THE OFFICIAL MAGAZINE OF THE OCEANOGRAPHY SOCIETY

Oceanography

CITATION

da Silva, J.C.B., J.M. Magalhães, T. Gerkema, and L.R.M. Maas. 2012. Internal solitary waves in the Red Sea: An unfolding mystery. *Oceanography* 25(2):96–107, http://dx.doi.org/10.5670/oceanog.2012.45.

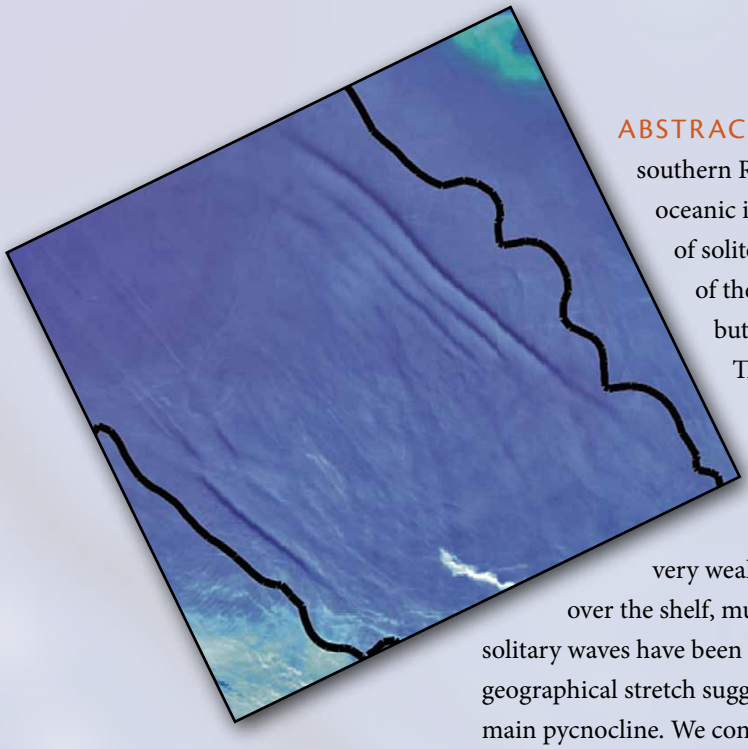
DOI

http://dx.doi.org/10.5670/oceanog.2012.45

INTERNAL SOLITARY WAVES IN THE RED SEA

AN UNFOLDING MYSTERY

BY JOSÉ C.B. DA SILVA, JORGE M. MAGALHÃES, THEO GERKEMA, AND LEO R.M. MAAS



ABSTRACT. The off-shelf region between 16.0° and 16.5°N in the southern Red Sea is identified as a new hotspot for the occurrence of oceanic internal solitary waves. Satellite observations reveal trains of solitons that, surprisingly, appear to propagate from the center of the Red Sea, where it is deepest, toward the continental shelf, but they do not survive as coherent structures over the shelf. These solitons are characterized by coherent crest lengths exceeding 80 km and crest-to-crest distances of more than 2 km, compatible with signatures of large-amplitude solitary waves. Despite the fact that these Red Sea solitons have large amplitudes, they appear to be generated by very weak surface tides. Tidal current velocity is only about 5 cm s^{-1} over the shelf, much weaker than over other ocean shelves where similar solitary waves have been reported. The appearance of these waves over this particular geographical stretch suggests generation by a locally amplified internal tide on the main pycnocline. We consider three possible explanations for soliton generation in the Red Sea: interfacial tide resonance, local generation by internal tidal beams generated at the shelf breaks, and local generation by internal tidal beams generated at the shelf breaks but first amplified by repeated focusing reflections.

INTRODUCTION

The coastal ocean and the continental margin (including the shelf and the slope) are often characterized by the presence of very large-amplitude internal waves of short period (typically 30 minutes or less). These waves occur in the ocean's interior and propagate horizontally, concentrating their energy around the oceanic pycnocline. They can have amplitudes up to 100 m and are usually generated by the interaction of the surface tide with seafloor topography. These waves are often referred to as internal solitary waves (ISWs) because they occur in isolated packets that are rank ordered in amplitude (the leading wave usually being the largest because nonlinear theory indicates faster propagation for the largest-amplitude wave). ISWs are equally ubiquitous in the atmosphere (usually confined to the lower troposphere), occurring when a wave duct permits their generation and existence, which usually occurs due to a strong thermal inversion layer aloft. Both oceanic and atmospheric ISWs can be easily detected in satellite imagery, provided adequate spatial resolution is attained. They can be identified in various wavebands of the electromagnetic spectrum because they produce currents (or winds and clouds, in the case of atmospheric waves) with sufficient surface strain to alter the surface roughness of the ocean. Thus, in many areas of (and over) the ocean, our knowledge of their existence comes from satellite imagery alone.

The Red Sea is an example of a region where our knowledge about ISWs is based on satellite measurements. Magalhães et al. (2011) conducted a Red Sea synthetic aperture radar (SAR) image survey that revealed a nearly year-round abundance of large atmospheric gravity waves with solitary-like wave train character, especially between April and September. The survey also revealed, in some instances, the existence of oceanic ISWs (Christopher Jackson, Global Ocean Associates, *pers. comm.*, 2010; see also Jackson, 2007). Here, we give a preliminary account of some of the available observations of oceanic ISWs in the Red Sea to elucidate three mysterious aspects. First, these ISWs occur repeatedly only over a particular 80 km stretch where the Red Sea is deepest (about ~ 1,500 m deep, along the sea's main axis). Second, they occur in spite of the surface tide—generally seen as their source—being very weak. Third, the ISWs originate from the deep parts of the Red Sea and not from the nearest shelf edge, as is common. Given these unusual characteristics, the source and generation mechanism of these ISWs are unclear at present.

We propose three possible explanations that we hope will motivate further studies focusing, in particular, on modeling and in situ measurements of these waves' temporal and spatial characteristics.

Large tidal-period internal waves (internal tides) are well known to result from the interaction of strong surface (or astronomic) tides with steep seafloor or shelf-break topography. Subsequently, these internal tides propagate as interfacial waves and steepen as they propagate away from the shelf break; ISWs then develop, usually phase locked to the internal tides' troughs (see, e.g., Pingree et al., 1986; Gerkema, 1996; Alford et al., 2010). When the maximum surface tidal current at the internal wave generation site is significantly less than the phase speed of the generated internal tides, it is generally accepted that linear theory should provide a good approximate description of the internal tides (see, e.g., Vlasenko et al., 2005). Tidal currents over the shelf of the Red Sea reach only about 5 cm s^{-1}. This estimate is obtained using the Oregon State University Tidal Inversion Software (OTIS), available with a high resolution of 1' (see Egbert and Erofeeva, 2002, for details). Such currents are very weak compared to shelves of other regions where ISWs have been reported, for example, over the Sofala Bank of the Mozambique Channel where tidal currents can easily reach 1 m s^{-1} (da Silva et al., 2009).

In principle, therefore, linear models should account for low-amplitude internal wave generation in the Red Sea. However, as we will show, this is not the case. Explaining the existence of powerful ISWs in a localized region of the Red Sea represents a challenge

José C.B. da Silva *is Senior Scientist, Department of Geosciences, Environment and Spatial Planning and Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), University of Porto, Portugal, and Guest Investigator, Woods Hole Oceanographic Institution, Woods Hole, MA, USA.* **Jorge M. Magalhães** *is PhD Candidate, Department of Geosciences, Environment and Spatial Planning, University of Porto, Portugal.* **Theo Gerkema** *is Senior Scientist, Royal Netherlands Institute for Sea Research (NIOZ), Texel, the Netherlands.* **Leo R.M. Maas** *(maas@nioz.nl) is Senior Scientist, NIOZ, Texel, the Netherlands, and Professor, Institute for Marine and Atmospheric Research Utrecht, Utrecht University, the Netherlands.*

to our current knowledge about internal waves. In this paper, we show that the off-shelf region between 16.0° and 16.5°N in the southern Red Sea is a hotspot for the occurrence of oceanic ISWs. Surprisingly, these ISWs appear to originate somewhere beyond the middle of the Red Sea's deep "trough," and then propagate onshore from this region. The pycnocline appears to be located inside of this deep region as a result of the continental shelf being generally shallower than the mixed-layer depth in this region. We first describe the observations of these shelf-inward propagating ISWs and then discuss possible generation mechanisms. The questions regarding why these ISWs are confined to that location (no other hotspot of oceanic ISWs is currently known in the Red Sea), how weak surface tidal currents can generate them, and why they first occur over the deep trough and then propagate shelfward all remain enigmatic. We investigate possible explanations for these observations as well as for the fate of the waves as they approach the shelf regions after apparently propagating tens of kilometers. It is important to stress that we observe no significant ISW activity over the shelves. At present, it is unknown what happens to the strong ISWs as they reach the shelf break.

OBSERVATIONS

Figure 1 is a typical MERIS (Medium Resolution Imaging Spectrometer) satellite image of the southern Red Sea. This full-resolution (300 m) image has been processed to simulate true color. ISWs are evident (in box with dashed white lines) because there are coherent sunglint patterns along wave crests that result from variable surface roughness produced by surface currents associated with the internal waves (e.g., Jackson and Alpers, 2010). The overall pattern of the ISW crests suggests that these waves propagate from the center of the Red Sea trough toward the continental shelf, but apparently not over the shelf and toward the coasts of the Arabia Peninsula to the east and the African continent to the west. It is possible that the ISWs partially reflect off the shelf when they propagate over the steep continental slopes. In fact, ISWs were observed to reflect off the shelf on two occasions in the satellite images. For instance, they were seen in two consecutive Moderate Resolution Imaging Spectroradiometer (MODIS) images near the slope with opposite propagation directions that were separated by a few hours.

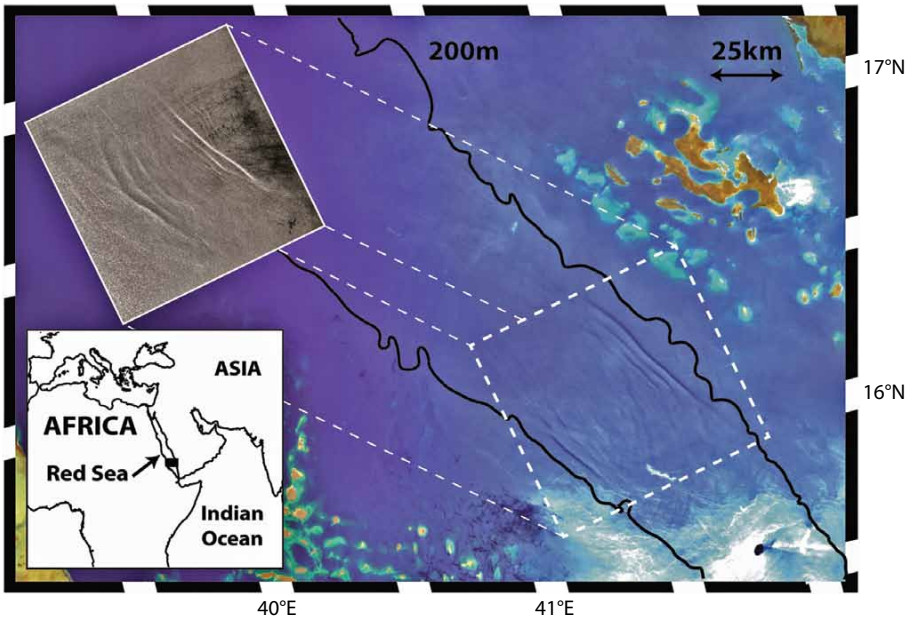


Figure 1. MERIS (Medium Resolution Imaging Spectrometer) satellite image from April 18, 2008, showing surface manifestations of internal solitary waves in the Red Sea (in box with dashed lines). The inset in the top left corner shows a high-resolution synthetic aperture radar (SAR) image of that box from April 20, 2008. The inset in the lower left corner shows the study region in relation to the Red Sea map. The black line gives the 200 m depth contour.

Figure 2a is a composite drawing of internal wave crests based on a sequence of MODIS and MERIS images acquired during the spring of 2008. Internal waves occur in a limited region, situated between the Saso Islands (S) to the east and the Dahlak Marine Park Islands (D) to the west. This location coincides with one of the widest zones of the southern Red Sea where the surface tides are strongest, but only slightly stronger than anywhere else in the Red Sea. Yet, these tidal currents are very weak compared to other continental shelves where ISWs are usually observed. ISWs observed in the satellite imagery are characterized by crest lengths exceeding 80 km (see Figures 1 and 2a) and crest-to-crest distances of more than 2 km (see SAR backscatter profile in Figure 3b). Internal waves were observed in images made during the boreal winter months, from November to May. No ISW signatures were apparent in satellite images corresponding to the summer months (June to October).

In addition to imaging spectrometers, such as MODIS and MERIS that operate in the visible wavelengths of the electromagnetic spectrum, ISWs are also detectable in radar images (e.g., Alpers, 1985), where they have higher-resolution signatures, showing more detail. Radars also are independent of cloud cover or aerosol contamination, and are capable of working equally well during the day and at night. Thus, we have undertaken a systematic survey of archived SAR data from the European Remote Sensing (ERS) satellites and their successor, Envisat (Environmental Satellite). The SAR analysis also revealed the preferential activity of ISWs between 16°N and 16.5°N in the Red Sea. The SAR data allowed us to build a time sequence of ISW crests in order to study their evolution in relation to the semidiurnal tides. Such a sequence of observations can be translated into a travel-time diagram that is useful for estimating the physical properties of these waves, such as propagation speed. Previous studies have successfully used travel-time graphs to retrieve phase speeds of ISWs (e.g., Pingree et al., 1986; da Silva et al., 2011). When we assume that ISW generation occurs at the same tidal phase for each semidiurnal cycle, the slope of the linear fit to the data (position versus time since a fixed tidal phase) is a good approximation to the average propagation speed of the waves.

Figure 3a is a travel-time plot based on all available SAR images from Envisat and ERS satellites (12 images), along with a selection of MODIS and MERIS images, relating position (distance from a reference point, whose geographic coordinates are 16.20°N, 41.35°E) and time (with respect to low water at Harmil Island; see Figure 2a) of each ISW packet detected in the satellite. The intent of the plot is to help locate the generation position and time (with respect to the semidiurnal tide phase) of the ISWs, and to assess the propagation speeds of the waves. The reference point chosen to measure the travel distance by the ISWs is arbitrary, and, in this case, it is located near the center of the Red Sea trough (deepest part). The travel distance is measured from that point, along a line inclined 50° from true north (i.e., the orientation is 40° north of east) to the leading ISW in a given packet of a given image. In the study region, the tidal currents are phase-locked to tidal heights, with a phase shift. The phase shift is 2.99 hours ± 5 min, which corresponds to approximately one-quarter of a semidiurnal tidal period, with an uncertainty of approximately 3% (standard deviation) relative to the semidiurnal period (according to the OTIS model). In Figure 3a, we used the time of low water (at Harmil) as reference, as this time can be readily related to the tidal current's phase (e.g., maximum flood is approximately three hours after low water). Figure 3a shows that the average propagation speeds of the SAR ISWs (slopes of dashed linear fit lines) are 1.3 m s^{-1} and 1.1 m s^{-1} for east- and west-propagating waves, respectively. Phase speed

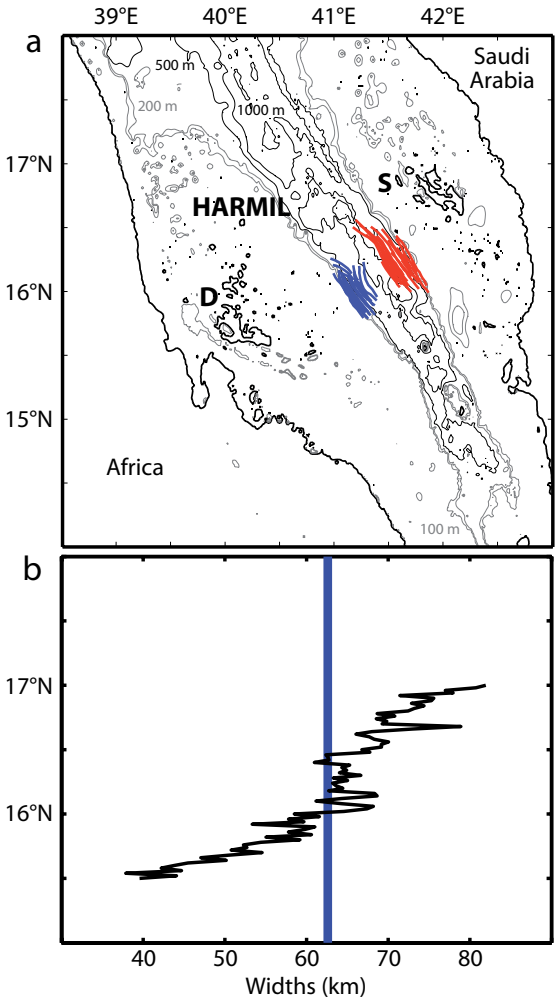


Figure 2. (a) Composite map based on a sequence of MODIS (Moderate Resolution Imaging Spectroradiometer) and MERIS images acquired April 17–22, 2008, showing internal solitary wave crests (in blue and red) in relation to bathymetry of the southern Red Sea. The innermost isobath is 100 m, followed by 200 m, 500 m, 1,000 m, and 1,500 m. The wave crests marked in red were identified as propagating with an eastward component, while those crests marked in blue propagate toward the west. This color code is also used in Figures 3a and 4b. The letter "S" denotes the Saso Islands and "D" the Dahlak Marine Park Islands (see text for details). The tide station of Harmil (see text) is also marked in the map. (b) Variation of Red Sea width (defined for depths below 200 m) with latitude, along transects inclined 50° from north taken approximately perpendicular to the 200 m depth contours on continental slopes of both margins. The blue vertical line represents the wavelength of a semidiurnal interfacial tidal wave that matches the width of the Red Sea approximately between latitudes 16° and 16.5°N, for the stratification of March in Figure 4a.

estimations based on a sequence of three subsequent images revealed speeds of about 1.3 m s^{-1}, consistent with the linear fit to the SAR data in Figure 3a. Note that the solid triangles represent space-time coordinates obtained from a sequence of images separated in time by approximately 24 hours on three consecutive days (May 4–6, 2004), while the slopes of the continuous thick lines linking the solid triangles provide their phase speeds. The filled circles in Figure 3a correspond to a time series based on a sequence of six consecutive days of MODIS/MERIS data. The MODIS/MERIS data points fit better linear regressions than the SAR data (they are less scattered than the SAR data, shown as open circles), perhaps because the stratification was more stable during these six consecutive days compared to the less-homogeneous and much longer time series of the SAR. Note, however, that the slopes of the linear fits to the filled circles are slightly less steep than those for the SAR data (which is particularly noticeable for east-propagating waves), meaning that, during the period of this sequence, the phase speeds were smaller (1.1 m s^{-1} and 0.9 m s^{-1} for east- and west-propagating waves, respectively).

DISCUSSION

We now investigate the generation mechanisms of ISWs observed in the Red Sea. It is important to stress that the ISWs observed in the SAR images exhibit large-amplitude signals (comparable to observed SAR ISW signatures in other regions, such as near Mascarene Ridge in the western equatorial Indian Ocean and Sofala Bank in the Mozambique Channel). Figure 3b shows a SAR backscatter profile obtained from an image recorded on April 20, 2008, at 19:13 UTC, also shown in the top Figure 1 inset. The radar intensity variations are normalized relative to background clutter, unperturbed by the ISWs, as is common practice for wave contrast analysis (da Silva et al., 1998). The SAR image contrasts produced by the Red Sea ISWs shown in Figure 3b are comparable, and actually quite close, to SAR image contrasts seen in other regions where very powerful ISWs have been documented (see, e.g., Figure 6 of da Silva et al., 2009, which provides documentation on ISWs in the Mozambique Channel, and Figure 3 of da Silva et al., 2011). Given the similar radar morphology of Red Sea ISWs compared to ISWs in other regions with similarly strong stratification and large depths, we expect comparable isotherm displacements (ISW amplitudes in the Mascarene Ridge region exceed 50 m; e.g., Konyaev et al., 1995). In the Mozambique Channel, ISWs are very intense because of strong tidal forcing near the Sofala Bank shelf break. We are left with a puzzle: how can tides an order of magnitude weaker and clearly subcritical (Froude number < 0.02 over the continental slope), such as those of the Red Sea, generate similarly strong ISWs? In the following discussion, we offer three suggestions: (1) resonance and disintegration of interfacial tides, (2) generation of interfacial tides by impinging, remotely generated internal tidal beams, and (3) as (2), for geometrically focused and amplified internal tidal beams.

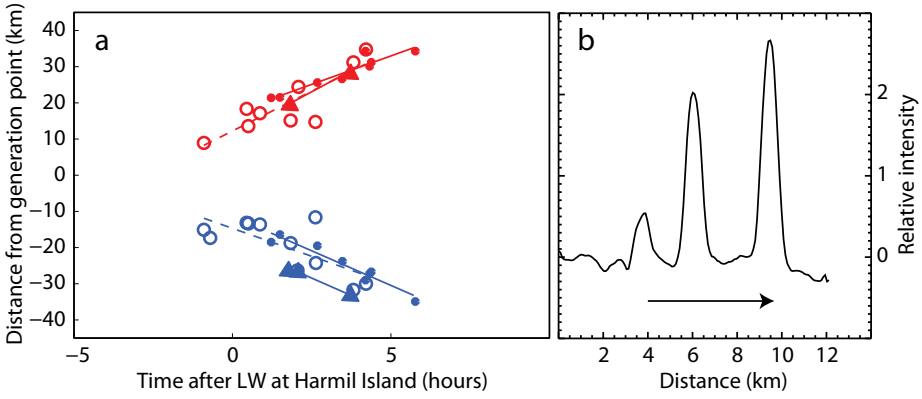


Figure 3. (a) Travel-time graph based on available satellite images for this study. The vertical axis represents the distance (in km) of leading internal solitary waves in each packet, measured from the assumed generation position along the propagation axis (see text for details). Times (in hours) refer to low water (LW) at Harmil Island. Open blue circles represent westward-propagating waves (negative distances), and open red circles show eastward-propagating waves (positive distances) observed in SAR images. Dashed lines are linear fits to the SAR data. The filled triangles, representing satellite synergy (see text for details), are linked by thick lines whose slopes give average propagation speeds (approximately 1.3 m s^{-1} for both eastward- and westward-propagating waves). Filled circles represent a time series of six consecutive days, while thin lines are linear fits to those data points. (b) SAR image backscatter profile across an eastward-propagating internal solitary wave train obtained from an Envisat (Environmental Satellite) Advanced SAR Wide-Swath (ASAR WS) image dated April 20, 2008, in the southern Red Sea. Note that image contrasts are normalized by unperturbed backscatter away from the influence of internal solitary waves. The arrow indicates the propagation direction of internal solitary waves.

Resonant Interfacial Tides

It is well known that (interfacial) internal tides can generate ISWs through nonlinear effects as they evolve in time

(e.g., Gerkema, 1996). Often, these waves occur on the abrupt interface between a buoyant surface layer and a denser layer beneath, and such waves are termed "interfacial" for this reason. This internal tide evolution begins by splitting a large depression (or lee wave) that forms during ebb (flow in offshore direction) near the continental shelf edge (e.g., Pingree et al., 1986; Vlasenko et al., 2005; Jackson et al., 2012, in this issue). In principle, the initial large depression produces two perturbations (pycnocline depression waves), one of which propagates toward the shelf and the other away from the shelf. In most cases, these depression waves, or ISWs, propagate onto the shelf. Sometimes, both on-shelf and off-shelf directed ISWs propagate in phase with the internal tides and are generally phase locked with them (e.g., in the northern Bay of Biscay; Pingree et al., 1986; New and da Silva, 2002). In some regions, however, such as Sofala Bank in the Mozambique Channel, ISWs are seen to propagate only in the off-shelf direction (see da Silva et al., 2009, their Figure 1). This also seems to be the case for the Red Sea ISWs, but with one fundamental difference: in the Red Sea, the off-shelf-propagating ISWs form quite a long distance from the shelf break. This fact unwraps a pertinent question: can Red Sea solitons be explained by the traditional view of nonlinear evolution of internal tides or, instead, does some other less-common and different generation mechanism come into play?

Here, we consider the possibility that the interfacial tide resonates with the basin scale of the Red Sea (at the latitudes where ISWs have been observed), possibly leading to nonlinear evolution of the interfacial tidal wave and subsequent disintegration into ISWs. Provided the wavelengths are long compared to ocean depth (as in the Red Sea), in a two-layer system, linear theory indicates that the phase speed, c, for interfacial tides with crests that are parallel and horizontal, and with periods long enough to be influenced by the effects of Earth's rotation (represented by Coriolis frequency, f), is given by

$$c = \left(g \frac{\delta\rho}{\rho} \frac{h_1 h_2}{H(1 - f^2/\sigma^2)} \right)^{1/2}, \qquad (1)$$

where g is the gravitational acceleration, $\delta\rho$ is the density difference between the upper layer of thickness h_1 and the lower layer of thickness h_2, ρ is average density, H is the total depth $h_1 + h_2$, and f and σ are the Coriolis and wave frequencies, respectively (e.g., Pingree et al., 1986). In Equation 1, we use a two-layer idealization of the stratification (see Figure 4a). The horizontal dashed line in Figure 4b represents the interface at depth h_1, where upper-layer thickness $h_1 = 75$ m, average total depth $H = 800$ m, and interfacial density difference $\delta\rho/\rho = 0.0029$. These numbers translate to a linear phase speed of approximately $c = 1.4$ m s^{-1} for the southern Red Sea and a characteristic wavelength of $\lambda = 62$ km.

The cross-sectional width of the southern Red Sea varies with latitude, generally increasing from south to north. In Figure 2b, we plot the length of the cross sections measured perpendicular to the main axis of the Red Sea as a function of latitude. The orientation of these cross sections is 50° from true north, to be consistent with the position data in Figure 3a and the cross section shown in Figure 4b. This width was measured for depths greater than 200 m, where we are confident that the waves can propagate (as in Figure 2a). It is then interesting to note that the internal tide wavelength ($\lambda = 62$ km) is quite close to the width of the Red Sea between 16° and 16.5°N (defined as indicated above, and marked with a vertical blue line in Figure 2b).

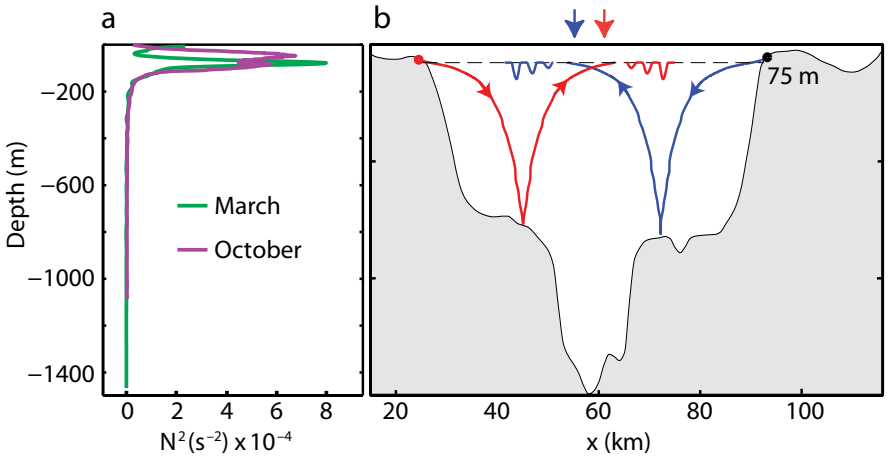


Figure 4. (a) Two vertical profiles of buoyancy frequency represent typical stratifications for two different seasons: purple is typical for October and green for March. (b) Vertical cross section of the Red Sea at the location where internal solitary waves are observed. Red and blue show two internal tide ray paths originating from the upper part of the slopes. Filled circles indicate the positions of origin. The vertical arrows point to the places where the rays reach the interface (dashed line) for the first time. *Buoyancy frequency provided by the US National Ocean Data Center World Ocean Database, http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html*

However, further north or further south of this region, the length scale of the basin differs significantly from the resonant wavelength (see Figure 2a,b). Furthermore, the theoretical two-layer model phase speed is in close agreement with the observed values from the synergy data in the travel-time graph (Figure 3a, filled triangles).

It is puzzling that the linear model phase speed is slightly above the measured ISW phase speeds based on SAR (Figure 3a); in theory, it should be the other way around, because nonlinear theory predicts slightly faster ISW propagation speeds relative to linear waves. Although we do not have an explanation for this at present, we can speculate that it has to do with slight changes in the mixed-layer depth, because Equation 1 is relatively sensitive to h_1. Thus, the theoretical value of the phase speed retrieved from analysis of the stratification in Figure 4a (c = 1.4 m s^{-1}) may not represent the true average value for the SAR observational period or the six-day sequence of MODIS/MERIS images (Figure 3a).

We reiterate that no ISWs were observed in the satellite images in summer (June–October), which may be related to significant changes in stratification and mixed-layer depth variability throughout the year. Such seasonal variability is probably due to the intrusion of Gulf of Aden Intermediate Waters into the Red Sea (e.g., Sofianos and Johns, 2007). Gulf of Aden Intermediate Waters are known to produce a kind of three-layer system (see buoyancy frequency profile in Figure 4a for October, in purple, and the development of a double pycnocline maximum). This three-layer system could, in principle, prevent resonance from occurring, because in the upper ocean it would alter the interfacial propagation of the internal tide, modifying the characteristic wavelength of the interfacial tide and thus changing the resonance length scale.

Some cases have been reported where resonant amplification of internal tides is quite significant. One of these cases is in the deep canyon at the mouth of Tokyo Bay, where the amplification mechanism of the semidiurnal internal tide was examined using a two-dimensional model with simplified bottom topography (Kitade et al., 2011). There, the primary mechanism for amplification of the semidiurnal internal tide during a period of strong stratification is resonance of the internal seiche inside the canyon along with the semidiurnal internal ocean tide. The results of Kitade et al. (2011) revealed that resonance occurs when stratification becomes "appropriate" and the period of the internal seiche approaches that of the semidiurnal tide. Another interesting and relevant case of resonant amplification occurs in the South China Sea (SCS). Powerful internal tides generated in Luzon Strait were modeled and observed propagating to both sides of a double ridge system there (e.g., Farmer et al., 2009). Buijsman et al. (2010) modeled SCS internal tides and found that the region between the two "Gaussian-type" ridges acts as a resonator in amplifying westward-propagating internal tides and solitons for a semidiurnal tidal forcing. These authors suggest that a co-oscillation between the local internal tide generated at both ridges and the incoming internal tides generated by the opposite ridge cause the amplification. In their model results, Buijsman et al. (2010) found that the resonant amplification is optimal for the actual distance of about 100 km between the ridges, which is very close to the internal tide wavelength (for the given stratification) in that location. These general ideas of resonance are corroborated by Alford et al. (2011), who also interpreted the ridge spacing in the SCS as (at least) partly responsible for the strong internal tides (and associated ISWs).

Although resonant amplification of internal tides offers an attractive and plausible explanation for the occurrence of strong internal tides in the southern Red Sea, and in particular explains why the waves would be confined between 16° and 16.5°N, there is still no explanation of why the ISWs would develop as nonlinear features of the resonant internal tide. In particular, the location of the first appearance of the ISWs, beyond the middle of the deepest parts of the trough, is essentially unclear at the present time. Also, it is unclear why these ISWs would appear every tidal period, as this requires a precise match of the growth rate of resonant (small-amplitude) linear internal waves and the tidal frequency.

Local Generation

Several physical mechanisms can generate ISWs in the ocean, and their generation in the Red Sea as a result of nonlinear evolution of an interfacial tide is just one possibility. Another mechanism is known as "local generation," in which tidal beams impinge on the pycnocline from below (see New and Pingree, 1992; Gerkema, 2001; New and da Silva, 2002). In a gently and continuously stratified ocean, the deep layers below the mixed layer exhibit a buoyancy frequency (N) that varies slowly with depth (z), which

permits internal tide energy to propagate as beams (or "rays") inclined to the horizontal. The slope of these rays to the horizontal (s), is determined by the strength of N, in a relation that also includes the wave frequency (σ) and the Coriolis frequency (f) that is related to Earth's rotation:

$$s = \pm \left(\frac{\sigma^2 - f^2}{N(z)^2 - \sigma^2} \right)^{1/2}. \quad (2)$$

In nature, these tidal beams have been observed to reflect off the seafloor and propagate upward, toward the surface. Near the sea surface, the beams may produce large interfacial solitons where they hit the pycnocline from below (New and Pingree, 1992). But, where do these beams form in the first place? They usually emanate from continental shelf edges, where the slope of the bottom topography matches the angle of the beam to the horizontal. These locations are called critical slopes; a downward-propagating ray can form there and, after reflecting off the seafloor, propagate upward and impinge upon the pycnocline quite a long distance away from the continental shelf (typically, some tens of kilometers or even more than 100 km away, depending on the local stratification and the depth where the ray reflects from the bottom). When the beam then encounters the pycnocline, it typically generates both a reflected (downward) beam and also an interfacial disturbance that (through nonlinearities, as explained in Gerkema, 2001, and Akylas et al., 2007) can evolve into a series of higher-frequency ISWs. Such ISWs are trapped in the upper layers of the ocean, where N is large enough to support them, and propagate in the same horizontal direction as the incident beam. New and Pingree (1992) first proposed this kind of ISW generation for the central region of the Bay of Biscay, where the stratification is sufficiently strong in summer to allow for ISW propagation. Our knowledge of regions where ISWs appear to have been generated by this same local mechanism has been increasing in recent years, now including the Iberian Peninsula (Azevedo et al., 2006; da Silva et al., 2007) and Mozambique Channel (da Silva et al., 2009). Numerical and laboratory experiments have been successful in reproducing the local generation of ISWs under controlled conditions (e.g., Grisouard et al., 2011; Mercier et al., in press), and we now believe that the local generation mechanism is more widely applicable in the ocean than previously thought.

> "ISWs ARE PRESENT ONLY OVER THE RED SEA'S CENTRAL, DEEPEST PART, SUGGESTING THEY EITHER BREAK AS THEY PROPAGATE ONTO THE SHELF OR DISSIPATE RAPIDLY."

Figure 4b shows two examples of rays emanating from locations near the upper parts of the (critical) slopes, together with typical stratification of the southern Red Sea (Figure 4a). Their local inclination follows from the theoretical formula (Equation 2). Thus, from a given initial location (the critical slopes, indicated by filled circles in Figure 4b), the trajectory can be calculated using this formula. (We should note that a slightly more complex form has, in fact, been used here because the stratification is very weak in the deeper part of the basin, which makes it important to include the full effects of Earth's rotation (see Gerkema et al., 2008). In this case, in addition to latitude, the angle of propagation in the horizontal plane is also needed (i.e., the orientation north of east, which is 40° in the present case). For each ray, a vertical arrow pointing downward in Figure 4b indicates the horizontal position where it encounters the pycnocline for the first time. We thus see that the ray originating from the western slope (i.e., at left in Figure 4b), marked in red, reaches the pycnocline *to the right* of the position where the blue ray (which comes from the right) surfaces for the first time (see blue ray and blue vertical arrow in Figure 4b). This would explain why the ISWs appear to originate from the center of the basin, yet are due to local generation. In this view, then, the (internal tide) rays lie at the origin of the ISWs.

Note that the three-layer, summer stratification in the Red Sea could potentially inhibit the local generation mechanism of ISWs. The reason is that the strength of the pycnocline is a critical

factor for the evolution of ISWs through the internal tide beam mechanism (e.g., Gerkema, 2001; Akylas et al., 2007; Grisouard et al., 2011; Mercier et al., in press). In addition, the lower pycnocline may distort the rays before they reach the upper one.

Geometric Focusing and Amplification

The weakness of the surface tide and the curious location of shelf-inward propagating ISWs suggest a third explanation. This explanation elaborates on the previous model in which ISW generation was supposed to be due to the internal tidal beam hitting the pycnocline from below. In that model, we supposed this beam was excited near the shelf edge, and that this beam, by virtue of the presence of continuous deep stratification, propagated downward and reflected off the bottom.

This explanation meets with two problems. First, the weak surface tide will generate a weak, downward-propagating internal tidal beam. Second, the reflection of this beam from a part of the bottom that also slopes downward will lead to a broadening of the reflected beam, making it even weaker. The broadening is caused by the anomalous nature of internal wave reflection in a continuously stratified sea. These waves reflect from a sloping bottom while preserving their inclination with respect to the *vertical*, and not, as is common for other types of waves, with respect to the bottom normal (Snell's law). The beam thus defocuses when the reflecting wall's inclination is in the same quadrant as that of the incident beam, while otherwise leading to (geometric) focusing.

The weakness of the reflected beam expresses the need for an internal tidal beam "booster," an amplifying mechanism rendering the internal tide strong enough that at some point it can hit the pycnocline sufficiently hard to excite ISWs. Such a boosting mechanism actually exists, as we illustrate in Figure 5. Here, we idealize the Red Sea by crudely modeling the trough cross section below the pycnocline as a parabolic basin. Moreover, we assume the basin to be filled with a uniformly stratified fluid, which guarantees that internal waves of tidal frequency propagate with a *fixed* inclination relative to the horizontal. In Figure 5a, we follow a single internal wave ray (a number of adjacent rays constituting a beam) from an arbitrary initial location, X_0. Using the simple rule that upon reflection the ray's inclination with the horizontal is preserved, we find that the internal tidal beam bounces back and forth in the deep, stratified part of the trough, while being subject to focusing reflections at the supercritically sloping, upper parts of the two sides of the trough. These are the parts of the bottom that are steeper than the rays, given by inclination *s*, defined in Equation 2. Suppose for the moment that the energy in an internal tidal beam is conserved during reflection. Upon the beam's reflection from a supercritical slope, the decrease in beam width accompanying such focusing reflections leads to an *increase* in its amplitude (i.e., to an increase in the particle velocity within the beam). While this focusing is partly offset by subsequent defocusing reflections at subcritical parts of the bottom, in enclosed fluid domains, focusing

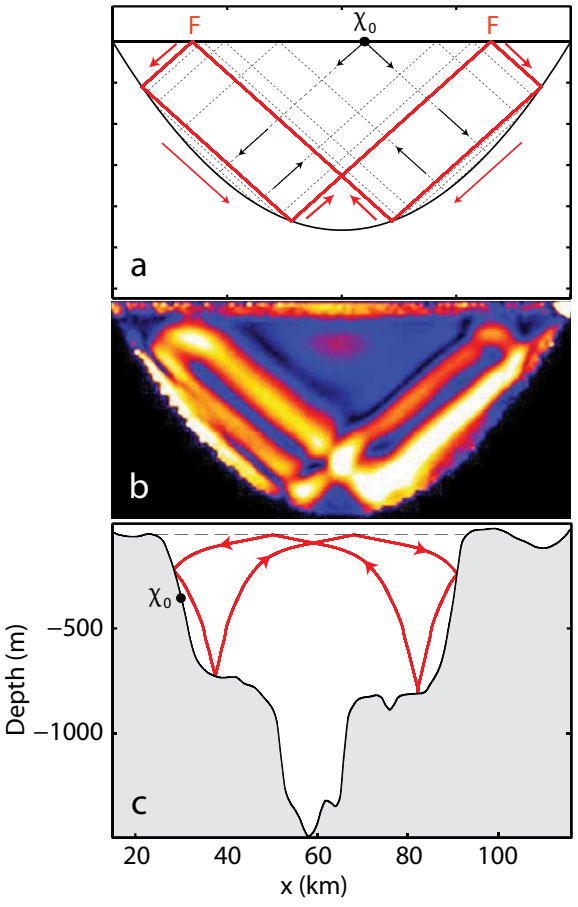


Figure 5. Uniformly stratified parabolic basin. (a) Rays along which internal waves of suitable frequency propagate. Rays (dashed) maintain fixed inclinations relative to the horizontal, and move rightward and leftward from X_0. Both rays approach the limit cycle (wave attractor), displayed in red, and eventually propagate in the *same* direction, indicated by long red arrows (from Maas and Lam, 1995). (b) Laboratory experiment showing observed spatial distribution of the amplitude of density perturbations due to internal waves generated by weakly rocking a tank sideways. Notice that in this experiment the surface (transition between blue and orange at the top) is a free surface. *Courtesy of Jeroen Hazewinkel.* (c) Simple wave attractor predicted by ray-tracing simulations in the frequency range $0.85–0.95 \times 10^{-4}$ rad s^{-1} for the stratification of the study region in winter. Rays are launched at X_0.

dominates defocusing. The amplification mechanism that we appeal to requires weak friction at the bottom and within the water column so that the internal tide may survive a few reflections from the pycnocline and the bottom. In theory, as Figure 5a shows, this ray and all other rays appear to approach a unique limit cycle (red lines), leading to a pileup of internal tidal energy. Internal waves are steered to this particular location regardless of where they are forced within the trough. For this reason, this limit cycle is referred to as a *wave attractor* (Maas and Lam, 1995; Maas, 2005). During the approach of the attractor, again in theory, the wave amplitude would grow indefinitely. Of course, other physical processes, neglected so far, will check this growth. The excitation of ISWs at the pycnocline, at locations labeled F in Figure 5a, is one such process (others being viscous damping and wave breaking). Note that internal waves propagating in the interior toward the attractor follow the direction of the red arrows in Figure 5a. At focusing points F, they propagate *toward* the nearest shelf, as observed. Figure 5b shows the result of a laboratory experiment in a configuration as sketched above (Jeroen Hazewinkel, Scripps Institution of Oceanography, *pers. comm*., 2012). In that experiment, however, the surface is the fluid's free surface, and not the pycnocline. Internal waves are excited by weakly rocking the tank sideways, and this figure displays in color the magnitude of density perturbations taking place in a wave period at each pixel. It shows localization and amplification of wave energy at an attractor. Further observations (Hazewinkel et al., 2010) also reveal the propagating nature of the internal waves, continuously approaching the attractor where, lacking a pycnocline, they are absorbed by viscous dissipation (Hazewinkel et al., 2008).

The ray pattern in Figure 5c shows that for the winter stratification and transect of Figure 4, such a wave attractor might indeed also exist in the Red Sea. Nonuniformity of stratification manifests itself again in curved rays. This simple attractor exists in the frequency range $0.85–0.95 \times 10^{-4}$ rad s^{-1}. This frequency is a little below the actual semidiurnal M_2 tidal frequency of 1.4×10^{-4} rad s^{-1}. Given the uncertainties in the actual stratification frequency profile, its assumed independence of the cross-trough coordinate, and the neglect of concomitant vertically sheared currents (all affecting ray paths), this result is encouraging.

While for any given ratio of basin aspect ratio (depth H, divided by half-width L) to ray slope s, wave attractors exist in nontrivially shaped, stratified basins, such as the parabola (Figure 5a) or Red Sea trough (Figure 5c), the shape of these ray pattern cycles can range from simple (as in Figure 5) to very complicated. We expect that only the simpler ones, discussed in the previous paragraph, are physically relevant (but see Hazewinkel et al., 2010). Although, theoretically, interfacial wave resonance, discussed earlier, occurs for discrete frequencies, wave attractors of a shape displayed in Figure 5 occur over a *finite* frequency σ-interval (see previous paragraph). These attractor intervals again define resonant situations, but the condition under which resonance occurs now appears fuzzy. This means that the resonance condition is more easily met in practice.

One final issue concerns ISW speed. Traditionally, the small-scale, large-amplitude disturbances of the interface visible in satellite imagery are interpreted as high-frequency ISWs. In the wave attractor view, discussed in this section, these ISWs are believed to result from the disintegration of interfacial displacements produced by the impinging amplified internal tide beam. But, ISWs are nonlinear waves and, in theory, retain their coherence by propagating *faster* than the fastest linear interfacial wave. Our estimates of propagation speeds, however, suggest this might not be the case for the interfacial disturbances that we analyzed. Another explanation for the appearance of these high-frequency ISWs is that geometric focusing leads directly to strong scale reduction of the cross-beam width of the internal tide beam, without the need for the broad incident beam to disintegrate. This would imply that these small-scale features might retain the long tidal period of the internal tide beam. Moreover, because the wave attractor, toward which the internal tide beam propagates, is a structure that is fixed in space, the expression of the waves at the interface can, in this view, acquire a paradoxical character. Their crests and troughs (i.e., their phase) might propagate toward the shelf, without the need for any corresponding propagation of their envelope (i.e., of their energy), because the wave *group* is fixed to the attractor. This would explain why SAR disturbances are observed only over the trough and not on the shelves. Because our satellite images were acquired on different days, the spatial spread in their location might, in that interpretation, be due to a changing attractor location owing to changes in stratification.

SUMMARY

Satellite observations reveal the frequent presence of strong ISWs in winter over a particular 80 km stretch of the Red Sea. It is puzzling that they appear to be generated by very weak surface tides somewhere near the middle of the sea, and they propagate from this location toward the shelf edge.

ISWs are present only over the Red Sea's central, deepest part, suggesting they either break as they propagate onto the shelf or dissipate rapidly. The appearance of ISWs over this particular geographical stretch indicates a locally amplified internal tide on the main pycnocline. We considered three possible explanations for their appearance: (1) interfacial tide resonance, (2) local generation of ISWs by internal tidal beams generated at the shelf breaks, and (3) the same as (2) but for internal tidal beams that are first amplified by repeated focusing reflections.

The discussion of the first mechanism idealizes the Red Sea using a two-layer stratification between the deep trough's sides. This model clearly allows for the possibility of resonance when the interfacial wavelength matches the trough width. It explains the geographical location (in latitude) at which the ISWs are seen; however, it remains unclear why the resonantly amplified interfacial tide develops into ISWs a bit beyond the middle of the trough (see Figure 2a) and why the waves would break every tidal period, which requires an exact match with the inverse growth rate.

The second mechanism involves local generation. The discussion considers the Red Sea stratification to be continuous and to extend below and above the main pycnocline. This model allows for obliquely propagating internal tidal beams, believed to originate at the shelf edges, in addition to the small-amplitude interfacial tide that remains trapped to the pycnocline. Local generation of ISWs assumes them to result from beams that reflect from the bottom and then hit the main pycnocline from below, somewhere beyond the middle of the trough. This mechanism explains why ISWs are found at the opposite sides of the trough and propagate toward the opposite shelf. It does not explain why the beams are stronger than the directly generated interfacial wave or why they can excite any response at all, given that they are generated by weak surface tides, weakened further by defocusing reflections from the bottom.

The third mechanism describes how the internal tidal beam might first be amplified by repeated focusing reflections before exciting a pycnocline disturbance. As demonstrated in laboratory experiments, this focusing and amplification mechanism explains why a weak surface tide can be converted into a strong internal tide beam. This boosting mechanism requires that the internal tide beam be subject to weak friction and that the main pycnocline act as a reflector during the first few boosting cycles. The pycnocline disturbance may consist of shelfward-propagating ISWs, although the relatively low observed ISW speed suggests it may also express the (non-propagating) attractor reflection itself.

Clearly, to address the properties of the interfacial disturbances (ISWs), and to address which of these mechanisms, or perhaps others, are at work, we recommend sea-going observations in wintertime in this particular region of the Red Sea.

> "CLEARLY, TO ADDRESS THE PROPERTIES OF THE INTERFACIAL DISTURBANCES (ISWs), AND TO ADDRESS WHICH OF THESE MECHANISMS, OR PERHAPS OTHERS, ARE AT WORK, WE RECOMMEND SEA-GOING OBSERVATIONS IN WINTERTIME IN THIS PARTICULAR REGION OF THE RED SEA."

ACKNOWLEDGEMENTS

The SAR and MERIS image data presented here were provided by the European Space Agency (ESA) under Project number AOPT-2423. MODIS image data in Level 1 (250 m) format were obtained from the NASA LAADS online archive. We are grateful to Christopher Jackson of Global Ocean Associates for providing critical information about a series of satellite observations in the Red Sea, which encouraged the progress of this investigation. Our interest in Red Sea ISWs dates to early 2007, when one of us was on sabbatical

at WHOI and had stimulating discussions with Jesus Pineda of the Biology Department about satellite observations in the Red Sea. JdS would like to thank the NIOZ Royal Netherlands Institute for Sea Research for warm hospitality received during a sabbatical there, when most of this work was carried out.

REFERENCES

Akylas, T.R., R.H.J. Grimshaw, S.R. Clark, and A. Tabaei. 2007. Reflecting tidal wave beams and local generation of solitary waves in the ocean thermocline. *Journal of Fluid Mechanics* 593:297–313, http://dx.doi.org/10.1017/S0022112007008786.

Alford, M.H., R. Lien, H. Simmons, J.M. Klymak, S. Ramp, Y.J. Yang, D. Tang, and M.-H. Chang. 2010. Speed and evolution of nonlinear internal waves transiting the South China Sea. *Journal of Physical Oceanography* 40:1,338–1,355, http://dx.doi.org/10.1175/2010JPO4388.1.

Alford, M.H., J.A. MacKinnon, J.D. Nash, H. Simmons, A. Pickering, J.M. Klymak, R. Pinkel, O. Sun, L. Rainville, R. Musgrave, and others. 2011. Energy flux and dissipation in Luzon Strait: Two tales of two ridges. *Journal of Physical Oceanography* 41:2,211–2,222, http://dx.doi.org/10.1175/JPO-D-11-073.1.

Alpers, W. 1985. Theory of radar imaging of internal waves. *Nature* 314:245–247, http://dx.doi.org/10.1038/314245a0.

Azevedo, A., J.C.B. da Silva, and A.L. New. 2006. On the generation and propagation of internal solitary waves in the southern Bay of Biscay. *Deep-Sea Research Part I* 53:927–941, http://dx.doi.org/10.1016/j.dsr.2006.01.013.

Buijsman, M.C., J.C. McWilliams, and C.R. Jackson. 2010. East-west asymmetry in nonlinear internal waves from Luzon Strait. *Journal of Geophysical Research* 115, C10057, http://dx.doi.org/10.1029/2009JC006004.

da Silva, J.C.B., S.A. Ermakov, I.S. Robinson, D.R.G. Jeans, and S.V. Kijashko. 1998. Role of surface films in ERS SAR signatures of internal waves on the shelf: Part 1. Short-period of internal waves. *Journal of Geophysical Research* 103(C4):8,009–8,031, http://dx.doi.org/10.1029/97JC02725.

da Silva, J.C.B, A.L. New, and A. Azevedo. 2007. On the role of SAR for observing "local generation" of internal solitary waves off the Iberian Peninsula. *Canadian Journal of Remote Sensing* 33:388–403, http://dx.doi.org/10.5589/m07-041.

da Silva, J.C.B., A.L. New, and J.M. Magalhães. 2009. Internal solitary waves in the Mozambique Channel: Observations and interpretation. *Journal of Geophysical Research* 114, C05001, http://dx.doi.org/10.1029/2008JC005125.

da Silva, J.C.B., A.L. New, and J.M. Magalhães. 2011. On the structure and propagation of internal solitary waves generated at the Mascarene Plateau in the Indian Ocean. *Deep-Sea Research Part I* 58:229–240, http://dx.doi.org/10.1016/j.dsr.2010.12.003.

Egbert, G.D., and S.Y. Erofeeva. 2002. Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology* 19:183–204, http://dx.doi.org/10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2.

Farmer, D., Q. Li, and J.-H. Park. 2009. Internal wave observations in the South China Sea: The role of rotation and non-linearity. *Atmosphere-Ocean* 47:267–280, http://dx.doi.org/10.3137/OC313.2009.

Gerkema, T. 1996. A unified model for the generation and fission of internal tides in a rotating ocean. *Journal of Marine Research* 54:421–450, http://dx.doi.org/10.1357/0022240963213574.

Gerkema, T. 2001. Internal and interfacial tides: Beam scattering and local generation of solitary waves. *Journal of Marine Research* 59:227–255, http://dx.doi.org/10.1357/002224001762882646.

Gerkema, T., J.T.F. Zimmerman, L.R.M. Maas, and H. van Haren. 2008. Geophysical and astrophysical fluid dynamics beyond the traditional approximation. *Reviews of Geophysics* 46, RG2004, http://dx.doi.org/10.1029/2006RG000220.

Grisouard, N., C. Staquet, and T. Gerkema. 2011. Generation of internal solitary waves in a pycnocline by an internal wave beam: A numerical study. *Journal of Fluid Mechanics* 676:491–513, http://dx.doi.org/10.1017/jfm.2011.61.

Hazewinkel, J., P. van Breevoort, S.B. Dalziel, and L.R.M. Maas. 2008. Observations on the wavenumber spectrum and decay of an internal wave attractor. *Journal of Fluid Mechanics* 598:373–382, http://dx.doi.org/10.1017/S0022112007000031.

Hazewinkel, J., C. Tsimitri, L.R.M. Maas, and S.B. Dalziel. 2010. Observations on the robustness of internal wave attractors to perturbations. *Physics of Fluids* 22, 107102, http://dx.doi.org/10.1063/1.3489008.

Jackson, C.R. 2007. Internal wave detection using the Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research* 112, C11012, http://dx.doi.org/10.1029/2007JC004220.

Jackson, C.R., and W. Alpers. 2010. The role of the critical angle in brightness reversals on sunglint images of the sea surface. *Journal of Geophysical Research* 115, C09019, http://dx.doi.org/10.1029/2009JC006037.

Jackson, C.R., J.C.B. da Silva, and G. Jeans. 2012. The generation of nonlinear internal waves. *Oceanography* 25(2):108–123, http://dx.doi.org/10.5670/oceanog.2012.46.

Kitade, Y., Y. Igeta, R. Fujii, and M. Ishii. 2011. Amplification of semidiurnal internal tide observed in the outer part of Tokyo Bay. *Journal of Oceanography* 67:613–625, http://dx.doi.org/10.1007/s10872-011-0061-0.

Konyaev, K.V., K.D. Sabinin, and A.N. Serebryany. 1995. Large-amplitude internal waves at the Mascarene Ridge in the Indian Ocean. *Deep-Sea Research Part I* 42:2,075–2,091, http://dx.doi.org/10.1016/0967-0637(95)00067-4.

Maas, L.R.M. 2005. Wave attractors: Linear yet nonlinear. *International Journal of Bifurcation and Chaos* 15:2,757–2,782.

Maas, L.R.M., and F.P.A. Lam. 1995. Geometric focusing of internal waves. *Journal of Fluid Mechanics* 300:1–41, http://dx.doi.org/10.1017/S0022112095003582.

Magalhães, J.M., I.B. Araujo, J.C.B. da Silva, R.H.J. Grimshaw, K. Davis, and J. Pineda, 2011. Atmospheric gravity waves in the Red Sea: A new hotspot. *Nonlinear Processes in Geophysics* 18:71–79, http://dx.doi.org/10.5194/npg-18-71-2011.

Mercier, M.J., M. Mathur, L. Gostiaux, T. Gerkema, J.M. Magalhães, J.C.B da Silva, and T. Dauxois. In press. Soliton generation by internal tidal beams impinging on a pycnocline: Laboratory experiments. *Journal of Fluid Mechanics*.

New, A.L., and J.C.B. da Silva. 2002. Remote-sensing evidence for the local generation of internal soliton packets in the central Bay of Biscay. *Deep-Sea Research Part I* 49:915–934, http://dx.doi.org/10.1016/S0967-0637(01)00082-6.

New, A.L., and R.D. Pingree. 1992. Local generation of internal soliton packets in the central Bay of Biscay. *Deep-Sea Research Part A* 39:1,521–1,534, http://dx.doi.org/10.1016/0198-0149(92)90045-U.

Pingree, R.D., G.T. Mardell, and A.L. New. 1986. Propagation of internal tides from the upper slopes of the Bay of Biscay. *Nature* 312:154–158, http://dx.doi.org/10.1038/321154a0.

Sofianos, S.S., and W.E. Johns. 2007. Observations of the summer Red Sea circulation. *Journal of Geophysical Research* 112, C06025, http://dx.doi.org/10.1029/2006JC003886.

Vlasenko, V., N. Stashchuk, and K. Hutter. 2005. *Baroclinic Tides: Theoretical Modeling and Observational Evidence*. Cambridge University Press, Cambridge, UK, 351 pp.