Seasonality in Transition Scale from Balanced to Unbalanced Motions in the World Ocean

BO QIU AND SHUIMING CHEN
Department of Oceanography, University of Hawai‘i at Mānoa, Honolulu, Hawaii

PATRICE KLEIN
LOPS-CNRS, Ifremer, France, and California Institute of Technology, Pasadena, California

JINBO WANG, HÉCTOR TORRES, LEE-LUENG FU, AND DIMITRIS MENEMENLIS
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

(Manuscript received 26 August 2017, in final form 10 January 2018)

ABSTRACT

The transition scale $L_t$ from balanced geostrophic motions to unbalanced wave motions, including near-inertial flows, internal tides, and inertia–gravity wave continuum, is explored using the output from a global 1/48° horizontal resolution Massachusetts Institute of Technology general circulation model (MITgcm) simulation. Defined as the wavelength with equal balanced and unbalanced motion kinetic energy (KE) spectral density, $L_t$ is detected to be geographically highly inhomogeneous: it falls below 40 km in the western boundary current and Antarctic Circumpolar Current regions, increases to 40–100 km in the interior subtropical and subpolar gyres, and exceeds, in general, 200 km in the tropical oceans. With the exception of the Pacific and Indian sectors of the Southern Ocean, the seasonal KE fluctuations of the surface balanced and unbalanced motions are out of phase because of the occurrence of mixed layer instability in winter and trapping of unbalanced motion KE in shallow mixed layer in summer. The combined effect of these seasonal changes renders $L_t$ to be 20 km during winter in 80% of the Northern Hemisphere oceans between 25° and 45°N and all of the Southern Hemisphere oceans south of 25°S. The transition scale’s geographical and seasonal changes are highly relevant to the forthcoming Surface Water and Ocean Topography (SWOT) mission. To improve the detection of balanced submesoscale signals from SWOT, especially in the tropical oceans, efforts to remove stationary internal tidal signals are called for.

1. Introduction

Advent of nadir-looking satellite altimetry in the 1990s has revolutionized our ability to measure with high precision the global sea surface height (SSH) field and to explore, through geostrophy, the upper-ocean circulation dynamics. This exploration has significantly advanced our understanding of mesoscale oceanic variability; see Fu et al. (2010) and Morrow and Le Traon (2012) for comprehensive reviews. A critical limitation of the nadir-looking altimeters is their 100–300-km spacing between the satellite ground tracks. Even with combined data merged from multiple altimeters, the spatial resolution in a two-dimensional SSH map is typically on the order of 200 km in wavelength (Ducet et al. 2000; Chelton et al. 2011). This length scale is inadequate to fully capture the mesoscale oceanic signals that contain 90% of the kinetic energy of the ocean (Ferrari and Wunsch 2009) and misses completely the submesoscales that have length scales shorter than the Rossby radius of deformation.

With the wide-swath radar interferometry, the next-generation Surface Water and Ocean Topography (SWOT) satellite will improve the measured SSH resolution potentially down to the spectral wavelength of 15 km, allowing us to investigate for the first time the global upper-ocean circulation variability at the short mesoscale and submesoscale ranges (Fu and Ubelmann...
Dynamically, upper-ocean processes with length scales of $O(15–200)$ km are important because they determine the equilibrium state of the upper ocean through the turbulent kinetic energy cascade and energy dissipation (e.g., Ferrari and Wunsch 2009; McWilliams 2016). They are also crucial to how the surface ocean communicates with the subsurface interior ocean, affecting the mixed layer evolution and upper-ocean thermal anomalies (e.g., Lapeyre et al. 2006; Capet et al. 2008; Thomas et al. 2008; Klein et al. 2008). Because of the correlation existing between surface and interior ocean potential vorticity anomalies, meso- and submesoscale SSH information in the 15–200-km range can potentially be used to reconstruct the three-dimensional, upper-ocean circulation structures, including the balanced vertical velocity field (Lapeyre and Klein 2006). A rich literature is available relating to this research topic and an up-to-date review of the topic and its application to the planned SWOT SSH measurements can be found in Qiu et al. (2016).

With the decrease in horizontal scale, the kinetic energy level of the geostrophically balanced motion has been observed to weaken proportional to $k_h^{-2} \sim k_h^{-3}$, where $k_h$ is the horizontal wavenumber (Callies and Ferrari 2013; Bühler et al. 2014; Rocha et al. 2016a). For the 15–200-km range pertinent to the SWOT mission, it is possible that the balanced motion can lose its dominance and be overtaken by the unbalanced wave motions, including near-inertial flows, internal tides, and inertia–gravity waves. In such a case, the altimeter-measured SSH information can no longer be used readily to infer the time-varying surface geostrophic flows.

The transition scale separating the balanced and unbalanced motions was explored in our recent analyses of repeat shipboard ADCP measurements along 137°E across the wind-driven tropical and subtropical gyres in the northwestern Pacific (Qiu et al. 2017). By analyzing the repeat ADCP surveys from 2004 to 2016, we found that the observed transition scale depends sensitively on the energy level of local mesoscale eddy variability. In the eddy-abundant western boundary current band of Kuroshio, the transition scale is short, <15 km, and it increases to ~50 km in the moderately unstable Subtropical Countercurrent (STCC) band in the central subtropical gyre. Along the path of relatively stable North Equatorial Current, the transition scale was observed to exceed 200 km. In addition to these geographical changes, the repeat ADCP measurements along 137°E further revealed that the transition scale is decreased in winter and lengthened in summer in the Kuroshio and STCC bands because of the occurrence of mixed layer instability in winter (Sasaki et al. 2014; Qiu et al. 2014). No apparent seasonal difference in transition scale was observed in the tropical gyre.

Information about the seasonally and geographically varying transition scale is important for the SWOT mission because it provides an important length scale measure below which the SSH data may no longer be used to accurately infer the surface geostrophic velocity. To expand our analyses on transition scale to the global ocean, we adopt in this study the state-of-the-art global ocean simulation output from the Massachusetts Institute of Technology general circulation model (MITgcm) llc4320. With its horizontal resolution at 1/48° and with the realistic tidal forcings, the output of llc4320 has been used recently to explore the submesoscale dynamics in different parts of the World Ocean (Rocha et al. 2016a,b), to compare internal gravity wave SSH spectra against globally available moored measurements (Savage et al. 2017a,b), and to design calibration and validation (CalVal) approaches for SWOT-measured SSH signals (Wang et al. 2018).

This paper is organized as follows: In section 2, we provide a brief description about the MITgcm llc4320; its performance in resolving the global mesoscale eddy variability is compared to that observed by available nadir-looking satellite altimeters. In section 3, we introduce the transition scale based on kinetic energy (KE) equality of the balanced and unbalanced motions and examine its geographical variations. Seasonal variations in transition scale are examined in section 4 where opposing seasonal contributions from the balanced and unbalanced motions are emphasized. Section 5 explores the transition scale that is based on the SSH variance equality of the balanced and unbalanced motions. Its differences from the KE-based transition scale will be contrasted. In section 6, we summarize the analysis results in the context of the planned SWOT mission and discuss the need to reduce the unbalanced motion KE by removal of stationary internal tidal signals.

2. MITgcm llc4320 simulation

The data product analyzed in this study is based on the state-of-the-art global ocean simulation of MITgcm (Marshall et al. 1997) in a latitude–longitude–cap (llc) configuration with a polar cap that has $4320 \times 4320$ grid cells. Referred to as llc4320, the model has a 1/48° horizontal resolution and 90 vertical levels. To better resolve the submesoscale upper-ocean variability, the model’s vertical levels have a ~1-m vertical resolution at the surface and increases to ~30 m near the 500-m depth. The llc4320 simulation is initialized from the output of the Estimating the Circulation and Climate of
the Ocean, Phase II (ECCO2), project (Menemenlis et al. 2008), with the subsequent model resolution increased to llc1080 (1/12°), llc2160 (1/24°), and the present llc4320 (1/48°). The llc4320 model is forced by the 6-hourly ERA-Interim atmospheric reanalysis as well as by a synthetic surface pressure field consisting of the 16 most dominant tidal constituents. For our analysis, we use the hourly output of the modeled surface height \( \eta \) and horizontal velocity \((u, v)\) data from 1 November 2011 to 31 October 2012 (366 days).

To verify the surface circulation changes modeled in the llc4320 simulation, we use the global SSH anomaly dataset generated by Ssalto/Duacs and distributed by AVISO with support from CNES (http://marine.copernicus.eu/). This dataset merges along-track SSH measurements from all satellite altimeter missions after October 1992 and has a 7-day temporal resolution and a 1/4° longitude Mercator spatial resolution (Ducet et al. 2000). The data period available for this study extends from January 1993 to December 2016.

Figure 1 compares the surface eddy kinetic energy (EKE) distribution simulated by llc4320 and that derived from the AVISO SSH product for the period of November 2011–October 2012. For comparison with the AVISO-derived EKE values, we average the llc4320 modeled SSH data into a weekly time series on the 1/4° longitude Mercator grid and compute the EKE assuming geostrophy:

\[
\text{EKE} = \frac{1}{2} \left( \left( \frac{g}{f} \frac{\partial \eta}{\partial y} \right)^2 + \left( \frac{g}{f} \frac{\partial \eta}{\partial x} \right)^2 \right),
\]

where \( g \) is gravitational acceleration, \( f \) is the Coriolis parameter, and \( \eta \) is the SSH anomalies from the mean SSH of November 2011–October 2012. By and large, llc4320 simulates faithfully the mesoscale eddy variability captured by the satellite altimeter missions that have spectral scales larger than \( O(150) \) km (e.g., Chelton et al. 2011). One noticeable discrepancy between the two EKE maps is within the equatorial band of \( \pm 10^\circ \), where the AVISO-derived EKE values in Fig. 1b appear consistently larger than those based on llc4320 (Fig. 1a). One likely reason for this discrepancy can be related to the measurement errors in the AVISO SSH data that are amplified when converting to EKE because of the small \( f \) values within the equatorial band.

Evaluating the llc4320’s skill in simulating the unbalanced wave motions is more challenging because of the lack of in situ high-frequency SSH measurements in the World Ocean. By utilizing the hourly dynamic height time series from nine McLane profiler moorings available in the Pacific, North Atlantic, and south Indian Oceans, Savage et al. (2017b) found recently that the wave frequency spectra of SSH variance simulated by llc4320 agreed reasonably well with those observed by the profiler moorings. Given these high-frequency SSH comparisons and those for the mesoscale eddy variability shown in Fig. 1, we believe the llc4320 results are adequate for quantifying the relative importance between the balanced versus unbalanced motions.

3. Balanced–unbalanced motion decomposition and \( L_t \)

Delineation of balanced and unbalanced motions can be conducted in several different ways. One common method is to select a temporal filter with the low-pass (high pass) filtered signals regarded as the balanced (unbalanced) motions. For example, Richman et al. (2012) used a 48-h filter in their study on KE spectral slopes relevant to the mesoscale variability. While straightforward in its implementation, such a single-frequency filter may not be ideal for decomposing the balanced and unbalanced motions on the global scale. Figure 2 compares the typical horizontal wavenumber–frequency \((k_h–\omega)\) spectra of
surface KE in four representative areas in the Pacific Ocean from the hourly output of llc4320: (i) the highly turbulent Kuroshio Extension, (ii) moderately unstable STCC, (iii) stable North Equatorial Current, and (iv) the tropical southwest Pacific. In the plots, elevated unbalanced motion KE is discernible in the near-inertial, diurnal, and semidiurnal tidal frequency bands as well as along the dispersion curves of inertia–gravity waves (IGWs) of discrete vertical modes: $\omega^2 = f^2 + c_n^2 k_n^2$ (see white solid lines), where $c_n$ ($n = 1$ to 10) are evaluated based on the World Ocean Atlas 2013 (WOA2013) climatological temperature–salinity profiles (Locarnini et al. 2013; Zweng et al. 2013). From Fig. 2, it is clear that the cutoff period delineating the balanced and unbalanced motions is highly sensitive to the local inertial period: near the local inertial frequency band, balanced and unbalanced motions can coexist with large and small wavenumbers, respectively. As a result, this renders the adoption of a single-frequency filter, such as the dashed black line showing the 36-h filter, undesirable.

In the present study, we adopt a more dynamically based method by defining the balanced (unbalanced)
motions below (above) the lower frequency of either the local tenth vertical-mode IGW dispersion curve or the permissible tides (see thick black lines in Fig. 2). To evaluate the transition scale \( L_t \) from balanced to unbalanced motion, we first select the hourly surface \((u, v)\) data from llc4320 in each fixed 500 km \( \times \) 500 km box. After removing the best-fit two-dimensional spatial slopes from the \((u, v)\) data and applying a two-dimensional Hanning window, we obtain the horizontal wavenumber–frequency spectrum of surface KE, \( \tilde{E}(k_h, \nu) \), such as those shown in Fig. 2. The transition scale \( L_t \) is defined as the wavelength \( k_h^{-1} \) at which KE integrated over the balanced motion frequencies;

\[
\text{KE}_{\text{bal}}(k_h) = \int_{\omega_{yr}}^{\omega_{cr}(k_h)} \tilde{E}(k_h, \omega) \ d\omega,
\]

is equal to that integrated over the unbalanced motion frequencies

\[
\text{KE}_{\text{unb}}(k_h) = \int_{\omega_{yr}}^{\omega_{Ny}} \tilde{E}(k_h, \omega) \ d\omega,
\]

where \( \omega_{yr} \) denotes the annual frequency, \( \omega_{cr}(k_h) \) is the dynamically delineating frequency defined above, and \( \omega_{Ny} \) is the Nyquist frequency at 0.5 cph. As an illustration, we plot in Fig. 3 \( \text{KE}_{\text{bal}} \) and \( \text{KE}_{\text{unb}} \) as a function of \( k_h \) in the same four representative boxes as those selected in Fig. 2. By our dynamical definition, the transition scale \( L_t \) is given by the vertical dashed lines where \( \text{KE}_{\text{bal}} = \text{KE}_{\text{unb}} \).

Figure 4 shows the global \( L_t \) distribution using the dynamical delineation method described above. Areas where \( L_t \) exceeds 500 km are blanked out. In accordance with the shipboard ADCP data results observed along 137°E in the northwestern Pacific (Qiu et al. 2017), \( L_t \) is short, \(<20\) km, in the western boundary current Kuroshio region south of Japan. It increases to \( \sim50\) km in the weakly baroclinically unstable STCC band of

![Figure 3](image-url)
The $L_t$ value exceeds 250 km along the dynamically stable North Equatorial Current band of 9°–14°N. In the North Equatorial Countercurrent (NECC) band of 2°–9°N, $L_t$ decreases to ~200 km. Notice the $L_t$ value in the NECC band is larger than that inferred from the shipboard ADCP measurements (~80 km). This difference could be because the 1-yr l4320 simulation is too short to account for the dominant interannual variability of the NECC, which was captured in the long-term shipboard ADCP data (Qiu and Joyce 1992; Chen et al. 2016).

These results in the northwestern Pacific are quite representative for the rest of the global ocean. The $L_t$ values smaller than 20 km can be seen in all other western boundary current regions: the Kuroshio Extension, Gulf Stream, Agulhas Current, and Brazil–Malvinas Confluence. One exception is in the East Australian Current (EAC) region, which will be commented on below. Another region with small $L_t$ is along the Antarctic Circumpolar Current (ACC) in the Southern Ocean, with the exceptions where prominent bottom topographic features exist, for example, the Kerguelen Plateau near 70°E, the East Scotia Basin east of the Drake Passage, and the Spiess Seamount near the southern tip of the Mid-Atlantic Ridge. In the Northern Hemisphere’s temperate latitudes of 15°–30°N, $L_t$ falls generally in the 50–100-km range. The corresponding latitudes in the Southern Hemisphere tend to have $L_t$ in the larger O(150) km range. In the tropics equatorward of 15°, the $L_t$ values often exceed 200 km with the unbalanced motion KE overwhelming the balanced motion KE. In the high latitudes, one area where $L_t$ exceeds 200 km is in the Alaskan Gyre, where the balanced oceanic mesoscale eddy variability is extremely weak (recall Fig. 1).
not coincidently, correspond to the major generation sites of internal tides in the World Ocean (Ray and Zaron 2016; Zhao et al. 2016; and references therein). In fact, it is this enhanced KE level associated with the unbalanced motions that is responsible for the elevated $L_t$ values at ~150 km in the Coral and Tasman Seas in Fig. 4, albeit the presence of the western boundary current EAC in the region. Likewise, the abyssal seamount and midocean ridge–enhanced unbalanced motions are responsible for the spatially dependent $L_t$ modulations along the ACC path in the Southern Ocean.

4. Seasonal variations in $L_t$

There exists increasing evidence in recent years that near-surface meso- and submesoscale variability in the 10–100-km range has a distinct annual cycle relating to the occurrence of mixed layer instability. Such evidence emerges from both in situ measurements in different parts of the World Ocean (e.g., Callies et al. 2015; Buckingham et al. 2016; Thompson et al. 2016; Qiu et al. 2017) and high-resolution ocean general circulation model simulations (e.g., Mensa et al. 2013; Sasaki et al. 2014; Qiu et al. 2014). Since this 10–100-km range overlaps extensively with the $L_t$ values obtained in the preceding section, it is natural to inquire if our above-defined transition scale $L_t$ is influenced by the seasonally modulating meso- and submesoscale variability. Our interest in the seasonal $L_t$ changes is further motivated by a recent study by Rocha et al. (2016b), who found that the unbalanced motion kinetic energy in the Kuroshio Extension region in the MITgcm lle4320 exhibited a seasonal maximum in summer, opposite to that of the balanced motion kinetic energy.

To examine the seasonal changes in $L_t$, we define winter (summer) as February–April (August–October) and evaluate $L_t$ according to Eqs. (2) and (3) with the year-long hourly $(u, v)$ data replaced by the seasonal ones (Figs. 6a,b). While the overall geographical dependence of the seasonal $L_t$ values is similar to the annual-mean pattern shown in Fig. 4, consistent changes with larger (smaller) $L_t$ values in the summer (winter) hemisphere are visually identifiable in Fig. 6. To better quantify the seasonal difference, we plot in Fig. 7 the $L_t$ difference of summer [August–October (ASO)] minus winter [February–April (FMA)]. In agreement with the visual changes seen between Figs. 6a and 6b, the ASO—FMA $L_t$ is by and large positive in the Northern Hemisphere (NH) and negative in the Southern Hemisphere (SH). Notice that the differential $L_t$ values are in the 50–150-km range in much of the off-equatorial ocean, and this range is on par with the magnitude of the annual-mean $L_t$ values presented in Fig. 4.

An interesting exception to the SH’s general negative ASO—FMA $L_t$ occurs in the Pacific and Indian sectors of the Southern Ocean where the SH winter-minus-summer $L_t$ is small but remains positive. To clarify the cause responsible for this regional exception, we evaluate the ASO—FMA balanced and unbalanced motion KE levels in the wavelength range $<100$ km of our interest (Figs. 8a,b). Because of the prevalence of winter mixed layer instability, the balanced motion KE level in ASO is clearly lower (higher) than in FMA in the NH (SH) oceans. This is no exception in the Pacific and Indian sectors of the Southern Ocean. For the unbalanced motion ASO—FMA KE level, Fig. 8b reveals that while it is overall positive in the NH and negative in the SH, the differential KE level remains weakly positive in the Pacific and Indian sectors of the Southern Ocean. In other words, it is the lack of summer enhancement by the unbalanced motion KE that causes ASO—FMA $L_t$ to be weakly positive in the Pacific and Indian sectors of the Southern Ocean.

In their study of submesoscale seasonality in the Kuroshio Extension region, Rocha et al. (2016b) argued that the enhancement in the unbalanced motion KE in summer is due to the shoaling of mixed layer and
strengthening of upper-ocean stratification. Both of these effects work to trap the inertia–gravity waves to the shallow surface ocean, enhancing the unbalanced motion’s projection onto the mixed layer (D’Asaro 1978). An inspection of the seasonal mixed layer depth changes in llc4320 reveals that the summer (FMA) mixed layer along the path of ACC (Fig. 9a) extends to below 100 m and this is particularly true in the Pacific and Indian Ocean sectors where the mixed layer can reach to deeper than 200 m. This 100–200-m mixed layer depth in the summer ACC should be contrasted to the 20-m mixed layer depth value observed typically in the summer NH oceans (see Fig. 9b) or the rest of the summer SH oceans. It is this deep summer mixed layer that is likely to be responsible for the ASO – FMA unbalanced motion KE to remain positive along the ACC shown in Fig. 8b.

It is important to emphasize that the seasonal difference in the unbalanced motion KE relating to the summer trapping to within the surface mixed layer does not penetrate deep into the subsurface ocean. At the main thermocline depth of 500 m, for example, the ASO – FMA unbalanced motion KE (Fig. 10b) shows an oppositely signed distribution to that of the surface map (cf. Fig. 8b; note that the color scales are different between Figs. 8 and 10 because of the shallower depth scales associated with the submesoscale flows). Below the surface mixed layer, the wintertime unbalanced motion KE exceeds that in summer because of the enhanced surface wind forcing that elevates the KE level of the wind-induced superinertial wave motions (e.g., Alford et al. 2016). For the ASO – FMA balanced motion KE with scales < 100 km at the 500-m depth, Fig. 10a reveals that the hemispheric contrast that exists in the surface layer (recall Fig. 8a) becomes less distinctive. The ASO – FMA difference shown in Fig. 10a is spatially patchy, although the general tendency for the wintertime balanced motion KE to surpass that in summer is still discernible.

In concluding this section, we plot in Fig. 11 the cumulative probability density function (PDF) distribution for $L_t$ in the NH and SH winter as a function of global ocean latitude. Taking the thick black lines as an example, it is clear from Fig. 11 that 80% of $L_t$ falls below 20 km in most of the NH oceans between 25° and 45°N and in all of the SH oceans south of 25°S. That the wintertime $L_t$ is shorter than 20 km in most of the extratropical NH and SH oceans indicates that the entanglement between balanced and unbalanced motions will likely pose less of a challenge for the SWOT mission in these locations to detect the balanced submesoscale features in winter.

---

2 Similar seasonally varying spatial patterns of the mixed layer depth can be found in the mixed layer depth climatology based on in situ measurements; see, for example, Fig. 5 in de Boyer Montégut et al. (2004).
5. SSH variance-based $L_{\text{SSH}}$

In the preceding sections, we have explored the global distribution of $L_t$ based on kinetic energy equality between the balanced and unbalanced motions. From a practical point of view, $L_t$ thus derived could be regarded as the length scale below which surface circulation inferred geostrophically from the SSH gradient becomes unreliable. With the high-resolution SSH information from the planned SWOT mission, it is instructive to also evaluate the transition scale that is based on the SSH variance equality between the balanced and unbalanced motions. To differentiate this transition scale based on the SSH variance equality from the $L_t$ defined in section 3, we denote hereafter the SSH variance-based $L_t$ as $L_{\text{SSH}}$.

To evaluate $L_{\text{SSH}}$ from the hourly output of llc4320, we adopt the same dynamical delineation method detailed in section 3 and replace the KE spectra by the SSH variance spectra in Eqs. (2)–(3). Figure 12a shows the distribution of $L_{\text{SSH}}$ calculated from the year-long llc4320 output. Geographically, $L_{\text{SSH}}$ exhibits a spatial pattern very similar to that of the KE-based $L_t$, shown in Fig. 4. Like for the balanced and unbalanced motion kinetic energy levels, the wavenumber-integrated SSH variance associated with the balanced and unbalanced motions are similarly spatially inhomogeneous (figures not shown). By and large, $L_{\text{SSH}}$ has a larger value than $L_t$ as compared in Fig. 12b and the general trend is such that their difference, $L_{\text{SSH}} - L_t$, decreases progressively from the equator toward higher latitudes.

It is of interest to consider here why $L_{\text{SSH}}$ exceeds in general $L_t$, and what contributes to the diminishing difference in $L_{\text{SSH}} - L_t$ from the equator toward higher latitudes. For simplicity, let us consider a linearized shallow-water ocean in which surface velocity $u$ and $v$ and SSH $\eta$ are governed by (Gill 1982)

$$-i\omega \hat{u} - f \hat{v} = -ig \hat{k} \hat{\eta},$$

$$-i\omega \hat{v} + f \hat{u} = -ig \hat{l} \hat{\eta},$$

and

$$-i\omega \hat{\eta} + iH(\hat{k} u + \hat{l} v) = 0,$$

where $(u, v, \eta) = (\hat{u}, \hat{v}, \hat{\eta}) \exp(ikx + ily - i\omega t)$, and $H$ denotes the water depth (or equivalent depth for a
baroclinic mode in the stratified ocean). Rearranging Eqs. (4)–(6), we obtain
\[ \hat{u} = g(\omega k + ilf)\hat{\eta}(\omega^2 - f^2), \quad \text{and} \quad (7) \]
\[ \hat{v} = g(\omega l - ikf)\hat{\eta}(\omega^2 - f^2), \quad \text{(8)} \]
and the wavenumber–frequency spectra for KE and SSH variance \( \hat{E}(k_h, \omega) \) and \( \hat{H}(k_h, \omega) \) are in this case related through
\[ \hat{E}(k_h, \omega) = \frac{g^2 k_h^2}{2} \frac{\omega^2 + f^2}{(\omega^2 - f^2)^2} \hat{H}(k_h, \omega), \quad \text{(9)} \]
where \( k_h^2 = k^2 + l^2 \). For the balanced motion, \( \omega \ll f \), and its KE spectrum defined in Eq. (2) can be approximated by
\[ \text{KE}_{\text{bal}}(k_h) = \int_{-\infty}^{\infty} g^2 k_h^2 \frac{\omega^2 + f^2}{2} \frac{\hat{H}(k_h, \omega)}{(\omega^2 - f^2)^2} d\omega \]
\[ \approx \frac{g^2 k_h^2}{2f^2} \text{HV}_{\text{bal}}(k_h), \quad \text{(10)} \]
where \( \text{HV}_{\text{bal}}(k_h) \) denotes the SSH variance spectrum associated with the balanced motion. That the KE spectral slope in Eq. (10) is shallower than the SSH variance spectral slope by \( k_h^2 \) has been extensively used in previous studies of mesoscale eddy variability based on satellite altimeter data (e.g., Stammer 1997).

For the unbalanced motion, combining Eqs. (3) and (9) leads to
\[ \text{KE}_{\text{unb}}(k_h) = \frac{g^2 k_h^2}{2f^2} \int_{-\infty}^{\infty} \frac{\omega^2 |f^2| + 1}{(\omega^2 |f^2| - 1)^2} \hat{H}(k_h, \omega) d\omega. \quad \text{(11)} \]
As demonstrated in Fig. 2, the unbalanced motion KE resides largely in the near-inertial, diurnal/semidiurnal tide and discrete IGW bands. Since the near-inertial motions have weak projections onto SSH (e.g., Alford et al. 2016), contributions to the RHS integration in Eq. (11) come mostly from the tidal and IGW motions.

If \( \hat{H}(k_h, \omega) \) is assumed to be dominated by the tidal or IGW motion with a particular frequency \( \omega_0 \), we may approximate \( \hat{H}(k_h, \omega) = \hat{H}(k_h) \delta(\omega - \omega_0) \), where \( \delta \) is the Dirac delta function. In this case, Eq. (11) becomes
\[ \text{KE}_{\text{unb}}(k_h) = \frac{g^2 k_h^2}{2f^2} \frac{\omega_0^2 |f^2| + 1}{(\omega_0^2 |f^2| - 1)^2} \text{HV}_{\text{unb}}(k_h), \quad \text{(12)} \]
and combining Eqs. (10) and (12) results in
\[ \frac{\text{KE}_{\text{unb}}(k_h)}{\text{KE}_{\text{bal}}(k_h)} = I \frac{\text{HV}_{\text{unb}}(k_h)}{\text{HV}_{\text{bal}}(k_h)}, \quad \text{(13)} \]
where \( I = \frac{(\omega_0^2 |f^2| + 1)/(|f^2| - 1)^2} \). From Eq. (13), it is clear that the difference between \( L_i \) and \( L_i^{\text{SSH}} \) depends on the magnitude of \( I \). If \( I > 1 \), \( L_i \) will surpass \( L_i^{\text{SSH}} \) and, if \( I < 1 \), \( L_i^{\text{SSH}} \) will exceed \( L_i \). For the semidiurnal tides, \( I \) is
spectra have similar contrast, the unbalanced motion KE and SSH variance versa. For a fixed mode $cn$, where

$$I(k_h) = \frac{c_n^2k_h^2/f^2 + 2}{(c_n^2k_h^2/f^2)^2},$$

(14)

where $c_n$ denotes the phase speed of the mode-$n$ internal gravity wave. In Eq. (14), $I < 1$ if $c_n^2k_h^2/f^2 > 2$ and vice versa. For a fixed mode $c_n$ and wavenumber $k_h$, the tendency is again for $L_i^{\text{SSH}}$ to exceed $L_i$ in lower latitudes. Figure 13 shows the threshold lines of $I = 1$ for different vertical mode IGWs averaged across the latitudinal bands of the World Ocean. As in Fig. 2, $c_n (n = 1$ to 10) are calculated based on the WOA2013 climatological temperature–salinity profiles (Locarnini et al. 2013; Zweng et al. 2013). Unlike the tidal effect noted above, the $I = 1$ threshold for IGWs not only depends on latitude but also on the wavenumber. Specifically, $I$ can remain less than unity in high-latitude oceans when $k_h$ is large in the 10–100-km wavelength range of our interest. It is in fact the contributions from the low-mode IGWs that keep $L_i^{\text{SSH}} > L_i$ in some of the high-latitude areas seen in Fig. 12b.

Like the seasonal changes in $L_i$, the SSH variance-based $L_i^{\text{SSH}}$ is also subject to distinct seasonal modulations. Figure 14 shows the $L_i^{\text{SSH}}$ difference between ASO and FMA, and its global pattern is spatially very similar to that for $L_i$, shown in Fig. 7. The seasonal amplitude in $L_i^{\text{SSH}}$, on the other hand, is about half that obtained in Fig. 7. One reason for this is that the balanced motion SSH variance spectral slope is steeper than the KE spectral slope; recall Eq. (10) and compare the spectra slopes for KE and SSH variance in Fig. 15 at a representative site of 37°N and 164°E in the North Pacific. In contrast, the unbalanced motion KE and SSH variance spectra have similar $k_h$ slopes (cf. dashed lines in Fig. 15). This renders the seasonally varying crossings by balanced and unbalanced motion SSH variance spectra more restricted in the $k_h$ space than those by balanced and unbalanced motion KE spectra.

6. Discussion and summary

We have in this study explored geographical and seasonal variations in the transition scale at which the geostrophically balanced motion loses its dominance over the unbalanced wave motions. As the metric for the transition scale, we adopted $L_i$ defined by the wavelength that has equal balanced and unbalanced motion KE spectral densities. By analyzing the hourly output from the 1/48° horizontal resolution MITgcm llc4320, we found that $L_i$ is geographically highly inhomogeneous. Broadly speaking, $L_i$ falls below 40 km in the western boundary current and ACC regions and increases to 40–100 km in the vast subtropical and subpolar world oceans. An exception to the latter case is in the Alaskan Gyre where $L_i$ can exceed 200 km. In the tropical oceans equatorward of 20° latitudes, $L_i$ is found, in general, to be larger than 150 km. These llc4320-derived, latitude-dependent $L_i$ values are, by and large, consistent with our recent estimates based on the repeat shipboard ADCP measurements along 137°E in the northwestern Pacific Ocean (Qiu et al. 2017).

On a regional scale, the $L_i$ value depends not only on the local level of mesoscale eddy variability, but also on the energy level set by the unbalanced motions consisting of near-inertial flows, internal tides, and inertia–gravity waves. The unbalanced motion energy level is spatially nonuniform, and high values are commonly detected where prominent bathymetric features exist. Examples of effects on $L_i$ by enhanced unbalanced motions include the East Australian Current region where $L_i$ reaches 100 km (rather than 20 km as in other western boundary current regions) and along the ACC path where increased $L_i$ values are found to be collocated with the major topographic features in the Southern Ocean.
The energy levels of both the balanced and unbalanced motions modulate seasonally and, as a result, they can alter the \( L_t \) value constructively. In winter, small-scale balanced motions are energized because of the ubiquitous occurrence of mixed layer instability. Shoaling of the mixed layer and increasing of density jump at the mixed layer base in summer, on the other hand, works to trap and elevate the surface unbalanced motion kinetic energy (Rocha et al. 2016b). The combined effect of these seasonal kinetic energy changes results in a seasonal contrast in \( L_t \) that can typically reach \( >100 \) km in significant parts of the global ocean. One important exception to this \( L_t \) change caused constructively by the seasonally varying balanced and unbalanced motions is seen in the Pacific and Indian sectors of the Southern Ocean. In these sectors, the regional mixed layer in summer remains deep, failing as a result to trap effectively the unbalanced motion kinetic energy to within the shallow surface layer.

In addition to the metric \( L_t \) based on KE equality, we have in this study also examined \( L_{SSH}^{t} \) defined by the wavelength at which balanced and unbalanced motions have equal SSH variance spectral densities. The \( L_{SSH}^{t} \) shows similar geographical and seasonal patterns as those of \( L_t \), highlighting the common dynamics controlling the spatial and temporal variations between these two transition-scale metrics. While exhibiting similar patterns, the \( L_{SSH}^{t} \) values are in general larger than \( L_t \), especially in low-latitude oceans. The difference \( L_{SSH}^{t} - L_t \) tends to diminish toward the high-latitude oceans, where \( L_{SSH}^{t} \) may become shorter than \( L_t \). The reason for this tendency is that in the wavenumber space, the KE and SSH variance spectra are functionally related differently for the balanced and unbalanced motions. While the balanced motions between the two are proportional to \( k^2 \), the unbalanced motions are connected via convolutions that are sensitive functions of latitude.

Since \( L_t \) signifies the transition scale below which the SSH data may no longer be used to accurately infer the geostrophically balanced flows, our findings about its geographical and seasonal variations are highly relevant for the forthcoming SWOT mission. With the 15-km spectral resolution expected for the SWOT mission (Fu and Ubelmann 2014), unbalanced motion signals are likely to pose little problems for detecting the balanced submesoscale flows in the mesoscale-rich regions like the ACC and the western boundary currents in the North Pacific, North Atlantic, South Atlantic, and south Indian Oceans. For the East Australian Current and in the regions of moderate mesoscale activities (e.g., the Subtropical Countercurrent bands in temperate latitudes), caution is likely warranted in diagnosing the surface geostrophic flows in the 50–100-km range.
It is worth emphasizing that the annual-mean $L_t$ values noted above can decrease significantly in winter because the seasonal balanced and unbalanced motion KE levels fluctuate out of phase, shortening $L_t$ constructively in winter. Indeed, our analysis revealed that 80% of the NH oceans between 25° and 45°N and all of the SH oceans south of 25°S will have $L_t$ shorter than 20 km during the winter season, including the EAC and the Subtropical Countercurrent bands cautioned above. Since winter is the season when the balanced submesoscale motions prevail, this implies that the SWOT-measured SSH signals can be used potentially to explore the winter-time submesoscale variability down to the 15–20-km range in a significant part of the extratropical world oceans.

The serious challenge for the SWOT mission to capture the small-scale geostrophically balanced motion will be in the tropical oceans within ±20° latitudes and in the Alaskan Gyre, where $L_t$ commonly exceeds 150 km in all seasons. This challenge stems from the fact that these regions have a high unbalanced motion KE level as compared to the balanced motion. One important source behind the high unbalanced motion KE is the internal tides generated by the interaction of barotropic tides with bottom topography. It is important to note that a significant portion of the low-mode internal tides have been known to exhibit phase stability, or coherence, in their SSH expressions (e.g., Ray and Mitchum 1996). By analyzing the internal tidal signals from the 17-yr combined record of TOPEX/Poseidon and Jason altimeters, Ray and Zaron (2011) estimated that the stationary variance is generally 75% or more of the average variance of mode-1 internal tides. More recent numerical modeling and data analysis studies by Shriver et al. (2014) and Zaron (2017) have revealed that the fraction explained by the stationary internal tides can be highly spatially inhomogeneous. In the context for $L_t$ determination, it is obvious that any removal of internal tidal signals that are stationary and predictable will result in a reduction in the $L_t$ value.

A judicious separation between the stationary and nonstationary internal tides requires multiyear in situ measurements or numerical model output. Rather than pursuing this separation based on the year-long lce4320 output, we attempt to quantify the reduction in $L_t$ by assuming that a fraction of the lce4320-modeled diurnal and semidiurnal internal tidal signals are stationary and could be, as a result, removed. Specifically, we follow Zaron (2017, his Fig. 9) and assume that the variance fraction explained by the stationary internal tides has a simple latitude-dependent form:

$$\alpha(y) = \begin{cases} 0.9 - 0.012|y| & |y| < 50^\circ \\ 0.9 & |y| \geq 50^\circ \end{cases},$$

(15)

where $y$ is the latitude in degrees. Using this form, we subjectively reduced the wavenumber KE spectra (recall Fig. 2) by $1 - \alpha(y)$ in the diurnal and semidiurnal frequency bands over the background continuum and reevaluate $\text{KE}_{\text{unb}}(k_h)$ in Eq. (3). Here, we define the diurnal (semidiurnal) frequency band as bounded by $\omega = 1/27.0–1/23.0$ ($1/13.5–1/11.5$) cph and the background continuum by linearly interpolating the frequency spectra within these chosen bounds by neighboring spectral values. Figure 16a shows the annual-mean $L_t$ distribution based on the newly evaluated $\text{KE}_{\text{unb}}(k_h)$, while its difference from the $L_t$ values shown in Fig. 4 is presented in Fig. 16b. From Fig. 16b, it is clear that removal of the stationary internal tidal signals can reduce $L_t$ by 50–100 km in many regions of the World Ocean; significant reduction in $L_t$ is seen in the Alaskan Gyre, the northwestern tropical Pacific, and the broad tropical and subtropical SH oceans. Because of the smaller fraction of the stationary internal
tides, the reduction in \( L_t \) is modest, generally less than 50 km, within the \( \pm 10^\circ \) equatorial band. For completeness, we present the same analysis results in Fig. 17 with respect to \( L_{SSH} \). Although having smaller amplitude when compared to \( L_t \), the reduction in \( L_{SSH} \) caused by the removal of stationary internal tidal signals exhibits a similarly geographical pattern.

The results shown in Figs. 16 and 17 are based on our current knowledge about the distributions of the stationary internal tides. Because of the background changes in ocean stratification, currents, and mesoscale eddies, isolating and removing the stationary internal tidal signals from the SWOT SSH measurements will be a formidable endeavor. Nevertheless, we believe the results of Figs. 16 and 17 suggest that this endeavor is critical and should be pursued in order to improve detection of the balanced submesoscale signals and maximize the science returns by the SWOT mission.

Acknowledgments. We thank Rosemary Morrow, Clement Ubelmann, and Ed Zaron for insightful discussions. Constructive comments made by two anonymous reviewers helped improve an early version of the manuscript. B. Q. and S. C. acknowledge support from NASA SWOT and OSTST missions (NNX16AH66G and NNX17AH33G). P. K. acknowledges the support of the Agence Nationale pour la Recherche [ANR-10-LABX-19-01 (LabexMER)] and of the SWOT project (JPL, NASA). J. W., H. T., L. F., and D. M.'s research presented in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. They acknowledge support from the SWOT projects.

REFERENCES


