Submesoscale Processes in the Upper Red Sea

Peng Zhan¹², Daquan Guo², George Krokos², Jihai Dong³, Rodrigo Duran⁴⁵, Ibrahim Hoteit²

Corresponding author: Peng Zhan, Department of Ocean Science and Engineering, Southern University of Science and Engineering, Shenzhen, China. (zhanp@sustech.edu.cn)

¹Department of Ocean Science and
Abstract.

Spatial-temporal submesoscale variabilities in the upper Red Sea and their generation mechanisms, including frontogenesis, mixed-layer instability (MLI), and symmetric instability (SI) are qualitatively investigated using high-resolution simulations. The results suggest that submesoscales are critical hydrodynamic components and stirring at submesoscale has a clear signal in the Red Sea, enhanced in winter, particularly in the central and northern basins, and in-
tensified toward the eastern coast. Frontogenesis and MLI energize submesoscales with winter peaks, when SI could also be triggered by the enhanced frontal gradients and buoyancy loss that reduce the surface potential vorticity. The MLI and SI have larger (smaller) scales in winter (summer). The seasonal submesoscale variability is governed by the vertical structure of the mixed layer forced by atmospheric conditions, significantly modulating the mesoscale eddies’ seasonality via an inverse cascade. This study offers new insights into understanding the Red Sea submesoscales and have potential applications to other marginal seas.
Key Points

The Red Sea submesoscales are enhanced in winter and enhanced toward the northern and eastern basin due to buoyancy flux and Ekman effects.

Submesoscales are forced by seasonal atmospheric conditions and modulate the seasonality of the mesoscales via an inverse energy cascade.

Submesoscales may play a role in chaotic stirring in the Red Sea, although they are dominated by two-dimensional structures.
Plain Language Summary Oceanic submesoscale dynamics have small dimensions (110s km), and they are important in the variability of physical, biological and chemical processes. We used a high-resolution numerical model to study the spatial-temporal submesoscale variabilities in the upper Red Sea, as well as their driving force and energy budget. We found that submesoscales are critical components in the Red Sea dynamics and make important contributions to stirring and transport in the basin. The seasonal submesoscale variability is essentially driven by atmospheric conditions and can modulate the eddies’ seasonality by converting energy from smaller scales to larger scales. This study offers new insights into understanding the Red Sea submesoscales and has potential applications to other marginal seas in the global oceans.
1. Introduction

Submesoscale processes at the surface mixed layer (SML) are prominent in the open oceans [Klein et al., 2008; Lvy et al., 2012; Gula et al., 2014; Qiu et al., 2014; Thompson et al., 2016; Sasaki et al., 2017; Jing et al., 2021], particularly in regions hosting rich mesoscale eddies [Capet et al., 2008a; Qiu et al., 2014; Molemaker et al., 2015; Renault et al., 2017; Damien et al., 2017; Sasaki et al., 2017; Su et al., 2018; Wang et al., 2018; Zhang and Qiu, 2018]. The SML submesoscales play a significant role in the kinetic energy (KE) cascade [Capet et al., 2008b; McWilliams, 2016; Dong et al., 2020a; Cao et al., 2021], heat budget [Sasaki et al., 2014; Su et al., 2018; Zhang et al., 2021], and SML restratification [Boccaletti et al., 2007; Yu et al., 2019; Plessis et al., 2019; Chrysagi et al., 2021]. In addition, they are believed to exchange surface and thermocline waters, which plays an important role in both physical and biogeochemical budgets [Lvy et al., 2001, 2012; Mahadevan et al., 2012; Huhn et al., 2012; Zhang and Qiu, 2020]. Although they exhibit scales smaller than the quasi-geostrophy framework, they are still strongly influenced by the Coriolis effect and the generally stable density stratification in the upper ocean [Thomas et al., 2008; McWilliams, 2019].

The Red Sea is a marginal sea known for its marine biodiversity [Dreano et al., 2016; Raitsos et al., 2017; Gittings et al., 2018, 2019a, b; Wang et al., 2019; Hoteit et al., 2020; Gittings et al., 2021], where distinct seasonally varying flow [Yao et al., 2014a, b] along with strong mesoscale eddies [Zhan et al., 2014, 2015, 2016, 2018, 2019; Krokos et al., 2019; Sanikommu et al., 2020] make it a unique sea in the global oceans. As many submesoscales can be created from mesoscale eddies, we therefore expect some degree of
correlated geographical and temporal variability for submesoscales in the Red Sea. This has been evidenced by the filament and vortex submesoscale structures of chlorophyll concentration captured by high-resolution ocean color images (Figures 1a to 1c). However, the dynamics of the submesoscales in the Red Sea are still poorly understood due to the lack of high fidelity in observations and previous models. As model resolutions continue to increase from mesoscale eddy-permitting to finer scales, submesoscale structures can now be (partially) resolved. This allows us, for the first time, to systematically investigate the spatial-temporal variabilities, the underlying mechanisms and the KE cascade processes of the Red Sea submesoscales. We also assess and interpret unique submesoscale features and their contribution in chaotic stirring by investigating the characteristics of their Lagrangian motion. The dynamics discussed here offer new insights into understanding submesoscales in the Red Sea and have potential applications in other marginal seas of the global oceans.

2. The Red Sea Simulation

The presented analysis is based on the outputs of a high-resolution MIT general circulation model [Marshall et al., 1997] developed for the Red Sea as an updated version of that of Yao et al. [2014a, b]. The model is configured with a horizontal resolution of \( \sim 1 \) km and 50 vertical \( z \) levels with a thickness ranging from 4 m at the surface to 300 m near the bottom. The model is forced with three-hourly surface atmospheric fields (10-m wind velocity, 2-m air temperature and specific humidity, downward long-wave and shortwave radiation, and precipitation) of a 5-km Weather Research Forecast product [Viswanadhapalli et al., 2017] downscaled from the ERA-Interim dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF, available at
https://www.ecmwf.int/en/forecasts/datasets/archive-datasets) [Dee et al., 2011]. The eastern open boundary in the Gulf of Aden was specified using daily temperature, salinity, and horizontal velocity fields provided by the Copernicus Marine Environment Monitoring Service (CMEMS) global ocean reanalysis (available at ftp://my.cmems-du.eu/Core/). K-profile parameterization (KPP) [Large et al., 1994] is used for vertical mixing. The model has been validated against different types of observations [Yao et al., 2014a, b; Papadopoulos et al., 2015; Toye et al., 2017; Zhan et al., 2018] and has been used to study various oceanic processes in the Red Sea, e.g., [Zhan et al., 2016, 2018; Guo et al., 2016, 2018; ]; Toye et al., 2018; Hoteit et al., 2020]. The simulation starts from 1990, and daily outputs from 2015 to 2018 are used in this study. As discussed in the following sections, due to the horizontal grid spacing and the hydrostatic configuration, the model cannot resolve most symmetric instability (SI). However, to a large extent, other associated submesoscales processes, including the strain-induced frontogenesis, mixed layer instability (MLI) and wind-front interactions, can largely be resolved, enabling a systematic investigation of submesoscales in the Red Sea.

3. Finite-time Lyapunov Exponent

The degree of stirring and dispersal of tracer patches can be quantified by calculating the finite-time Lyapunov exponent (FTLE). Regions of maximum separation rates produce ridges in the FTLE field, which approximate the Lagrangian coherent structures (LCSs) [?]. Because fluid particles that separate backward (forward) in time grow closer together in forward (backward) in time, when the FTLE is computed by integrating particle trajectories backward (forward) in time, a ridge in the FTLE field corresponds to an attracting (a repelling) LCS. The FTLE is related to the Lagrangian strain rate experi-
enced by the particles, and can be interpreted as the lengthening rate of a small patch of tracers within the study period [Waugh et al., 2006]. If the integration period is short, the FTLE can be approximated by the Eulerian strain rate [Serra and Haller, 2016]. As the period increases, the Eulerian method often fail to reveal material patterns and likely overestimates the actual stirring effect [Waugh et al., 2006; Waugh and Abraham, 2008; Olascoaga et al., 2013].

The FTLE is a scalar quantity that represents the separation rate of initially neighboring particles over a finite-time window \([t, t + T]\). Considering an arbitrary point \(x_t\) at time \(t\), the FTLE is the average principal-axis Lagrangian strain due to advection of the flow from time \(t\) and to time \([t, t + T]\). Maximal stretching occurs in the direction of the principal axis, given by the eigenvector associated with the largest eigenvalue \(\lambda_{max}\) of the Cauchy-Green deformation tensor \(\Delta = M^T M\), where \(M : d(x_t \mapsto x_{t+T})/dx_t\) is the derivative of the flow map that advances points \(x_t\) in the domain at time \(t\) to their new locations \(x_{t+T}\) at time \(t + T\) with respect to \(x_t\), with stretching given by the largest eigenvalue. The forward FTLE (i.e. \(T > 0\)) is then defined as \(FTLE = \log(\lambda_{max})/2T\). The readers are referred to [Haller, 2015] for more details about the computation of the FTLE discussed above.

The FTLE here is computed by integrating the flow map of particles forward in time on a 1-km grid using the code provided by AVISO and available at https://bitbucket.org/cnes_aviso/lagrangian/src/master/. The 10-day calculations have been performed once every five days using daily velocities from 2015 to 2018.
4. Results and Discussion

4.1. Spatiotemporal Characteristics of Submesoscales

Mesoscale eddies in the Red Sea exhibit a pronounced seasonality with favorable presence in the central and northern basin during winter [Zhan et al., 2014, 2016, 2019]. Similar variabilities are therefore expected for the submesoscales. To illustrate the scenarios of the Red Sea submesoscales, close-up snapshots of associated quantities on a typical winter day (9 January 2018) and summer day (6 September 2016) are presented in Figure 1. The figures reveal a distinct seasonal pattern of submesoscale structures that is more energetic in winter when patterns are identified as vortices and elongated filaments near the edges of mesoscale eddies or frontal regions when the horizontal density gradient is large, and as wiggles near the coastal regions with complex topographic conditions (Figures 1d and 1i). The submesoscales can be estimated using the Rossby number \( Ro = \frac{\zeta}{f} \) defined as the vertical component of relative vorticity \( \zeta \) normalized by the Coriolis frequency \( f \). The surface \( |Ro| \) frequently reaches order of one. In addition, the associated submesoscales are characterized by a strong positive skewness towards the dominance of cyclones (Figures 1e and 1j), as the asymmetric vortex stretching process favors cyclonic vorticity dominance for magnitudes larger than \( f \) and the anticyclonic vorticities are quickly destroyed by inertial instability [Capet et al., 2008a; Shcherbina et al., 2013; McWilliams, 2016]. The mesoscales are quasigeostrophic and to a large extent divergent-free, however, significant vertical velocity can be produced by restoring of the thermal wind balance within the submesoscale density filaments elongated by the horizontal stirring processes [Hakim et al., 2002; Callies et al., 2015]. Therefore, this process induces a strong correlation between density anomalies and vertical velocity \( w \), with a positive
$w$ in lighter filaments and negative in denser filaments within high-strain regions. Thus, the locations with a larger $|Ro|$ are usually accompanied by vigorous vertical motions (possibly reaching $>300$ m/day) within the SML (Figures 1f and 1k), whose depth varies following the mesoscale structures yet can jump locally as a response to strong fronts (as illustrated by the spike in the cross-section in Figure 1i). In comparison, the distributions on a typical summer day display a shallower SML with a much flatter base, a noticeably less dense population of submesoscales with weaker vertical motions (Figures 1n to 1p, 1s to 1u).

The seasonally averaged surface $|Ro|$ (Figures 2a and 2f) suggests that, in general, most prominent submesoscales in the Red Sea occur during winter (December to February) rather than summer (July to September). Corresponding to the distribution of a larger $|Ro|$, the submesoscale flows co-occur with more surface heat loss (Figures 2b and 2g), deeper SML (Figures 2c and 2h), stronger divergence and convergence that produce noticeable subsurface vertical velocities (Figures 2d and 2i), and submesoscale upward vertical heat flux ($VHF_{sma}$, Figures 2e and 2j). Here, $VHF_{sma} = \rho C_p w' T'$ is estimated by the vertical velocity anomaly $w'$ and temperature anomaly $T'$ from a spatially highpass filtered daily average reference with a cutoff scale of $50 - km$ at a depth of $50$ m, where $\rho$ and $C_p$ are the density and specific heat capacity, respectively. On average, the upward $VHF_{sma}$ in winter tends to produce a surface warming and a deeper cooling, highlighting the important role of submesoscales in modulating the near-surface heat budget with a magnitude of about one-third the surface air-sea flux. The wintertime upward $VHF_{sma}$ in the Red Sea can reach $80$ W/m$^2$, and this is even comparable to the hotspots in the global oceans where the submesoscale upward heat flux is strong, e.g., the Kuroshio Extension,
Gulf Stream, and the western Arabian Sea [Su et al., 2018; Jing et al., 2020]. Moreover, both $|\text{Ro}|$ and $|w|$ off the eastern coast during winter are enlarged, which suggests enhanced submesoscales and more active dynamical conditions in both the horizontal and vertical motions toward the eastern basin. Additionally, the strong seasonal variability of submesoscale processes can be observed in the spatially averaged time series of $|\text{Ro}|$, with the peak and trough occurring in winter and summer, respectively (black line in Figure 2k). In general, the fields in Figure 2 are highly nonuniform in space and in time, indicating submesoscales that are favorably developed in the northern Red Sea during winter.

4.2. Mechanisms and Characteristic Scales of Submesoscales

Submesoscale processes can be generated via mesoscale-driven surface frontogenesis and baroclinic MLIF that usually co-occur in the real ocean [Fox-Kemper and Ferrari, 2008; Callies et al., 2015; McWilliams, 2016; Srinivasan et al., 2017]. The former is strongly surface trapped, and it occurs when velocity shear and deformation of mesoscale currents intensify the horizontal density gradient through frontogenesis, inducing a submesoscale secondary circulation that resists the strengthening of the fronts and restores the geostrophic balance [Hoskins, 1982; Capet et al., 2008a; McWilliams et al., 2009; Gula et al., 2014; McWilliams, 2017]. The intensity of this process on the premise of horizontal density gradient can be depicted by the frontogenesis rate,

$$F = \left(-\frac{\partial \mathbf{u}}{\partial x} \cdot \nabla_h \rho, -\frac{\partial \mathbf{u}}{\partial y} \cdot \nabla_h \rho\right) \cdot \nabla_h \rho,$$

where $\mathbf{u}$ denotes the two-dimensional horizontal velocity, and $\nabla_h$ is the gradient operator. $F$ represents the rate of increase for the horizontal density gradient through mesoscale
straining [Capet et al., 2008a; Gula et al., 2014]. As depicted in Figures (1g) and (1l), $F$ is spatially intermittent and localized within the edges of mesoscale eddies or coastal regions where the velocity field is conducive to frontogenesis. A good consistency is found between the distribution of $F$ and $|Ro|$ extrema along the frontal filaments that usually appears at the edges of mesoscale eddies (Figures 1e and 1j), suggesting that the submesoscales are actively interacting with the fronts induced by the mesoscales in the Red Sea. This is different from another marginal sea, the Baltic Sea, where a strong and persistent frontal structure is maintained by the seasonal general circulation of the basin [Chrysagi et al., 2021]. The time series of spatially averaged $F$ corresponding to the $|Ro|$ exhibits a correlation coefficient of 0.70 (Figure 2k), indicating that frontogenesis acts as a key element in the basin to generate the submesoscales.

Moreover, MLI occurs when strained flows generate a horizontal density gradient and vertical shear in a low-stratification structure, and it extracts available potential energy (APE) and energizes the submesoscales throughout the SML [Boccaletti et al., 2007; Fox-Kemper et al., 2008; Mensa et al., 2013], as evidenced by the patterns across the section in Figures 1 (i-m). Although one cannot exclusively rule out those submesoscales generated by frontogenesis discussed above, the MLI processes can be diagnosed by the conversion rate of the APE to submesoscale KE within the SML

$$BC = \frac{1}{MLD} \int_{-MLD}^{0} \frac{w' b'}{dz},$$

where MLD is the mixed layer depth (Appendix A), and $w$ and $b$ represent the vertical velocity and buoyancy, respectively, and the prime, again, denotes the submesoscale fraction based on a spatial highpass filtering of the daily average with a cutoff scale of 50 – km. In

This article is protected by copyright. All rights reserved.
general, BC exhibits larger values at the flanks of the fronts where stronger vertical velocities drive larger buoyancy fluxes, and the spatially averaged BC is positive throughout the column with a clear maximum at the subsurface (not shown). As displayed in Figure 2 (l), the spatially averaged BC time series exhibits a close correlation (0.91) with the MLD with larger (almost zero) values in winter (summer). Its positive sign indicates a net source for submesoscale KE. This term increases when SML deepens from October to January, providing more space to store the APE, and then starts to decrease as the production declines during springtime restratification. Both BC and MLD are in phase with the $|Ro|$ (Figure 2k, with a correlation coefficient of 0.76 and 0.81, respectively), although their time series exhibit an abrupt decay after the $|Ro|$ peak from January to February.

Similar phenomena were also reported by previous studies [Capet et al., 2008c; Sasaki et al., 2014; Zhang and Qiu, 2020; Dong et al., 2020b], probably owing to the generation of submesoscales that causes strong restratification in the SML and hence the shoaling of the SML in spring [du Plessis et al., 2017; Su et al., 2018]. The submesoscales are seasonally forced by the MLI, which is amplified via SML deepening in winter and essentially induced by the strong seasonal atmospheric forcing over the Red Sea [Yao et al., 2014a, b; Viswanadhapalli et al., 2017; Langodan et al., 2017]. Nevertheless, a recent study of the Bolitic Sea suggests that the submesoscale restratification could occur rapidly (within less than a day) after strong convective events (e.g., storm), yielding localized regions with shallower SML that compensates the destratifying effect of convection [Chrysagi et al., 2021].

In comparison, the barotropic conversion term

$$BT = \frac{1}{\text{MLD}} \int_{-\text{MLD}}^{0} -\rho_0 \left( \bar{u}' \bar{v}' \frac{\partial \bar{u}}{\partial y} + \bar{u}' \bar{u}' \frac{\partial \bar{u}}{\partial x} + \bar{v}' \bar{v}' \frac{\partial \bar{v}}{\partial y} + \bar{u}' \bar{w}' \frac{\partial \bar{u}}{\partial z} + \bar{v}' \bar{w}' \frac{\partial \bar{v}}{\partial z} \right) dz$$

This article is protected by copyright. All rights reserved.
This suggests that submesoscale KE generated via barotropic instability, which is induced by mesoscale sheared flow, is weak and plays a minor role in the submesoscale generations in the Red Sea. Statistically, compared to the BC and BT estimated from the present model, the vertical shear production term is about two orders smaller (not shown), and thus negligible.

The scale of the fastest-growing MLI is related to the local deformation radius within the SML \( (Rd_{ML} = N_{ML} \cdot \text{MLD}/f, \text{Figures 3a and 3f}), \) where \( N_{ML} \) represents the buoyancy frequency within the SML. The scale of the most unstable perturbation in the SML can be approximated by \( L_{ML} = Rd_{ML} \sqrt{(1 + Ri^{-1})}, \) where \( Ri = N^2 f^2/|\nabla h b|^2 \) is the Richardson number [Boccaletti et al., 2007; Dong et al., 2020a]. Generally, \( N_{ML} \) is weaker (stronger) when MLD is opposingly deeper (shallower) in winter (summer); nevertheless, our analysis suggests that \( L_{ML} \) is primarily governed by MLD, indicating a seasonally varying MLI scale that enlarges in winter. Specifically, the winter and summer \( Rd_{ML} \) are about 12.4 and 6.1 \( km \) averaged over the northern Red Sea, and 9.6 and 8.0 \( km \) over the southern Red Sea (Figures 3a and 3f), with the resulting most unstable wavelength \( (\lambda_{ML} = 2\pi L_{ML}) \) in winter of about 77.4 \( km \) and 57.6 \( km \) in the northern and southern Red Sea, respectively. In winter, the \( Rd_{ML} \) is larger toward the northern basin, where SML is deeper, whereas in summer, the \( Rd_{ML} \) is smaller and generally decreases with latitude. Despite this, most submesoscale eddies generated by MLI have a scale several times larger than the present model resolution and are therefore expected to be resolved by the model. Moreover, the linear growth rate of MLI estimated as \( \nabla h b/N \) and presented in Figures 3b and 3g is generally faster in winter and is spatially corroborated with the...
distribution of the large frontogenesis rate where the horizontal buoyancy gradient is large (not shown). Generally, the MLI are ageostrophic baroclinic instabilities and differ from the deep mesoscale instabilities spanning the entire depth in their smaller scales $O(1\text{km})$ and faster growth rates $O(1/d) \approx f$.

Additionally, the submesoscales can be generated as a result of SI, which usually occurs, under the thermal wind balance and positive $f$, where the three-dimensional Ertel potential vorticity (PV)

$$q_{\text{Ertel}} = (f\mathbf{k} + \nabla \times \mathbf{u}) \cdot \nabla b \approx (\zeta + f)N^2 - |\nabla b|^2 < 0$$

[Hoskins, 1974; Haine and Marshall, 1998; Taylor and Ferrari, 2009; Thomas et al., 2013; Dong et al., 2021], where $\mathbf{u}$ and $\nabla$ represent the three-dimensional velocity and gradient operator, respectively. When stratification is weakened by vertical convection or mixing, this term can become negative and trigger SI [Bachman et al., 2017; Dong et al., 2021]. The SI can occur at spatial scales of $o(100 \text{ m})$ in the mixed layer [Dong et al., 2021; Jing et al., 2021], which is smaller than the present regional model grid; thus, SI is not expected to be resolved at the vast majority of locations. Nevertheless, the resolved structures and associated diagnostics enable us to estimate some potential characteristics of SI.

Snapshots of the simulation reveal that negative $q_{\text{Ertel}}$ occurs in company with filaments affected by strong frontal instabilities, particularly in winter (Figures 1h and 1m). Compared to the analysis of an approximately $2 - \text{km}$ global simulation where the negative surface $q_{\text{Ertel}}$ is more likely to be overestimated [Dong et al., 2021], our results appear to underestimate the occurrence of negative surface PV. This is partially because the model of Dong et al. [2021] is configured with a thinner surface layer (1 m) that enables the generation of more negative PV due to the surface buoyancy loss. Another possible reason is
that our 1–km model resolves relatively more MLI and (possibly) SI that tends to re-
stratify and mix so as to eliminate the negative PV. Nevertheless, the forcing that tends
to reduce the surface PV can be estimated. Budget analysis suggests that the surface PV
is largely modulated by the buoyancy flux. Specifically, the $q_{Ertel}$ can be reduced to yield
a negative PV when

$$\text{EBF} + B_0 > 0,$$

where EBF is the Ekman buoyancy flux

$$\text{EBF} = \frac{\tau \times k}{\rho_0(f + \zeta)} \cdot \nabla h b,$$

and $B_0$ is the buoyancy flux

$$B_0 = g\alpha \frac{Q_{\text{net}}}{\rho_0 C_p} + g\beta(EP)S,$$

where $\tau$ is the wind stress, $\alpha$ is the thermal expansion coefficient, $Q_{\text{net}}$ is the net surface
heat flux, $\beta$ is the saline contraction coefficient, EP is the net freshwater flux, and $S$ is the
surface salinity [Thomas, 2005; Bachman et al., 2017; Dong et al., 2021]. When EBF and
$B_0$ are positive, which occurs under down-front winds and buoyancy loss due to surface
cooling or evaporation, they tend to reduce the stratification in SML and promote verti-
cal convection. This process generates a smaller $Ri$ and potentially favors a negative PV,
so the flow configurations are preconditioned for SI in the SML. As illustrated in Figure
3, $B_0$ exhibits a strong seasonal variability, with a significant positive flux in the central
and northern Red Sea during winter, and an almost negative flux throughout the basin
during summer (Figures 3d and 3i). The $B_0$, therefore, influences the stratification and
serves as the dominant forcing to reduce PV in winter. In comparison, EBF is generally
smaller in magnitude, yet it displays significant positive values along the eastern coast and
near the strait of Bab-el-Mandeb. These values imply that the coastal regions in the eastern basin are potential areas where SI is likely to occur. Additionally, the northwesterly winds blowing along the Red Sea axis in the central and northern basins [Viswanadhapalli et al., 2017; Langodan et al., 2017] induce Ekman effects that pump colder water and transport it offshore, which may enhance the horizontal buoyancy gradient and precondition the frontogenesis and MLI, which could also explain the eastern intensification of the submesoscale described in Section 3.1.

The SI spans a band of unstable wavelengths over which the growth rate increases slowly with the wavenumber [Hoskins, 1974; Stamper and Taylor, 2017], yet one can still estimate the lower bound scale of the unstable modes that determines the resolution needed to resolve SI [Dong et al., 2021]. Based on Stone [1966]’s theory, the length scale of SI can be estimated using

$$L_{SI} = \frac{0.45}{f^2} \int_{-H}^{0} |\nabla_b| dz$$

[Dong et al., 2021], where $H$ is the depth where the bulk PV becomes positive, indicating the deepest penetration depth of the unstable SI modes [Haney et al., 2015]. However, due to the underestimation of the negative surface PV discussed above, the estimation of $H$ was not straightforward. As a compromise, the $H$ here is roughly approximated by half of the MLD (by comparing the magnitude in Figure 4 by Dong et al. [2021] and Figure 4 by Dong et al. [2020a]). As depicted in Figures 3c and 3h, the estimated $L_{SI}$ also exhibits a pronounced seasonality with larger scales in winter when the seasonally averaged scale could reach about 1 km, yet much smaller than the length scale of MLI. In winter, it can be inferred that the positive buoyancy flux (Figures 3d and 3e) preconditions the potential negative PV that favors larger SI. In summer when the buoyancy flux is dominated by
negative values (Figures 3i and 3j), the \( L_{SI} \) is latitude dependent with smaller scale in the northern basin, where the seasonality of the SI scale is more noticeable. As a result, to resolve the summertime SI in the Red Sea which has a minimum scale of about 0.2 km, a grid resolution of about 25 m (\( L_{SI}/8, [Dong et al., 2021] \)) is required.

4.3. Energy Cascade

Energetics analysis is an important approach to understanding submesoscales and their interactions with mesoscales. First, the seasonally averaged spectra of the surface KE in the northern and southern Red Sea are illustrated in Figures 4a and 4f, respectively, within a range up to the characteristic width of the basin. The velocity is first interpolated into a rectangular domain in Cartesian coordinates before detrending and multiplying with a two-dimensional Hanning window. The one-dimensional spectrum is then obtained by radially averaging the corresponding two-dimensional spectrum calculated using the fast Fourier transformation (FFT), assuming that signal is isotropic and homogeneous. Selecting a rectangular region artificially imposes the largest length scale that can be studied, yet the box we choose is sufficiently large enough (the width of which is about eight times the local Rossby radius of deformation) to investigate the range of submesoscales in the Red Sea. In the northern basin (Figure 4a), the spectrum of KE in winter exhibits larger amplitudes than that in summer throughout the resolved scale length, yet their slope does not significantly vary between seasons. In the mesoscale range larger than 100 km in wavelength, the slope is close to \( k^{-5/3} \), while in the range between about 20 and 70 km, a steeper slope of about \( k^{-3} \) occurs, where \( k \) is the isotropic horizontal wavenumber. This follows a typical 2-d turbulence characteristic, in which energy is cascaded both upscale and downscale [Vallis, 2006; Rocha et al., 2016]. The corresponding
frequency-wavenumber spectra in the two-dimensional Fourier space (the x and y axes correspond to the spatial wave number $k$ and temporal frequency $\omega$ represented by the wavelength and period, respectively) is displayed in Figure 4 (b). The KE spectral signals are most energetic at mesoscale over a longer period, in particular, the mesoscales larger than 100 km are likely exhibiting a period of longer than two months, whereas the small submesoscales tend to have a much smaller period of a few days. In the southern Red Sea (Figure 4f), the KE spectra exhibit a comparable magnitude within ranges smaller than 100 km in both winter and summer (similar to the northern Red Sea with a $k^{-3}$ slope down to about 10 km). However, the summer spectrum exhibits greater strength at the mesoscale range (larger than 150 km in wavelength), which may be linked to the wind-driven eddies and the Red Sea surface outflow [Zhai and Bower, 2013; Zhan et al., 2018]. The two-dimensional frequency-wavenumber spectra in the southern basin (Figure 4g) displays a pattern similar to that in the northern basin, but the KE across all analyzed frequency scales fluctuates more with larger intraseasonal oscillations as indicated by the contour spikes at the corresponding periods.

To further assess the interaction between the submesoscale and mesoscale processes, we investigated two key terms of the KE budget in spectral space. The KE budget spectrum can be represented by the equation

$$\frac{\partial \text{KE}(k)}{\partial t} = \Pi_{\text{Adv}}(k) + \Pi_{\text{PK}}(k) + \text{Res}(k),$$

[Scott and Wang, 2005], where the advective KE flux

$$\Pi_{\text{Adv}}(k) = -\int_k^k \text{Re} \left( \hat{u}^* \cdot (u \cdot \nabla_H u) \right)(k) dk$$
is estimated as the integral of the local horizontal advective term in the KE equation from the wave number $k$ to the largest wave number $k_s$ corresponding to the grid size, with $u$ denoting the horizontal velocity vector, $\hat{()}$ and $()^*$ representing the FFT and complex conjugate, respectively, and $\text{Re}()$ indicating the real part [Scott and Arbic, 2007; Klein et al., 2008; Schubert et al., 2020]. The advective spectral flux underlies the energy cascade and transfers energy conservatively between different scales, and a positive (negative) value corresponds to a direct (inverse) KE cascade. Similarly, the spectral KE production from the vertical buoyancy flux can be estimated as

$$\Pi_{PK}(k) = \int_{k}^{k_s} \text{Re}(\hat{w^*b})(k)dk,$$

accounting for the conversion from APE to PE, of which a positive (negative) value corresponds to a source (sink) of KE [Scott and Arbic, 2007; Klein et al., 2008]. These terms are then vertically integrated over the upper 50 m for the following analysis. The snapshots at an instant time are erratic, therefore, statistically significant signals are uncovered by temporally averaging the spectral results over a 4-year simulation.

The spectral KE flux exhibits a pronounced seasonal variability. In the northern Red Sea, the advective KE flux is positive over a wavelength lower than 70 km in winter, transferring KE toward smaller scales. This forward cascade changes to a significant inverse KE cascade towards larger scales across the wavelength ranging from 70 to 200 km with the peak at about 150 km (solid blue line in Figure 4c). This outcome highlights the significant dynamical influence that KE produces at submesoscales to mesoscales. The inverse cascade is strongly intensified at the surface and exhibits a belt-like structure in depth (Figure 4d), with the deeper layers exhibiting a direct cascade at mesoscale ranges that projects the KE back toward smaller scales. The flux divergence with respect to
$k$ can be further estimated as $\partial \Pi_{KE}/\partial k$. Within the range of wavelength between 50 to 150 km where $\Pi_{KE}$ has a positive slope to $k$, the flux is divergent. This implies the KE produced within this scale range is actively transferred across scales (either upscale or downscale depending on the sign), and the most active KE cascade occurs with a characteristic wavelength of about 75 km where the slope of $\Pi_{KE}$ is maximized. This scale coincides with the approximated wavelength of MLI estimated by $2\pi Rd$ [Zhan et al., 2011; Dong et al., 2020a], which provides evidence that MLI extracts the background APE and transfers KE toward larger scales via an inverse cascade, and to smaller scales via a weaker direct cascade.

Moreover, the buoyancy flux term is positive (dotted blue line in Figure 4c), implying a continuous downscale cascade of APE across the scales that is intensified in the subsurface (Figure 4e). The enhancement in $\Pi_{PK}$ at the wavelength toward 200 km is associated with the generation of mesoscale eddies in the Red Sea, as has been reported by Zhan et al. [2016]. In general, the APE in the mixed layer is converted to submesoscale KE through MLI, and is subsequently transferred back to mesoscale KE via an inverse cascade that also energizes the thermocline below, providing an efficient pathway of the energy cycle in the Red Sea. During summer, both $\Pi_{Adv}$ and $\Pi_{PK}$ have small positive values over the mesoscale range yet are almost zero at smaller scales (red lines in Figure 4c), suggesting a much weaker forward cascade compared to that in winter due to the limited baroclinic production under a well-stratified structure [Zhan et al., 2016].

The southern Red Sea also exhibits distinct seasonality in the spectral KE fluxes (Figure 4h). The $\Pi_{Adv}$ is not significant at smaller scales, but it features an inverse cascade in winter (negative values, solid blue line) and a direct cascade in summer (positive values,
solid red line) at scales larger than 50 km in wavelength. The $Π_{Adv}$ is comparable to $Π_{PK}$ in winter, almost balancing out the KE production in the upper 50 m. It intensified at the surface (Figure 4i), yet is about three times weaker compared to that in the northern basin. Further, the direct cascade in summer is even larger than the KE production, implying an additional KE source at scales of about 150 km in wavelength. Again, this result coincides with the presence of wind-driven eddies in summer due to strong Tokar wind jets [Zhai and Bower, 2013; Zhan et al., 2018]. The residual term $Res(k)$ covers other effects including wind-work, internal pressure work, and dissipation, which are reported to be at least one order less important than the two dominant terms in the energy budget [Dong et al., 2020b] and therefore is not discussed here.

5. Stirring by Submesoscales

Quantifying horizontal stirring and dispersal features in the ocean surface is critical for understanding a wide range of problems, such as the spatial distribution of temperature and biogeochemical tracers (e.g., chlorophyll, larval, nutrients, etc.) [Mezi et al., 2010; Huhn et al., 2012; Olascoaga et al., 2013; Gough et al., 2019]. Eulerian statistics cannot describe the behavior of Lagrangian tracers, yet the Lagrangian coherent structures (LCSs) may provide fundamental information about stirring and transport properties of the flow and thus can be considered as good indicators for the diagnostics [Boffetta et al., 2001; Haller, 2015; Duran et al., 2018; Gouveia et al., 2021]. The LCSs can be used to investigate variations in stirring, such as different dispersals of tracer patches (e.g., growth of phytoplankton patches) and spatial heterogeneities of tracers [Neufeld et al., 2000; Abraham and Bowen, 2002; Martin, 2003]. To investigate how submesoscales contribute to these effects, we compute the forward-time finite-time Lyapunov exponent (FTLE) for a
10-day period, which is long enough to characterize the effects of submesoscale LCSs in the Red Sea.

The forward FTLE reveals regions with different stirring features, and an extreme value indicates a location where two nearby particles rapidly separate. Their snapshots are characterized by narrow filaments of high values intermingled with coherent regions of low values. As depicted by the snapshots displayed in Figure 5a and 5b, typically, an eddy core has low values of FTLEs (i.e., low dispersion rates), whereas large values are found near the eddy edges, where the stretching of the fluid parcels is particularly important. At the edge of the mesoscale eddies, a dense tangle of line intersections appears, relating variations in the stirring and particle dispersion to the coherent vortex structures. The FTLE field can be considered as concentration of passively advected passive tracers \cite{Beron-Vera and Olascoaga, 2009}. Rather than exhibiting excessive small-scale wiggling, the FTLE exhibits long thin filaments in typically large smooth patterns. These organized structures are characteristic of chaotic stirring, which seems dominating over turbulent mixing on the time scales of interest. The winter FTLE is more vigorous and markedly different from the summer case when strong submesoscale motions are absent. The seasonally averaged distribution of the FTLE (Figures 6a and 6b) suggests that the submesoscales-induced strain is highly varied in space and that the most active dispersion occurs in the northern Red Sea during winter, in particular offshore along the eastern coastline corresponding to the regions with high $|Ro|$ (Figure 2a) and positive EBF (Figure 3e). The eastern intensification suggests an asymmetric dispersion conditions in the Red Sea with more active stirring in the eastern basin. In summer, larger values are observed at the edge of the wind-driven eddy in the southern basin, as well as the offshore coastal regions along
both sides. These variations have possible implications for understanding the distribution of biogeochemical tracers and for predicting potential oil spill trajectories in the Red Sea.

The FTLE is a material quantity, it behaves as a passive tracer. The spectrum of the variance of passive tracer concentration follows

\[ T(k) \sim k^{-1} \left[ \int_{k_i}^{k} \xi^2 E(\xi) d\xi \right]^{-1/2}, \]

where \( k \) and \( k_i \) are the total and energy containing horizontal wavenumber, respectively, and \( E(k) \sim k^{-\alpha} \) [Kraichnan, 1971]. The turbulent mixing can be identified with local dynamics, in which both velocity and tracer variance dissipate at similar length scales \((T(k) \sim k^{(\alpha-3)/2-1} \text{ when } 1 < \alpha < 3)\); while chaotic stirring refers to nonlocal dynamics, in which tracer evolution is governed by velocity features at a larger scale \((T(k)\) is flatter than the local case when \( \alpha > 3 \)) [Vallis, 2006; Beron-Vera and Olascoaga, 2009]. The spectrum of the FTLE show power laws of a typical 2-d turbulence prediction, roughly adhering to the forms with a slope of \( k^{-1} \) in the range below 50 km toward the resolvable grid scale, and a slope of \( k^{-5/3} \) toward the mesoscale (Figure 5c). This diagnostic reaffirms that passive tracer evolution on the Red Sea surface is governed by spectrally nonlocal dynamics, consistent with chaotic stirring but not with turbulent mixing on typical submesoscale time frames. (Note that the contribution by smaller submesoscales generated through SI is not included here, which generally extracts KE from geostrophic fronts and transfers it to smaller scales that tends to strengthen local turbulent mixing [McWilliams, 2019].) We then further analyze the FTLE along the central basin axis using spatial wavelet spectra (Figures 6c and 6d) to reveal and distinguish the spatial scale of FTLE. Stronger stirring is favorable in the central and northern Red Sea (north of 21 °N) in winter within wavelength ranges of less than 50 km, suggesting active submesoscale processes that contribute to the
intensity of the FTLE field. This effect is reflected by the conspicuously close correlation (0.85) between the temporal evolution of the basin-averaged FTLE and the submesoscale KE (KE = 0.5(u'^2 + v'^2)), Figure 6e), exhibiting a higher (lower) rate of dispersion during winter (summer).

Mesoscale circulation has been reported to play a dominant role in material transport in the Red Sea [Zhan et al., 2015; Raitos et al., 2017; Wang et al., 2019; Kheireddine et al., 2020; Zhan et al., 2021]. The submesoscale motions emerging from frontogenesis and MLI coincide spatially with the mesoscale frontal dynamics. They may add a significant range of structures to the mesoscale LCS field, where submesoscale LCSs are interwined or intersected with those of the mesoscales (Figure 5). This raises the possibility that submesoscale LCS exerts mesoscale transport barriers [Beron-Vera et al., 2019]. However, this effect is a topic for future research. Beron-Vera and LaCasce [2016] analyzed drifters designed to sample submesoscale processes and synthetic trajectories from a 1 – km resolution model, and concluded that stirring was nonlocal in the latter (model) case and inconclusive in the former (observations). Future work should continue to investigate high-resolution Lagrangian observations and high-resolution simulations to understand this critical aspect of the ocean transport.

6. Summary

The submesoscale processes in the upper Red Sea are investigated using a 1 – km MITgcm simulation and the associated diagnostic analyses, which is however not possible with a coarser-resolution product from altimetry dataset or global reanalysis. To the best of our knowledge, this is the first systematic study of submesoscale dynamics in a typical marginal sea. The submesoscales act as vital dynamics in the Red Sea circulation, and...
we address the necessity of using high-resolution models to at least partially resolve the submesoscales when studying the circulation dynamics in other regions of similar scales. The Red Sea submesoscales exhibit a pronounced seasonal variability that peaks in winter, with favorable occurrence in the central and northern basin, and an intensification toward the offshore regions along the eastern coast. Three generation mechanisms, including frontogenesis, MLI and SI, and their characteristics, are quantitatively discussed. The frontogenesis and MLI enhanced by strain, occur where a favorably aligned horizontal buoyancy gradient in a background deformation flow has a rapidly growing magnitude on the surface and upper layers. These processes are significantly strengthened in winter, particularly in the central and northern basin. Due to the small scale, a vast majority of the SI is unresolved by the present model, thus the associated forward cascade and mixing in the submesoscale may stem from the real ocean versus the simulated scenario. However, the SI characteristics are diagnosed from associated preconditions. The scale of MLI and SI is generally larger in winter and smaller in summer. Consequently, the grid resolution to resolve the seasonality is set by the smallest possible MLI and SI scales in summer, at about 8 km and 0.2 km for in the Red Sea, respectively.

The spectral KE budget analysis reveals a significant seasonal modulation of mesoscale dynamics through large-scale atmospheric forcing, via the deepening of the SML during winter, generating KE production at the submesoscale and an inverse KE cascade toward the mesoscale. In the real ocean, the convective and mechanical mixing driven by atmospheric forcing in winter is opposed by the restratification processes of the submesoscales. An integrated numerical study at different resolutions (from eddy-resolving to submesoscale-permitting) is beyond the scope of this study; however, it has been reported...
that when submesoscale starts to be resolved with a higher resolution, the mixed layer
formed during the surface cooling is significantly shallower [Couvelard et al., 2015; Whitt
and Taylor, 2017; Chrysagi et al., 2021], which is consistent with the submesoscale upward
heat flux as demonstrated in this study. We also address the role of the submesoscale cir-
culation in shaping surface stirring and dispersal effects in the Red Sea, which are highly
non-uniform in space and in time, adding a range of structures to the horizontal transport
induced by the mesoscale eddies. In this model, 2-d turbulence is ultimately the main
driver of Lagrangian dispersion, however, future observational and modeling experiments
are needed to fully understand, in Lagrangian terms, the effect of submesoscale structures
and their interaction with mesoscale structures. Moreover, the submesoscales largely de-
pend on APE, (i.e., thermal wind shear of fronts within the mixed layer). As they are
energized throughout the mixed layer, they could also make vital contributions to vertical
heat and material transports. The strong horizontal stirring and the associated vertical
motion (usually associated with the entrainment and detrainment of water masses) could
have vital biological consequences in the Red Sea, for instance, in the connectivity among
different coral complexes and the enhanced vertical transport of nutrients from the mixed
layer base, which deserves further study in the future.

Appendix A: Mixed Layer Depth

The mixed layer depth (MLD) is a diagnostic output variable of the MITgcm model,
computed according to the density-based criterion as the depth at which the water density
is equal to the density of water that is 0.8 °C colder than the surface water.
Acknowledgments. The research reported in this manuscript was supported by the Virtual Red Sea Initiative (REP/1/3268-01-01) sponsored by the Office of Sponsored Research (OSR) at King Abdullah University of Science and Technology (KAUST), the Saudi Aramco Marine Environment Research Center at KAUST, and the National Natural Science Foundation of China (42176023). Model runs were conducted on the KAUST supercomputing facilities. The data supporting the presented analysis can be found at Figshare (https://figshare.com/articles/dataset/RS_submesoscale/16597997).
References


Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. Beljaars, L. van de Berg,


Duran, R., F. J. Beron-Vera, and M. J. Olascoaga (2018), Extracting quasi-steady lagrangian transport patterns from the ocean circulation: An application to the gulf of


Gouveia, M. B., R. Duran, J. A. Lorenzzetti, A. T. Assireu, R. Toste, L. P. de, and D. F. Gherardi (2021), Persistent meanders and eddies lead to quasi-steady la-


Papadopoulos, V. P., P. Zhan, S. S. Sofianos, D. E. Raitsos, M. Qurban, Y. Abualnaja, A. Bower, H. Kontoyiannis, A. Pavlidou, M. A. T.T., N. Zarokanellos, and I. Hoteit (2015), Factors governing the deep ventilation of the red sea, *Journal of Geophysical Research: Oceans*, 201, doi:10.1002/2015JC010996, a variety of data based on hydrographic measurements, satellite observations, reanalysis databases, and meteorological observations are used to explore the interannual variability and factors governing the deep water formation in the northern Red Sea. Historical and recent hydrographic data consistently indicate that the ventilation of the near-bottom layer in the Red Sea is a robust feature of the thermohaline circulation. Dense water capable to reach the bottom
layers of the Red Sea can be regularly produced mostly inside the Gulfs of Aqaba and Suez. Occasionally, during colder than usual winters, deep water formation may also take place over coastal areas in the northernmost end of the open Red Sea just outside the Gulfs of Aqaba and Suez. However, the origin as well as the amount of deep waters exhibit considerable interannual variability depending not only on atmospheric forcing but also on the water circulation over the northern Red Sea. Analysis of several recent winters shows that the strength of the cyclonic gyre prevailing in the northernmost part of the basin can effectively influence the sea surface temperature (SST) and intensify or moderate the winter surface cooling. Upwelling associated with periods of persistent gyre circulation lowers the SST over the northernmost part of the Red Sea and can produce colder than normal winter SST even without extreme heat loss by the sea surface. In addition, the occasional persistence of the cyclonic gyre feeds the surface layers of the northern Red Sea with nutrients, considerably increasing the phytoplankton biomass.


Wang, S., Z. Jing, H. Liu, and L. Wu (2018), Spatial and seasonal variations of submesoscale eddies in the eastern tropical pacific ocean, *Journal of Physical Oceanography,*


Zhan, P., F. Yao, A. R. Kartadikaria, Y. Viswanadhapalli, G. Gopalakrishnan, and I. Hoteit (2015), Far-field ocean conditions and concentrate discharges modeling along the saudi coast of the red sea, doi:10.1007/978-3-319-13203-7{21}.


Figure 1. Chlorophyll concentrations in the Red Sea observed using (a) the Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua) on 7 February 2015, revised from Zarokanellos and Jones [2021], (b) MODIS-Aqua on 29 March 2010, revised from Raitos et al. [2017], (c) Sea-viewing Wide Field-of-View Sensor (SeaWiFS) on 16 March 2000, revised from NASA earth observatory (https://earthobservatory.nasa.gov/images/8039/chlorophyll-and-currents-in-the-red-sea). (d-h) Snapshots of $\rho$, $|Ro|$, $w$, frontogenesis tendency $F$ and $q_{Ertel}$ and (i-m) cross-section indicated by the red line in (d) on a typical winter day (9 January 2018), where the solid line represents the mixed layer depth. (n-w) Same as above but for a typical summer day (6 September 2016).
Figure 2. Seasonally averaged distribution of $|Ro|$, surface net heat flux $Q_{\text{net}}$, mixed layer depth MLD, and vertical velocity $|w|$ and vertical heat flux VHF at a 50–m depth, in (a-e) winter and (f-j) summer. The spatially averaged time series of $|Ro|$ and $F$ (k), and BC, BT and MLD (l).
Figure 3. Seasonally averaged distribution of the Rossby radius of deformation in the mixed layer $Rd_{ML}$, linear growth rate, scale length of SI $L_{SI}$, buoyancy flux $B_0$, and Ekman buoyancy flux EBF in (a-e) winter and (f-j) summer.
Figure 4. (a, f) Spectra of the surface kinetic energy (KE) in winter and summer; (b, g) frequency-wavenumber spectra of the surface KE; (c, h) spectral flux of KE: advective flux $\Pi_{Adv}$ with solid line and production $\Pi_{PK}$ with dotted line in winter (blue) and summer (red); (d, i) the vertical distribution of $\Pi_{Adv}$ in winter; (e, j) the vertical distribution of $\Pi_{PK}$ in winter. The upper (lower) panel displays the results of a typical rectangular region of the northern (southern) Red Sea as indicated by the box area in (a) and in (f).
Figure 5. Snapshots of the forward FTLE in the northern Red Sea on (a) a typical winter day (9 January 2018) and (b) a summer day (6 September 2016). (c) The FTLE spectra as a function of wavelength.
Figure 6. Surface FTLEs averaged for (a) winter and (b) summer, seasonal average of the spatial wavelet scalogram of the surface FTLE along the central axis of the Red Sea indicated by the dashed line in (a) for (c) winter and (d) summer. (d) Time series of FTLE and submesoscale EKE (EKE obtained using a high-pass filter with a cutoff scale of 50 km).