Generation Mechanism of Tropical Instability Waves in the Equatorial Pacific Ocean

YUKI TANAKA AND TOSHIYUKI HIBIYA
Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan

(Manuscript received 22 April 2019, in final form 28 August 2019)

ABSTRACT
Tropical instability waves (TIWs) are prominent features in the equatorial Pacific, propagating westward at a speed of \(-0.5 \text{ m s}^{-1}\) with a wavelength of \(-1000 \text{ km}\). In this study, we show that a linear stability analysis using a 1.5-layer shallow water model can predict successfully an unstable mode whose wavelength, phase speed, growth rate, and meridional structure are all consistent with those of the TIWs simulated by an eddy-resolving ocean general circulation model (OGCM). This unstable mode can be interpreted as resulting from the coupling of two Rossby waves, namely, one trapped just north of the equator (\(-1^\circ-3.5^\circ \text{N}\)) and the other trapped farther north (\(-3.5^\circ-8^\circ \text{N}\)). Although these two Rossby waves have opposite intrinsic phase propagation directions reflecting the negative and positive local meridional potential vorticity (PV) gradients, respectively, their actual propagation direction can be adjusted through the advection by the South Equatorial Current and the North Equatorial Countercurrent such that they might propagate westward at the same speed and with the same zonal wavenumber yielding the largest growth rate of TIWs. The unstable mode does not appear during the period in which the negative PV gradient is absent, which demonstrates its essential role in generating TIWs. Indeed, the seasonal and interannual variability of the TIWs simulated by the OGCM is shown to be significantly controlled by the strength of the negative PV gradient just north of the equator, suggesting that it could be a key parameter toward a dynamically based parameterization of the heat and momentum transfer associated with TIWs in coarse-resolution OGCMs.

1. Introduction
Tropical instability waves (TIWs) are prominent mesoscale features observed in the eastern to central equatorial Pacific Ocean. They are most distinctly identified as meanders in the sea surface temperature (SST) front just north of the equator and are also seen in terms of temperature, salinity, and velocity fluctuations in the depth range from the sea surface down to the thermocline and in the latitude range from \(-5^\circ \text{S}\) to \(-10^\circ \text{N}\) (Legeckis 1977; Qiao and Weisberg 1995; Flament et al. 1996; Chelton et al. 2000; Kennan and Flament 2000; Lyman et al. 2007; Farrar 2008). TIWs typically have a wavelength of \(-1000 \text{ km}\) and a period of \(-25 \text{ days}\), and propagate westward at a phase speed of \(-0.5 \text{ m s}^{-1}\). The intensity of TIWs significantly varies both seasonally and interannually, usually reaching maximum amplitudes in August–December and during La Niña periods when the equatorial cold tongue and associated equatorial zonal currents fully develop under the strong easterly trade winds.

It is well known that TIWs play important roles in both small- and large-scale processes in the equatorial Pacific. For example, previous observational and numerical studies showed that TIWs can modulate the vertical shear of the mean flow inducing vigorous turbulent mixing at their leading edges, which may explain larger vertical heat flux in the active TIW periods than in other periods (Dutrieux et al. 2008; Lien et al. 2008; Moum et al. 2009; Inoue et al. 2012; Holmes and Thomas 2015; Liu et al. 2016; Warner et al. 2018). TIWs can also radiate internal waves downward, which might create a possible energy pathway to turbulent mixing in the thermocline as well as deeper layers (Jing et al. 2014; Tanaka et al. 2015; Holmes et al. 2016; Holmes and Thomas 2016). Furthermore, TIWs are thought to play important roles in maintaining the equatorial current system and affecting air–sea interactions related to the global climate (Philander et al. 1986; Jochum et al. 2004; Jochum and Murtugudde 2006; Menkes et al. 2006; Farrar 2011). In particular, it has recently been shown
that the active meridional heat transport during La Niña compared with El Niño periods creates an asymmetric negative feedback to the El Niño–Southern Oscillation, which may contribute to the observed asymmetric feature with a larger-amplitude El Niño than a La Niña (An 2008; Imada and Kimoto 2012). Nevertheless, the momentum and heat transport by TIWs must often be parameterized in existing general circulation models that do not have sufficient spatial resolution to reproduce them. A better understanding of the generation mechanism of TIWs is therefore prerequisite to formulate a physically based parameterization that can improve the capability of models to predict climate variations.

Philander (1976, 1978) carried out a linear stability analysis using a 1.5-layer shallow water model to show that the meridional shear between the westward-flowing South Equatorial Current (SEC) and the eastward-flowing North Equatorial Countercurrent (NECC) becomes barotropically unstable, generating waves with wavelengths and periods similar to the observed ones. The importance of the meridional shear between the SEC and NECC (or the meridional shear in the northern flank of the SEC) has been confirmed by subsequent analytical, observational, and numerical studies (Cox 1980; Hansen and Paul 1984; Seigel 1985; Flament et al. 1996; Proehl 1996; McPhaden 1996; Seo et al. 2007; Imada et al. 2012), although some of them pointed out additional contributions from barotropic instability of the meridional shear between the SEC and the eastward-flowing Equatorial Undercurrent (EUC) (Luther and Johnson 1990; Qiao and Weisberg 1998) and/or baroclinic instability of the meridional density front between the equatorial cold tongue and the northern warm water (McCreary and Yu 1992; Yu et al. 1995; Masina et al. 1999).

A useful approach to investigate an unstable wave is to regard it as arising from the coupling between two neutral waves. This approach allows an intuitive understanding of the physical mechanism of the instability and provides reasonable explanations for various properties of the generated unstable wave such as the spatial structure, propagation speed, and growth rate, as well as their dependence on the background field (e.g., Kubokawa 1986; Hayashi and Young 1987; Iga 2013). Lyman et al. (2005) applied this approach to TIWs and concluded that TIWs arose from the coupling between the first- and the second-meridional-mode equatorial Rossby waves both with the first-vertical-mode structure. This curious result, however, may have arisen from the method they used to identify the coupling waves; they changed the background field in a manner that strongly alters the Rossby wave characteristics. It is generally two waves propagating in the opposite directions in a sheared mean flow that can resonantly interact to generate an unstable wave (Baines and Mitsudera 1994; Iga 1999a; Heifetz et al. 2004); the equatorial Rossby waves (Matsumo 1966; De-Leon et al. 2010) all propagating westward relative to the mean flow are inadequate to generate TIWs.

In this study, to investigate the generation mechanism of TIWs, we carry out a linear stability analysis using a 1.5-layer shallow water model on the equatorial beta plane with a zonally uniform background field. The background field is constructed from an eddy-resolving ocean general circulation model (OGCM) that has successfully reproduced realistic TIWs. The analysis clarifies the decisive parameters that control the activity of the TIWs and provides a physical explanation of the role of each equatorial current (SEC, NECC, and EUC) and the associated thermocline structure in generating TIWs.

2. TIWs reproduced in an OGCM

First, we briefly describe the characteristics of TIWs reproduced in a global eddy-resolving OGCM [OGCM for the Earth Simulator (OFES); Masumoto et al. 2004; Sasaki et al. 2008; Masumoto 2010]. OFES solves the three-dimensional primitive equations under the Boussinesq and hydrostatic approximations with a grid resolution of 0.1° in the horizontal direction and 54 levels in the vertical direction of which 20 levels are in the upper 200 m. OFES is driven by daily mean surface wind stresses obtained from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data, and surface heat flux calculated using a bulk formula incorporating the model-predicted SST and the necessary atmospheric variables obtained from the NCEP–NCAR reanalysis data.

Figures 1a–d show the distributions of the sea surface velocity and density anomalies in the eastern to central equatorial Pacific from the summer of 2010 to the spring of 2011, where the anomalies are all defined as the deviations from the 30-day running mean. We can confirm that TIWs with a wavelength of ~1000 km are well reproduced immediately north of the equator, which propagate westward at a speed of ~0.5 m s\(^{-1}\). They develop in boreal summer with a time scale of 1–2 months, become most energetic in fall to winter, and decay by the next spring. It is also confirmed that TIWs exhibit significant interannual variability; they are hardly generated even in summer to winter during El Niño periods (e.g., Fig. 1e). All these features are confirmed by observed results.

The meridional cross section of the zonal velocity and density averaged over 110°–150°W and July–August 2010 (corresponding to the longitude range and time period in which TIWs rapidly grow) is shown in Fig. 2a.
As characteristic of the equatorial Pacific, strong zonal jets such as the northern branch of the SEC flowing westward with a maximum speed of $\sim 1 \text{ m s}^{-1}$ (1°S–4°N, 0–80 m), the NECC flowing eastward with a maximum speed of $\sim 0.5 \text{ m s}^{-1}$ (4°–9°N, 0–100 m), and the EUC flowing eastward with a maximum speed of $\sim 1 \text{ m s}^{-1}$ (3°S–2°N, 40–200 m), are well reproduced. A sharp pycnocline at $\sim 23.5–25.0 \sigma_0$ can also be clearly recognized especially in the Northern Hemisphere, which slopes downward to the north between 0° and 5°N and then upward between
and 11°N in geostrophic balance with the mean zonal currents. All these features are consistent with CTD and ADCP observations repeatedly carried out along 150°W (see Figs. 6a and 8 of Luther and Johnson 1990).

Both the zonal and meridional velocity anomalies associated with TIWs are large and nearly uniform in the vertical direction above the pycnocline whereas they are fairly small below it, and the large density anomaly associated with TIWs is mostly confined within the pycnocline (Figs. 2b,c). This allows us to employ a 1.5-layer model as a first step to understand the generation mechanism of TIWs. The anomalies of the upper-layer thickness from the sea surface to the 24.375σ0 isopycnal surface (middle of the pycnocline) and of the layer-averaged velocity are shown in Fig. 3a, which will be compared with those predicted from the linear stability analysis in the next section. We can see that TIWs lean westward to the north against the meridional shear between the SEC and NECC (0°–8°N), which is favorable for barotropic instability.

3. Linear stability analysis

a. Formulation

We next carry out a linear stability analysis for the physical interpretation of the generation mechanism of TIWs. The governing equations for the linearized 1.5-layer shallow water ocean with a zonally uniform background field on the equatorial beta plane under the hydrostatic and rigid-lid approximations are expressed as

\[ \frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fu = -g \frac{\partial \eta}{\partial x}, \]  
\[ \frac{\partial \eta}{\partial t} + H \left( \frac{\partial u}{\partial x} + \frac{\partial \eta}{\partial y} \right) + U \frac{\partial \eta}{\partial x} + v \frac{\partial \eta}{\partial y} + \frac{dH}{dy} = 0, \]  

where \( t \) is time, \( x \) and \( y \) are horizontal coordinates defined as positive eastward and northward, respectively, \( f = \beta y \) is the Coriolis frequency with \( \beta = 2.2 \times 10^{-11} \text{m}^{-1}\text{s}^{-1} \), \( g' = 0.0334 \text{m s}^{-2} \) is the reduced gravity corresponding to the density difference of 3.5 kg m\(^{-3}\) between the upper and lower layers, \( H(y) \) and \( U(y) \) are the mean upper-layer thickness and layer-averaged zonal velocity, respectively, and \( \eta \) and \( (u, v) \) are the anomalies of the upper-layer thickness and the layer-averaged velocity in the \( (x, y) \) directions, respectively.

Assuming a wave-like solution for the perturbation field,

\[ u = \tilde{u}(y) \exp(ik(x - ct)), \]  
\[ v = \tilde{v}(y) \exp(ik(x - ct)), \]  

Fig. 2. (a) Meridional cross section of the zonal velocity (color) and density (contours) averaged over the longitude range of 110°–150°W and the time period of July–August 2010. Meridional cross sections of the anomalies on 18 Jul 2010 of (b) the zonal velocity (color) and density (contours) at 128°W and (c) the meridional velocity (color) and density (contours) at 130°W. All data are obtained from OFES. The contour intervals are 0.5σ0 in (a) and 0.25σ0 in (b) and (c). The dashed contours in (b) and (c) indicate negative values.
where $k$ is the zonal wavenumber assumed to be positive and $c$ is the complex phase speed, we can rewrite Eqs. (1)–(3) as

$$k(U - c)\tilde{u} + \left(f - \frac{dU}{dy}\right)\tilde{v} + g'k\tilde{h} = 0,$$

(7)

$$k(U - c)\tilde{v} + f\tilde{u} + g\frac{d\tilde{h}}{dy} = 0,$$

(8)

$$k(U - c)\tilde{h} + Hk\tilde{u} - \frac{d}{dy}(H\tilde{v}) = 0.$$  

(9)

A no-normal flow boundary condition, $\tilde{v} = 0$, is imposed at $y = L_S$ and $y = L_N$ with $L_S = 8^\circ$S and $L_N = 12^\circ$N the southern and northern boundaries of the model domain, respectively, both located far away from the southern and northern edges of the numerically reproduced TIWs. Equations (7)–(9), together with the boundary conditions, form a generalized eigenvalue problem for a fixed value of $k$, which can be rewritten in matrix form as

$$Aq = \lambda Bq,$$

(10)

where $A$ and $B$ are coefficient matrices for each term in Eqs. (7)–(9) discretized by a second-order central difference method with a meridional grid resolution of 0.1$^\circ$, and $\lambda = \lambda_c + ic$, and $q = (\tilde{u}, \tilde{v}, \tilde{h})^T$ are the eigenvalue and the corresponding eigenvector, respectively. The eigenvalue problem is solved numerically with a standard matrix method based on a QR decomposition (Anderson et al. 1999) for the range of $0 < k \leq 2 \times 10^{-5}$ m$^{-1}$. Assuming that the upper layer extends from the sea surface to the 24.375$\sigma_b$ isopycnal surface, $H(y)$ and $U(y)$ are both determined from OFES results by averaging over the longitude range of 110$^\circ$–150$^\circ$W and the time period of July–August 2010 and smoothing meridionally with a 0.2$^\circ$ running mean so as to remove small-scale features associated with TIWs. Figure 4 shows the distributions of $H(y)$ and $U(y)$, together with the distribution of the background potential vorticity (PV) defined as

$$Q = -\frac{f - \frac{dU}{dy}}{H},$$

(11)

an essential parameter controlling the wave properties in a rotating shallow water system. Note that the meridional gradient of $Q(y)$ is generally positive due to the beta effect but becomes negative between 1$^\circ$ and 3.5$^\circ$N because of the sharp change in both $H(y)$ and $U(y)$.

b. Unstable mode solution

Figure 5a shows the growth rate ($\kappa\lambda_c$) and phase speed ($\kappa c_c$) of the fastest growing mode as a function of $k$. The equatorial zonal flow becomes most unstable for the disturbance with the zonal wavelengths of $\approx$1700 km ($k \approx 3.7 \times 10^{-6}$ m$^{-1}$). Although this is larger than the wavelength of the TIWs reproduced in OFES ($\approx$1000 km), this discrepancy may be tolerable, considering that the background field in the linear stability
analysis is highly simplified due to the employment of the 1.5-layer shallow water model. Furthermore, the phase speed of this unstable mode is $0.44 \text{ m s}^{-1}$, very close to that of the numerically reproduced TIWs ($0.5 \text{ m s}^{-1}$).\footnote{We can also find a second unstable mode at $k = 9 \times 10^{-4} \text{ m}^{-1}$, which will be briefly discussed later (see footnote 2).}

The spatial structure of the fastest growing mode is shown in Fig. 5b, where we can see that the anomaly of the upper-layer thickness distributes between $0^\circ$ and $8^\circ$N with the maximum located at $\sim 5^\circ$N, while the zonal and meridional velocity anomalies are located slightly to the south. Furthermore, this unstable mode leans westward to the north against the meridional shear between the SEC and NECC, which enables barotropic conversion from the mean to the eddy kinetic energy through the horizontal Reynolds stress. All these features are more or less recognized in Fig. 3a, showing that the unstable mode obtained in the linear stability analysis corresponds to the TIWs reproduced in OFES.

The energy equation in the 1.5-layer model can be written from Eqs. (1)–(3) as

$$\frac{\partial}{\partial t} \left[ \frac{1}{2} H (u^2 + v^2) + \frac{1}{2} g \eta^2 \right] + \frac{\partial}{\partial y} (g H \eta \eta) = -H \frac{\partial U}{\partial y},$$

(12)

where an overbar denotes the $x$ average, and the right-hand side represents the energy conversion from the mean flow to the waves. Figure 5c shows the energy conversion rate of the fastest growing mode as a function of $y$, which can be compared to the corresponding one diagnosed from OFES (Fig. 3b). We can confirm that active energy conversion takes place at $2^\circ$–$6^\circ$N in both the unstable mode and the OFES simulation. Furthermore, the growth rate of the TIWs in OFES can be estimated as $\sigma = C/E$ or $\sigma = (1/E)(dE/dt)$, where $E$ and $C$ are the area-integrated TIW energy [the first term on the left-hand side of Eq. (12)] and the area-integrated energy conversion rate [the right-hand side of Eq. (12)], respectively. Both estimates yield the $e$-folding time of $\sim 100$ days in this period, consistent with that of the unstable mode ($\sim 87$ days, Fig. 5a).

c. Unstable mode as a coupling between two neutral modes

One of the well-known methods to identify a pair of neutral modes that form an unstable mode is to examine the dispersion diagram to find two curves that merge into the unstable mode. However, the applicability of this method is limited to special cases in which either the background flow or background PV is uniform. This is due to the so-called “continuous modes” (Case 1960).
FIG. 5. (a) Growth rate \( k_c \) (black dots and right axis) and phase speed \( c_r \) (red dots and left axis) of the fastest growing mode as a function of zonal wavenumber \( k \). The phase speed is plotted only when the growth rate exceeds \( 1.0 \times 10^{-8} \text{ s}^{-1} \). (b) Spatial structure of the anomalies of the upper-layer thickness \( \eta \) (color) and the layer-averaged velocity \((u, v)\) (arrows) associated with the fastest growing mode with the zonal wavenumber \( k = 3.7 \times 10^{-6} \text{ m}^{-1} \). (c) Meridional distribution of the energy conversion rate from the mean flow to the unstable mode in (b). (d) Phase speed of all the eigenmodes as a function of \( k \). Unstable modes are shown in red.
Iga (1999b) has shown that Eqs. (7)–(9) have a singularity at a critical latitude \( y_c \) [which satisfies \( c = U(y_c) \)] unless \( dQ(y_c)/dy = 0 \), and the solution in this case can satisfy any two boundary conditions. This means that any value of \( c \) within the range of \( U \) can continuously be an eigenvalue of Eq. (10) unless \( Q \) is uniform, so that numerous neutral modes (continuous modes) spread widely over the dispersion diagram. We can actually confirm this in Fig. 5d.

One of the well-known necessary conditions for instability in a midlatitude quasigeostrophic system is that the meridional PV gradient changes sign somewhere in the

**FIG. 6.** (a) Latitudinal distribution of \( Q(y) \) and (b) the growth rate \( k c_i \) (black dots and right axis) and phase speed \( c_i \) (red dots and left axis) of the fastest growing mode as a function of \( k \) for the background field in which the negative PV gradient between 1° and 3.5°N is smoothed out. The phase speed is plotted only when the growth rate exceeds \( 1.0 \times 10^{-3} \) s\(^{-1}\). (c), (d) As in (a) and (b), but for the background field in which the positive PV gradient south of 1°N is smoothed out. (e), (f) As in (a) and (b), but for the background field in which the positive PV gradient north of 3.5°N is smoothed out.
domain (e.g., Pedlosky 1987; Vallis 2017). Indeed, Fig. 4c clearly shows that there exists a local negative PV gradient just north of the equator (1°–3.5°N) in the ambient positive PV gradients. To test this hypothesis, the linear stability analysis is applied to the background field in which the negative PV gradient is artificially smoothed out by slightly changing $f(y)$ in a narrow latitude range including 1°–3.5°N while $H(y)$ and $U(y)$ are kept unchanged (Fig. 6a). The dispersion relation obtained from this modified background field is shown in Fig. 6b, where we can find that this slight change in the local PV distribution brings about the complete elimination of the unstable mode corresponding to TIWs ($k \sim 3.7 \times 10^{-6} \text{m}^{-1}$), demonstrating the validity of the abovementioned hypothesis.

We also investigate the importance of the adjacent positive PV gradients south of 1°N and north of 3.5°N. First, when the positive PV gradient south of 1°N is smoothed out by slightly changing $f(y)$ (Fig. 6c), we find that the dispersion relation and the spatial structure of the unstable mode corresponding to TIWs are both hardly changed ($k \sim 3.7 \times 10^{-6} \text{m}^{-1}$ in Fig. 6d). Next, when the positive PV gradient north of 3.5°N is smoothed out by slightly changing $f(y)$ (Fig. 6e), in contrast, we find that the unstable mode corresponding to TIWs again almost completely disappears ($k \sim 3.7 \times 10^{-6} \text{m}^{-1}$ in Fig. 6f). This indicates that the combination of the negative PV gradient just north of the equator (1°–3.5°N) and the positive PV gradient further north (3.5°–11°N) plays an essential role in generating TIWs.

The second unstable mode ($k \sim 9 \times 10^{-6} \text{m}^{-1}$), in contrast, owes its existence to the combination of the negative PV gradient at 1°–3.5°N and the positive PV gradient south of 1°N (Figs. 6a–f). Although this unstable mode might be related to another type of TIWs with shorter periods and wavelengths found closer to the equator (Halpern et al. 1988; McCreary and Yu 1992; Yu et al. 1995; McPhaden 1996; Lyman et al. 2007), it will not be discussed further because we cannot find corresponding disturbances in the OFES simulation.
To understand the role of these negative and positive PV gradients, we further investigate neutral waves that can exist corresponding to each of the PV gradients. Figure 7b shows the dispersion relation of the neutral waves obtained when only the negative PV gradient just north of the equator is retained by smoothing out the ambient positive PV gradient using the same method as described above (Fig. 7a). Although this PV distribution is expected to allow the existence of trapped Rossby waves discretized in a manner similar to equatorially or topographically trapped Rossby waves (e.g., Staniforth et al. 1993; Iga 1995; Cohen et al. 2010), we cannot identify them in the dispersion relation (Fig. 7b) because of the superposition of the continuous modes described above. To remove these continuous modes, we homogenize the background zonal velocity by averaging it over the latitude range with the negative PV gradient while keeping the PV distribution unchanged. A similar approach was also employed by Taniguchi and Ishiwatari (2006). The dispersion relation obtained from this modified background field (i.e., negative PV gradient with uniform zonal flow) is shown in Fig. 7c, where we can clearly identify a discrete set of Rossby waves with eastward intrinsic phase propagation superposed on the SEC flowing westward at \( \sim 0.65 \) m s\(^{-1}\).

Figure 8b, on the other hand, shows the dispersion relation of the neutral waves obtained when only the positive PV gradient farther north (3.5°–11°N) is retained (Fig. 8a). Again, although the existence of a discrete set of Rossby waves is expected, we cannot identify them in the dispersion relation because of the existence of numerous continuous modes. Also in this case, by averaging the background zonal velocity over the latitude range from 3.5° to 6.5°N (roughly corresponding to the equatorial radius of deformation), we can retrieve the discrete set of Rossby waves with westward intrinsic phase propagation superposed on the NECC flowing eastward at \( \sim 0.1 \) m s\(^{-1}\) (Fig. 8c).

Of special interest is that we can reconstruct a spatial structure fairly close to that of the unstable mode by superposing the two first-meridional-mode Rossby waves obtained above with the northern one trapped in the positive PV gradient \( (k = 3.7 \times 10^{-6} \text{ m}^{-1} \text{ and } c = -0.43 \text{ m s}^{-1}; \text{ Fig. 8d}) \) leading the southern one trapped in the negative PV gradient \( (k = 3.7 \times 10^{-6} \text{ m}^{-1} \text{ and } c = 0.65 \text{ m s}^{-1}) \).
and $c = -0.46 \text{ m s}^{-1}$; Fig. 7d) by $\sim \pi/3$ (compare Fig. 9b with Fig. 5b). This indicates that TIWs can be interpreted as resulting from a coupling of two Rossby waves with opposite intrinsic phase propagation directions. Note that this coupling becomes possible because the Rossby waves just north of the equator are advected westward by the strong SEC, whereas the other Rossby waves farther north propagate westward against the weak NECC, so that they both propagate westward at the same speed keeping $k = 3.7 \times 10^{-6} \text{ m}^{-1}$ which yields the largest growth rate of TIWs (see Fig. 9a).

It should be noted that, although the EUC does not seem to play a direct role in generating TIWs since it does not appear in $U(y)$, it is still important in sustaining the negative PV gradient north of the equator through $H(y)$, which rapidly increases northward from the equator to satisfy the thermal wind balance with the vertical shear between the SEC and EUC (see Fig. 2a). This means that, although the employment of the 1.5-layer model precludes the occurrence of baroclinic instability, the effects of baroclinicity of the background field are still partially taken into account.

d. Importance of the negative PV gradient

The linear stability analysis is also applied to other periods: November–December 2010 when TIWs become most energetic (Figs. 10a,b), April–May 2011 when TIWs disappear almost completely (Figs. 10c,d), and July–August 2015 when TIWs do not develop even in summer (Figs. 10e,f). During the periods of weak TIWs, unstable modes with a significant growth rate and a phase speed of $\sim -0.5 \text{ m s}^{-1}$ can hardly be found (Figs. 10d,f), consistent with the PV gradient north of the equator being mostly positive (Figs. 10c,e). Furthermore, the growth rate of the unstable mode in winter is much smaller than that in summer (Fig. 10b versus Fig. 5a) although TIWs are more active in winter, again consistent with the PV gradient north of the equator becoming nearly flat by this season (Fig. 10a versus Fig. 4c).

Figure 11 shows the divergence of the meridional eddy PV flux \( \left( \frac{\partial}{\partial y} \eta \right) q \), where \( q = (1/H)[(\partial u/\partial x) - (\partial u/\partial y)] - Q \eta \) is the upper-layer eddy PV} averaged from July 2010 to March 2011 (over one TIW season) diagnosed in OFES. A strong divergence can be found around the peak latitude of the background PV ($\sim 0.5^\circ$N; compare Fig. 11 with Fig. 4c), indicating that TIWs can act to homogenize the background PV once they grow to significant amplitudes. This suggests that the weakened negative PV gradient in winter (Fig. 10a) may be partially explained in terms of the homogenization of the background PV by eddies (e.g., Rhines and Young 1982).

Figure 12 shows the time series of the strength of the negative PV gradient just north of the equator and the eddy energy associated with TIWs, both obtained from OFES. It is confirmed that they are significantly correlated with each other with the former leading the latter by 1–2 months except for some occasional large discrepancies (e.g., March–May 2007). Although the existence of the negative PV gradient is merely a necessary condition for the generation of TIWs, this result suggests that it could serve as a good indicator of both the seasonal and interannual variability of the TIWs. The most noticeable difference can be found in March–May 2007. The large negative PV gradient in this period is
associated with an extremely thin upper-layer thickness (less than 10 m) confined within a narrow zonal region of $\sim 140^\circ$–150$^\circ$W near the equator (not shown). The weak TIW energy during this period suggests that this large negative PV gradient is too narrow to generate TIWs. Another possibility that can cause the discrepancy is baroclinic energy conversion, which cannot be fully taken into account in the present 1.5-layer framework.

4. Summary and discussion

In this study, we have investigated the generation mechanism of TIWs by carrying out a linear stability
analysis using a 1.5-layer shallow water model on the equatorial beta plane with a zonally uniform background field. The background field is constructed from the eddy-resolving OGCM that has successfully reproduced realistic TIWs. The linear stability analysis has been shown to predict successfully an unstable mode whose wavelength, phase speed, growth rate, and meridional structure are all in reasonable agreement with those of the TIWs simulated by the OGCM. This unstable mode can be interpreted as resulting from the coupling of two Rossby waves, namely, one trapped just north of the equator (−1°–3.5°N) and the other trapped farther north (−3.5°–8°N). Although these two Rossby waves have opposite intrinsic phase propagation directions reflecting the negative and positive local meridional PV gradients, respectively, their actual propagation direction can be adjusted through the advection by the SEC and the NECC such that they might propagate westward at the same speed and with the same zonal wavenumber yielding the largest growth rate of TIWs. The unstable mode does not appear during the period in which the negative PV gradient is completely absent, suggesting its essential role in generating TIWs. In fact, the seasonal and interannual variability of the TIWs simulated by the OGCM is shown to be significantly controlled by the strength of the negative PV gradient just north of the equator.

The coupling of two neutral waves with opposite intrinsic phase propagation directions is a well-known interpretation of an unstable wave. As described in section 3c, however, previous application of this interpretation was limited to highly idealized situations, and the application to TIWs has not yet been satisfactory (e.g., Lyman et al. 2005). The unstable mode obtained in this study is similar to the one previously obtained by Philander (1978) and Hansen and Paul (1984), who carried out a linear stability analysis for a similar background field taking into account the negative PV gradient [whose importance was pointed out by Philander (1976)] but did not provide a physical interpretation of the generation mechanism. This study thus extends their study by explicitly demonstrating that TIWs actually result from a coupling of two Rossby waves, one of which is sustained by the negative PV gradient.

There are some limitations in the linear stability analysis carried out in this study. The most serious one may be the employment of the highly idealized 1.5-layer model, which precludes the occurrence of baroclinic instability. Although the effects of baroclinicity of the background field has partially been taken into account through the meridional variation of the upper-layer thickness, a more realistic stability analysis using a continuously stratified model is indispensable to obtain a definite conclusion about the importance of baroclinic instability. The second is the use of the linear theory. A nonlinear theory that can take into account the interaction between the mean and eddy fields is indispensable to investigate both the mature and decay stages of TIWs. In spite of these limitations, this study is the first to provide a simple and clear physical interpretation of...
the generation mechanism of TIWs in the complicated equatorial current system. We believe that this study provides a useful first step toward a more dynamically based parameterization of heat and momentum transfer in the equatorial oceans where TIWs are the dominant mesoscale processes, which could lead to an improvement in the predictability of future climate models.

Acknowledgments. This work was supported by the Japan Society for the Promotion of Science through the Grants-in-Aid for Scientific Research (JSPS KAKENHI Grant Number 17K14389). The authors thank Yukio Masumoto who kindly provided the OFES data. The LAPACK Library was used to solve the eigenvalue problem, and the GFD-DENNOU Library was used to produce the figures.

REFERENCES