2, associated with a median wavenumber $m = 1/(200\,\text{km})$ characteristic of the north–south distribution of wind-stress curl across the equatorial zone, is around 1 m/s. Some low wavenumber components of the signal are propagated faster, but show a somewhat greater tendency to fan out in other directions adjacent to the westward direction. The distance over which the influence of the Southwest Monsoon has to be propagated before affecting the Somali Current is variable, but a median value is around 2000 km, corresponding to a propagation delay (at the speed noted above) of about 3 weeks. The total time scale for barotropic response (adding 1 week as noted already for the process of concentration on the western boundary) becomes, in summary, about 1 month.

† This statement, and equation (8), represent the theory in its simplest form, neglecting among other things the angle (about 40°) of the Somali coast to the north–south direction, but the more detailed and accurate discussion in §2, taking this angle into account, leads to identical conclusions with the predicted time lag revised to about 10 days.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Baroclinic response, at least in oceanic regions *away* from the Equator as studied in the classic paper of Veronis & Stommel (1956), is expected to be far slower. This is because the effective depth H for baroclinic modes is at most about 1 m, which means that in (3) the third term is at least as important as the fourth. When all significant forcing components have frequencies small compared with f (typically about $10^{-4}\,\mathrm{s}^{-1}$ in such regions), and wavenumbers small compared with $f(gH)^{-\frac{1}{2}}$ (typically about $0.03\,\mathrm{km}^{-1}$, or more for the higher baroclinic modes), the first two terms are much smaller than the third, and (3) becomes approximately

$$f^2 v_t - gH\beta v_x = G_{tt} - fF_t - gH(G_{xx} - F_{xy}), \tag{10}$$

a simple first-order equation whose solutions all propagate to the west with a uniform velocity $gH\beta/f^2$. These are Veronis & Stommel's 'non-dispersive baroclinic Rossby waves'.

Even without any restriction on wavenumber, the westward component of group velocity cannot exceed this value $gH\beta/f^2$. For example, waves excited by wind stresses with a predominantly zonal distribution have group velocity directed nearly westward, with magnitude

$$\frac{\beta}{m^2 + (f^2/gH)}, \tag{11}$$

which as m decreases makes a transition from the barotropic value β/m^2 to the uniform value $gH\beta/f^2$ but never exceeds the latter. This maximum speed of propagation of signals to the west (attained for signals of small wavenumber) is plotted as a function of latitude in figure 2 for a typical value $H = 1$ m corresponding to the first baroclinic mode (speeds for higher modes would be much less).

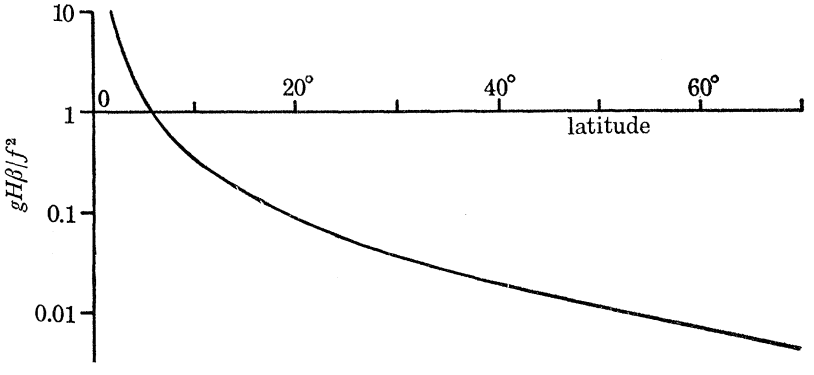

FIGURE 2. Maximum value, $gH\beta/f^2$, of the westward component of group velocity of baroclinic waves of sinusoidal form, plotted in m/s on a logarithmic scale as a function of latitude, for the $n = 1$ mode with effective depth $H = 1$ m.

At high latitudes values are seen to cluster around 0.01 m/s, already two orders of magnitude less than the value quoted above for barotropic modes of wavenumber $1/(200\,\mathrm{km})$. This leads to a characteristic timescale of baroclinic response in such latitudes that is typically a decade rather than a month. By contrast, the speed rises above 0.1 m/s only for latitudes less than 19°, and attains 1 m/s only at latitude 6°. The last figure indicates a possibility (to be followed up below) that rather rapid response of the first baroclinic mode may occur in equatorial regions, where however the wave velocity (11) deduced for constant f is far too variable with latitude across a wavelength to have any real significance even as an approximation.

Before considering the refinements necessary to overcome *this* difficulty, it is of interest in relation to non-equatorial regions to identify the physical significance of the simple first-order equation (10) for low wavenumber response in the first baroclinic mode. The quantity v, either in (3) or in its approximate form (10), represents the coefficient of $c_1(z)$ in the northward velocity. Here, $c_1(z)$ is a velocity mode with only one zero, negative in the denser part of the ocean and positive in the upper less dense part; the shear region between them is predominantly (as the 'string' analogy given earlier implies) where the density gradient is greatest. Thus, positive v means northward flow near the surface, balanced by southward flow near the bottom.

Propagation of such a positive value of v is shown in § 3 to occur as follows. The β-effect tends to lower the level of the region of greatest density gradient if v is positive, so that absolute vorticity may be conserved, as vortex lines in the upper northward moving fluid stretch vertically to compensate for the decreasing angle they make with the vertical, while those in the lower southward moving fluid, whose angle with the vertical is increasing, decrease their vertical extent. If to the west, however, the level of the region of greatest density gradient has not yet fallen, there must result a tilting of density contours, which produces positive v on that side as demanded by the relations for baroclinic steady flow; and so the region of positive v gradually propagates towards the west.

Application of this theory to the North Atlantic, say, may give only a rough order-of-magnitude time scale of response, because the wave velocity predicted (of the order of 0.01 m/s) is so small that any westerly advection of vorticity might significantly supplement it (although hardly increasing its order of magnitude). Bottom topography might also be significant, although less so than in the barotropic case since the ratio of bottom to surface velocities is small in the baroclinic mode if the region of greatest density gradient is much nearer the surface than the bottom. It is reasonable to infer from the theory, however, that response of the North Atlantic to changes in wind stress from month to month is such that in the Gulf Stream some associated barotropic changes, but practically no baroclinic changes, would result.

In an attempt to increase understanding of westward intensification by extending the transient motion reasoning of Pedlosky (1965) to baroclinic components, equation (10) may be used with a zonal distribution of negative wind-stress curl 'switched on' at time $t = 0$. The solution, to close approximation (§ 3), is a southward flow pattern satisfying the baroclinic analogue of the Sverdrup relation, plus an equal and opposite northward flow pattern which propagates steadily to the west at speed $gH\beta/f^2$. When the whole northward flow pattern has finally reached the western boundary, effects similar to those in the barotropic case cause it to cease propagation and form a concentrated stream, reinforcing in the upper part of the ocean (but reducing in the lower part) the barotropic concentrated stream formed much earlier.

On this simple linear theory, higher baroclinic modes would, at much later times, successively appear also. When they had all appeared, the different current modes would be found each in proportion to (6), just as in the representation of surface forcing by wind stress, so that their sum would reconstitute a function concentrated very close to the surface. But the conclusion that currents would be so concentrated cannot be drawn for the real ocean; to set up those higher modes would take an exceedingly long time, before which the first mode would have generated major nonlinear disturbances; for example, the steady baroclinic motion (southward near the surface) set up by wind stress of the order of magnitude found in the North Atlantic demands such a tilt of density contours from east to west as to give them a height variation of order 1 km over the width of the ocean. Such considerations make it not too surprising that the observed

form of wind-driven current is rather more like the sum of a barotropic and a first baroclinic mode.

Difficulties of this nature are much less when similar considerations are applied to the Somali Current; in the time scale on which it appears and disappears the higher baroclinic modes do not have time to be more than marginally excited; furthermore, in such an equatorial region velocities of propagation of the first baroclinic mode are not nearly so small (for reasons already hinted at), although the disturbance amplitudes they carry are about the same, and therefore nonlinear effects are less likely to dominate. Accordingly, a simple linear theory of baroclinic response of an ocean at rest to the onset of a wind-stress pattern similar to that of the Southwest Monsoon is given in §§ 3 and 4, for the purpose of estimating what current changes are produced by that onset.

As already stated, the possibility of fast westward propagation at low latitudes is limited by the fact that the velocity (11), whose value (figure 2) for $m = 0$ increases near the equator to such high levels, can attain them only for rather small m, such that the speed (11) does not remain sensibly constant over a north–south distance even of order $1/m$. Actually, (11) remains sensibly constant only in an equatorial belt whose width is a modest fraction of $m(gH)^{\frac{1}{2}}/\beta$, since $\mathrm{d}f/\mathrm{d}y = \beta$. This is already less than $1/m$ (so that a theory approximating locally by using a wave velocity calculated with f constant has become quite untenable) if $1/m$ exceeds a fraction of $(gH)^{\frac{1}{4}}/\beta^{\frac{1}{2}}$ (which we may therefore expect to be the fundamental length scale for baroclinic propagation near the equator). A fraction of $(gH)^{\frac{1}{2}}$ is suggested as the maximum possible westward propagation velocity (11) by these very crude arguments.

For a more satisfactory study of baroclinic propagation near the equator, it is essential to allow for the variability of f in (3). Nevertheless, β may be taken constant, and f replaced by βy, since their departures from these values over distances of the order $(gH)^{\frac{1}{4}}/\beta^{\frac{1}{2}}$ (which for $H = 1$ m, approximately its maximum value, is 370 km) are negligibly small. With these substitutions, (3) becomes

$$v_{ttt} - gH\nabla^2 v_t + \beta^2 y^2 v_t - gH\beta v_x = G_{tt} - \beta y F_t - gH(G_{xx} - F_{xy}). \tag{12}$$

Equation (12) is solved in § 4, for a wind-stress distribution representing the Southwest Monsoon, after the fundamental modes of propagation (namely, the solutions proportional to $\exp\{-\mathrm{i}\omega t + \mathrm{i}lx\}$ of equation (12) with the right-hand side replaced by zero) have first been studied in § 3.

These fundamental modes satisfy

$$v_{yy} + \left(\frac{\omega^2}{gH} - l^2 - \frac{\beta l}{\omega} - \frac{\beta^2}{gH}y^2\right) v = 0. \tag{13}$$

A solution of (13) is concentrated near the equator (rather than becoming exponentially large as y tends to either $+\infty$ or $-\infty$ or both) only if

$$\frac{\omega^2}{gH} - l^2 - \frac{\beta l}{\omega} = (2M+1)\frac{\beta}{\sqrt{(gH)}} \quad (M = 0, 1, 2, \ldots). \tag{14}$$

The associated solutions of (13) are the parabolic cylinder functions (otherwise well known as the harmonic-oscillator wave functions, taking the form of Hermite polynomials multiplied by $\exp\{-\frac{1}{2}\beta y^2 (gH)^{-\frac{1}{2}}\}$). Equations (13) and (14) were noted already by Blandford (1966) in a general survey of baroclinic modes of propagation. He referred to a peculiarity of the $M = 0$ mode (which in westward propagation would have a velocity of $(gH)^{\frac{1}{2}}$ exactly) but did not

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

observe the important fact (§ 3) that propagation in this mode is actually not possible.† This is because, although a solution for v that is concentrated near the Equator exists for $M = 0$, (1) and (2) show that no such solution for u exists. Rather, in the $M = 0$ case, u must become exponentially large away from the Equator.

The only true propagating modes concentrated near the Equator, then, satisfy (14) for $M = 1, 2, 3, \ldots$. Each has a westward group velocity that is greatest when the east–west wavenumber l is small compared with $\beta^{\frac{1}{2}}(gH)^{-\frac{1}{4}} \geqslant 1/(370\,\mathrm{km})$, as is probably mainly the case for forcing by the monsoon type of disturbance. In such a case, non-dispersive westward propagation occurs at a velocity

$$(gH)^{\frac{1}{2}}/(2M+1), \tag{15}$$

which is given by neglecting the first two terms in (14) (corresponding to neglect of the v_{ttt} and v_{txx} terms in (12)). Thus, a first signal from any disturbance arrives at a speed $\frac{1}{3}(gH)^{\frac{1}{2}}$ (corresponding to $M = 1$), which with the estimated value $H = 0.75\,\mathrm{m}$ obtained in the Appendix for the $n = 1$ baroclinic mode in the equatorial Indian Ocean is 0.9 m/s, matching the typical barotropic propagation velocity quoted earlier; a second signal arrives at a speed $\frac{1}{5}(gH)^{\frac{1}{2}} \simeq 0.55\,\mathrm{m/s}$, and so on. These speeds suggest the possibility of good results on time-scale of baroclinic response, and encourage us to pursue to model further.

Returning to the equation (12) for forced motion, then, we begin by expanding the forcing term on the right-hand side as a function of y in a series of normal-mode solutions of (13), that is, the solutions satisfying (14) for $M = 0, 1, 2, \ldots$. The forced velocity component v is also expanded in such a series. Each coefficient in this series satisfies a certain partial differential equation with x and t as independent variables. We find that the solution for $M = 0$ incorporates no westward-propagating wave, in agreement with the earlier conclusion that no such wave exists for $M = 0$. The solutions for $M > 0$, however, each incorporate a westward-propagating wave.

The equations are made nondimensional by use of the following units of length and time:

$$\text{length } (gH)^{\frac{1}{4}}(2\beta)^{-\frac{1}{2}}, \quad \text{time } (gH)^{-\frac{1}{4}}(2\beta)^{-\frac{1}{2}}. \tag{16}$$

For the first baroclinic mode ($n = 1$), typical values of these units, using $H = 0.75\,\mathrm{m}$, are 250 km and 1 day, respectively. Provided that in these units the frequencies and east–west wavenumber of significant forcing terms are small, as they probably are for the onset of the Southwest Monsoon, the equations for the coefficients v^M of the Mth normal mode in the north–south distribution of v, in terms of the corresponding coefficients F^M for the eastward component of wind stress F, take the simple forms

$$v^0_t - v^0_x = F^1_x - F^1_t, \tag{17}$$

$$3v^1_t - v^1_x = -F^0_x - F^0_t + 2(F^2_x - F^2_t), \tag{18}$$

$$5v^2_t - v^2_x = -F^1_x - F^1_t + 3(F^3_x - F^3_t), \tag{19}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Evidently, (17) has the solution $v^0 = -F_1$, representing a purely local response to the wind-stress field without any propagating component. However, (18), (19), etc., have solutions much more similar to that found in the barotropic case, involving a local response and a balancing motion that is propagated westward at a constant speed, whose value is $\frac{1}{3}$, $\frac{1}{5}$, etc. of the basic unit

† By contrast, Matsuno (1966), in a paper to which the author's attention has been drawn by a referee, has made the corresponding calculations for a barotropic ocean without missing this point. See also Longuet-Higgins (1968).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

of speed $(gH)^{\frac{1}{2}} = 2.7$ m/s. The first signal (term in v^1) arrives, as predicted earlier, at a speed of 0.9 m/s, close to that which we regarded as typical of barotropic propagation; and, as in that case, all the signals on reaching the coast are predicted to cease propagation and become concentrated, with a time lag of about a week to 10 days, in a narrow region.

The sign of the first ($M = 1$) signal is positive, essentially because the F^2 terms in (18) dominate over the F^0 terms. The F^0 terms would be more important only if the region, where the eastward component of wind stress is strong, extended to less than 150 km north of the equator, whereas all the evidence suggests that it starts considerably farther north than this. The earliest arriving signal ($M = 1$) in the first baroclinic mode ($n = 1$) is thus predicted to be symmetrically distributed with northward flow north of the equator and southward flow south of it. The barotropic signal ($n = 0$), however, arrives at about the same time and is entirely northward, enhancing (as far as surface currents are concerned) northward ones and reducing southward ones.

Only a little later, the second equatorial mode ($M = 2$) should arrive and further disturb the symmetry, helping to increase the correspondence with the high degree of asymmetry actually observed. Its sign is negative because the F^1 terms in (19) dominate over the F^3 terms. The surface boundary current generated by this signal is northward from 3° S to 3° N and southward outside this interval.

The reasonable accord with observation as regards time scale and type of boundary-current response suggests that it may be worth while to attempt an amplitude prediction for comparison with Somali Current data. This might seek to establish whether, during the month needed for the propagated effects of wind stress at large distances (of the order of 500 to 2000 km) to generate substantial boundary currents, the predicted current magnitudes could become comparable with those observed. An attempt to do this is made in § 4.

This does not include any attempt to calculate the effects of purely local wind stress, acting within 500 km of the coast, for which the present method of normal mode expansion would probably be much less suitable. It concentrates, in fact, on the effect of onset of SW winds within a region north of 2° N and east of 50° E. The conclusion is that, if the speed of those winds is as much as 14 m/s, then the surface element (including the baroclinic, and smaller barotropic, components) of boundary-current flux will already have reached 40 % of the strength typically observed after 1 month (of which 22 days represents time needed for propagation, and the remainder time for concentration, of the flux).

This amplitude prediction is thought reasonably satisfactory; error by a factor of 2.5 is hardly excessive in ocean-current predictions, and much of the shortfall could well be made up by the effect of local winds. Furthermore, the most reliable flux observations have all been made after a time lag considerably exceeding 1 month. The current pattern predicted (figure 6 below) is also reasonable, giving baroclinic flow that is northward on the surface for $-0.4 < y < 2.7$, corresponding to latitudes from about 1° S to 6° N.

It is perhaps a merit that this model predicts a western boundary current whose baroclinic component is northward at the surface up to a certain latitude (6° N) and southward beyond, implying separation of the current there, even though no geographical feature is present to promote such separation. Actually, this sort of separation occurs, but at 9° N; however, two features can be regarded as likely to move the point of separation to such a greater northern latitude. One is the barotropic component, which is predicted, on the other hand, as relatively weak. Another is inertial overshoot; computations like those of Bryan (1963) and Veronis (1966) indicate that nonlinear inertial terms, when included, cause boundary currents to continue in

concentrated form over rather greater distances than when they are omitted (owing, presumably, to propagation effects being somewhat supplemented by advection).

The author has studied one more possible extension of the theory, to include the influence of any equatorial undercurrent, preliminary work being based on the incorrect hypothesis that the undercurrent might, in the Indian Ocean, be assumed generally similar to that which has been studied much more thoroughly (see, for example, Knauss 1960) in the Pacific. It seemed important to check whether such an undercurrent might alter the propagation of the baroclinic equatorial modes in such a way as to increase the estimates of response time or in other ways substantially alter the conclusions. Because such an undercurrent is an eastward flow concentrated near the level of the thermocline it might especially affect the first baroclinic mode ($n = 1$) by advection of those horizontal vorticity patterns distributed around the thermocline whose propagation is the main activity of that mode.

Dr J. C. Swallow, has, however, pointed out to the author that present evidence (Swallow 1967) indicates no significant equatorial undercurrent in the Indian Ocean during the northern-hemisphere summer. For this reason, no detailed investigation of the influence of such an undercurrent is included in this paper. It may, nevertheless, be worthwhile to conclude this Introduction with a brief indication why even an undercurrent on the scale of that in the Pacific, with a peak velocity a little over 1 m/s, would not significantly alter the present conclusions. Eastward advection, at a velocity rising to such a peak, of a horizontal vorticity pattern situated around the thermocline, might be thought to reduce greatly the speed of westward propagation (at a little less than 1 m/s) of the equatorial baroclinic modes. Because, however, the undercurrent is concentrated in a narrow stream, about 200 km wide, estimates appear to indicate that propagation of these modes would be only slightly altered by such advection.

For such estimates, we may define $U(y)$ as the effective undercurrent speed for transport of horizontal vorticity on a parallel of latitude a distance y north of the equator. Then the left-hand side of equation (12), simplified by omitting v_{ttt} and v_{txx} terms when characteristic east–west lengths and times are much greater than the units (16), is modified by the inclusion of an advection term to become

$$-gH\left(\frac{\partial}{\partial t}+U(y)\frac{\partial}{\partial x}\right)v_{yy}+\beta^2y^2v_t-gH\beta v_x. \tag{20}$$

On this approximation, equation (15) for the fundamental modes becomes

$$\left[1-\frac{U(y)\,l}{\omega}\right]v_{yy}+\left(\frac{\beta l}{\omega}-\frac{\beta^2y^2}{gH}\right)v=0. \tag{21}$$

Here, because l/ω is negative, the coefficient of v_{yy} everywhere exceeds 1; accordingly, there is no singularity of the Orr–Sommerfeld critical-layer type, due to vanishing of this coefficient. Physically, this is because the direction of local streaming is opposite to the direction of propagation; thus, as the vorticity pattern undergoes propagation to the west, the advection of individual elements of vorticity to the east simply redoubles the importance of the first term in (21), which represents the discrepancy between them that must be balanced by planetary vorticity shifts.

As far as the shape of the function $v(y)$ representing an equatorial mode is concerned, we can conclude from (21) that its curvature $|v_{yy}|$ is reduced substantially, in a small interval around $y = 0$, below what it would be without advection. This makes little difference to the modes with $M = 1, 3, 5, \ldots$, for which v_{yy} vanishes on the equator in any case. In particular, the response time based on the mode ($M = 1$) with greatest propagation speed is practically unaltered.

The effect on even modes $v(y)$, however, is to flatten them very considerably for small y. This brings about a somewhat reduced speed of propagation; also, it pushes farther north the zero of $v(y)$ where northward flow changes to southward. The total changes to earlier conclusions are, however, small even quantitatively, and the fact that nothing like so pronounced an undercurrent is actually observed during the northern hemisphere summer suggests that still smaller changes would really arise due to this cause.

To summarize this long Introduction, then, in two sentences, it is adequate in the present state of observational knowledge of winds and currents in the Indian Ocean to use simple linear theory to estimate what changes in current field can arise from wind-stress changes associated with monsoon onset, and to suppose that associated changes in flux reaching the Somali coast become concentrated in a boundary current. Measurements of the changes are still not complete enough to permit detailed comparison with the results of this theory, but a preliminary rough comparison is encouraging.

2. Barotropic response

It was noted in § 1, and explained in more detail in the Appendix, that the linearized barotropic response (u, v) of an ocean to an external force (F, G) per unit mass is effectively uniform with respect to depth, and satisfies equations (1) and (2) (with H equal to the actual depth of the ocean); furthermore, these imply (4) and (5), that is

$$\psi_x = v, \quad \psi_y = -u, \quad \nabla^2\psi_t + \beta\psi_x = G_x - F_y, \tag{22}$$

under conditions which exclude very rapid changes (with characteristic radian frequencies greater than about $0.1\,\mathrm{h}^{-1}$) or very low wavenumbers (less than about $0.0005\,\mathrm{km}^{-1}$ in the equatorial case). Since in an arbitrary response of the ocean the barotropic component is the integral of the horizontal velocity with respect to depth, divided by the depth, (22) can be regarded as a statement of how fast the total vorticity (integrated from surface to bottom) changes as a result of vorticity generation by wind stress and of convection of planetary vorticity.

The properties of the Rossby waves, which satisfy

$$\nabla^2\psi_t + \beta\psi_x = 0, \tag{23}$$

were briefly recalled in § 1. The group-velocity property (Longuet-Higgins 1964; Pedlosky 1965) is particularly important for waves generated by forcing within a limited region. Lighthill (1965, 1966, 1967) described methods of calculating the Rossby waves generated far from an arbitrary forcing region, in the absence of boundaries. He emphasized that characteristic frequencies and wavenumbers of the forcing effect are very influential in determining what kinds of waves are produced, and where; a travelling forcing effect satisfies certain relationships between frequencies and wavenumbers which are particularly influential in this sense.

The present work uses on (22) methods of this general type, but modified slightly because its right-hand side is significantly large in a region whose east–west dimension is much greater than its north–south dimension. Particularly for transient problems like the subject of this paper, such a solution (which we call ψ_{U}) for an unbounded ocean is a valuable step towards obtaining the solution for a bounded ocean, which we write

$$\psi = \psi_{\mathrm{U}} + \psi_{\mathrm{B}}. \tag{24}$$

Since ψ_{U} as well as ψ satisfies the equation for forced Rossby waves, (22), it follows by subtraction

that ψ_B (representing the 'reflexion' of the wave group from the coast) satisfies the equation for free Rossby waves, (23). It is indeed the solution of (23) satisfying the condition

$$\psi_B = -\psi_U \quad \text{on the boundary,} \tag{25}$$

which expresses that there can be no barotropic flow across the boundary.

The wind-stress curl in (22), $G_x - F_y$, tends to be dominated by the second term because, although the wind-stress components F and G are comparable in the Southwest Monsoon, y derivatives (gradients in the north–south direction) greatly exceed x derivatives. The region of high wind-stress curl is where the eastward component of wind stress changes from its moderate negative values south of the equator to large positive values somewhat north of it. It is a zonal region whose characteristic dimensions are of the order of 30° of longitude by 5° of latitude.

For obtaining the oceanic response to such a pattern of wind-stress curl, Fourier analysis in terms of m, the y component of wavenumber, is useful, especially because we are interested in propagation of effects over distances large compared with $1/m$, which permits the use of approximations of the group-velocity type. By contrast, the large characteristic values of $1/l$, the reciprocal of the x component of wavenumber, make the idea of the location of a wave packet with respect to x within the Indian Ocean somewhat too fuzzy. Accordingly, the waves generated are Fourier analysed only with respect to y, by putting

$$\psi_U = \int_{-\infty}^{\infty} e^{imy}\, \Psi(x,m,t)\, dm, \quad G_x - F_y = \int_{-\infty}^{\infty} e^{imy}\, C(x,m,t)\, dm, \quad \psi_B = \int_{-\infty}^{\infty} e^{imy}\, \Phi(x,m,t)\, dm. \tag{26}$$

In fact, because we are concerned mainly with response to the change in wind-stress curl associated with monsoon onset, we take $C(x,m,t)$ as the Fourier transform of the difference between the value of $G_x - F_y$ at some time t and its value at a time $t = 0$ just before monsoon onset; and similarly with Ψ. Then the equation satisfied by Ψ, namely,

$$\frac{\partial}{\partial t}(\Psi_{xx} - m^2\Psi) + \beta\Psi_x = C, \tag{27}$$

can be solved by the Heaviside p-operator method appropriate to a transient problem, where the forcing term C vanishes for $t < 0$ and we look for the solution that vanishes for $t < 0$, representing the change in ocean-current pattern arising from the changes in wind-stress pattern.

Thus, (27) is rewritten

$$\Psi_{xx} + (\beta/p)\,\Psi_x - m^2\Psi = C/p, \tag{28}$$

in Heaviside notation (those more familiar with Laplace transforms can regard the Ψ in (28) as p time the Laplace transform with respect to t of the Ψ in (27)). We are especially interested in the solution of (28) with C independent of p, representing step-function onset of the monsoon winds, because this case can give useful information on time-scales of response, and because, if this case is worked out, the response to wind stresses changing more generally with time than in step-function fashion can be obtained by convolution.

Complementary functions for (28) are $e^{\lambda_1 x}$ and $e^{\lambda_2 x}$ where

$$\lambda_1 = -\frac{\beta}{2p} + \left(\frac{\beta^2}{4p^2} + m^2\right)^{\frac{1}{2}}, \quad \lambda_2 = -\frac{\beta}{2p} - \left(\frac{\beta^2}{4p^2} + m^2\right)^{\frac{1}{2}}, \tag{29}$$

and in terms of these the solution for Ψ (corresponding to the unbounded ocean) is

$$\Psi = -\left[e^{\lambda_1 x}\int_x^{\infty} C(X)\, e^{-\lambda_1 X}\, dX + e^{\lambda_2 x}\int_{-\infty}^{x} C(X)\, e^{-\lambda_2 X}\, dX\right] \Big/ [p(\lambda_1 - \lambda_2)]. \tag{30}$$

For understanding time scale of response, we need to know how this solution develops after an extended period of time, which may be generally expected in a Heaviside treatment to be given by the behaviour of (30) for small p. Since, as $p \to 0$, equations (29) give

$$\lambda_1 \sim pm^2/\beta, \quad \lambda_2 \sim -\beta/p, \tag{31}$$

the corresponding behaviour of (30) for small p is

$$\Psi = -\beta^{-1}\int_x^\infty C(X)\,\mathrm{e}^{-(pm^2/\beta)(X-x)}\,\mathrm{d}X - \beta^{-1}\int_{-\infty}^x C(X)\,\mathrm{e}^{-(\beta/p)(x-X)}\,\mathrm{d}X, \tag{32}$$

but the second term in (32) appears likely to be progressively more insignificant (since it tends to zero as $p \to 0$).

On the other hand, the Heaviside interpretation of the first term in (32) is

$$\Psi = -\beta^{-1}\int_x^\infty C(X)\,H[t-(m^2/\beta)(X-x)]\,\mathrm{d}X, \tag{33}$$

where $H(t)$ is Heaviside's unit step function, and (33) gives

$$\Psi = -\beta^{-1}\int_x^{x+(\beta/m^2)t} C(X)\,\mathrm{d}X, \quad \Psi_x = \beta^{-1}C(x) - \beta^{-1}C[x+(\beta/m^2)\,t]. \tag{34}$$

Here, Ψ_x is the Fourier component of the northward velocity v which has north–south wavenumber m; in its expression the term $\beta^{-1}C(x)$ represents a steady local response related to the wind-stress curl by Sverdrup's law. On the other hand, the term $-\beta^{-1}C[x+(\beta/m^2)\,t]$ represents an equal and opposite response propagated to the west with velocity β/m^2, which, as expected, is the group velocity for wave groups with east–west wavenumber l small compared with m. A physical interpretation of these conclusions in terms of vorticity was given in § 1.

When Ψ and its Fourier transform ψ_{U} have been obtained, by the methods briefly introduced above and further refined below, ψ_{B} must be obtained as the solution of (23) satisfying the boundary condition (25). If, as the above arguments indicate, non-zero values of ψ_{U} are found only within the part of the ocean where wind stress operates and (owing to propagation) to the west of it, this condition $\psi_{\mathrm{B}} = -\psi_{\mathrm{U}}$ will specify non-zero values for ψ_{B} only on the western boundary. The calculation of the solution from those boundary values is easiest of all if the western boundary is a meridian (whose equation without loss of generality may be taken as $x = 0$); the calculation is given first in this case, and later extended to boundaries at an angle to the north–south direction.

By the equations (26), the Fourier transform Φ of ψ_{B} must satisfy

$$\Phi = -\Psi \quad \text{on} \quad x = 0, \tag{35}$$

as well as a differential equation which is (28) with zero on the right side. Solutions of this differential equation which do not increase exponentially as x increases (that is, in the eastward direction) are all proportional to $\mathrm{e}^{\lambda_2 x}$ by (29). Hence, for all $x \geqslant 0$,

$$\Phi = -(\Psi)_{x=0}\,\mathrm{e}^{\lambda_2 x}. \tag{36}$$

Evidently, the Heaviside interpretation of $\mathrm{e}^{\lambda_2 x}$ is of great importance; the value of Φ for a step-function change in $(\Psi)_{x=0}$ is proportional to this, while more generally Φ is minus a convolution of $(\Psi)_{x=0}$ with the Heaviside interpretation of $\mathrm{e}^{\lambda_2 x}$.

The former simplified arguments, which included the asymptotic forms (31), indicate that this Heaviside interpretation for large t behaves like that of $e^{-\beta x/p}$, which is known to be $J_0[2\sqrt{(\beta x t)}]$. This falls from 1 at $x = 0$ to zero at $x = 1.4/\beta t$ in a progressively thinning boundary layer, outside which it oscillates about zero and its gradients are much less. These simple arguments, if they are reliable, give an answer independent of m; they imply, therefore, that the functions ψ_B and ψ_U, of which Φ and Ψ are Fourier transforms, possess a similar relationship; namely, that ψ_B is minus the convolution of $(\psi_U)_{x=0}$ with a function which for large t is asymptotically $J_0[2\sqrt{(\beta x t)}]$.

We can write this relation

$$\psi_B \sim -\int_{t_0=0}^{t_0=t} J_0\{2\sqrt{[\beta x(t-t_0)]}\}\, \mathrm{d}(\psi_U)_{x=0,\, t=t_0}, \tag{37}$$

meaning as explained in § 1 that each new element of northward flux ψ_0, which would have reached $x = 0$ at time $t = t_0$ had the boundary been absent (so that ψ_0 represents a change in $(\psi_U)_{x=0}$) produces a reaction ψ_B which is asymptotically $-\psi_0 J_0\{2\sqrt{[\beta x(t-t_0)]}\}$. Within the thinning boundary current (once its thickness is small compared with any east–west distances in which ψ_U changes significantly) we may add ψ_0 to this to obtain the total contribution to the stream function $\psi = \psi_U + \psi_B$ in the form written in § 1 as (8) and depicted in figure 1.

Thus, all the flux which is propagated to the western boundary 'piles up' there in a continually thinning boundary current, and figure 1 shows the process of concentration of an element of flux ψ_0 arriving at time t_0. Essentially, the whole flux is confined, when $t - t_0$ has become large enough, within a current of width $1.4/[\beta(t-t_0)]$. The physical interpretation of this process in terms of vorticity was given in § 1. Here, some further critique of the approximations involved is given.

The accurate form of the Heaviside interpretation of $e^{\lambda_2 x}$ is

$$\frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} e^{pt+\lambda_2 x} \frac{\mathrm{d}p}{p}, \tag{38}$$

where $c > 0$. To obtain rigorously its asymptotic form as $t \to \infty$, it is necessary to bend the path of integration into the form of loops around the singularities of the integrand, which by (29) are at $p = 0$ and at $p = \pm \mathrm{i}\beta/2m$. The asymptotic form is in three parts, therefore:

$$\frac{1}{2\pi \mathrm{i}} \int_{-\infty}^{(0+)} e^{pt+\lambda_2 x} \frac{\mathrm{d}p}{p} + \frac{1}{2\pi \mathrm{i}} \int_{-\infty}^{(\mathrm{i}\beta/2m+)} e^{pt+\lambda_2 x} \frac{\mathrm{d}p}{p} + \frac{1}{2\pi \mathrm{i}} \int_{-\infty}^{(-\mathrm{i}\beta/2m+)} e^{pt+\lambda_2 x} \frac{\mathrm{d}p}{p}. \tag{39}$$

The first part of (39) is asymptotically evaluated in terms of the behaviour of the integrand as $p \to 0$, which to the approximation given in (31) yields

$$\frac{1}{2\pi \mathrm{i}} \int_{-\infty}^{(0+)} e^{pt-\beta x/p} \frac{\mathrm{d}p}{p} = J_0[2\sqrt{(\beta x t)}]. \tag{40}$$

To the next approximation λ_2 is $-(\beta/p) - (pm^2/\beta)$ which puts an extra factor $e^{-pm^2 x/\beta}$ into the first part of (39), which simply shifts the t-origin to give

$$J_0\{2\sqrt{[\beta x(t - m^2 x/\beta)]}\} \tag{41}$$

as a still closer approximation for large t. Once the boundary current has become thin, this is a small correction within it. For example, if we are interested in the time taken for the thickness of the boundary current (value of x for which (41) first vanishes) to be reduced to 100 km, the first approximation (40) gives $1.4/\beta x = 6 \times 10^5$ s and the second approximation (41) adds xm^2/β,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

which for the median wavenumber $m = 1/(200\,\mathrm{km})$ (see below) is 1.1×10^5 s. This confirms that variation in the boundary response with m is only slight after about a week. Essentially, a boundary layer is then formed, within which the x-derivative terms in (27) greatly exceed the m^2 or y-derivative term.

The second and third loop integrals in (39) make asymptotic contributions that are relatively unimportant compared with (41). Because the loops embrace the singularities $p = \pm \mathrm{i}\beta/2m$, each represents waves of radian frequency $\beta/2m$. From this value of frequency, or from the fact that $\lambda_2 = \pm \mathrm{i}m$ at the singularities, we infer that these waves have $l = \pm m$ (crests at 45° to parallels of latitude), and group velocity in the north–south direction (parallel to the coast). By applying Watson's lemma to the integrand (approximated in the neighbourhood of each singularity in the usual way), we deduce that the amplitude of this wavy contribution to the Heaviside interpretation of $\mathrm{e}^{\lambda_2 x}$ is

$$(2/\pi)^{\frac{1}{2}}\,(mx)\,(2m/\beta t)^{\frac{3}{2}}. \tag{42}$$

Its radian frequency, for a median wavenumber $m = 1/(200\,\mathrm{km})$, is $\beta/(2m) = 1/(5\text{ days})$.

The relative unimportance of this wavy contribution, compared with the contribution (40) which falls from 1 at $x = 0$ to 0 at $x = 1.4/\beta t$, is clear from two considerations. In the first place, once t exceeds 5 days, the term $(2m/\beta t)^{\frac{3}{2}}$ is already less than 1 and continues to decrease fairly quickly thereafter, while within the thinning boundary current mx is already less than 1 and also decreasing, and the factor $(2/\pi)^{\frac{1}{2}} \doteq 0.8$ further reduces the amplitude for large t. In the second place, the fact that this contribution is wavy in nature reduces its total effectiveness by destructive interference. Thus, whereas different elements of flux with the same sign that reach the boundary at different times make contributions to the integral (37) that add up constructively, the analogous integral (an integral with respect to both t_0 and m) for the wave contribution must be subject to a great deal of destructive interference.

To sum up then, a linear theory neglecting turbulent transport of momentum yields the quite reasonable result, that any northward barotropic flux reaching the coast must beome concentrated after about a week in a boundary current of thickness about 100 km, but adds to it the unacceptable result, that the thickness will continue to decrease thereafter like $1.4/\beta(t - t_0)$. Real boundary currents do not become indefinitely thin, and we know from the steady theory of western boundary currents that their thickness may be forced in the steady state to take a fixed value by nonlinear effects or horizontal turbulent transport or both. We may expect that these influences, when present in the unsteady problem, limit the continual reduction of thickness and cause it to tend to a definite non-zero value.

It would be difficult to demonstrate this for the nonlinear effects, and hardly correct to attempt such a demonstration in relation to the barotropic component by itself; but a demonstration can easily be given in the linear case with horizontal mixing specified by a uniform eddy viscosity ν. Then ψ_{B} satisfies (23) with an extra term $\nu\nabla^4\psi$, representing horizontal diffusion of vorticity, on its right-hand side, and its Fourier transform Φ therefore satisfies, in Heaviside notation,

$$p(\Phi_{xx} - m^2\Phi) + \beta\Phi_x = \nu(\partial^2/\partial x^2 - m^2)^2\Phi. \tag{43}$$

Horizontal mixing allows the satisfaction of the no-slip condition on the boundary, so that condition (35) is replaced by two conditions

$$\Phi = -\Psi \quad \text{and} \quad \Phi_x = -\Psi_x \quad \text{on} \quad x = 0. \tag{44}$$

Solutions of (43) proportional to $\mathrm{e}^{\lambda x}$ exist when

$$p(\lambda^2 - m^2) + \beta\lambda = \nu(\lambda^2 - m^2)^2. \tag{45}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

For all p in the right-hand half-plane, there are always just two roots of (45) with negative real part, which will be called λ_3 and λ_4. This is obvious when $|p|$ is large enough, these two roots being close to $-m$ and to $-(p/\nu)^{\frac{1}{2}}$ (and the others close to $+m$ and $+(p/\nu)^{\frac{1}{2}}$). Furthermore, if any curve in the right-hand half-plane existed where the number changed from two to some other value, then at any point on that curve λ would be able to take some pure imaginary value il, whence by (45) p must be equal to

$$-\nu(l^2+m^2)+(\beta il)/(l^2+m^2). \tag{46}$$

This is a contradiction, since (46) is in the left-hand half-plane.

This pair of solutions $e^{\lambda_3 x}$ and $e^{\lambda_4 x}$, which remain finite as x increases, can be combined into a solution of (43) satisfying the conditions (44), namely,

$$\Phi=[(\Psi)_{x=0}(\lambda_4 e^{\lambda_3 x}-\lambda_3 e^{\lambda_4 x})+(\Psi_x)_{x=0}(e^{\lambda_4 x}-e^{\lambda_3 x})]/(\lambda_3-\lambda_4). \tag{47}$$

This is a regular function of p in the right-hand half-plane. When $|p| \gg (\beta^2\nu)^{\frac{1}{3}}$, one root λ_3 of (45) is close to λ_2 (see (29)), while the other, λ_4, is much larger, being near $-\sqrt{(p/\nu)}$. Under these circumstances Φ is close to the value (36) without horizontal transport. Since at all times $t \ll (\beta^2\nu)^{-\frac{1}{3}}$ a steepest descent contour for the Heaviside interpretation of (47) remains in the part of the plane where $|p| \gg (\beta^2\nu)^{\frac{1}{3}}$, the conclusion deduced earlier about the form of the thinning boundary current are still approximately valid at those times.

At later times t, however, the Heaviside interpretation of (47) depends more and more on its behaviour for very small p, when the roots λ_3 and λ_4 of (45) are complex conjugates, close to

$$(\beta/\nu)^{\frac{1}{3}} e^{\pm\frac{2}{3}\pi i}, \tag{48}$$

provided that this is much bigger than m. Asymptotically for very large t, the Heaviside interpretation becomes equal to the value of (47) for $p=0$, which is a steady boundary current of the type first studied by Munk (1950); on the approximation (48), it is given by

$$\Phi=-(\Psi)_{x=0}\exp[-\tfrac{1}{2}(\beta/\nu)^{\frac{1}{3}}x]\{\cos[(\tfrac{1}{2}\sqrt{3})(\beta/\nu)^{\frac{1}{3}}x]+(1/\sqrt{3})\sin[(\tfrac{1}{2}\sqrt{3})(\beta/\nu)^{\frac{1}{3}}x]\}. \tag{49}$$

(The terms in $(\Psi_x)_{x=0}$ have not been written down here because they would be insignificant in the usual case when this quantity is small compared with λ_3 or λ_4 times $(\Psi)_{x=0}$.) As (49) is independent of both t and m, the same relationship holds asymptotically for the Fourier transforms ψ_B and ψ_U.

With horizontal transport accounted for, then, the western boundary current continues to thin like $1.4/\beta t$ until this begins to be comparable to $(\nu/\beta)^{\frac{1}{3}}$ (that is, until t is comparable with $(\beta^2\nu)^{-\frac{1}{3}}$), and then it asymptotes to a constant value, which according to (49) is about $2.4(\nu/\beta)^{\frac{1}{3}}$. Although the thinning ceases, the total flux incident on the western boundary remains concentrated thereafter in this narrow current. It seems reasonable to suppose that, if other mechanisms productive of a definite boundary-current thickness in steady flow were taken into account, the thinning boundary current in unsteady flow would similarly asymptote to that thickness, with all the incident flux again trapped within it.

With this background, a return to the problem of estimating ψ_U through its Fourier transform Ψ is now useful. The important thing to estimate correctly is how ψ_U varies on the western boundary itself, since the effect of ψ_B is mainly to cause the flux represented by ψ_U to be concentrated (with only a relatively short time lag) into a boundary current. For estimating ψ_U or Ψ on the western boundary $x=0$, it is exact to neglect the second integral in equation (30) or its

approximate form (32), because $C(X)$ is non-zero only for $X > 0$ (there is forcing only to the east of the boundary).† Thus,

$$(\Psi)_{x=0} = -[p(\lambda_1 - \lambda_2)]^{-1} \int_0^\infty C(X)\, \mathrm{e}^{-\lambda_1 X}\, \mathrm{d}X. \tag{50}$$

The simplest estimate of time-scale of response comes from the asymptotic form of (50) given in (34), which with $x = 0$ becomes

$$(\Psi)_{x=0} = -\beta^{-1} \int_0^{\beta t/m^2} C(X)\, \mathrm{d}X. \tag{51}$$

This has reached half its ultimate value when the integral of the wind-stress curl's component of wavenumber m is just half as much from 0 to $\beta t/m^2$ as it is from 0 to ∞. Thus, $\beta t/m^2$ is, in this sense, a median distance from the western boundary of those components of wind-stress curl. Such a median distance for characteristic Southwest Monsoon winds in general would appear from climatic charts to be about 2000 km. The low wavenumber components will traverse this distance soonest; the high wavenumber components much later.

But we can consider also (as mentioned already) a median wavenumber m_0, in the sense that half of the distribution of wind-stress curl has wavenumbers greater than m_0, and half less. The definition is ambiguous unless it is specified whether the half is to be measured on an amplitude or on an energy basis. For example, a distribution of wind-stress curl proportional to $\exp\{-y^2/y_0^2\}$ (with a suitable parallel of latitude chosen as $y = 0$) has Fourier transform proportional to $\exp\{-\frac{1}{4}y_0^2 m^2\}$, and the median wavenumber is deduced from error-function tables as $0.95/y_0$ on an amplitude basis or $0.67/y_0$ on an energy basis. A suitable value might lie in between, because as we shall see the waves partly combine constructively (favouring amplitude) and partly interfere (favouring energy).

The intermediate choice $m_0 = 0.8/y_0$, with y_0 chosen so that the width of the zone within which the wind-stress curl exceeds one-sixth of its central value is 5° of latitude, gives $m_0 = 1/(200\,\mathrm{km})$. This is proposed, however, merely as a one-significant-figure estimate of median wavenumber. It suggests that half of the signal might have reached the western boundary when with $m = m_0$ the propagation distance $\beta t/m^2$ has reached its median value 2000 km. This gives $t = 25$ days, to which the addition of about a week's time lag for concentration of this half signal into a narrow boundary current yields an overall time-scale for the current's formation of about a month.‡

A better idea of the distribution of western boundary current with respect to y can be obtained by calculating the Fourier transform of (51) for some plausible, relatively simple distribution of wind-stress curl. It is important to discover, in particular, how the fact that low wavenumber components arrive soonest affects the current's form. For this purpose there is no need for exact representation of the wind-stress curl distribution, which in any case is not known. Instead, a simple distribution is used whose median distance from the western boundary $x = 0$ is $x_0 \ln 2$, namely,

$$G_x - F_y = -A \exp\{-(x/x_0) - (y^2/y_0^2)\}. \tag{52}$$

† At other locations, actually, this second integral becomes negligible fairly quickly, as we can now see from its approximate form (32), whose interpretation is

$$\beta^{-1} \int_{-\infty}^{x} C(X)\, J_0\{2\sqrt{[\beta t(x-X)]}\}\, \mathrm{d}X,$$

tending to 0 as $t \to \infty$ because the integrand is only significant for $x - X$ less than about $1.4/\beta t$.

‡ Consideration of the inclination of the Somali coast to the north–south direction, which follows, increases the estimated time lag for the concentration process to 10 days, giving 35 days as the total time-scale of response, but the difference from a month is not significant.

Its Fourier transform is $$C = -\tfrac{1}{2}(Ay_0/\sqrt{\pi})\exp\{-(x/x_0)-\tfrac{1}{4}y_0^2m^2\}, \tag{53}$$

so that (51) becomes

$$(\Psi)_{x=0} = \frac{Ax_0y_0}{2\beta\sqrt{\pi}}\exp\{-\tfrac{1}{4}y_0^2m^2\}\left[1-\exp\left\{-\frac{\beta t}{x_0m^2}\right\}\right]. \tag{54}$$

The required total flux $(\psi_U)_{x=0}$ in the barotropic part of the western boundary current must, after its Fourier transform has become asymptotic to (54), take the form (Ax_0/β) times a function of y/y_0 and $y_0\sqrt{(\beta t/x_0)}$. The value for $y/y_0 = 0$, where the flux is greatest, is exactly

$$(\psi_U)_{x=0} = (Ax_0/\beta)\,[1-\exp\{-y_0\sqrt{(\beta t/x_0)}\}], \tag{55}$$

depending, in a manner unusual in physics, on an exponential of the square root of the time. This maximum flux has risen to half its ultimate value Ax_0/β when

$$t = (x_0/\beta y_0^2)\,(\ln 2)^2; \tag{56}$$

this agrees well with the previous estimate, because $\beta t/m_0^2$ equals the median propagation distance $(x_0 \ln 2)$ when $m_0 = 0.83/y_0$ (compared with the previous $0.8/y_0$).

As to the distribution of current as a function of y/y_0, the ultimate behaviour for very large $y_0\sqrt{(\beta t/x_0)}$ is

$$(\psi_U)_{x=0} = (Ax_0/\beta)\exp\{-y^2/y_0^2\}, \tag{57}$$

varying with y exactly as does the wind-stress curl itself. But at earlier times the current is spread over a broader range of values of y/y_0, because the high wavenumber parts of the signal have not yet arrived. Indeed, evaluation of the Fourier transform of the product of two exponentials in (54) by steepest descent gives

$$(\psi_U)_{x=0} = (Ax_0/\beta)\,[\exp\{-y^2/y_0^2\}-[\exp\{-\tfrac{1}{4}y^2/y_0^2-y_0\sqrt{(\beta t/x_0)}\}]\cos\{(y/y_0)\,[2y_0\sqrt{(\beta t/x_0)}]^{\frac{1}{2}}\}], \tag{58}$$

where the $\exp(-\tfrac{1}{4}y^2/y_0^2)$ factor indicates the extra breadth for moderate values of $y_0\sqrt{(\beta t/x_0)}$; for example, at the time (56) when the maximum flux has reached half its ultimate value, figure 3 shows that (58) has significant magnitude for almost twice the range of values of y/y_0.

At much earlier times still, such a steepest descent evaluation of the Fourier transform of (54) would be inadequate, but accurate evaluation is not worthwhile because (54) is an inadequate representation of (50) at such times. The physical difficulty is that for the components of very low m which predominate at small times typical values of l may no longer be small compared with m, and therefore the group velocity may not be directed even approximately westward. Equation (50) takes this into account, and indicates for what values of m this difficulty becomes important.

With C given by (53), the Heaviside interpretation of (50) is

$$(\Psi)_{x=0} = \frac{Ay_0}{\sqrt{\pi}}\exp\{-\tfrac{1}{4}y_0^2m^2\}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\exp\{pt\}\,dp}{(\beta^2+4p^2m^2)^{\frac{1}{2}}[2px_0^{-1}-\beta+(\beta^2+4p^2m^2)^{\frac{1}{2}}]}. \tag{59}$$

The integrand of (59) has branch points at $p = \pm i\beta/2m$ and a pole at $p = 0$. There may be a second pole where the expression in square brackets is zero, namely at $p = -\beta x_0/(m^2x_0^2-1)$, but with the positive sign outside the square root this is found to be so only if $|m| > x_0^{-1}$. The total contribution to (59) from the residues at the poles is found to be

$$(Ax_0y_0/2\beta\sqrt{\pi})\exp(-\tfrac{1}{4}y_0^2m^2)\,[1-H(|m|-x_0^{-1})\exp\{-\beta x_0t/(m^2x_0^2-1)\}], \tag{60}$$

and, as in the earlier discussion of (39), $(\Psi)_{x=0}$ differs from (60) by the sum of two loop integrals over the branch points.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Evidently, (60) differs from the previous approximation (54) only when m is of order x_0^{-1}, a value relevant only at times of order $(\beta x_0)^{-1} \simeq 4$ h. For ratios x_0/y_0 as big as we are concerned with (say, 18), the steepest-descent estimate of (60) is, furthermore, indistinguishable from that of (54). The relative importance of the loop integrals is harder to assess, but as in the case of (39) their contribution is strictly wavy with frequencies at least $\beta/2m$, and seems likely to be reduced in effect very substantially by destructive interference. The slowly varying background to these relatively rapid fluctuations is the boundary-current distribution given by (58) and plotted in figure 3, which would be expected to be the main observable effect.

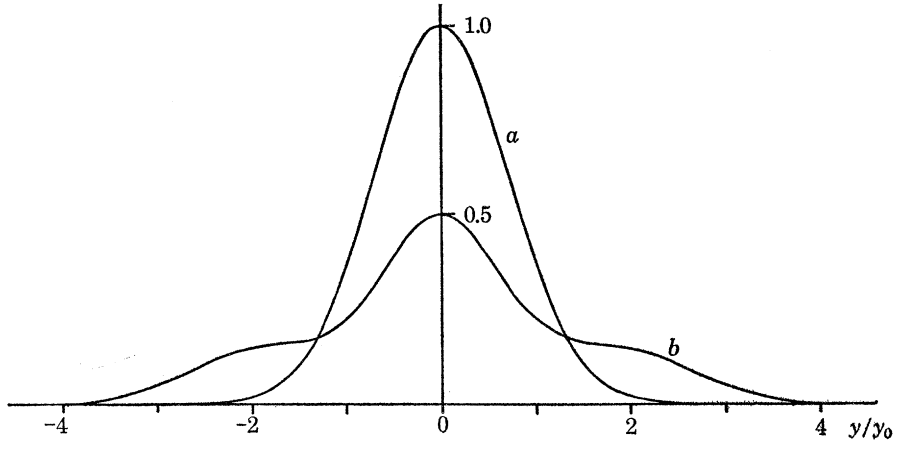

FIGURE 3. Predicted barotropic component of boundary-current flux, $(\psi_U)_{x=0}$, associated with onset of a distribution of wind-stress curl given by (52). The predicted flux is measured on a scale Ax_0/β; curve (a) indicates steady-state values after a long time, while curve (b) indicates values after a time given by (56), when the peak has reached half its ultimate value.

The conclusions of this section are finally tested further by varying two of the hypotheses which underlie them: the north–south alinement of the western boundary, and the uniformity of the ocean depth. The mean inclination (to the north–south direction) of the African coast in the region of the Somali Current (say, from 2° S to 9° N) is 40°; this inclination is maintained fairly constant, only a slight bulge occurring around 3° N. The continental shelf is narrow in this region and remains closely parallel to the coastline. This suggests trying a modification of the theory in which the western boundary runs at a constant angle α east of north, with α taken as 40° in application to the Somali Current problem.

No change is required in the evaluation of ψ_U, but the boundary condition (25) which determines ψ_B must now be applied on the line $x = y \tan\alpha$. For this, it is convenient to use rotated axes in which the coordinates are

$$X = x\cos\alpha - y\sin\alpha, \quad Y = x\sin\alpha + y\cos\alpha. \tag{61}$$

The boundary value of ψ_U is supposed Fourier analysed with respect to Y as

$$(\psi_U)_{x=0} = \int_{-\infty}^{\infty} e^{iMY}\,\Psi^{(\alpha)}(M,t)\,dM, \tag{62}$$

so that if

$$\psi_B = \int_{-\infty}^{\infty} e^{iMY}\,\Phi^{(\alpha)}(X,M,t)\,dM \tag{63}$$

then $\Phi^{(\alpha)}$ satisfies

$$\Phi^{(\alpha)}(0,M,t) = -\Psi^{(\alpha)}(M,t). \tag{64}$$

Because east-west wavenumbers l in ψ are an order of magnitude smaller than the north–south wavenumbers m, it is a good approximation for a moderate inclination such as $\alpha = 40°$ to write

$$\Psi^{(\alpha)}(M,t) = (\sec\alpha)\,\Psi(0, M\sec\alpha, t), \tag{65}$$

in terms of the quantity $(\Psi)_{x=0}$ studied intensively above.

In the new axes, equation (23) satisfied by ψ_B becomes

$$\left(\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial Y^2}\right)\frac{\partial\psi_B}{\partial t}+\beta\left(\cos\alpha\frac{\partial\psi_B}{\partial X}+\sin\alpha\frac{\partial\psi_B}{\partial Y}\right)=0, \tag{66}$$

and so $\Phi^{(\alpha)}$ satisfies, in Heaviside notation,

$$p(\Phi^{(\alpha)}_{XX}-M^2\Phi^{(\alpha)})+\beta[(\cos\alpha)\,\Phi^{(\alpha)}_X+(\mathrm{i}M\sin\alpha)\,\Phi^{(\alpha)}]=0. \tag{67}$$

Independent solutions of (67) are $e^{\Lambda_1 X}$ and $e^{\Lambda_2 X}$, where

$$\left.\begin{matrix}\Lambda_1\\ \Lambda_2\end{matrix}\right\} = -\frac{\beta\cos\alpha}{2p}\pm\left(\frac{\beta^2\cos^2\alpha}{4p^2}+M^2-\frac{\mathrm{i}\beta M\sin\alpha}{p}\right)^{\frac{1}{2}}. \tag{68}$$

The solution of (67) which satisfies (64) and does not increase exponentially as $x\to\infty$ is

$$\Phi^{(\alpha)} = -\Psi^{(\alpha)}e^{\Lambda_2 X}, \tag{69}$$

where $\Psi^{(\alpha)}$ stands for the Heaviside representation of $\Psi^{(\alpha)}(M,t)$. As was done following (36), we must now evaluate the Heaviside interpretation of $e^{\Lambda_2 X}$, representing the boundary response to a step-function change in the flux arriving at it.

We may go straight to the approximation parallel to (41), based on an expansion of Λ_2 in ascending powers up to and including the term in p itself:

$$\Lambda_2 \simeq -(\beta/p)\cos\alpha+\mathrm{i}M\tan\alpha-(pM^2/\beta)\sec^3\alpha. \tag{70}$$

On this approximation, the interpretation of $e^{\Lambda_2 X}$ is

$$\exp\{\mathrm{i}MX\tan\alpha\}\,J_0\left\{2\sqrt{\left[(\beta X\cos\alpha)\left(t-\frac{XM^2\sec^3\alpha}{\beta}\right)\right]}\right\}. \tag{71}$$

Three differences from the earlier solution (41) are apparent. First, the approximate time required for concentration of the boundary current in a width X (say, 100 km) is not $1.4/\beta X$ but $1.4/(\beta X\cos\alpha)$. Secondly, this time increase by a factor $\sec\alpha$ appears also from the correction $(XM^2\sec^3\alpha)$, which by (65) can be written $Xm^2\sec\alpha$ in terms of a median wavenumber m of the incident wave. For $\alpha = 40°$ this factor $\sec\alpha = 1.3$ increases the previous estimate of the time lag to about 10 days.

Thirdly, there is a phase difference between the boundary response, carrying the factor $\exp\{\mathrm{i}MX\tan\alpha\}$, and the incident wave, which (for typical $l \ll m$) carries a factor

$$\exp\{-\mathrm{i}MX\tan\alpha\}.$$

Combining these with the factor $\exp\{\mathrm{i}MY\}$, we may say that the lines of constant phase of the incident wave make an angle approximately $+\alpha$ with the normal to the boundary (that is, they are nearly parallels of latitude), but those of the boundary response, on the approximation (71), make an equal and opposite angle $-\alpha$ with the normal. This fact, noted in a rather general context already by Longuet-Higgins (1964), directs parallel to the boundary the inner part of the current, where J_0 is near 1. Farther out, the current direction veers as the westward feed flow

turns round to stream along the boundary. The estimate of boundary-current thickness is unaffected by these phase considerations, being simply the distance from the boundary at which the Bessel function factor in the response (71) first falls to zero.

The other modification to the theory of barotropic response, that due to non-uniformity of ocean depth, was outlined briefly in § 1, noting especially the interaction term (9), added to the left-hand side of (5) or (22) to allow for changes in vertical vorticity (integrated from surface to bottom) produced by vertical stretching when a current moves into a region of greater depth. This makes possible energy exchange between different Rossby wave modes through 'resonant interaction' (see, for example, Hasselmann 1967) with some wavy component of variation of depth H. Such resonant interactions of three wave modes are subject to the rule that the frequency and wavenumber components of one mode are each equal to the sum of the frequency and wavenumber components (respectively) of the other two modes.

Because the depth variation has zero frequency, this requires the two interacting Rossby waves to have identical frequency. Now, the wavenumbers (l, m) of Rossby waves of fixed frequency ω lie on a circle of radius $\beta/2\omega$, which is centred on the point $l = -\beta/2\omega$, $m = 0$, and passes through the origin. The group velocity of waves corresponding to a point on that circle is directed along the inward radius from that point to the centre. Points with $|l| \ll |m|$ are close to the origin, and the associated group velocity is approximately westward. Evidently, if the (l, m) values of such a point on the circle are known, an approximate value $m^2/2|l|$ can be deduced for the radius of the circle. This is important because a wavy component of bottom topography, that can transfer energy from such a wave to one with group velocity *not* approximately westward must have wavenumber whose magnitude is (within a factor of 2) equal to that radius.

Specifically, the wavenumber vector of such a component of bottom relief must be represented by a line joining two points on the circle, one of which is near the origin. If the group velocity of the other makes an angle of more than 30° with the westward direction, the magnitude of the vector is that of a chord subtending some angle between 30° and 180° at the centre, and this is equal to the radius within a factor of two. For a median m of $1/(200\,\mathrm{km})$, and a median l of $1/(3000\,\mathrm{km})$ corresponding to the distribution (52) with a median distance $x_0 \ln 2$ equal to 2000 km, this radius is $m^2/2l = 0.04\,\mathrm{km}^{-1}$.

We may infer that attenuation of the westward travelling barotropic signal by energy transfer to waves travelling in other directions occurs mainly by resonant interaction with components of depth variation with wavenumbers of order $0.04\,\mathrm{km}^{-1}$. Because the more prominent, larger-scale variations would not be effective, and because the interaction term (9) is of substantially reduced importance relative to the beta effect near the Equator, attenuation by this mechanism is likely to be only moderate. Furthermore, it would not affect the estimates of time-scale of barotropic response that have been given in this section.

3. Baroclinic modes of propagation

Baroclinic response, for reasons mentioned in § 1 and the Appendix, is expected to be most significant in the first baroclinic mode $n = 1$. The effective depth H in this mode is normally of the order 1 m and, for the equatorial region of the Indian Ocean, is estimated more closely in the Appendix as 0.75 m. Because interest in the following studies of baroclinic propagation and response mainly centres on the $n = 1$ mode, and this alone is used in illustrations and applications of the theory, no suffixes or other distinguishing marks are placed on H, u, v, etc., to signify which

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

baroclinic mode is being studied. Much of the theory to be described can be applied equally to higher baroclinic modes, but is not so applied to any significant extent in this paper.

Equation (3) is, as with barotropic response, the foundation of the theory, but because H is so much smaller in the baroclinic case the third term in (3) plays a far more crucial role. It was explained in § 1 why the north–south variation of f^2 in this term must be allowed for in equatorial regions, and the study of baroclinic propagation near the equator with this taken into account is the main concern of the present section. That study is preceded, however, by introductory notes on baroclinic propagation at higher latitudes, where it is locally permissible to treat f^2 as well as β in equation (3) as a constant.

Such propagation is governed by equation (3) with the right-hand side replaced by zero. Wave-like solutions of this equation, proportional to $\exp\{-i\omega t+ilx+imy\}$, satisfy the dispersion relation

$$-\omega^3+gH\omega(l^2+m^2)+\omega f^2+gH\beta l=0. \tag{72}$$

Solutions of (72) in the baroclinic case are much more radically separated than in the barotropic case into 'tidal' solutions with ω slightly exceeding the 'inertial' frequency $f\doteq 10^{-4}\,\mathrm{s}^{-1}$ and solutions with enormously lower frequency. For the former solutions, which are not of interest in ocean current dynamics, the last term in (72) is completely negligible and effectively

$$\omega^2=f^2+gH(l^2+m^2),$$

which gives only a moderate correction to the inertial frequency except at rather high wave-numbers.

As ω decreases below f, however, there is an extensive gap without any solution until

$$\omega<\tfrac{1}{2}\beta(gH)^{\frac{1}{2}}/f, \tag{73}$$

an upper limit which at (say) a typical Gulf Stream latitude of 40° is about $3\times10^{-7}\,\mathrm{s}^{-1}$. When (73) is satisfied, the first term in (72) is negligible and effectively

$$\left(l+\frac{\beta}{2\omega}\right)^2+m^2=\left(\frac{\beta}{2\omega}\right)^2-\frac{f^2}{gH}, \tag{74}$$

which in the (l,m) plane represents a circle with centre $(-\beta/2\omega,0)$ and radius

$$[(\beta/2\omega)^2-(f^2/gH)]^{\frac{1}{2}}.$$

Wind stresses of approximately zonal distribution should preferentially excite waves with $|l|\ll|m|$, a condition which can be satisfied on the circle only if its radius is very close to $\beta/2\omega$, requiring that (73) be satisfied with a large factor to spare.

An alternative form of (74), solved for ω in terms of l and m, is

$$\omega=-\frac{\beta l}{l^2+m^2+(f^2/gH)}, \tag{75}$$

which shows that, under these conditions with $|l|\ll|m|$, the group velocity $(\partial\omega/\partial l,\partial\omega/\partial m)$ is approximately westward with magnitude given by equation (11), which is greatest when m is small compared with $f(gH)^{-\frac{1}{2}}$. The maximum which it then approaches is $gH\beta/f^2$, a value independent of m, so that propagation under these circumstances is non-dispersive. It is governed, in fact, by equation (10), which, when propagation without local forcing is being considered (giving zero on the right-hand side), becomes

$$f^2v_t-gH\beta v_x=0. \tag{76}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

This equation (76), for Veronis & Stommel's 'non-dispersive baroclinic Rossby waves', indicates so clearly limitations on the time-scale of baroclinic response of oceans away from the Equator, as well as being a prototype model of the equation to be used near the Equator, that it is desirable to elucidate its physical meaning. This can be done with especial precision in the simple two-layer model of the ocean (see Appendix). We begin the discussion, however, with a consideration that applies to an ocean with any number of layers.

Equations (4) and (9) of the Appendix show that, in any baroclinic mode, we can write for the disturbed value h^k of the vertical extent of the kth layer (which in the undisturbed state is H^k)

$$h^k = H^k(1 + c^k h), \tag{77}$$

where the variable h (the same for each layer) satisfies

$$h_t = -(u_x + v_y). \tag{78}$$

In terms of h, equations (10) and (11) of the Appendix without any forcing term become

$$u_t - fv = -gHh_x, \tag{79}$$

$$v_t + fu = -gHh_y, \tag{80}$$

while the equation for $\zeta = \partial v/\partial x - \partial u/\partial y$, the curl of (u, v), is obtained from (79) and (80), using (78), as

$$\zeta_t + \beta v - fh_t = 0. \tag{81}$$

In the special case of a two-layer model, we may take $c^1 = 1$, so that (u, v) represents the velocity in the upper layer (of depth H^1), and then $c^2 = -H^1/H^2$, so that the velocity in the lower layer is $-(H^1/H^2)(u, v)$. The thermocline sinks, under the disturbed conditions (77), by an amount $H^1 h$, so that equation (81) can be interpreted as a vorticity equation for the upper layer which takes into account not only the beta effect but also the rate of increase of vorticity due to stretching of vortex lines when the thermocline sinks. Alternatively, it can be interpreted as $(-H^1/H^2)$ times a vorticity equation for the lower layer, including a rate of decrease due to contraction of vortex lines when the thermocline is being lowered.

For the low-frequency changes involved in baroclinic propagation, it appears probable that equation (76) represents a limiting case when the local rates of change u_t and ζ_t of momentum and vorticity in equations (79) and (81) can be neglected, giving

$$fv = gHh_x \quad \text{and} \quad \beta v = fh_t, \tag{82}$$

and hence (76) itself. The second of equations (82) simply balances the convective rate of increase of vertical vorticity in the upper layer against a rate of increase due to thermocline sinking. When v is positive, it says that reduction of relative vorticity, due to northward motion in the upper layer, must be balanced by the stretching effect of thermocline sinking.

This idea can also be expressed in terms of absolute vorticity (the sum of the planetary vorticity 2Ω, parallel to the Earth's angular velocity Ω, and the vorticity of ocean movements relative to the Earth's rotation), as indicated in § 1. The second of equations (82) makes the rate of thermocline sinking when v is positive such that the vertical component of absolute vorticity in both layers is conserved, as vortex lines in the upper northward-moving fluid stretch vertically to compensate for the decreasing angle they make with the vertical, while those in the lower southward-moving fluid, whose angle with the vertical is increasing, decrease their vertical extent.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

At the same time, the first equation is the familiar equation (derived by Margules for the meteorology of cold fronts) which relates the slope of a density discontinuity to the velocity change across it under conditions when local rates of change are negligible. The velocity discontinuity is $(H^0/H^2)\,v$, where $H^0 = H^1 + H^2$ is the ocean depth, and the thermocline slope is $H^1 h_x$, so the Margules relation between them would be

$$f(H^0/H^2)\,v = g(\Delta\rho/\rho^0)\,H^1 h_x, \tag{83}$$

which agrees with the first of equations (82) since, for a two-layer model, by equation (20) of the Appendix, $H = (\Delta\rho/\rho^0)\,(H^1H^2/H^0)$.

Westward propagation of a positive value of v occurs then, in a two-layer model, by (i) the thermocline sinking demanded by absolute-vorticity conservation, (ii) the tilting of the thermocline that then results if to the west it has not yet sunk, (iii) the new positive values of v which the Margules relation demands to be produced in that region of tilt to the west. For propagation of the first baroclinic mode, even in a many-layered oceanic model, the mechanism seems likely to be similar, requiring sinking of the region of greatest density gradient where $v > 0$ so that absolute vorticity changes are minimized; the analogy to the Margules relation in that general case is the general baroclinic relationship between angle of tilt of density contours and vertical gradient of horizontal velocity.

When forcing by wind stress is present, equation (76) contains additionally various right-hand side terms as in (10), because generation of additional vorticity by wind-stress curl alters (81), while generation of eastward momentum alters also (79). Although this linear equation (10) for forced baroclinic response of oceanic regions away from equator is rather unsatisfactory, in that for various reasons noted in § 1 nonlinear effects must be important, it does give significant insight into limitations on time-scale of response. It is therefore worthwhile to note briefly the solution of (10) appropriate when in an ocean at rest a distribution of negative wind-stress curl is 'switched on' at time $t = 0$.

The left-hand side is proportional to a single derivative taken along a path P, defined by

$$y = \text{constant}, \quad x + (gH\beta/f^2)\,t = \text{constant}. \tag{84}$$

Such a path P can be used to join any point (x, y, t) to a point $(x + (gH\beta/f^2)\,t, y, 0)$ corresponding to the instant before the wind-stress curl was switched on. The path contains a main part P_1 during which the wind stress is steady and (10) takes the form

$$-gH\beta(\mathrm{d}v/\mathrm{d}x)_{P_1} = -gH(G_{xx} - F_{xy}), \tag{85}$$

and a short stretch P_2 near $t = 0$ where the wind stress is rising to its steady value quickly enough for the time-derivatives on the right of (10) to dominate over space-derivatives, so that (10) can be written

$$f^2(\mathrm{d}v/\mathrm{d}t)_{P_2} = [\mathrm{d}(G_t - fF)/\mathrm{d}t]_{P_2}. \tag{86}$$

The solution of (85) on P_1, starting from the initial condition given by the solution of (86) on P_2, is

$$v = \left[(G_x - F_y)/\beta\right]^{x}_{x+(gH\beta/f^2)t} - f^{-1}F(x + (gH\beta/f^2)\,t, y). \tag{87}$$

At early times t, while the first term is still small because the values of wind-stress curl at (x, y) and $(x + (gH\beta/f^2)\,t, y)$ are still close and nearly cancel, the geostrophic second term dominates (87), but it becomes negligible compared with the first later, when such cancelling ceases. That dominant first term then implies, for northern-hemisphere mid-latitude zones with negative wind-stress curl, that a southward motion satisfying the baroclinic analogue of the Sverdrup

relation is accompanied by an equal and opposite northward flow pattern, which propagates steadily to the west at a speed $gH\beta/f^2$, which in such zones is of the order of 0.01 m/s.

This baroclinic analogue of the Sverdrup relationship demands, for a given wind-stress distribution, currents exactly proportional to the barotropic ones produced by the Sverdrup relation proper. However, their surface value is greater by a factor (expression (21) of the Appendix) which for the $n = 1$ mode may be much greater than 1. As soon as the propagating component of (87) has moved right away from the steady residual current, the latter combines with the steady residual barotropic current (see § 2) to form a strong response, concentrated mainly in the upper layer of the ocean above the regions of greatest density gradient.

Such a combination is not, however, predicted as occurring to the propagating parts of the barotropic and baroclinic response until after a time interval of many years, because as was seen in § 2 the barotropic flux field propagates to the west at speeds of the order of 1 m/s and piles up on the boundary when it arrives there. Corresponding baroclinic signals, travelling at around 0.01 m/s, arrive at an enormously later time. Nevertheless, when they do, a method similar to that used in § 2 in the barotropic case, or § 4 in a different baroclinic case, predicts again that horizontal mixing, at any rate, would cause them to pile up on the boundary in a concentrated stream.

There can be little doubt that nonlinear effects would, in reality, play a bigger role in regulating the western boundary current; furthermore, reasons were given in § 1 why they can hardly fail to be important in the main body of a mid-latitude ocean. The linear discussion has been included above, however, not only because it brings out limiting factors for those oceans, but also because it forms a useful introduction to the more complicated considerations necessary in an equatorial ocean.

Especially near the equator, the physical description given above needs modification already in the dispersive case governed by equation (75), when the simple balance (76) is complicated by local rates of change of vorticity (mainly, of its u_y component), although, in a mid-latitude ocean, these are important only when m is at least of order $f(gH)^{-\frac{1}{2}} = 1/(30\,\mathrm{km})$. Near the equator, indeed, as explained in § 1, this critical wavenumber $f(gH)^{-\frac{1}{2}}$ takes much lower values, and the u_y component of vorticity must in all cases be taken into account; furthermore the variation of f causes large changes in the group velocity (11) in a distance even as small as $1/m$, so that the variability of the f^2 coefficient in (3) needs to be taken into account.

These arguments lead to the representation of (3) near the equator by (12), which when propagation in regions without local forcing is being studied takes the form

$$v_{ttt} - gH\nabla^2 v_t + \beta^2 y^2 v_t - gH\beta v_x = 0. \tag{88}$$

The same group of arguments, or alternatively a simple dimensional analysis of (88), indicate that baroclinic propagation near the equator takes place on fundamental scales of length $(gH)^{\frac{1}{4}}(2\beta)^{-\frac{1}{2}}$ and time $(gH)^{-\frac{1}{4}}(2\beta)^{-\frac{1}{2}}$ as in (16). When all lengths and times are measured on these scales, (88) becomes

$$v_{ttt} - \nabla^2 v_t + \tfrac{1}{4} y^2 v_t - \tfrac{1}{2} v_x = 0. \tag{89}$$

For the value 0.75 m of H derived for the equatorial Indian Ocean in the Appendix, these scales are close to 250 km and 1 day respectively.

The fundamental wave-like solutions of (89) which are concentrated near the Equator (actually, all those which do not tend exponentially to infinity with distance from it) take the form

$$v = \exp\{-\mathrm{i}\omega t + \mathrm{i}lx\} D_M(y) \tag{90}$$

for $M = 0, 1, 2, \ldots$, where $$\omega^2 - l^2 - (l/2\omega) = M + \tfrac{1}{2}. \tag{91}$$

Here $D_M(y)$ is the parabolic cylinder function

$$D_M(y) = [\exp\{\tfrac{1}{4}y^2\}]\,(-\mathrm{d}/\mathrm{d}y)^M[\exp\{-\tfrac{1}{2}y^2\}], \tag{92}$$

which (Whittaker & Watson 1927) is the only solution of

$$w''(y) + (M + \tfrac{1}{2} - \tfrac{1}{4}y^2)\,w(y) = 0, \tag{93}$$

which tends to zero (rather than becoming exponentially large) both as $y \to +\infty$ and as $y \to -\infty$. The parabolic cylinder functions arise naturally near the equator, since the Earth's surface is locally developable into a cylindrical shape of locally parabolic section.

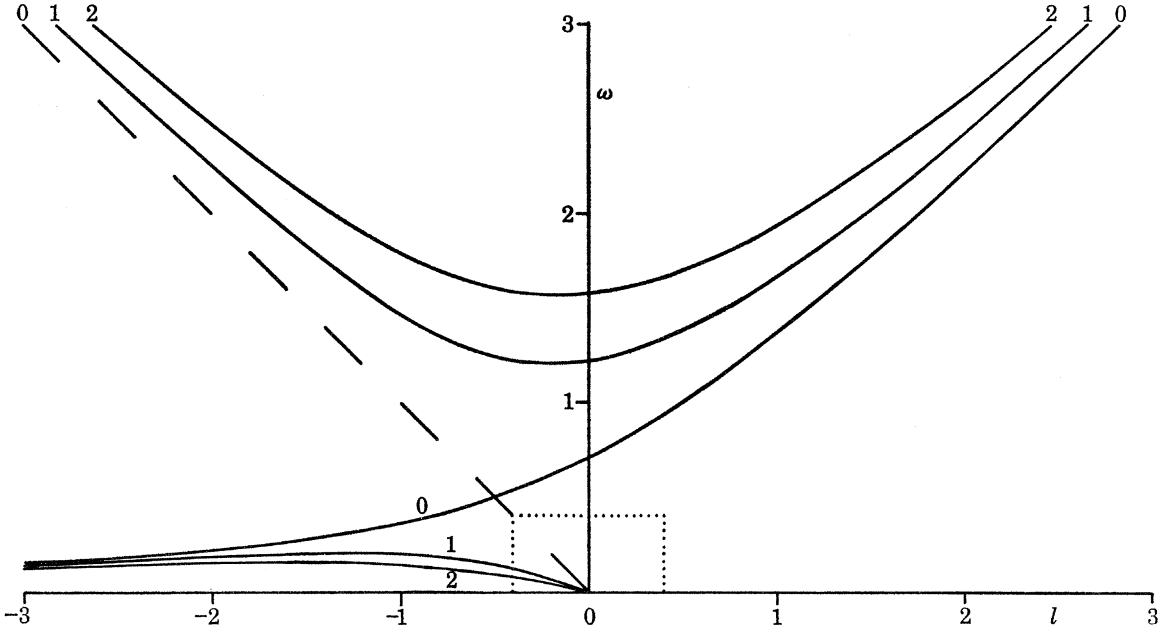

FIGURE 4. Equatorial baroclinic modes; dispersion relation (91) between frequency ω and eastward component of wavenumber l, where the number labelling each curve is the associated value of M, and the scales (16) are used for length and time. Only the half-plane $\omega > 0$ is shown; rotation through 180° about the origin gives the form of the curves for $\omega < 0$. Broken line: inadmissible branch of the dispersion-relation curve for $M = 0$. Dotted-line box: region $|l| < 0.4$, $|\omega| < 0.4$, relevant to the monsoon problem, where the simplified dispersion relation (99) holds.

The dispersion relation (91) is plotted in the (l, ω) plane in figure 4 for $M = 0, 1$ and 2. The case $M = 0$ is special, in that (91) factorizes into two separate solution curves

$$\omega + l = 0 \quad \text{and} \quad l = \omega - (1/2\omega). \tag{94}$$

These have group velocity $\mathrm{d}\omega/\mathrm{d}l$ exclusively negative (westward) and exclusively positive (eastward), respectively. Thus, the former solution curve (whose group velocity, relative to the fundamental scale of velocity $(gH)^{\frac{1}{2}}$, is -1) might be supposed important for the propagation of baroclinic signals to form western boundary currents. That curve is drawn as a broken line on figure 4, however, because it is not in fact possible for a mode concentrated near the equator to be propagated according to this dispersion relation $\omega + l = 0$.

This conclusion follows as soon as we consider the deduction of the other velocity component u from equation (90) for v. This must use (1) and (2) which, for propagation in regions without local forcing, become (when f is replaced by βy and all lengths and times are referred to the fundamental scales (16))

$$u_{tt} - u_{xx} = v_{xy} + \tfrac{1}{2}yv_t, \tag{95}$$

$$u_{xy} - \tfrac{1}{2}yu_t = v_{tt} - v_{yy}. \tag{96}$$

When $\omega^2 \neq l^2$, (95) enables us to deduce that, corresponding to the mode (90) for v, there is a solution

$$u = \exp\{-i\omega t + ilx\}[ilD'_M(y) - \tfrac{1}{2}yi\omega D_M(y)]/(l^2 - \omega^2) \tag{97}$$

for u, which like that for v tends to zero both as $y \to +\infty$ and as $y \to -\infty$.

It is easily verified that this solution satisfies also (96), as it must since v satisfies an equation found by eliminating u from (95) and (96). Those equations have a different relative status, however, when $\omega = -l$, and when (90) includes the factor $D_0(y) = \exp\{-\frac{1}{4}y^2\}$; then (95) gives no information at all about u because both sides are identically zero. Accordingly, (96) must be used; its solution with $u \to 0$ as $y \to -\infty$ is

$$u = (i/l)\exp\{il(t+x) + \tfrac{1}{4}y^2\}\int_{-\infty}^{y} \exp\{-\tfrac{1}{4}y^2\}[D''_0(y) + l^2 D_0(y)]\,dy; \tag{98}$$

this becomes exponentially large as $y \to +\infty$, unless in this limit the integral takes the value zero. But the value which it takes, $(l^2 - \frac{1}{4})\sqrt{(2\pi)}$, is in general not zero.

For $M = 0$, then, the first of the two alternative dispersion relations (94), namely, $\omega + l = 0$, does not correspond to any equatorially concentrated wave-like mode. On the contrary, the eastward component of velocity u becomes exponentially infinite (either as $y \to -\infty$ or as $y \to +\infty$, or both), unless the integral in (98) vanishes when taken from $-\infty$ to ∞, which is true only for $l = \pm\frac{1}{2}$, so that $\omega = \mp\frac{1}{2}$ and therefore l and ω satisfy the second of the dispersion relations (94) in this particular case. Thus, an equatorially concentrated mode satisfies (90) for $M = 0$ *only* when that second dispersion relation $l = \omega - (1/2\omega)$, corresponding to exclusively eastward group velocity, is satisfied.

Blandford (1966) obtained the solutions (90) (though there is a trivial difference in his method of non-dimensionalizing, and he expresses them in terms of Hermite polynomials instead of parabolic cylinder functions) but did not discard the solution $M = 0$, $\omega + l = 0$ because he calculated u incorrectly in this case. Essentially, he substituted $M = 0$ in (97) and then took the limit as $\omega \to -l$. This gives an incorrect answer because M would not be exactly zero, as (91) shows, if ω differed slightly from $-l$. The difference between $D_M(y)$ and $D_0(y)$ in this case would give a non-zero additional contribution to the limit of (97); furthermore, because $D_M(y)$ for non-integral M becomes exponentially infinite as y tends to at least one out of $+\infty$ and $-\infty$, the said limit (excluded from Blandford's expression for u) has the same property.

The $M = 0$ mode governed by $l = \omega - (1/2\omega)$, or in the original units by $l = \omega(gH)^{-\frac{1}{2}} - (\beta/\omega)$, is interesting, but likely to be excited to only a small extent by phenomena on the monsoon's characteristic scales of time or east–west breadth. In the non-dimensional (l, ω) plane of figure 4, where the scales for l and ω are $1/(250\,\mathrm{km})$ and $1/(1\ \mathrm{day})$, the $M = 0$ dispersion relation (plain line) contains no points with l and ω both small; the shortest distance from the origin to the curve is 0.64. Zonal disturbances with $l < \frac{1}{4}$ (that is, $l < 0.001\,\mathrm{km}^{-1}$) can generate waves only in the frequency range $0.59 < \omega < 0.84$ (on a scale of day^{-1}), within which disturbances of the monsoon type would have relatively little energy. The waves, furthermore, would have an eastward group

velocity (of between 1.1 and 1.6 m/s) and therefore no influence on the formation of a western boundary current.†

By contrast, figure 4 shows that for the higher modes $M = 1, 2, \ldots$ there are waves with both l and ω small, represented by points near the origin where the curves are nearly straight lines. Equation (91) shows that near the origin the slope of these is

$$\omega/l = -1/(2M+1), \tag{99}$$

so that the waves propagate non-dispersively to the west with non-dimensional velocity $1/(2M+1)$; in the original units, with the velocity (15). These non-dispersive waves (whose existence was noted also by Longuet-Higgins (1968, p. 527)) are analogous to those found by Veronis & Stommel (1956) in the case f constant, but typical velocities are greater, namely, 0.9, 0.55, 0.4 m/s and so on.

For $M \geqslant 1$, branches of the dispersion curves in figure 4 which pass through the origin satisfy (99) to reasonable approximation (within 10% for $M = 1$ and more accurately for $M > 1$) when $l < 0.4$ (that is, $l < 1/(600\,\mathrm{km})$ in the original units), so that for studying response to the monsoon (99) is an adequate approximation. The branches which do not pass through the origin have, on the other hand, much too high frequency to be relevant (in every case more than 1.2). All this suggests that for studying dynamic response to monsoon onset equation (99) may be a satisfactory first approximation to the dispersion relation for the Mth mode. Going to a second approximation (involving a slight reduction in group velocity for the higher wavenumbers, but far less than in the barotropic case) would not be justified in the present state of knowledge of the forcing wind stresses.

Accordingly, the first approximation, adequate when l^2 and ω^2 are both small compared with 1 (or, in practice, when both $|l|$ and $|\omega|$ are less than 0.4), is used in what follows. It involves the neglect of the $\omega^2 - l^2$ term in (91), which in turn means neglect of v_{ttt} and v_{txx} (but not v_{tyy}) in (88). That equation therefore becomes

$$-gHv_{tyy} + \beta^2 y^2 v_t - gH\beta v_x = 0. \tag{100}$$

The second and third terms here can be interpreted in terms of the absolute-vorticity conservation principle outlined above in the case f constant, while the first represents the effect of shear (that is, of additional vorticity due to changes in the gradient of eastward velocity u in the north–south direction).

Equation (100) has the same fundamental solutions (90) as does (88), but the associated dispersion relation is simplified to (99). There is now no solution with $M = 0$, which has $\omega + l = 0$ as sole dispersion relation on this approximation, leading necessarily to an exponential infinity of u in at least one of the limits $y \to -\infty$ and $y \to +\infty$. (The alternative curve $l = \omega - (1/2\omega)$ is on this approximation (100) excluded, quite properly because it contains no points with $|l|$ and $|\omega|$ both less than 0.4.) Accordingly, the first two propagating modes, obtained by substituting for $D_1(y)$ and $D_2(y)$ in (90) and (97), are

$$M = 1: \quad (u, v) = [(3\mathrm{i}/8l)\,(3 - y^2), y] \exp\{\mathrm{i}l(x + \tfrac{1}{3}t) - \tfrac{1}{4}y^2\} \tag{101}$$

and

$$M = 2: \quad (u, v) = [(5\mathrm{i}/12l)\,(6y - y^3), y^2 - 1] \exp\{\mathrm{i}l(x + \tfrac{1}{5}t) - \tfrac{1}{4}y^2\}. \tag{102}$$

† Dr Bretherton has pointed out to the author that there is also one mode representing a solution of (95) and (96) with v identically zero, so that it is not included in the sequence (90). This Kelvin-wave mode (referred to again in § 4) has both phase velocity *and* group velocity eastward and equal to +1, and is given by $u = D_0(y)\,\mathrm{e}^{\mathrm{i}l(x-t)}$.

Studies in § 4 indicate that the initial baroclinic response of the western equatorial part of the Indian Ocean to monsoon onset may be particularly influenced by these two modes of lowest order (101) and (102).

4. Baroclinic response in an equatorial ocean

In this section, with the work on modes in § 3 as background, an attempt is made to tackle in the baroclinic case the problem investigated in the barotropic case in § 2; that is, to find the changes in current distribution from some supposedly steady state at time $t = 0$, that may result from changes in the wind-stress field beginning at that time. As in § 2 we divide the response into two parts: a part calculated for an unbounded ocean, which we distinguish by suffix U, and a part subsequently added to it in order to satisfy boundary conditions.

If the special units (16) for length and time are used in an equatorial ocean, the basic equation (12) for baroclinic response becomes

$$v_{ttt} - \nabla^2 v_t + \tfrac{1}{4}y^2 v_t - \tfrac{1}{2}v_x = G_{tt} - G_{xx} - \tfrac{1}{2}yF_t + F_{xy}. \tag{103}$$

Here F and G have the dimensions of force per unit mass, so that the unit in terms of which they are measured is

$$(gH)^{\frac{3}{4}}(2\beta)^{\frac{1}{2}} \doteq 3 \times 10^{-5}\,\mathrm{m\,s^{-2}} \quad \text{for} \quad H = 0.75\,\mathrm{m}. \tag{104}$$

The form (90) of solutions to (89), the special case of (103) with zero on the right-hand side, may suggest that (103) be solved by expanding the quantities v, F and G each in a series of parabolic cylinder functions $D_M(y)$ for $M = 0, 1, 2, \ldots$. This is the method used in this section, but it is carried out only on an approximate simplified form of (103).

In fact, it was argued in § 3 that changes in wind-stress field associated with monsoon onset are overwhelmingly characterized by frequencies ω and east–west wavenumbers l that are small in the units here used. The response of, at least, the unbounded ocean should accordingly take the form of modes such that the v_{ttt} and v_{txx} terms in (103) are negligible (as was indicated in connexion with (100) to be so if $|\omega|$ and $|l|$ are both less than 0.4).

Furthermore, as in § 2, the G_{xx} term is neglected in relation to the F_{xy} term because the wind stress varies more steeply in the y direction (north–south) than in the x direction. Similarly, G_{tt} is neglected because for small ω it is small compared with $\frac{1}{2}yF_t$ (except in that narrow part of the region of interest where $\frac{1}{2}y$ is small). Thus, for calculating the response of an unbounded ocean, the simplified equation

$$-v_{tyy} + \tfrac{1}{4}y^2 v_t - \tfrac{1}{2}v_x = -\tfrac{1}{2}yF_t + F_{xy} \tag{105}$$

is used, although the methods are in principle applicable to the more complicated equation (103).

The solution of (105) for an unbounded ocean, v_{U}, is expanded, together with the eastward component of wind stress F, in series

$$v_{\mathrm{U}} = \sum_{M=0}^{\infty} v^M(x,t)\,D_M(y), \quad F = \sum_{M=0}^{\infty} F^M(x,t)\,D_M(y), \tag{106}$$

as is possible because the $D_M(y)$ form a complete set. The coefficients in these equations are given, for example (Whittaker & Watson 1927, art. 16.7), by

$$F^M(x,t) = \frac{1}{M!\sqrt{(2\pi)}}\int_{-\infty}^{\infty} D_M(y)\,F(x,y,t)\,\mathrm{d}y. \tag{107}$$

The inclusion of the $M = 0$ term is essential, but the work of § 3 indicates that its coefficient $v^0(x,t)$ cannot on this approximation contain any propagating component.

When the series (106) are substituted in (105), and the fact is used that $D_M(y)$ is a solution of (93), we obtain

$$\sum_{M=0}^{\infty} [(M+\tfrac{1}{2})\, v_t^M - \tfrac{1}{2} v_x^M]\, D_M(y) = \sum_{M=0}^{\infty} [-\tfrac{1}{2} F_t^M y D_M(y) + F_x^M D'_M(y)]. \tag{108}$$

The recurrence formulae (Whittaker & Watson 1927, art. 16.61),

$$\left.\begin{aligned} y D_M(y) &= D_{M+1}(y) + M D_{M-1}(y), \\ D'_M(y) &= -\tfrac{1}{2} D_{M+1}(y) + \tfrac{1}{2} M D_{M-1}(y), \end{aligned}\right\} \tag{109}$$

which hold for all $M > 0$, and also for $M = 0$ provided $D_{-1}(y)$ is taken as zero, allow the right-hand side of (108) to be written

$$-\frac{1}{2} \sum_{M=1}^{\infty} (F_t^{M-1} + F_x^{M-1})\, D_M(y) + \frac{1}{2} \sum_{M=0}^{\infty} (M+1)\, (F_x^{M+1} - F_t^{M+1})\, D_M(y), \tag{110}$$

so that coefficients of $D_M(y)$ can be separately equated to zero, giving

$$v_t^0 - v_x^0 = F_x^1 - F_t^1, \tag{111}$$

while, for $M > 0$,

$$(2M+1)\, v_t^M - v_x^M = -F_x^{M-1} - F_t^{M-1} + (M+1)\, (F_x^{M+1} - F_t^{M+1}). \tag{112}$$

Particular cases $M = 1$ and 2 of (112) were given in (18) and (19).

Equation (111), with its simple solution $v^0 = -F^1$, shows that the coefficient of $D_0(y)$ in v involves a purely local response to the wind-stress field without any propagating component. This agrees with the impossibility, argued in § 3, of any $M = 0$ mode being propagated with $\omega = -l$. The same conclusion would have been found for the more exact equation (103), since Fourier components of the forcing wind stress with $\omega = -l$ would produce zero effect on the right-hand side (as is evident for the terms in G, and has just been calculated for the terms in F), implying that waves satisfying this relationship cannot be generated.

Equation (112) is rather easy to solve by integrating along characteristics. For the important case of a near-step-function response (a wind-stress field F rising to a steady state in a period of only a few days), the method of solution used for the mid-latitude problem in § 3 can be used. The value of v^M at a point (x, t) is obtained by integrating along a straight path P in (x, t) plane, joining that point to the point $(x + t/(2M+1), 0)$, and divided into two parts P_1 and P_2. On the main part P_1 of the path P, F has reached its steady value, and (112) can be written

$$-(\mathrm{d}v^M/\mathrm{d}x)_{P_1} = -F_x^{M-1} + (M+1)\, F_x^{M+1}, \tag{113}$$

but, on the short stretch P_2 where F rises to its steady value, (112) takes the different approximate form

$$(2M+1)\, (\mathrm{d}v^M/\mathrm{d}t)_{P_2} = -\{\mathrm{d}[F^{M-1} + (M+1)\, F^{M+1}]/\mathrm{d}t\}_{P_2}. \tag{114}$$

The solution of (113) on P_1, starting from the initial condition given by the solution of (114) on P_2, is

$$\begin{aligned} v^M &= \left[-F^{M-1} + (M+1)\, F^{M+1}\right]_x^{x+t/(2M+1)} \\ &\quad - (2M+1)^{-1} \{F^{M-1}[x + t/(2M+1)] + (M+1)\, F^{M+1}[x + t/(2M+1)]\} \\ &= [F^{M-1}(x) - (M+1)\, F^{M+1}(x)] \\ &\quad + [2(M+1)/(2M+1)]\, \{M F^{M+1}[x + t/(2M+1)] - F^{M-1}[x + t/(2M+1)]\}. \end{aligned} \tag{115}$$

In this last expression, the first term represents a localized response to wind stress, and the second a propagating response. The latter is especially important, partly because it allows the ocean to respond outside the actual region of forcing by wind stress, but mainly because, as in the barotropic case, it appears that the propagated motions must largely cease propagation on reaching a western boundary, where the flux they represent must become concentrated in a narrow region.

The method of calculation of the boundary response is similar to that used in § 2 for the barotropic case, but in detail is slightly complicated by the fact that the incident flux in the baroclinic case is not given by a stream function, because the solution (u_U, v_U) for an unbounded ocean represents waves propagated at speeds as much as $\frac{1}{3}$, $\frac{1}{5}$, etc. of $\sqrt{(gH)}$, and so does not satisfy the two-dimensional equation of continuity to adequate accuracy. The calculation is performed first without taking any turbulent mixing into account, so that the appropriate boundary condition is one of zero velocity component normal to the boundary.

To express this, we must write down the velocity field for the unbounded ocean (u_U, v_U) at the boundary, where only the propagating part of (115) is non-zero. The appropriate value of v_U is obtained from (106), in which v^0 has no propagating part, as

$$v_U = \sum_{M=1}^{\infty} [2(M+1)/(2M+1)]\, D_M(y)\, \{MF^{M+1}[x+t/(2M+1)] - F^{M-1}[x+t/(2M+1)]\}. \tag{116}$$

To obtain the corresponding value of u_U, we note, from (97) with ω/l substituted from (99), that the v given by equation (90) is accompanied by

$$u = -(\mathrm{i}l)^{-1} \exp\{-\mathrm{i}\omega t + \mathrm{i}lx\}[(2M+1)/4M(M+1)]\,[(2M+1)\, D'_M(y) + \tfrac{1}{2}yD_M(y)]. \tag{117}$$

Hence the v given by (116) is accompanied by

$$u_U = -\sum_{M=1}^{\infty} (1/2M)\,[(2M+1)\, D'_M(y) + \tfrac{1}{2}yD_M(y)] \int_0^{x+t/(2M+1)} [MF^{M+1}(X) - F^{M-1}(X)]\, \mathrm{d}X, \tag{118}$$

where the lower limit can be taken as $X = 0$ because the origin of x is chosen so that there is no wind stress to the west of it.

It might be thought somewhat arbitrary in (118) to take the lower limit of integration in the extreme west rather than in the extreme east, but the knowledge that no signal propagates westward along characteristics $\mathrm{d}x/\mathrm{d}t = -1$ of the equation for u (which in the forced case is (95) with an extra term F_t on the right-hand side) indicates that this must be correct. Furthermore, direct solution of that equation for u, with v and F given by (106) and (115), which is straightforward though tedious, confirms this conclusion.†

To the baroclinic wave given by (116) and (118) a further wave field (u_B, v_B) must be added so that the sum has zero normal velocity at the boundary. This reflected wave field cannot satisfy the approximate form (99) of the general dispersion relationship, since that form involves westward energy propagation. Furthermore, that form requires small l, and there is no reason

† Actually, for general values of x, the value of u_U so deduced is equal to expression (118), with the lower limit replaced by x, plus a term

$$\tfrac{1}{2}D_0(y) \int_{x-t}^{x} F_0(X)\, \mathrm{d}X$$

representing an eastward propagating Kelvin wave of the type described in the footnote to § 3, p. 76.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

to suppose that the east–west wavenumber for the wave field (u_B, v_B) will be small; that field, rather, is expected to assume the form of a concentrated boundary current.

On the other hand, the frequencies associated with (u_B, v_B) must be low, as low as those associated with the field (u_U, v_U) whose normal component it must cancel on the boundary. If l_0 is a typical wavenumber associated with the forcing wind stress F, the frequency characteristic of (u_U, v_U) is $l_0/(2M+1)$, as is clear from the wave-velocity value (99) or from the form of (116) or (118). The only baroclinic modes of low frequency without westward group velocity, by (91), are those which approximately satisfy

$$-l^2-(l/2\omega)=0, \tag{119}$$

that is, with $l=-1/(2\omega)$, corresponding to the bottom left of figure 4. The error in retaining from (91) only the terms shown in (119) is a numerical overestimate of l by not more than 4% if $l_0 < 0.001\,\mathrm{km}^{-1}$ (giving $l_0 < \frac{1}{4}$ when the length scale (16) is used).

The dominance in (91) of the terms shown in (119) means that only the v_{txx} term in (3) is of crucial importance, along with the β-effect term, in the boundary response. These are precisely the terms which dominated in the barotropic case. Furthermore, the fact that $|\omega/l|$ is extremely small in the boundary response means that the two-dimensional equation of continuity $u_x+v_y=0$, implying the existence of a stream function, can be used to excellent approximation for the boundary response field (u_B, v_B). The theory is therefore almost identical with that used in the barotropic case, except for the slight complication that no stream function exists for the incident wave field (u_U, v_U).

For a western boundary taking the form $x=0$ first discussed in § 2, the Rossby-wave equation (23) for the stream function ψ of the boundary response should accordingly be solved with

$$(\psi_y)_{x=0} = -u_B = +u_U. \tag{120}$$

This boundary condition can, however, be thrown into the form (25) used in § 2 by integration with respect to y, giving by (118)

$$(-\psi)_{x=0} = \sum_{M=1}^{\infty} \psi_M(y) \int_0^{t/(2M+1)} [MF^{M+1}(X) - F^{M-1}(X)]\,\mathrm{d}X, \tag{121}$$

where

$$\psi_M(y) = \int_{y_M}^{y} (1/2M)\,[(2M+1)\,D'_M(y) + \tfrac{1}{2}yD_M(y)]\,\mathrm{d}y. \tag{122}$$

The appropriate lower limit in (122) depends on whether M is even or odd. Components of u_U with M odd are symmetrical about the equator $y=0$, and we can therefore expect a symmetrical boundary response, with ψ an odd function of y; hence $y_M=0$ for M odd. Components of u_U with M even, on the other hand, are antisymmetric about $y=0$, and the boundary response may be expected to relieve this by flow across the equator; this indicates that $y_M=\pm\infty$ for M even.

A rigorous explanation of why these choices of y_M are needed, if the theory of § 2 is to be used to infer the boundary response, derives from the fact that only they cause $\psi_M(+\infty)+\psi_M(-\infty)$ to be zero, and so avoid the need for any δ-function component $\delta(m)$ in the Fourier transform of $\psi_M(y)$. The method of § 2 could not be applied to the constant term in $(\psi)_{x=0}$ corresponding to such a δ-function component (although it can be applied to the rest of the Fourier representation) because $\lambda_1=0$ for $m=0$ and so an appropriate boundary response to a constant value of $(\psi)_{x=0}$ is simply a constant value of ψ.

The whole theory of boundary response goes through as in the barotropic case with $(\psi_U)_{x=0}$ replaced by $(-\psi)_{x=0}$, that is, by the right-hand side of (121). The answer (corresponding to (37)) is

$$-\psi = \sum_{M=1}^{\infty} \psi_M(\eta) \int_0^t J_0\{2\sqrt{[\beta x(t-t_0)]}\} \frac{MF^{M+1}[t_0/(2M+1)] - F^{M-1}[t_0/(2M+1)]}{2M+1} \, dt_0. \quad (123)$$

The physical interpretation of (123) is simply that every element of flux (121) reaching the boundary becomes concentrated in a gradually thinning boundary current. The part of the current corresponding to a given quantity of arriving flux takes about a week to be concentrated in a thickness of about 100 km. As with the barotropic case, the influence of nonlinear effects or horizontal transport or both can be expected to prevent further reduction of thickness beyond a steady-state value; for horizontal transport this is proved by the identical mathematics (leading to (49)).

If a time lag of about a week is not too significant, we may infer that the important quantity predicted by the theory is expression (121), namely, the total flux reaching the western boundary in the westward propagating wave. This becomes a measure of the baroclinic component of total flux in the boundary current, namely the integral of the baroclinic component of the northward velocity v with respect to the eastward coordinate x across the current. We are not speaking here of 'total transport', which involves an integral of flux from surface to bottom and is essentially barotropic; actually, a flux that is from some points of view more important, as well as easier to measure, is the upper-layer transport (or transport in the current within the upper 200–300 m) which would be influenced principally by this baroclinic component of flux (121).

These conclusions are not much altered if we take into account the slope of the Somali coast at an angle α of about 40° to the north–south direction. The time-lag is slightly lengthened (by a factor of $\sec\alpha \simeq 1.3$) as in § 3. The boundary condition is altered, depending now on the integral along the boundary of the normal velocity $u\cos\alpha - v\sin\alpha$. However, for the modes such as (101) and (102) with which we are concerned, with l small, u is so much greater than v that the term $v\sin\alpha$ is relatively insignificant. Integration with respect to $y\sec\alpha$ along the boundary then yields the same boundary value (right-hand side of (121)) for the boundary-response stream function as in the case $\alpha = 0$.

The importance of equation (121) for the baroclinic flux delivered to the western boundary makes it desirable to plot the functions $\psi_M(y)$ representing the contributions made by different modes, and this is done in figure 5. Positive ψ_M corresponds to northward flow in the boundary current, and negative to southward. Thus, the $\psi_1(y)$ mode with positive coefficient represents a boundary current flowing outwards from the equator, whereas the $\psi_2(y)$ mode with negative coefficient represents a northward flow near the equator balanced by a southward flow far from it; we write here of a positive coefficient for $\psi_1(y)$ and a negative coefficient for $\psi_2(y)$ because these are the signs which, as we shall see, appear most probable from (121).

It is the $\psi_M(y)$ with low values of M, plotted in figure 5, that are mainly important in the problem of this paper. They represent fluxes resulting from modes concentrated quite close to the equator, with relatively fast speeds of westward baroclinic propagation (between about 0.3 and 0.9 m/s). The modes with higher values of M penetrate to higher latitudes, but propagate more slowly. They would be relevant to any attempt to get a complete picture of dynamical response to monsoon onset, but their appearance in early stages of the response could at most represent response to nearby winds. This paper is directed, rather, towards understanding to

what extent western boundary currents can be set up within, say, a month by meteorological changes many hundreds, or even thousands, of kilometres away.

In order to estimate that response qualitatively, we need to use the form of the coefficients in (121) to obtain estimates of their magnitudes and signs. This requires estimation of the normal-mode expansion (106) of the function F (which itself is the first baroclinic component of the eastward wind force per unit mass, given by (17) of the Appendix with $n = 1$, and referred to (104) as a unit).

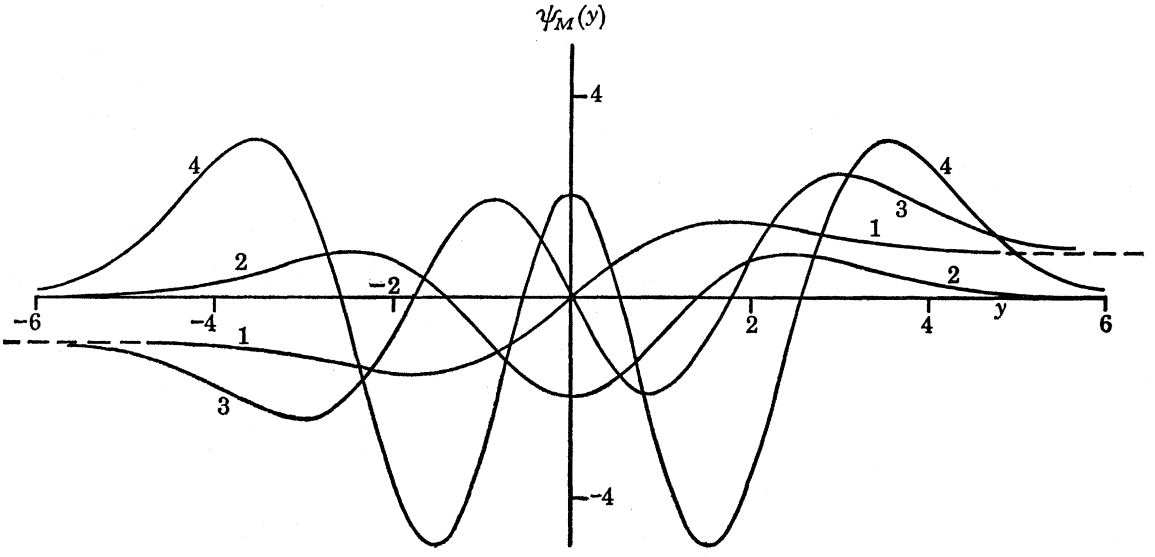

FIGURE 5. Baroclinic flux functions $\psi_M(y)$ given by (122), where M takes the values 1 to 4 attached to the curves.

To obtain that normal-mode expansion, in the case of what was described earlier as a near-step-function change in the pattern of eastward wind stress arising from onset of the monsoon, we seek to evaluate (107) with $F(x, y, t)$ independent of t for $t > 0$ and zero for $t < 0$. Furthermore, $F(x, y, t)$ is supposed to vary much more slowly with x than with y. The wind changes associated with the monsoon are mainly concentrated north of the equator, and we may take the boundary between small and large changes as approximately a parallel of latitude, say, $y = y_1$.

This suggests a tentative model with

$$F = F_c(x)\, H(t)\, H(y - y_1), \tag{124}$$

where the function $F_c(x)$ represents the change in F occurring at about the time $t = 0$ for $y > y_1$. Then (107) gives

$$F^M(x, t) = \frac{F_c(x)\, H(t)}{M!\sqrt{(2\pi)}} \int_{y_1}^{\infty} D_M(y)\, \mathrm{d}y, \tag{125}$$

and so (121) becomes

$$(-\psi)_{x=0} = \sum_{M=1}^{\infty} A_M \psi_M(y) \int_0^{t/(2M+1)} F_c(X)\, \mathrm{d}X, \tag{126}$$

where we can write the coefficient A_M as

$$A_M = \frac{1}{(M-1)!\sqrt{(2\pi)}} \int_{y_1}^{\infty} \left[\frac{D_{M+1}(y)}{M+1} - D_{M-1}(y) \right] \mathrm{d}y, \tag{127}$$

and also as

$$A_M = \frac{1}{(M-1)!\sqrt{(2\pi)}} \frac{2}{M+1} [\psi_M(y_1) - \psi_M(\infty)], \tag{128}$$

if we use the recurrence formulas (109) to express it in a form whose value can conveniently be read off from figure 5.

We can now estimate how strong a western boundary current would be created within, say, a month by wind changes in regions not very close to the boundary. By contrast, effects of local winds are not considered because the present method is probably not the most suitable for dealing with those. Specifically, we consider how large the flux (121) might become in 22 days (allowing the rest of the month for concentration of that flux into a boundary current). During this period the fastest mode ($M = 1$) would travel about 1800 km.

To make such an estimate, we calculate (126) with $F_c(x)$ replaced by some typical average value F_a for $x_E < x < x_W$ and zero for other values of x. Here, $F_c(x)$ is taken to be zero for $x < x_E$ to exclude effects of local wind changes, where in practice x_E is taken equal to 2 in the units here used (that is, to 500 km). On the other hand, x_W is taken large enough so that no signals from $x > x_W$ reach the boundary in the time involved (this means $x_W > 1800$ km). We are essentially calculating the effect at the boundary of wind changes at distances between 500 and 1800 km from it.

With these assumptions for $F_c(x)$, there are contributions to (126) only from the first four modes $M = 1$ to 4 (because only those can travel 500 km in 22 days), the coefficient of $\psi_M(y)$ when $t = 22$ being $A_M[22/(2M+1) - 2] F_a$ in each case. If the parallel of latitude to the north of which there is substantial eastward wind stress is taken as $y_1 = 1$ (roughly 2° N) then

$$(-\psi)_{x=0}/F_a = 0.75\psi_1(y) - 0.50\psi_2(y) - 0.21\psi_3(y) - 0.04\psi_4(y). \tag{129}$$

This function, plotted in figure 6, estimates the ratio of baroclinic flux in the western boundary current to the baroclinic component of wind stress per unit mass after monsoon winds have begun to blow in a region north of 2° N and east of 50 °E.

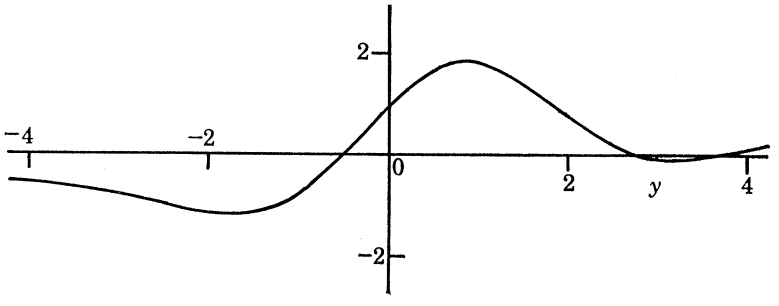

FIGURE 6. Predicted baroclinic component of boundary-current flux, $(-\psi)_{x=0}$, associated with onset of a distribution of wind stress given by (124) with $F_c(x)$ equal to F_a for $2 < x < 7.2$ and zero elsewhere. The function plotted is given by (129), so that the flux is measured on the scale F_a if the units (16) are used for length and time.

The general form of figure 6 does represent some sort of rough picture of the Somali Current as described in § 1. The ratio takes a maximum value of about 2 and exceeds half this maximum in the region $0 < y < 2$. The general pattern of northward flow, beginning at $y = -0.4$ (that is, at 1° S) and continuing to $y = +2.7$ (that is, to 6° N) is reminiscent of the Somali Current pattern. The observed current extends farther up the coast, however, to about 9° N (Swallow &

Bruce 1966). This could well result from nonlinear effects neglected in the present theory, since numerical computations with and without nonlinear inertia terms have shown that they cause boundary currents to 'overshoot', that is, to continue in a concentrated form for a longer distance than in their absence. That may be because linear theory is concerned only with propagation effects, whereas nonlinear theory combines these with effects of advection.

If we are to compare with observations the predicted absolute values of the flux that becomes concentrated in the western boundary current (in 1 month after the changes of wind stress occur more than 500 km away), we must take a particular value for F_a. Here, F_a can be expressed in terms of an average value τ_a of the eastward wind stress by equation (17) of the Appendix for $n = 1$, in which c_1^1 is taken equal to 1. Using the fact that expression (21) of the Appendix was calculated as equal to about 14 for $n = 1$, we obtain $F_a = 14\tau_a/\rho^0 H^0$. With $c_1^1 = 1$, the physical significance of $(-\psi)_{x=0}$ is surface current (strictly, its baroclinic component, which however is expected to be considerably bigger than the barotropic), integrated across the width of the boundary current. Figure 6 shows this to rise to about $2F_a$ in non-dimensional units, which in dimensional units means that

$$(-\psi)_{x=0} = F_a/\beta = 14\tau_a/\rho^0 H^0 \beta. \tag{130}$$

Actually, the corresponding barotropic component can be estimated from the results of § 2. The assumption (52) is consistent with an eastward wind stress increasing by τ_a across the equatorial region provided that the corresponding increase $\tau_a/\rho^0 H^0$ in the barotropic component of the wind force F is equal to $Ay_0\sqrt{\pi}$. The scale for $(\psi_U)_{x=0}$ in figure 3 is therefore

$$Ax_0/\beta = (x_0/y_0\sqrt{\pi})\,(\tau_a/\rho^0 H^0 \beta), \tag{131}$$

which for values of x_0 and y_0 proposed in § 2 is about $10\tau_a/\rho^0 H^0 \beta$. Hence in figure 3 the graph (b), appropriate to the boundary-current distribution after 1 month, gives a maximum of $5\tau_a/\rho^0 H^0 \beta$, realized at the centre of the zone of transition of wind stress between the values 0 and τ_a; that is, at $y = 0$ in the coordinates used in § 2, corresponding to $y = 1$ in the coordinates used in this section. This is about the same position as that of the maximum of the baroclinic component (figure 6); the sum of the two maxima is $19\tau_a/\rho^0 H^0 \beta$.

It is interesting that the 14 : 1 advantage of the $n = 1$ mode over the $n = 0$ mode (see Appendix) in strength of forcing term is reduced, in an equatorial ocean, by a factor of 5 as far as the resulting predicted peak boundary current is concerned, because barotropic propagation is more concentrated, and less prone to effects of destructive interference. Thus, these predicted barotropic components of surface current are all positive, and tend to increase somewhat the predicted surface strength of the Somali Current, as well as to reduce the significance of the negative baroclinic component of northward flux change for $y < -0.4$. They are, however, mainly confined to the region $|y| < 3y_0$ in the coordinates of § 2, which for those of § 4, with the value $y_0 = 160$ km proposed in § 2, is the region $-0.9 < y < 2.9$. In latitude terms this is the region between 2° S and 7° N, so that the barotropic contribution cannot make much difference to the predicted point of separation.

Observed peak values of surface flow, $(-\psi)_{x=0} + (\psi_U)_{x=0}$, are about 2×10^5 m²/s (whether we obtain this by integrating a concentrated current of 2 m/s over a width of 100 km or by dividing an observed volume flow of 5×10^7 m³/s by a mean depth of a little over 200 m). The theory predicts that 40 % of that value, namely 0.8×10^5 m²/s, is reached in 1 month (taking distant winds only into account) if $\tau_a = 0.4$ N/m² (note that 1 N/m² = 10 dyn/cm²).

This appears to be a reasonable value of τ_a for monsoon winds. It corresponds to southwest

winds of velocity 14 m/s, or force 6 on the Beaufort scale, a typical enough figure (Anantha-krishnan 1964), producing a wind-stress 0.55 N/m² with eastward component 0.4 N/m². The data on winds over the ocean are not complete enough to allow much weight to be attached to this crude test of the theory's amplitude predictions. The test does appear to indicate, however, that the equatorial baroclinic modes not only are propagated fast enough to account for the rate of response of the Indian Ocean, but also can be sufficiently excited by monsoon winds to account for much of the amplitude of response near the western boundary.

Appendix. Linear theory of long waves in a horizontally stratified ocean of uniform depth

In this Appendix, a simplified account of the linear theory of long waves in a horizontally stratified ocean of uniform depth, with turbulent momentum transfer neglected, is given. Because a stratification including some well-mixed layers, as well as continuously stratified portions, may be of interest, the analysis is given first for N discrete well-mixed layers of depth H^j, where j goes from 1 at the top to N at the bottom, and then extended (by letting $N\to\infty$) to cases of continuous or partly continuous stratification. Finally, the characteristics of modes for simplified density distributions representing the parts of the Indian Ocean near the Equator are calculated.

As stated in the Introduction, we use ρ to indicate, for water of given temperature and salinity, the density which that water would have at atmospheric pressure, p_a say. (Thus, ρ is the quantity whose value in kg/m³ is written $1000+\sigma_t$ by oceanographers.) When the actual density at a different pressure p has to be written, $\rho^{(p)}$ is used. The equations of motion in each layer take somewhat simplified forms in terms not of the pressure p but of the integral (for fixed temperature and salinity)

$$P=\int_{p_a}^{p}(\rho/\rho^{(p)})\,\mathrm{d}p, \tag{A 1}$$

representing a pressure reduced by the modifying action of compressibility so that within each well-mixed layer $(\mathrm{grad}\,p)/\rho^{(p)}$, the acceleration due to pressure gradient, can be written $(\mathrm{grad}\,P)/\rho$. In particular, since for long waves vertical accelerations can be neglected,

$$P_z=-\rho g. \tag{A 2}$$

Furthermore, if in the jth layer ρ^j is the value of the pressure-corrected density ρ, while (F^j, G^j) is the external horizontal force (if any) per unit mass, averaged through the depth of the layer, and (u^j, v^j) the horizontal fluid momentum per unit mass similarly averaged, the linearized equations of rate of change of momentum are

$$u^j_t-fv^j=F^j-(P_x/\rho^j),\quad v^j_t+fu^j=G^j-(P_y/\rho^j). \tag{A 3}$$

Here, P must be obtained by solving (A 2), given that in the disturbed state of the ocean the vertical extent of the kth layer may have changed from its undisturbed value H^k to a disturbed value h^k which by mass continuity satisfies

$$h^k_t/H^k=-(u^k_x+v^k_y). \tag{A 4}$$

If we ignore any slight discontinuity of P at the interface between two layers,† the solution of (A 2) satisfying $P=0$ at the free surface takes at a point in the jth layer the form

$$P=g[\rho^1h^1+\rho^2h^2+\ldots+\rho^{j-1}h^{j-1}+\rho^j(h^j+h^{j+1}+\ldots+h^N-z)]. \tag{A 5}$$

† Such a discontinuity is possible, even though p itself is continuous, owing to slight variations in the compressibility with temperature and salinity, but its effect on the final equations is negligibly small.

Here, z is measured from the bottom, so that the term in round brackets represents the thickness of the part of the jth layer above the point in question. Equations (A 3), with expression (A 5) substituted for P, can be differentiated with respect to t to eliminate the h^k through use of (A 4), giving

$$u^j_{tt} - fv^j_t = F^j_t + g\sum_{k=1}^{N} a_{jk}(u^k_{xx} + v^k_{xy}), \tag{A 6}$$

$$v^j_{tt} + fu^j_t = G^j_t + g\sum_{k=1}^{N} a_{jk}(u^k_{xy} + v^k_{yy}), \tag{A 7}$$

where the matrix element a_{jk} has the value

$$a_{jk} = H^k\rho^{\min(j,k)}/\rho^j. \tag{A 8}$$

Solutions of (A 6) and (A 7) can be expressed as a combination of normal modes; in each of these modes the distribution of horizontal force and momentum per unit mass between the different layers is in fixed proportions c^j ($j = 1$ to N); that is

$$(u^j, v^j) = c^j(u, v), \quad (F^j, G^j) = c^j(F, G). \tag{A 9}$$

In order that (A 6) and (A 7) take the form of equations for (u, v) alone, namely,

$$u_{tt} - fv_t = F_t + gH(u_{xx} + v_{xy}), \tag{A 10}$$

$$v_{tt} + fu_t = G_t + gH(u_{xy} + v_{yy}), \tag{A 11}$$

which are (1) and (2) of the main paper, the c^j must satisfy

$$\sum_{k=1}^{N} a_{jk}c^k = Hc^j. \tag{A 12}$$

This means that c^j must be an eigenvector of the matrix a_{jk}, with H as associated eigenvalue.

If (A 10) is operated on by $gH\,\partial^2/\partial x\,\partial y - f\partial/\partial t$, and then added to the result of operating on (A 11) by $\partial^2/\partial t^2 - gH\,\partial^2/\partial x^2$, the terms in u cancel (even though $\mathrm{d}f/\mathrm{d}y = \beta \neq 0$) and the equation after one integration with respect to t becomes the fundamental equation (3) of the main paper. Conclusions from that equation are fully worked out there. In this Appendix, however, some further properties of the eigensolutions of (A 12) are elucidated.

Equation (A 12) possesses altogether N linearly independent eigenvectors c^j, all real; each of these, with its associated real eigenvalue H, will be distinguished by a suffix n, where $n = 0, 1, 2, \ldots, N-1$. The need for eigenvectors and eigenvalues to be real is proved by making the simple substitution $c^j/\sqrt{(\rho^j H^j)} = C^j$ in (A 12), from which H is seen to be an eigenvalue (with associated eigenvector C^j) of the real symmetric matrix $\rho^{\min(j,k)}(H^jH^k/\rho^j\rho^k)^{\frac{1}{2}}$. This last fact implies also the orthogonality of different eigenvectors C^j_n and $C^j_{n'}$. In terms of the c^j, this can be written

$$\sum_{j=1}^{N} \rho^j H^j c^j_n c^j_{n'} = 0 \quad (n \neq n'). \tag{A 13}$$

Furthermore, the N different eigenvectors C^j form a complete basis, and so therefore do the c^j. This is important when a given distribution of force per unit mass (F^j, G^j) is applied to the layers $j = 1$ to N. Whatever the distribution, it can be expanded in normal modes, for example as

$$(F^j, G^j) = \sum_{n=0}^{N-1} c^j_n(F_n, G_n), \tag{A 14}$$

where the coefficients (F_n, G_n) are determined using the orthogonality relation (A 13) as

$$(F_n, G_n) = \frac{\sum_{j=1}^{N} \rho^j H^j c_n^j (F^j, G^j)}{\sum_{j=1}^{N} \rho^j H^j (c_n^j)^2}. \tag{A 15}$$

This is important because the nth normal mode of external force distribution can excite only the nth normal mode of the distribution of momentum per unit mass; that is, if

$$(u^j, v^j) = \sum_{n=0}^{N-1} c_n^j (u_n, v_n), \tag{A 16}$$

the (u_n, v_n) are related to the (F_n, G_n) by equations (A 10) and (A 11) with H replaced by H_n.

The special case of surface forcing by wind stress (τ_x, τ_y) is important to this paper. If the Ekman layer is not thicker than the uppermost well-mixed layer, then the momentum increase directly generated by wind stress appears solely in that uppermost layer $j = 1$. Thus, (F^j, G^j) is $(\tau_x, \tau_y)/\rho^1 H^1$ for $j = 1$ and zero for $j > 1$, and the coefficients (A 15) in its normal-mode expansion (14) are

$$(F_n, G_n) = \frac{c_n^1 (\tau_x, \tau_y)}{\sum_{j=1}^{N} \rho^j H^j (c_n^j)^2}. \tag{A 17}$$

This is the formula which, in a more general case of stratification that may be partly continuous, can in the limit be written with the sum replaced by an integral as in (6) of the main paper.

This oceanographically relevant case of partly continuous stratification demands a limiting process in which certain layer thicknesses remain constant (representing thicknesses of well-mixed layers) while the others decrease in magnitude and increase in number (representing continuous stratification). In the limit, it is possible to expand every function which is constant in each well-mixed layer (but may vary arbitrarily elsewhere) in a series of the eigenfunctions $c_n(z)$.

The need to make a distinction between a barotropic mode $n = 0$ and $(N-1)$ baroclinic modes $n = 1, 2, \ldots, N-1$ appears from a particular degenerate case of (A 8), in which the variations in ρ (amounting at most to three or four parts per thousand) are neglected altogether. That is a case when a_{jk} is a matrix of rank 1 (the kth column having all its elements equal to H^k), and so has only one non-zero eigenvalue (the total depth $H^0 = \sum_{k=1}^{N} H^k$), corresponding to an eigenvector satisfying $c^j = 1$ for all j. At the same time, it has the $(N-1)$-fold degenerate eigenvalue zero, corresponding to a class of eigenvectors limited only by the condition $\sum_{k=1}^{N} c^k H^k = 0$. Perturbation by the very small density variations can be expected to alter only slightly the non-degenerate eigenvector $c^j = 1$, with eigenvalue equal to the actual depth H^0; but the degenerate eigenvalue zero should be split into $(N-1)$ separate ones, which can be expected to be far smaller than H^0, but possibly varying in magnitude considerably among themselves.

The perturbation analysis for the barotropic mode $n = 0$, starting from the eigenvector $c^j = 1$ and eigenvalue H^0, is elementary, but the result is hardly worth writing down as the relative changes in eigenvector and eigenvalue are actually smaller than the total relative change in ρ, which is at most three or four parts per thousand. Within the accuracies in which we are interested, we can take the effective depth H_0 of the barotropic mode to be equal to the actual depth

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

H^0, and its distribution of momentum per unit mass to be uniform in the different layers. Owing to the degeneracy, however, more detailed analysis is needed for the baroclinic modes $n = 1$ to $N-1$.

For this purpose, (A 12) may be multiplied by ρ^j/H and subtracted from the same result with $j+1$ substituted for j. We obtain.

$$\rho^{j+1}c^{j+1} - \rho^j c^j = \frac{1}{H}\sum_{k=1}^{N} H^k c^k [\rho^{\min(j+1,k)} - \rho^{\min(j,k)}]$$
$$= \frac{1}{H}(\rho^{j+1} - \rho^j) \sum_{k=j+1}^{N} H^k c^k. \tag{A 18}$$

This is the equation which, to a very close approximation, as stated in the Introduction, we can interpret in terms of the vibrations of a string stretched to unit tension between $z = 0$ and $z = H^0$. In this interpretation, c^j is proportional to the inclination of the string in the jth interval (of width H^j) to its unstretched direction, $1/H$ is equal to the square of the radian frequency of vibration, and the mass concentrated at the join between the jth and $(j+1)$th intervals is $(\rho^{j+1} - \rho^j)/\rho^0$, where ρ^0 is an average value of ρ (no mass being present except at such joins).

This mechanical interpretation is exact for a modified form of equation (A 18), where both sides are divided by ρ^0 and the left-hand side is then approximated by $c^{j+1} - c^j$. For this difference represents the component of tension at right angles to the unstretched direction of the string acting at the join between the jth and $(j+1)$th interval, while the right-hand side becomes the mass at that join times the square of the frequency times the perpendicular displacement

$$\sum_{k=j+1}^{N} H^k C^k$$

of the join in question; thus, the equation states correctly the rate of change of momentum of that mass. For the baroclinic normal modes, c^{j+1}/c^j is not close enough to unity to invalidate the accuracy of replacing $(\rho^{j+1}c^{j+1} - \rho^j c^j)$ by $\rho^0(c^{j+1} - c^j)$.

This mechanical interpretation of the eigenvalues or 'effective depths' $H_1, H_2, \ldots$, of the baroclinic modes, as inverse squares of normal frequencies of vibration of such a stretched string, is valuable because it can be immediately extended to the case of partly continuous stratification. Then the string analogy is correct provided that the relative fall in density, $(-\Delta\rho)/\rho^0$, in every interval Δz of the vertical coordinate z, is equal to the mass on the corresponding length of string. The nth baroclinic mode of variation of (u, v) with z, namely $c_n(z)$, the continuous limit of c_n^j, is still represented by the inclination of the string to its unstretched direction in the nth mode of vibration.

This statement is more general than a formulation as a classical Sturm–Liouville problem

where
$$\left.\begin{aligned} c_n(z) &= \phi'(z), \\ \phi''(z) + (1/H_n)\,(-\rho'(z)/\rho^0)\,\phi(z) &= 0, \\ \phi(0) = \phi(H^0) &= 0, \end{aligned}\right\} \tag{A 19}$$

because that makes no explicit provision for discontinuities in $\rho(z)$; formulations in terms of a Väisälä–Brunt frequency

$$N(z) = \sqrt{[-g\rho'(z)/\rho^0]}$$

are still more unsatisfactory, because delta functions lack square roots.... By contrast, the statement in terms of inverse squares of normal frequencies of vibration of a stretched string gives

useful physical feel for how the eigenvalues will vary. Furthermore, from knowledge regarding normal modes of vibration we can infer the useful general result that, if $H_1, H_2, H_3, \ldots$, are in descending order, the nth baroclinic mode $c_n(z)$ must change sign just n times between $z = 0$ and $z = H^0$.

To apply the general theory to the part of the Indian Ocean near the Equator, it is necessary to take into account distributions of ρ (usually written, in kg/m^3, as $1000 + \sigma_t$) typical of that area. Like other tropical regions, it has a rather sharp thermocline at quite a moderate depth: a change in σ_t from about 24.0 in the upper well-mixed layer (corresponding to a temperature about 24 °C and a salinity about 35.2‰) to about 26.6 (corresponding to a temperature about 12 °C and a salinity about 34.9‰) occurs within a few tens of metres around a depth of 200 m. Beyond this, values of σ_t increase slowly with depth to around 27.7, taking a value close to this at depths exceeding about 2000 m (Warren *et al.* 1966; Hamon 1967).

We consider the character of baroclinic modes, first as resulting from the thermocline by itself, and then from its combined action with the slow density rise below it. The situation when ρ varies only in a thermocline region of rather small thickness can most simply be approximated by a two-layer model. This has just one baroclinic mode $n = 1$, whose effective depth H_1 is easily calculated as

$$H_1 = \frac{\rho^2 - \rho^1}{\rho^0} H^1 \left(\frac{H^2}{H^0}\right), \tag{A 20}$$

where H^1 is the depth of the upper layer and the last factor (H^2/H^0) is very close to 1. For typical values $\rho^2 - \rho^1 = 0.0026\rho^0$, $H^1 = 200$ m and a depth $H^0 = 4000$ m (to whose value, however, the results are not very sensitive), this gives $H_1 = 0.50$ m. The mode has the fluid velocity in the upper layer 19 times as much as in the lower, and in the opposite sense.

The actual thermocline, even in the tropics, is not perfectly sharp; furthermore, its thickness varies considerably. Fortunately the theory of the baroclinic modes for a not perfectly sharp thermocline indicates that the first mode remains essentially like that just described, although with the velocity discontinuity smoothed out over the region of the thermocline. A perturbation method shows, furthermore, that the formula (A 20) remains correct if the depth H^1 to which the 'upper layer' is deemed to extend is the depth of a certain level within the thermocline region, which is higher than the level of the centroid of the density-gradient distribution by a difference equal (for typical distributions) to around one-eighth of the thickness of that region.

This correction is at most 10 m for typical thicknesses, so that the first baroclinic mode is expected to be essentially that for a sharp discontinuity in ρ at about 200 m. Admittedly, with the smoothed-out thermocline higher baroclinic modes exist, but the effective depth of the second baroclinic mode is already very small, being given by a formula like (A 20) with H^1 replaced by at most a tenth of the width of the thermocline region. These higher modes, with effective depths not more than about 0.02 m, have such slow response times that their effect in the problem of this paper is negligible; furthermore, for reasons to be described, they are negligibly excited by surface wind stress.

An adequate representation of baroclinic response in the Indian Ocean needs, in fact, to take into account the gradual density rise below the thermocline region more than these minor effects of the region's thickness. The various forms taken in various places by this density rise to about $\sigma_t = 27.7$ can all be crudely approximated by a linear rise. Accordingly, the first baroclinic mode has been calculated for a sharp density rise $\Delta\rho$ at depth H^1, together with a gradual linear increase in density by a further amount $0.4\Delta\rho$ between depths H^1 and H^*.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 1 gives the calculated effective depth H_1 of the first baroclinic mode as a fraction of $(\Delta\rho/\rho^0)\,H_1$. We shall use the central value, corresponding to $H^1 = 200\,\mathrm{m}$, $H^0 = 4000\,\mathrm{m}$ and $H^* = 1600\,\mathrm{m}$; this latter depth, the limit of linear rise in density beyond which uniform density is assumed, is taken rather less than the depth 2000 m beyond which density rise becomes practically negligible because the rate of rise is declining rapidly for some distance above that level; but the table shows the ratio not to depend critically on the exact values assumed.

TABLE 1. VALUES OF $H_1/[(\Delta\rho/\rho^0)\,H^1]$ FOR VARIOUS VALUES OF H^1/H^0 AND H^*/H^0

	H^*/H^0		
H^1/H^0	0.3	0.4	0.5
0.04	1.514	1.570	1.602
0.05	1.420	1.443	1.450
0.06	1.363	1.370	1.363

Accordingly, the effective depth of the first baroclinic mode in the equatorial zone of the Indian Ocean will be taken as (1.443) (0.0026) (200 m) = 0.75 m. The associated distribution of velocity with depth, $c_1(z)$, is shown in figure 7, where the plain line shows the calculated distribution with a sharp thermocline and the broken line the expected modification in the presence of a more diffuse thermocline.

From this distribution, it is possible to calculate for $n = 1$ the coefficient

$$H^0[c_n(H^0)]^2 \Big/ \int_0^{H^0} [c_n(z)]^2\,\mathrm{d}z. \qquad (\mathrm{A}\ 21)$$

This represents non-dimensionally the relative effect on the surface itself (or, more strictly, within the upper well-mixed layer) of the term in $c_n(z)$ (whose coefficient is given by (6) of the main paper) in the expansion of a surface-acting force. Expression (A 21) is 1 for the barotropic mode $n = 0$, but for the first baroclinic mode $n = 1$ as depicted in figure 7 it takes the value 13.9, a much larger value because $[c_1(z)]^2$ is comparable with its surface values only in regions quite near the surface.

A surface stress can, accordingly, force surface currents through the baroclinic mode $n = 1$ about 14 times as strongly as through the barotropic mode $n = 0$. The extent of response to the forcing is not necessarily in this proportion, however, because solutions of equation (3) are quite different for different values of H. Barotropic modes propagate much more efficiently than baroclinic modes in mid-latitude oceans, while even in equatorial oceans § 4 shows that the above 14 : 1 disadvantage is somewhat reduced.

For higher baroclinic modes, however, the corresponding coefficient is much smaller than even the barotropic value 1, and effective depths H_n are much smaller than for the first mode $n = 1$. Table 2 gives values of the largest of them, H_2, in the same form (and for the same oceanic model) as in table 1. Typically, H_2 is about 8 % of H_1, so that propagation speeds (proportional to $\sqrt{H_n}$) are for $n = 2$ less than 30 % of what they are for $n = 1$. The motion is mainly confined within the region of gradual density rise below the thermocline; surface effects are very small, with values of the coefficient (A 21) around 0.1 or less.

The response of the northern Indian Ocean in the higher baroclinic modes, then, is less than 0.3 times as rapid as in the first mode; in addition, it is quantitatively irrelevant to any

explanation of surface-current response. Only the first baroclinic mode $n = 1$, and to a minor extent the barotropic mode $n = 0$, with times of response which are shown in the main paper to be comparable, need to be taken into account.

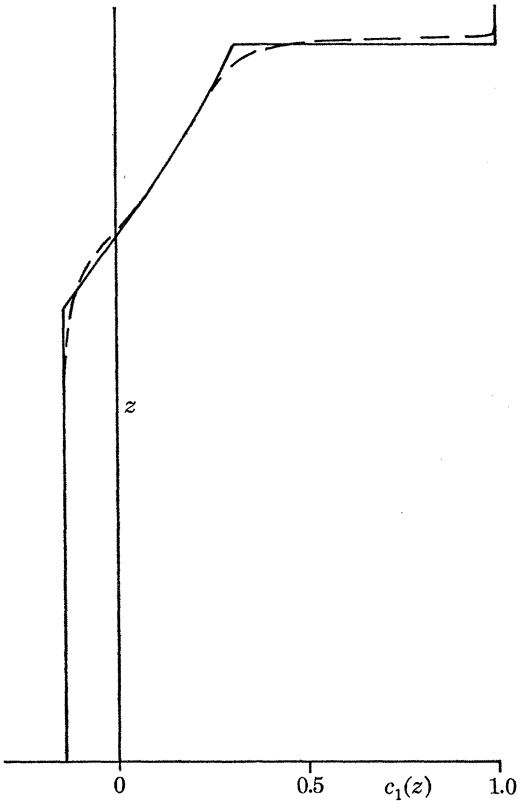

FIGURE 7. Distribution, $c_1(z)$, of velocity with depth in the first baroclinic mode $n = 1$. Plain line: velocity distribution for an ocean model with

$$\rho = \begin{cases} \rho^0 & \text{for} \quad 0 < z/H^0 < 0.6, \\ \rho^0 - (\frac{8}{7})\,(z/H^0 - 0.6)\,\Delta\rho & \text{for} \quad 0.6 < z/H^0 < 0.95, \\ \rho^0 - 1.4\Delta\rho & \text{for} \quad 0.95 < z/H^0 < 1. \end{cases}$$

Broken line: suggested velocity distribution for a density distribution identical except that discontinuities are somewhat smoothed out.

TABLE 2. VALUES OF $H_2/[(\Delta\rho/\rho^0)\,H^1]$ FOR VARIOUS VALUES OF H^1/H^0 AND H^*/H^0

	H_0^*/H^0		
H^1/H^0	0.3	0.4	0.5
0.04	0.109	0.148	0.185
0.05	0.084	0.116	0.145
0.06	0.067	0.094	0.118

The author is most grateful to Dr H. Stommel, Dr J. C. Swallow and Dr G. Veronis for helpful discussions on the subject of this paper.

References

Ananthakrishnan, R. 1964 *Tracks of storms and depressions in the Bay of Bengal and the Arabian Sea, 1877–1960*. Delhi: India Meteorological Department.

Blandford, R. 1966 *Deep Sea Res.* **13**, 941.

Bryan, K. 1963 *J. Atmos. Sci.* **20**, 594.

Eckart, C. 1960 *Hydrodynamics of oceans and atmospheres*. London: Pergamon Press.

Hamon, B. V. 1967 *Deep Sea Res.* **14**, 169.

Hasselmann, K. 1967 *Proc. Roy. Soc.* A **299**, 77.

Knauss, J. A. 1960 *Deep Sea Res.* **6**, 265.

Lighthill, M. J. 1965 *J. Inst. Math. Applics.* **1**, 1.

Lighthill, M. J. 1966 *J. Fluid Mech.* **26**, 411.

Lighthill, M. J. 1967 *J. Fluid Mech.* **27**, 725.

Longuet-Higgins, M. S. 1964 *Proc. Roy. Soc.* A **279**, 446.

Longuet-Higgins, M. S. 1965 *Proc. Roy. Soc.* A **284**, 40.

Longuet-Higgins, M. S. 1968 *Phil. Trans.* A **262**, 511.

Matsuno, T. 1966 *J. met. Soc. Japan* **44**, 25.

Munk, W. H. 1950 *J. Meteorol.* **7**, 79.

Pedlosky, J. 1965 *J. mar. Res.* **23**, 207.

Schott, G. 1935 *Geographie des Indischen und Stillen Ozeans*. Hamburg: Boysen.

Stommel, H. 1958 *The Gulf Stream* (1st edition; see also 2nd edition 1965). University of California Press.

Swallow, J. C. 1967 *Stud. Trop. Oceanogr. Miami*, no. 5, p. 15

Swallow, J. C. & Bruce, J. G. 1966 *Deep Sea Res.* **13**, 861.

Veronis, G. 1966 *Deep Sea Res.* **13**, 31.

Veronis, G. & Stommel, H. 1956 *J. mar. Res.* **15**, 43.

Warren, B. A. 1966 *Deep Sea Res.* **13**, 167.

Warren, B. A., Stommel, H. & Swallow, J. C. 1966 *Deep Sea Res.* **13**, 825.

Whittaker, E. T. & Watson, G. N. 1927 *Modern analysis*. Cambridge University Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES