

Enhancing Ocean Model Performance with Wave-Informed, Neural Network-Enhanced *K*-Profile Parameterization for Vertical Mixing: A Test Case in the Gulf of Mexico

JIANGUO YUAN,^{a,b} JUN-HONG LIANG,^{a,b,c,d} ERIC P. CHASSIGNET,^{c,d} ALEXANDRA BOZEC,^c AND ALAN J. WALLCRAFT^c

^a *Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana*

^b *Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana*

^c *Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida*

^d *Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida*

(Manuscript received 12 September 2025, in final form 20 March 2026, accepted 27 March 2026)

ABSTRACT: Turbulent mixing in the ocean surface boundary layer is unresolved in ocean models and is typically parameterized using physics-based formulas with coefficients estimated empirically. In this study, deep neural networks (DNNs) are employed to learn turbulent mixing characteristics in the ocean surface boundary layer from turbulence-resolving large-eddy simulations (LES) driven by diverse wind, wave, and buoyancy conditions over the Gulf of Mexico (GOM). The extracted knowledge is used to constrain parameters in the *K*-profile parameterization (KPP) scheme, which is widely adopted in ocean models. The DNN-informed KPP (KPP_DNN) is improved and implemented into the Hybrid Coordinate Ocean Model (HYCOM) to simulate the evolution of oceanic properties in the GOM, and its performance is evaluated against both observations and HYCOM simulations using the original KPP. Results indicate that the KPP_DNN improves sea surface temperature (SST) simulations, particularly in winter, effectively mitigating the warm SST biases by up to 70% in some west GOM regions compared to simulations using the original KPP. Additionally, it better captures the distribution of mixed layer depth (MLD), especially wintertime MLD events shallower than 50 m. Sensitivity experiments demonstrate that including ocean surface waves in the DNN input is critical, and omitting them degrades model performance.

KEYWORDS: Deep learning; Ocean models; Oceanic mixed layer; Vertical motion; Parameterization

1. Introduction

The ocean surface boundary layer (OSBL), typically spanning a few tens to hundreds of meters below the ocean surface, is a critical ocean layer dominated by various turbulent mixing processes such as wind-driven shear, buoyancy forcing, and wave-induced Langmuir turbulence (Belcher et al. 2012; D'Asaro 2014; Fox-Kemper et al. 2022). These mixing processes play a vital role in regulating air–sea exchanges of heat, momentum, and gases, thereby influencing the climate system and ocean environment (Toole 1998; McPhee 2008). Turbulent mixing in the OSBL also facilitates the entrainment of colder, nutrient-rich water from the ocean interior into the surface layer, affecting not only the thermal structure of the upper ocean but also biological productivity and the marine carbon cycle (e.g., Gargett and Marra 2002; Schmittner et al. 2005; Lévy et al. 2013; Liang et al. 2013; Smith et al. 2018). Due to its small spatial scale and highly dynamic nature, explicitly resolving turbulent mixing in global ocean models remains computationally prohibitive in the foreseeable future (Fox-Kemper et al. 2019). As a result, turbulent mixing in the

OSBL must be represented through subgrid-scale parameterizations, and the fidelity of these schemes strongly influences the accuracy of ocean model simulations (e.g., Large et al. 1994; McPhee 1999).

The development of mixing parameterizations often relies on high-fidelity simulations of turbulent mixing under controlled forcing conditions generated by turbulence-resolving models such as large-eddy simulations (LES). While observational data were directly utilized to develop OSBL mixing parameterizations in early studies (e.g., Large et al. 1994), they are impacted by submesoscale, mesoscale, and large-scale processes (Thompson et al. 2016). Though observations are essential, the limited coverage of observations and the difficulty in isolating small-scale turbulence contributions from larger-scale circulations make them less suitable than turbulence-resolving simulations for developing new or improving existing turbulent mixing parameterization schemes. Properties extracted from LES solutions have been widely used to derive or calibrate structure functions in parameterization schemes, such as stability-dependent mixing efficiency functions, vertical shape functions of turbulent diffusivity and viscosity, and flux–gradient relationships (e.g., Skillingstad et al. 1996; Large and Gent 1999; Harcourt and D'Asaro 2008; Noh et al. 2016). These schemes are then typically tested in one-dimensional (1D) ocean models, such as the General Ocean Turbulence Model (GOTM), which serve as efficient testbeds for structural comparisons and sensitivity analyses (He and Chen 2011; George et al. 2025; Wang et al. 2025). Ultimately, parameterization schemes must

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/AIES-D-25-0084.s1>.

Corresponding author: Jun-Hong Liang, junhong.liang@fsu.edu

be integrated into three-dimensional (3D) ocean models, where their numerical stability within a 3D framework can be assessed and their performance in predicting key state variables, such as sea surface temperature (SST) and mixed layer depth (MLD), can be rigorously evaluated against observational datasets across a realistic ocean region under a wide range of forcing, geographic, and hydrographic conditions (e.g., [Li et al. 2016](#); [Ali et al. 2019](#); [Sane et al. 2023, 2026](#)).

A variety of vertical mixing parameterization schemes have been developed to represent turbulent mixing processes in the OSBL in realistic ocean models ([Kantha and Clayson 1994](#); [Large et al. 1994](#); [Umlauf and Burchard 2003](#); [Reichl and Hallberg 2018](#)). However, they remain inaccurate under certain conditions because some physical processes modulating turbulent mixing were not originally included in the parameterizations. One such example is the effect of nonbreaking surface waves and the associated Langmuir turbulence. Many LES simulations (e.g., [Skylningstad et al. 1996](#); [McWilliams et al. 1997](#); [McWilliams and Sullivan 2000](#)) and observations (e.g., [Weller and Price 1988](#); [D'Asaro and Dairiki 1997](#)) have shown that non-breaking wave-induced Langmuir turbulence can intensify turbulent mixing, deepen MLDs, and enhance entrainment fluxes through wave–current interactions that generate Langmuir circulation, thereby enhancing vertical transport even in the absence of wave breaking. However, when winds and nonbreaking waves are misaligned, turbulent mixing can be suppressed (e.g., [Van Roekel et al. 2012](#)). Motivated by these findings, various studies have attempted to incorporate nonbreaking wave effects into baseline parameterization schemes (e.g., [Qiao et al. 2004](#); [Harcourt 2015](#); [Li and Fox-Kemper 2017](#); [Reichl and Li 2019](#); [Wang et al. 2025](#)). Despite differences in implementation, most studies report enhanced mixing, deeper MLDs, and cooler SSTs compared to baseline runs without nonbreaking wave effects (e.g., [Li et al. 2019, 2021](#); [Yuan et al. 2024](#)). In addition, no consensus has been reached on the optimal approach to implementing nonbreaking wave effects ([Li et al. 2019](#)).

Wave-informed K -profile parameterization (KPP) schemes have been incorporated into 3D ocean models and evaluated against both baseline models without nonbreaking wave effects and observational data. The outcomes vary depending on the wave-informed scheme, geographic region, and season. For example, [Li et al. \(2016\)](#) found that incorporating both wave-enhanced velocity scales and unresolved shear leads to improved MLD representation in both summer and winter in the Southern Ocean. [Ali et al. \(2019\)](#) reported that the wave effect tends to have a much stronger influence in winter than in summer across six wave-informed schemes for the North Atlantic Ocean, except for the [Li and Fox-Kemper \(2017\)](#) scheme, which incorporates wave-enhanced unresolved shear and the surface-layer-averaged Langmuir number. It has been shown that some of the schemes tend to systematically overestimate MLDs across much of the global ocean under various conditions (e.g., [Fan and Griffies 2014](#); [Li et al. 2016](#); [Ali et al. 2019](#)). In certain regions and seasons, such as at midlatitudes in the Northern Hemisphere in winter, simulated MLDs may have already been overestimated in baseline models. In such cases, further adding wave enhancement can degrade MLD and SST simulations. Further improving traditional turbulent

mixing parameterization schemes is challenging. In recent years, there has been a growing interest in leveraging machine learning techniques to address this issue, particularly deep neural networks (DNNs). Some studies have used DNNs to directly predict turbulent fluxes (e.g., [Gentine et al. 2018](#); [Rasp et al. 2018](#); [Liang et al. 2022](#); [Iyer et al. 2025](#)), although numerical stability over long integration periods remains a challenge ([Brenowitz et al. 2020](#); [Rasp 2020](#); [Chattopadhyay and Hassanzadeh 2023](#)). Other approaches have used DNNs to predict empirical parameters within well-established schemes, while maintaining the framework of the physically based parameterization structure (e.g., [Reichl and Hallberg 2018](#); [Zhu et al. 2022](#); [Sane et al. 2023](#)).

[Yuan et al. \(2024\)](#) utilized DNNs to predict two key uncertain enhancement coefficients in the KPP scheme [DNN-informed KPP (KPP_DNN)]: the coefficients for the velocity scale and unresolved shear. The trained DNNs were implemented into the 1D GOTM and compared against traditional deterministic models. Although the study by [Yuan et al. \(2024\)](#) was carried out at Ocean Station Papa, the LES training dataset spans a wide range of atmospheric and wave forcing conditions representative of open-ocean regimes. While the DNNs were trained using data from a single geographic location, they were shown to generalize across a broad parameter space of background forcing conditions. Their results show that the GOTM with KPP_DNN is not only computationally efficient but also reduces errors in both SST and MLD. Furthermore, their results revealed more complex yet structured relationships between coefficients and a wide range of physical variables, such as MLD and turbulent Langmuir number, that are not captured by traditional wave-informed deterministic schemes.

This study builds upon and extends [Yuan et al. \(2024\)](#)'s work by adopting their approach of using DNNs to predict the two coefficients in the KPP scheme. Instead of testing the method in a 1D model, the well-tuned DNNs are integrated into the 3D Hybrid Coordinate Ocean Model (HYCOM) to simulate the evolution of temperature, salinity, MLD, and regional-scale circulation in the Gulf of Mexico (GOM). First, the DNN input array structure is modified to better capture spatial heterogeneity across a broader region rather than at a single fixed location. This allows the DNN-based parameterizations to operate across a wide range of forcing regimes, as well as spatial variations related to Coriolis effects and bathymetric variation in the GOM. Second, embedding the DNN-based parameterizations into the fully 3D HYCOM makes it possible to evaluate the numerical stability, physical consistency, and predictive skill of the DNN-informed KPP scheme under realistic oceanic conditions, as well as its interaction with regional-scale circulation and mesoscale dynamics. The model outputs are compared against both observational data and HYCOM simulations using the baseline KPP scheme without wave effects, to evaluate whether the improved KPP_DNN reduces the biases observed in the traditional parameterization and better aligns with observations. While the physical interpretability of the DNN-predicted coefficients has been explicitly examined and validated in the 1D GOTM framework of [Yuan et al. \(2024\)](#), the present work focuses on evaluating their performance and robustness when embedded into a fully 3D regional circulation model.

The remainder of this paper is organized as follows: [Section 2](#) describes the generation of the training data, the configuration of the DNN models, and their implementation into the HYCOM framework. [Section 3](#) presents HYCOM simulation results, along with comparisons to observations from multiple perspectives, including SST and MLD. [Section 4](#) discusses the improvements achieved by HYCOM simulations using DNN over the baseline KPP, as well as the remaining challenges and areas requiring further investigation. Finally, [section 5](#) provides concluding remarks.

2. The neural network–informed KPP for the GOM and its implementation in HYCOM

a. The KPP scheme with enhancement coefficients

This subsection provides an overview of the baseline KPP scheme and the two coefficients introduced to the baseline model to account for nonbreaking wave effects. A more comprehensive description can be found in [Yuan et al. \(2024\)](#).

In ocean models, the turbulent flux of a variable x (e.g., momentum, temperature, salinity, or other scalars) is expressed as

$$\overline{w'x'} = -K_x \left(\frac{\partial \bar{x}}{\partial z} - \gamma_x \right). \quad (1)$$

Here, z is the vertical coordinate and w is the vertical velocity. Overbars indicate ensemble averages, while primes denote turbulent fluctuations. The K_x is diffusivity or viscosity. The γ_x represents the nonlocal transport term, which is not proportional to local gradients. This term typically applies only to scalar fluxes (e.g., temperature and salinity) and is not included for momentum. In this study, only K_x is modified.

In the KPP_{LMD} framework ([Large et al. 1994](#)), K_x is given by

$$K_x(\sigma) = w_x(\sigma)hG_x(\sigma), \quad (2)$$

where h is the OSBL depth and $\sigma = z/h$ is the water depth normalized by the OSBL depth. The w_x is a velocity scale determined by both surface forcing and the Monin–Obukhov similarity theory, and $G_x(\sigma)$ is a nondimensional shape function.

The OSBL depth is diagnosed using the bulk Richardson number:

$$\text{Ri}_b(z) = \frac{z[b_r - \overline{b(z)}]}{[\mathbf{u}_r - \mathbf{u}(z)]^2 + U_t^2(z)}, \quad (3)$$

where b is the buoyancy and $\mathbf{u}$ is the ocean current vector. The subscript r represents the reference value. The effect of unresolved shear is parameterized as

$$U_t^2(z) = \frac{C_v N(z) w_x(z) |z|}{\text{Ri}_c}, \quad (4)$$

where C_v is a dimensionless constant and $N(z)$ is the Brunt–Väisälä frequency.

To account for additional physical processes that are absent (e.g., nonbreaking waves) or poorly represented (e.g., buoyancy stability) in the baseline KPP scheme (KPP_{LMD}), the

KPP_{LMD} scheme is modified by introducing two additional coefficients: the velocity scale coefficient ϵ , which modifies vertical mixing, and the unresolved shear coefficient η , which modifies entrainment at the base of the OSBL. Then, Eqs. (1) and (4) become

$$K_x(\sigma)_{\text{new}} = \epsilon w_x(\sigma) |h| G_x(\sigma), \quad (5)$$

$$U_t^2(z)_{\text{new}} = \eta U_t^2(z)_{\text{LMD}}. \quad (6)$$

Here, the subscript Large–McWilliams–Doney (“LMD”) indicates the terms in the baseline KPP scheme in [Large et al. \(1994\)](#), while the subscript “new” represents terms after modifications.

In previous studies (e.g., [McWilliams and Sullivan 2000](#); [Reichl et al. 2016](#); [Li and Fox-Kemper 2017](#)), the two coefficients were modeled using semiempirical equations tuned with LES solutions of a limited set of forcing conditions and were only used to quantify effects due to nonbreaking waves. In contrast, in [Yuan et al. \(2024\)](#) and in this study, these parameters are predicted using DNNs trained on LES data, driven by a much broader spectrum of wind, nonbreaking wave, and buoyancy conditions in the real ocean.

b. Data generation using large-eddy simulations

The DNNs are trained using solutions generated by the National Center for Atmospheric Research large-eddy simulation (NCAR LES) model for the OSBL ([Sullivan and McWilliams 2010](#)). The NCAR LES model has been applied to study mixing processes under various external atmospheric forcing (e.g., [Kukulka et al. 2009](#); [Liang et al. 2017](#)), turbulent mixing and entrainment enhancement due to breaking and nonbreaking waves (e.g., [McWilliams et al. 2012](#)), mixing and entrainment associated with a frontal zone and submesoscale instabilities (e.g., [Hamlington et al. 2014](#); [Yuan and Liang 2021](#)), tracer dispersion and transport driven by turbulence (e.g., [Liang et al. 2018](#); [Kukulka 2020](#)), and ocean response to extreme forcing such as hurricanes (e.g., [Rabe et al. 2015](#); [Liang et al. 2020](#)). Turbulence-resolving models such as the NCAR LES model offer an economical approach to generating extensive training datasets for machine learning–based parameterizations.

In this study, a suite of NCAR LES simulations are conducted at various locations in the GOM and adjacent regions ([Fig. 1](#)) to generate turbulence-resolving solutions for DNN training. Over 700 locations within the domain ([Fig. 1](#)) are selected. To ensure that our LES ensemble captures turbulent mixing under a wide range of meteorological, water column, and geographic conditions across the GOM, the study area is divided into two subdomains: one coastal subdomain shallower than 300 m and one deep-sea subdomain deeper than 300 m. Sixty percent of the selected sites are located in the coastal subdomain, while the remaining 40% are in the deep-sea subdomain. In each subdomain, the exact locations for LES simulations are randomly selected. A higher proportion of sites is intentionally allocated to the coastal subdomain to better cover the effects of water depth on turbulence. At each location, a 5-day LES simulation is performed, with the first day of each simulation treated as spinups and discarded, while solutions from days 2 to 5 are processed for DNN training.

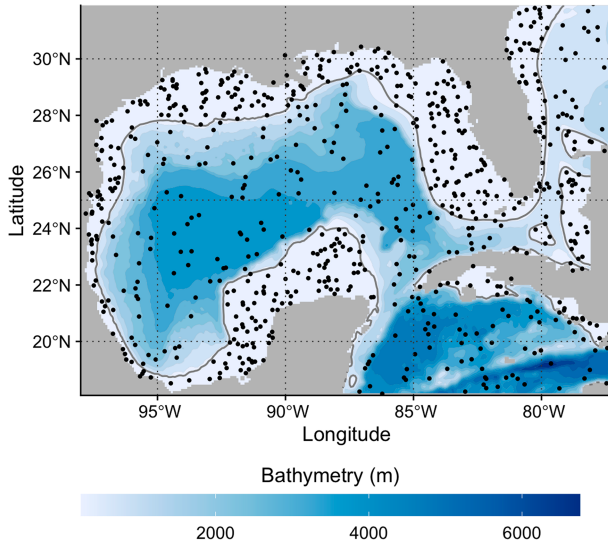


FIG. 1. Map of GOM study region. The blue shading indicates bathymetry, with gray contour lines at 300 m. Black dots are the selected locations where LES are conducted. The land mask is shown in gray.

All LES simulations are forced by atmospheric and oceanic conditions derived from the ERA5 reanalysis (Hersbach et al. 2020), including surface winds, net shortwave radiation, net heat flux excluding shortwave components, rates of evaporation minus precipitation, and Stokes drift vector profiles computed from wave spectra. Initial temperature and salinity profiles are extracted from the 41-layer HYCOM + NCODA Global 1/12° Reanalysis (GOFS 3.1). Both initial conditions and forcing fields are interpolated to the location and time window of each LES run, with forcing fields provided at an hourly resolution.

While the wind stress, net heat flux excluding shortwave radiation, and rates of evaporation minus precipitation are applied only at the ocean surface, shortwave radiation penetrates the water column. The Jerlov water type 1 (Jerlov 1976) is used to calculate the shortwave radiation at depth z , given by

$$q_{sw}(z) = q_{sw}(0)[re^{z/\mu_1} + (1-r)e^{z/\mu_2}], \quad (7)$$

where $q_{sw}(0)$ is the net shortwave radiation at the surface and the constants $r = 0.58$, $\mu_1 = 0.35$, and $\mu_2 = 23$ correspond to Jerlov water type 1. In the HYCOM, to avoid unrealistic heating at the ocean floor in shallow coastal regions, HYCOM redistributes the bottom-reaching shortwave radiation throughout the water column. We apply the same adjustment in LES. The modified shortwave radiation profile $Q_{sw}(z)$ then becomes

$$Q_{sw}(z) = \begin{cases} 0, & z = -H, \\ q_{sw}(z) - q_{sw}(H)|z|/H, & z > -H, \end{cases} \quad (8)$$

with $q_{sw}(H)$ being the shortwave radiation flux on the seafloor using equation (7). This ensures that the shortwave radiation at the seafloor is always zero, while preserving the net shortwave radiation value at the ocean surface. The adjustment reduces $Q_{sw}(z)$ relative to the original $q_{sw}(z)$, with larger differences in

shallower waters. In deep waters, the impact of this adjustment is negligible.

Training data were generated by horizontally averaging the three-dimensional LES solutions to obtain one-dimensional profiles. These 1D profiles do not resolve turbulence but incorporate turbulent effects. They are, therefore, representative of the true profiles with turbulent effects in operational models such as the HYCOM where turbulent mixing in the OSBL is parameterized.

c. Data standardization and DNN training

Feedforward deep neural networks are employed in this study. The inputs include profiles of the squared Brunt–Väisälä frequency, Stokes drift profile in the zonal and meridional directions, zonal and meridional wind components, net shortwave radiation, net heat flux excluding shortwave components, rates of evaporation minus precipitation, latitude, OSBL depth, and bathymetry. A simplified illustration of the DNN architecture used in this work is shown in Fig. 2. To assess the role of Stokes drift, an additional set of DNNs are trained with the same inputs but without Stokes drift profiles, to highlight the importance of including Stokes drift information for achieving sufficient predictive accuracy. The DNN-based KPP schemes with and without Stokes drift profiles are referred to as $KPP_{DNN,WV}$ and $KPP_{DNN,NW}$, respectively, while KPP_{DNN} collectively refers to both DNN-based schemes.

Compared with Yuan et al. (2024), the DNN models in this study introduce three additional variables into the input array: the latitude to indicate Coriolis effects, the OSBL depth to indicate the absolute depth range of input profile variables, and the bathymetry depth to indicate the data sparsity passed from HYCOM solutions to DNN models. It also replaces temperature and salinity profiles with profiles of the Brunt–Väisälä frequency squared, to avoid potential out-of-distribution inputs when temperature and salinity fields from long-term 3D ocean simulations locally exceed the range represented in the LES-based training dataset.

Three DNN models are trained to predict two scalar coefficients: Models D_{η_0} and D_{η} predict the coefficient to the unresolved shear contribution η , while the model D_{ϵ} predicts the velocity scale coefficient ϵ in the KPP scheme. In shallow-water regions, h often reaches the seafloor even when $\eta = 0$, resulting in a high occurrence of $\eta = 0$ in the data. To better capture this, we adopt a two-step prediction strategy: Model D_{η_0} first predicts whether $\eta = 0$, and if not, D_{η} predicts the actual non-zero value of η . For predicting η , h from the previous time step is included as an input. When predicting ϵ in model D_{ϵ} , h at the current time step is used because ϵ depends directly on h at the same time step.

All input variables are normalized before being fed into DNNs. Each input variable is standardized by first subtracting its mean and then being normalized by its standard deviation. For the profile variables (e.g., the Brunt–Väisälä frequency squared, Stokes drift profiles in x and y directions), the data are interpolated to 33 evenly distributed vertical levels from the surface to h . For scalar input variables, an additional sigmoid transformation $f(x) = 1/(1 + e^{-x})$ is then applied to the normalized value. This use of the sigmoid function is heuristic, introduced

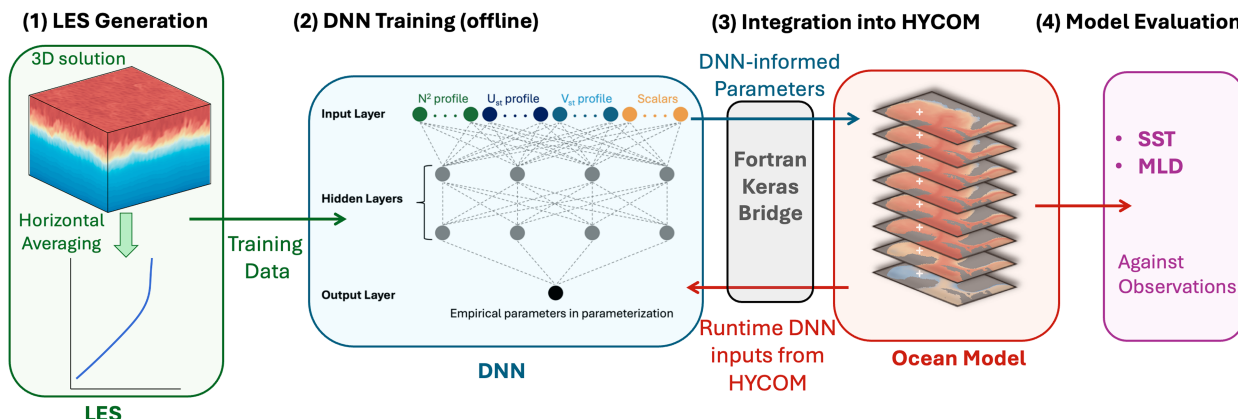


FIG. 2. Conceptual framework of the modeling system. (1) Turbulence-resolving LES simulations are performed and horizontally averaged to generate training data. (2) DNNs are trained offline using LES-derived profiles and scalar inputs. (3) The trained DNNs are integrated into the 3D HYCOM through an FKB to provide DNN-informed parameter values during runtime. (4) The resulting ocean simulations are evaluated against observations using diagnostics such as SST and MLD.

after we observed that it improves prediction accuracy during model training. For the outputs, η in D_η and ϵ in D_ϵ are transformed using a base-10 logarithm, i.e., $\eta_{\text{norm}} = \log_{10}\eta$ and $\epsilon_{\text{norm}} = \log_{10}\epsilon$. To remove extreme and potentially nonphysical samples, data points with η or ϵ larger than 10 (less than 5% of data samples) are excluded from the training dataset. No additional outlier filtering is applied.

Data are randomly separated into training, validation, and testing datasets at a ratio of approximately 6:2:2. To avoid spurious correlations and ensure that the datasets reflect distinct climatological conditions, we follow a block sampling strategy (Schultz et al. 2021), separating datasets by LES simulations rather than by individual samples.

To identify the best-performing models, we evaluate architectures ranging from 1 to 10 hidden layers and from 4 to 128 neurons in each hidden layer. Each model is trained for 1000 epochs, with a batch size of 4096. A leaky ReLU activation, $f(x) = \max(x, \alpha x)$ with $\alpha = 0.1$, is applied in all hidden layers, and Adam is chosen as the optimizer. Model weights are initialized using random normal noise with a standard deviation of 0.05, while all biases are initialized to zero. The learning rate is set to be 0.001 for all DNN models initially, which is reduced by a factor of 0.1 if no improvement in validation loss is observed over five consecutive epochs. The binary cross-entropy is used as the loss function in D_{η_0} , while the mean absolute error is used in D_η and D_ϵ . The DNN models are trained and tested using TensorFlow and Keras in R. Models with the lowest validation losses (for D_η and D_ϵ) and the highest classification accuracy (for D_{η_0}) are selected for implementation in the HYCOM. Configurations of the best-trained DNN models are shown in Table 1. DNN models smaller than the best trained models are not complex enough to capture the true range of the label values, while DNN models more complex than the best-trained models typically result in overfitting. Model performance, including training and validation losses over epochs, and the comparison between predicted and target value distributions, are shown in Figs. S1 and S2 in the online supplemental material. Offline tests confirm

that the trained models reliably capture the essential data distributions, justifying their implementation into ocean models.

In $\text{KPP}_{\text{DNN,WV}}$, Stokes drift is included as an additional input to explicitly represent wave-induced Lagrangian transport and its vertical shear, which directly contribute to Langmuir turbulence through the Craik–Leibovich mechanism. This information cannot be uniquely reconstructed from wind stress or bulk wave parameters alone, as it depends on the full wave spectrum and its depth-dependent decay. By contrast, $\text{KPP}_{\text{DNN,NW}}$ relies solely on nonwave inputs, allowing a controlled assessment of the added value of Stokes drift in the DNN-enhanced mixing parameterization.

d. The implementation of DNN-enhanced KPP in HYCOM

The HYCOM (Bleck 2002; Chassignet et al. 2003, 2009) is a 3D ocean model designed to simulate the ocean’s general circulation across a wide range of spatial and temporal scales. It employs a hybrid vertical coordinate system that combines terrain-following (sigma), isopycnic (density-following), and fixed-depth (z-level) layers, allowing it to flexibly represent both shallow coastal regions and the deep open ocean. HYCOM is

TABLE 1. Configurations of the best-trained DNNs used in KPP_{DNN} schemes. The $\text{KPP}_{\text{DNN,NW}}$ excludes Stokes drift, while $\text{KPP}_{\text{DNN,WV}}$ includes it. Model D_{η_0} classifies whether the unresolved shear coefficient η equals 0. If not, model D_η is activated to predict its value. Model D_ϵ predicts the velocity scale coefficient ϵ .

Scheme	Model	Hidden layers	Neurons in each hidden layer
$\text{KPP}_{\text{DNN,NW}}$	D_{η_0}	2	16
	D_η	5	8
	D_ϵ	5	24
$\text{KPP}_{\text{DNN,WV}}$	D_{η_0}	5	4
	D_η	3	16
	D_ϵ	3	12

particularly well suited for regional and global ocean forecasting applications, as well as climate simulations. Its vertical layering approach enables good representation of stratification and vertical mixing processes, making it a powerful tool for studying upper ocean dynamics and their interactions with larger-scale circulation.

This study modifies the KPP-related subroutines in HYCOM to incorporate well-tuned DNNs for predicting η and ϵ . The Fortran-Keras Bridge (FKB; Ott et al. 2020) is used to utilize weights and biases from trained DNNs to reconstruct the network structures in HYCOM. The incorporation process involves the following steps. First, the trained weights stored in Hierarchical Data Format, version 5 (HDF5), format files are converted into specifically organized ASCII files. These files serve as inputs to FKB, allowing it to reconstruct the DNN architecture and weights within the Fortran environment. Input arrays required by the DNNs, such as model state variables, various surface forcing fields, latitudes, OSBL depths, and water depths, are read from HYCOM, organized, and normalized according to the format expected by the DNNs. The DNN-predicted coefficients are then denormalized and incorporated into the KPP scheme within HYCOM. A conceptual framework of information exchange between HYCOM and DNNs through FKB is shown in Fig. 2.

e. HYCOM configuration

The HYCOM domain, depicted in Fig. 1, spans from 18° to 32°N in latitude and from 77° to 98°W in longitude. This domain covers the entire GOM, the northwestern Caribbean Sea, and parts of the Atlantic Ocean off the southeastern U.S. coast. Bathymetry within the domain is derived from the 2019 version of the General Bathymetric Chart of the Oceans (GEBCO Bathymetric Compilation Group 2019).

The model uses a Mercator grid with 293×193 grid points in the zonal and meridional directions, respectively, and the horizontal resolution of the model is $1/12^\circ$. Vertically, the model is configured with 41 hybrid layers. The upper 83 m of the water column consists of 14 sigma Z layers, which are terrain-following where the seafloor depth is less than 83 m. In water deeper than 83 m, layers 15–41 are isopycnal unless they outcrop into the near-surface Z (fixed depth) regime. Four sets of HYCOM simulations are conducted in this study. The HY_{LMD} employs the KPP_{LMD} scheme to parameterize the effect due to vertical mixing in the OSBL. The HY_{LF} uses a deterministic KPP scheme with wave enhancement introduced by Li and Fox-Kemper (2017). The $\text{HY}_{\text{DNN,WV}}$ uses the $\text{KPP}_{\text{DNN,WV}}$ scheme, with Stokes drift profiles included in the input features incorporated into HYCOM to predict coefficients, η and ϵ . The $\text{HY}_{\text{DNN,NW}}$ uses the $\text{KPP}_{\text{DNN,NW}}$ scheme, excluding Stokes drift information from the input arrays. The HY_{LMD} serves as a baseline control simulation, where both η and ϵ are fixed at one.

All simulations are initialized from the 3D HYCOM + NCODA Global $1/12^\circ$ Reanalysis on 1 January 2001 (GLBu0.08, https://tds.hycom.org/thredds/catalogs/GLBu0.08/expt_19.1.html) and integrated for 22 years from the year 2001 to the year 2022, with the first year treated as model spinup. The HYCOM

simulations employ open lateral boundary conditions using a one-way nesting, in which boundary information is provided by external parent-model archives (HYCOM GLBb0.08), following the standard HYCOM nesting approach. Atmospheric forcing required by HYCOM is provided by the hourly Climate Forecast System Reanalysis (CFSR) for the years 2001–10, and from the hourly Climate Forecast System, version 2 (CFSv2), for the years 2011–22. Stokes drift is one of the inputs in $\text{HY}_{\text{DNN,WV}}$. To avoid the high computational cost of coupled ocean wave modeling (Li et al. 2016), Stokes drift profiles are computed offline from ERA5 wave spectra (Hersbach et al. 2020). These wave forcing fields are treated as forcing, as in Ali et al. (2019).

3. Results

In this section, we compare simulated SSTs and MLDs with both observations and the HY_{LMD} model runs. We also construct regime diagrams to reveal how DNN-enhanced modifications systematically affect SSTs and MLDs under different forcing regimes. We focus on SST and MLD in the interior GOM (bathymetry deeper than 300 m), where the influence of coastal processes is reduced, making the findings more broadly applicable to other open ocean regions. Although our primary goal is to evaluate the performance of mixing parameterizations, it is important to note that biases in a 3D ocean model may stem from other processes as well, such as the misrepresentation of horizontal circulations at larger spatial scales, not just vertical mixing alone.

a. Sea surface temperature

In this subsection, the simulated SSTs are compared with reprocessed observational SST data. The observational SST product used is the Global Ocean Sea Surface Temperature and Sea Ice Reprocessed dataset, produced by the Met Office as part of the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system and distributed by the Copernicus Marine Environment Monitoring Service (CMEMS). It provides daily global SST fields from October 1981 to May 2022, at a horizontal resolution of $0.05^\circ \times 0.05^\circ$. The dataset is generated using a data assimilation scheme that combines in situ observations (e.g., from ships and buoys) with satellite radiometer measurements (e.g., AVHRR, SEVIRI) via optimal interpolation (Good et al. 2020). In this study, both observed and simulated SST fields are averaged over the period 2002–21, separately for summer months (June–August) and winter months (December–February).

Figure 3a shows the spatial distribution of observed SST in the GOM, averaged over summer months. During summer, the entire GOM is warm, with mean SSTs exceeding 28°C across most of the basin, except in certain narrow coastal regions such as the westernmost GOM off the coast of Texas and Mexico, as well as north of the Yucatan Peninsula. The cold bands along the coast, characterized by SSTs that are $3^\circ\text{--}4^\circ\text{C}$ lower than offshore waters and by strong horizontal temperature gradients, result from summertime coastal upwelling (e.g., Merino 1997; Zavala-Hidalgo et al. 2003; Zavala-Hidalgo et al. 2006; Jouanno et al. 2018).

Figures 3b–e display the SST differences between model simulations and observations in summer, for HY_{LMD} , HY_{LF} ,

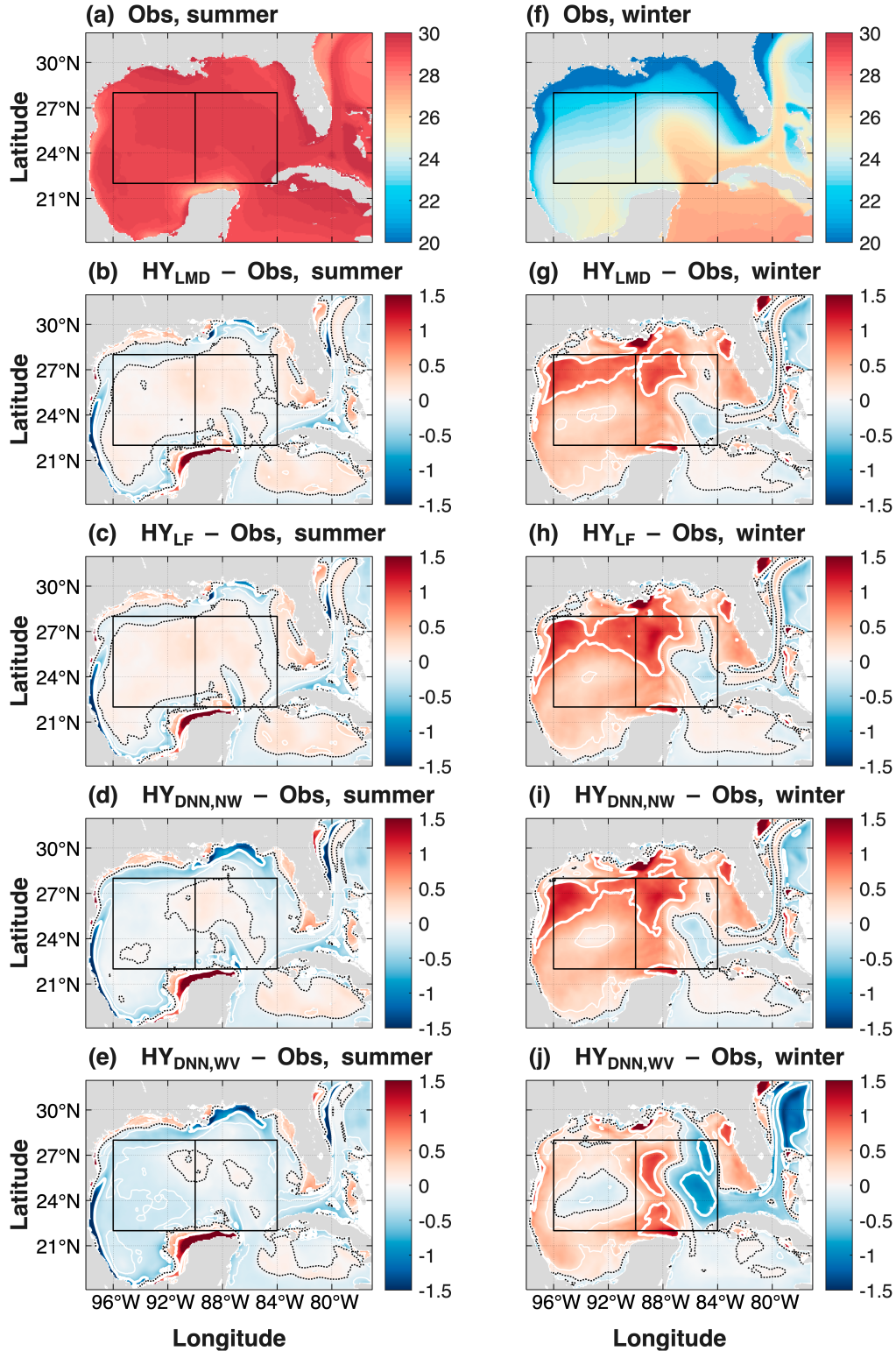


FIG. 3. Climatological mean SST in the GOM from 2002 to 2021. (a)–(e) Summer (June–August) SST and (f)–(j) winter (December–February) SST. (a),(f) Observed mean SST. (b),(g) SST biases between the HY_{LMD} and observations. (c),(h) SST biases between HY_{LF} and observations. (d),(i) SST biases between $HY_{DNN,NW}$ and observations. (e),(j) SST biases between $HY_{DNN,WV}$ and observations. The black dashed lines in (b)–(h) indicate the 0°C bias contour, thin white solid lines indicate $\pm 0.25^{\circ}\text{C}$ bias, and thick white lines indicate $\pm 0.75^{\circ}\text{C}$ bias. The two black boxes designate the west and east GOM regions, and the combined area of these two subdomains is referred to as the GOM region.

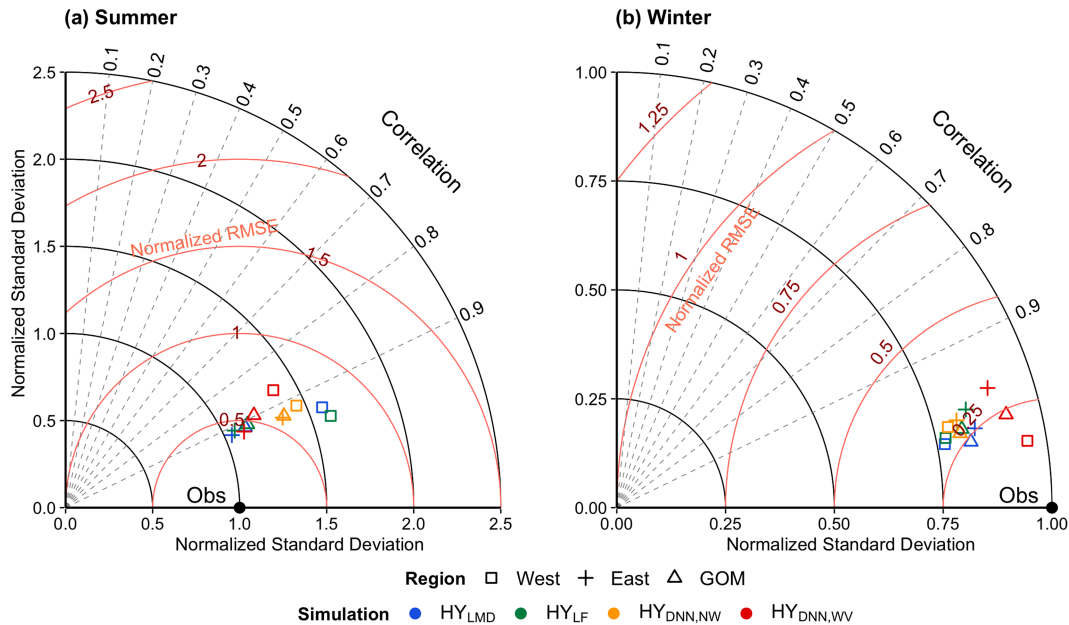


FIG. 4. Taylor diagrams comparing the mean SST in the GOM from 2002 to 2021. Results (a) for summer (June–August) and (b) for winter (December–February). In each diagram, the azimuthal position indicates the correlation coefficient between model simulations and observations, the radial distance indicates the normalized standard deviation (relative to the observed SST), and the red contours represent the normalized RMSE. The spatial locations of the regions (west, east, and GOM) are shown as black boxes in Fig. 3.

HY_{DNN,NW}, and HY_{DNN,WV}, respectively. In the open GOM, simulated SST deviations are small, remaining within $\pm 0.25^{\circ}\text{C}$ of the observations. Positive deviations dominate in HY_{LMD} and HY_{LF}, while negative deviations prevail in HY_{DNN,NW} and HY_{DNN,WV}. However, all models exhibit larger errors along the coasts, particularly in regions with sharp SST gradients, such as the upwelling zones north of the Yucatan Peninsula, highlighting the persistent challenge of resolving coastal dynamics at $1/12^{\circ}$ resolution.

Figure 3f shows the spatial distribution of observed SSTs in the GOM averaged over winter months. In contrast to summer, winter SSTs exhibit a pronounced meridional gradient, with SSTs increasing from north to south. SSTs along the northern coastal margin of the GOM shelf are significantly cooler than those in the open ocean, and sharp cross-shelf gradients are evident. The influence of the Loop Current (LC) is also clearly apparent, characterized by a northwestward intrusion of warm water from the Yucatan Channel into the east-central GOM, before exiting through the Florida Strait.

Figures 3g–j show the winter SST deviations from observations for HY_{LMD}, HY_{LF}, HY_{DNN,NW}, and HY_{DNN,WV}, respectively. Except for the west GOM north of Cuba, SSTs in HY_{LMD} are dominantly overestimated by at least 0.25°C across the GOM. Especially in the north GOM, a prominent band of positive bias exceeding 0.75°C is centered around 27°N . The only exception is the western GOM north of Cuba, where the bias is relatively small. The HY_{LF} and HY_{DNN,NW} do not mitigate this domainwide warm bias much. In fact, significant positive biases persist, particularly in the northern GOM. In contrast, HY_{DNN,WV} shows substantial improvement: SST

biases in the western and northern regions are markedly reduced, with widespread areas showing biases within $\pm 0.25^{\circ}\text{C}$. The pronounced warm bias belt evident in HY_{LMD} and HY_{LF} becomes less distinct in HY_{DNN,WV}, and the SST biases there are reduced by around 70%. However, near the LC core region, a bias dipole pattern emerges, characterized by strong positive biases ($>0.75^{\circ}\text{C}$) to the west (north of the Yucatan Peninsula) and negative biases ($<-0.75^{\circ}\text{C}$) to the east (northwest of Cuba). Strong biases are also evident in several narrow coastal regions across all simulations. North of the Yucatan Peninsula, pronounced positive SST biases ($>0.75^{\circ}\text{C}$) appear consistently in both summer and winter, while similar positive biases emerge outside the mouth of the Mississippi River during winter. In contrast, strong negative SST biases ($<-0.75^{\circ}\text{C}$) are present during summer along the narrow coastal belt of the western GOM in all simulations. SST biases in these coastal areas are heavily affected by local processes, such as smaller-scale coastal dynamics and river discharge, and therefore cannot be effectively reduced by improving vertical mixing parameterization schemes alone.

The performance of HYCOM simulated SSTs relative to observations is further evaluated using Taylor diagrams (Fig. 4), which enable simultaneous comparison of three statistical metrics: normalized standard deviation, correlation coefficient, and centered root-mean-square error. All standard deviations and root-mean-square errors are normalized by dividing by the observed standard deviation. To contrast bias characteristics between the western and eastern GOM, two subregions (west and east) are defined, as indicated by the black boxes in Fig. 3. The combined area of these two subdomains is referred to as

the GOM region. In summer, simulations within each region occupy distinct positions in the Taylor diagram, reflecting differences in overall performance. For example, $HY_{DNN,WV}$ exhibits a slightly smaller root-mean-square error (RMSE), where some other simulations show slightly higher correlation. In winter, although distinctions among simulations remain ambiguous in the east GOM and over the GOM as a whole, a notable improvement is evident in the west GOM for $HY_{DNN,WV}$. In this case, all three metrics—standard deviation, RMSE, and correlation—show substantial improvement relative to the other simulations.

It should be noted that the results in Fig. 4 are intended to compare model performance within a given region and season and should not be compared across different regions or seasons, as both the standard deviation and RMSE are normalized by the observed standard deviation specific to each region and season. For example, in summer, basinwide warming leads to reduced spatial variability in observed SSTs (Fig. 3a). Consequently, even when normalized variability in the HYCOM simulations is not minimal (Fig. 4a), the absolute deviation of simulated summer SSTs from observations remains small. In contrast, during winter, enhanced variability associated with a stronger Loop Current signal and increased latitudinal contrasts results in a larger observed SST standard deviation (Fig. 3b), such that modest reductions in RMSE or standard deviation correspond to more pronounced improvements in model performance (Fig. 4b).

To assess the temporal behavior and long-term stability of simulated SSTs, we computed daily climatological mean SSTs over a 20-yr simulation period at each location for both observations ($T_{obs,i,j}$) and model outputs ($T_{x,i,j}$), where x denotes a specific HYCOM simulation, i denotes a specific location, and j denotes the j th day of a year. Here, T represents the 20-yr average SST for a given calendar day. Then, the SST bias and RMSEs were calculated over all grid points within a selected region with n locations:

$$\begin{aligned} \text{Bias}_{x,j} &= \frac{1}{n} \sum_{i=1}^n (T_{x,i,j} - T_{obs,i,j}), \\ \text{RMSE}_{x,j} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{x,i,j} - T_{obs,i,j})^2}. \end{aligned} \quad (9)$$

Seasonally resolved daily SST biases and RMSEs, averaged over a 20-yr simulation period, are shown in Fig. 5. Positive SST biases accompanied by elevated RMSEs dominate the months from October to March. These biases are more pronounced in HY_{LMD} , HY_{LF} , and $HY_{DNN,NW}$. In contrast, in $HY_{DNN,WV}$, positive biases during these months are substantially reduced. Consistently, RMSEs are also reduced in $HY_{DNN,WV}$ in both the whole GOM region and the western GOM, except from mid-spring to midsummer (April–July). From April to July, strong freshwater discharge from the Mississippi River together with intense surface heating enhances upper-ocean stratification and leads to shallower OSBLs. Consequently, the OSBL is less well resolved by the vertical grid which contains fewer vertical points within the OSBL than other months. As a result, the improvement provided by the DNN-based parameterization is less evident in summer months than in winter

months. In the East GOM subregion, even though the SST biases in winter in $HY_{DNN,WV}$ are small, RMSEs are higher than in other simulations, consistent with the SST error dipole pattern observed in Fig. 3j.

b. Mixed layer depth

In this subsection, the simulated MLDs are compared with observations. Here, the MLD is diagnosed as the depth at which water density increases by 0.03 kg m^{-3} relative to the density at a reference depth of 10 m (de Boyer Montégut et al. 2004). Observed MLDs are calculated using vertical profiles of temperature, salinity, and pressure, obtained from Argo floats (Wong et al. 2020) in the GOM. Argo floats have been extensively deployed and have continuously sampled the global ocean since 2000 (Holte et al. 2017). However, the floats were mainly deployed in the open ocean basin before 2009 (Roemmich and The Argo Steering Team 2009), with increased deployment in marginal seas such as the GOM starting after 2010 (Jayne et al. 2017). Argo coverage in the GOM was sparse prior to 2010 but increased markedly thereafter (Fig. S3a). Furthermore, the spatial distribution of Argo floats in the GOM is uneven. Most floats sample the open ocean where the water exceeds 1000 m, while observations in shallower regions ($<1000 \text{ m}$) are scarce (Fig. S3b). Even within the deep ocean regions of the GOM, Argo profiles are not uniformly distributed, as data are more abundant in the northern and eastern portions, whereas the southwestern area, including the Bay of Campeche, has significantly fewer observations. Given the spatially sparse and uneven distribution of Argo-derived MLD observations in the GOM, our analysis focuses primarily on spatially aggregated statistics.

Figure 6 presents a detailed comparison of seasonal mean MLDs in the GOM during summer (June–August, left column) and winter (December–February, right column), based on Argo float observations and the four HYCOM simulations. The analysis is conducted within a set of hexagonal spatial bins, each summarizing MLD statistics from all Argo profiles that fall within the region between 2002 and 2021. To ensure a fair and consistent comparison with Argo observations, model outputs are interpolated to the exact time and location of each Argo profile. Only hexagons containing at least 50 Argo profiles are retained for analysis.

During summer (Fig. 6a), observed MLDs across the GOM are relatively shallow, mostly shallower than 30 m. Spatial variations in summer MLDs are small, with the deepest MLD near the Yucatan Channel ($\sim 35 \text{ m}$), near the core region of LC, while the shallowest MLDs are found offshore of Louisiana ($\sim 15 \text{ m}$), where freshwater flux from the Mississippi River helps stabilize the upper ocean. Mean MLD biases from all four HYCOM simulations are small (Figs. 6b–e). Both HY_{LMD} and HY_{LF} underestimate MLD over most of the GOM, with the largest negative biases ($\sim 10 \text{ m}$) in the Yucatan Channel. In contrast, simulated MLDs are generally deeper in both $HY_{DNN,NW}$ and $HY_{DNN,WV}$, with positive biases more pronounced in the west GOM and reduced shallow MLD biases in the Yucatan Channel and west of the Florida Peninsula.

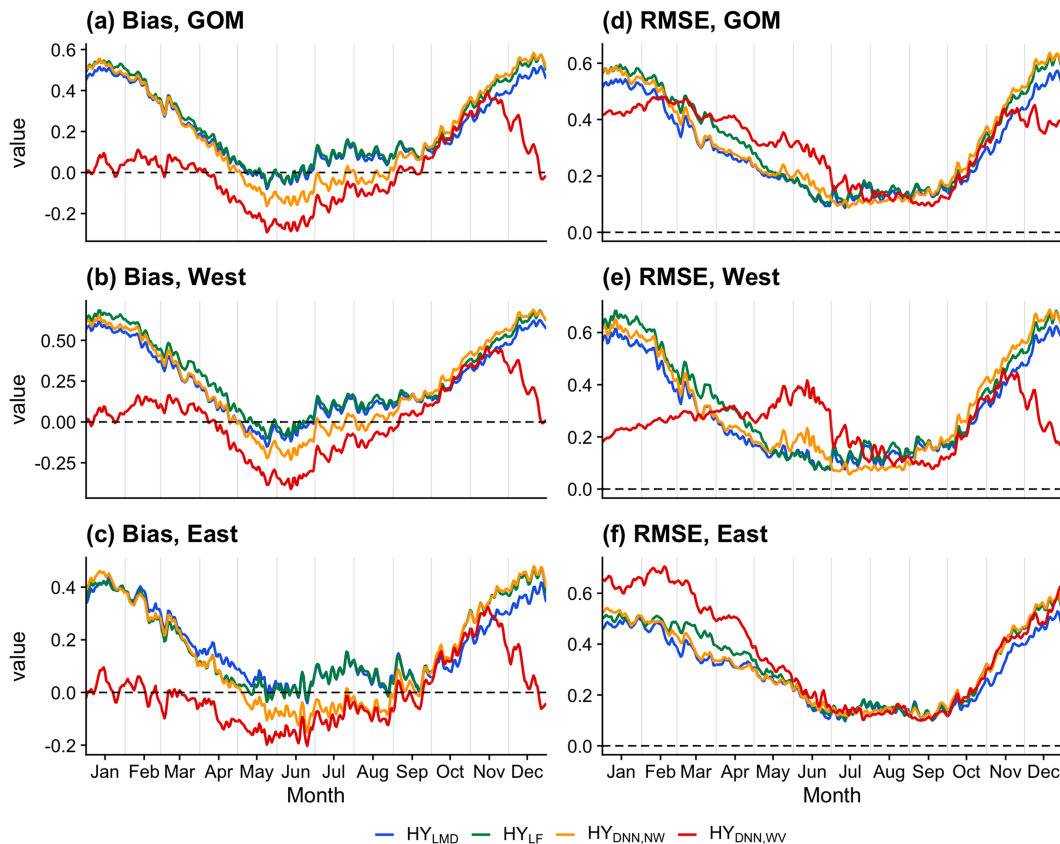


FIG. 5. Seasonal evolution of daily SST biases and RMSE, averaged over 20 years, for different regions of the GOM (entire GOM, western GOM, and eastern GOM). A 5-day running mean is applied to smooth out high-frequency variability and emphasize seasonal signals. The spatial extents of the regions are indicated by black boxes in Fig. 3.

During winter, the observed MLDs are substantially deeper and exhibit greater spatial variability across the GOM (Fig. 6f). Two distinct regions exhibit particularly deep winter MLDs: one around the LC and the other in the west-central GOM. All four HYCOM simulations overestimate winter MLDs across much of the central-western GOM (Figs. 6g–j). Among the evaluated regions, hexagon c exhibits the largest positive MLD biases, with mean winter MLD overestimated by ~ 25 – 30 m in HY_{LMD} . MLD biases are further increased in $\text{HY}_{\text{DNN,NW}}$ and $\text{HY}_{\text{DNN,WV}}$ compared to HY_{LMD} , reaching ~ 30 and ~ 40 m, respectively. This suggests that while wave-induced mixing enhances vertical mixing in certain areas, it may also lead to excessive deepening of the MLD in regions where background stratification is already weak.

To gain a more complete picture of model behavior, boxplots of winter MLD distributions are constructed for six representative winter hexagonal regions (c, d, e, h, i, and j in the right column of Fig. 6), as shown in Fig. 7. The corresponding sample sizes and Wilcoxon p values are reported in Table S1. These regions span a range of open-ocean dynamical regimes in the GOM, including areas influenced by the LC, regions of enhanced mesoscale activity, and relatively quiescent regions, and are therefore suitable for evaluating winter MLD climatology. Table S1 shows that winter MLD distributions exhibit

statistically significant climatological biases relative to observations in four of the six selected hexagonal regions ($p < 0.05$ in hexagons c, d, h, and i), indicating that systematic winter MLD biases persist regardless of the vertical mixing scheme implemented in HYCOM. Only hexagons e and j show no statistically significant differences from observations. Nevertheless, boxplot diagnostics reveal that $\text{HY}_{\text{DNN,WV}}$ reduces MLD median biases and shows increased distributional overlap with observations compared to the HY_{LMD} , HY_{LF} , and $\text{HY}_{\text{DNN,NW}}$ in several hexagons. For example, a large fraction of observed winter MLDs is shallower than 50 m, representing shallow mixing events in winter. These shallow MLD events are underrepresented in HY_{LMD} , HY_{LF} , and $\text{HY}_{\text{DNN,NW}}$, whose interquartile ranges are consistently biased toward deeper values. In contrast, $\text{HY}_{\text{DNN,WV}}$ not only better reproduces the observed median MLD values but also produces more consistent interquartile ranges, better capturing both the shallow and deep MLD events in most of the selected hexagons. An exception is found in hexagon c, where none of the simulations adequately capture the observed MLD distribution. In this region, the interquartile range of all four models is substantially deeper than the observations, suggesting that factors other than vertical mixing, such as horizontal advection or mesoscale processes, may play a more dominant role in

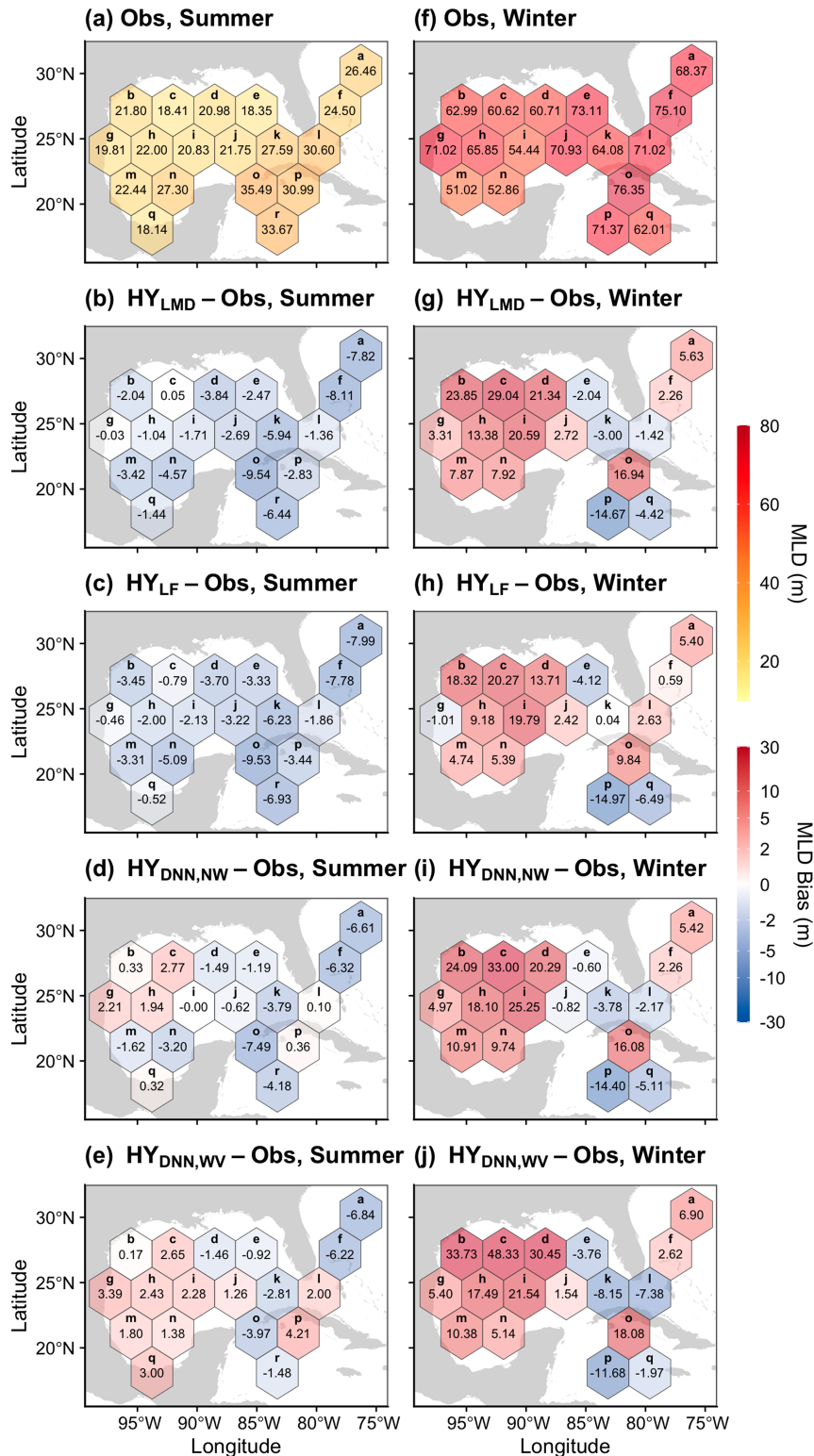


FIG. 6. Comparison of seasonal mean MLD between HYCOM simulations and observations. MLD climatology (left) for summer (June–August) and (right) for winter (December–February). (a),(f) The observed mean MLDs are shown. (b)–(e),(g)–(j) Mean MLD biases (simulation minus observation) for HY_{LMD} , $\text{HY}_{\text{DNN,NW}}$, and $\text{HY}_{\text{DNN,WV}}$ are displayed. Each hexagon summarizes MLD statistics from 2002 to 2021 for all Argo profiles located within the hexagonal region, along with corresponding HYCOM outputs interpolated to the same times and locations. Only hexagons containing at least 50 Argo profiles are shown.

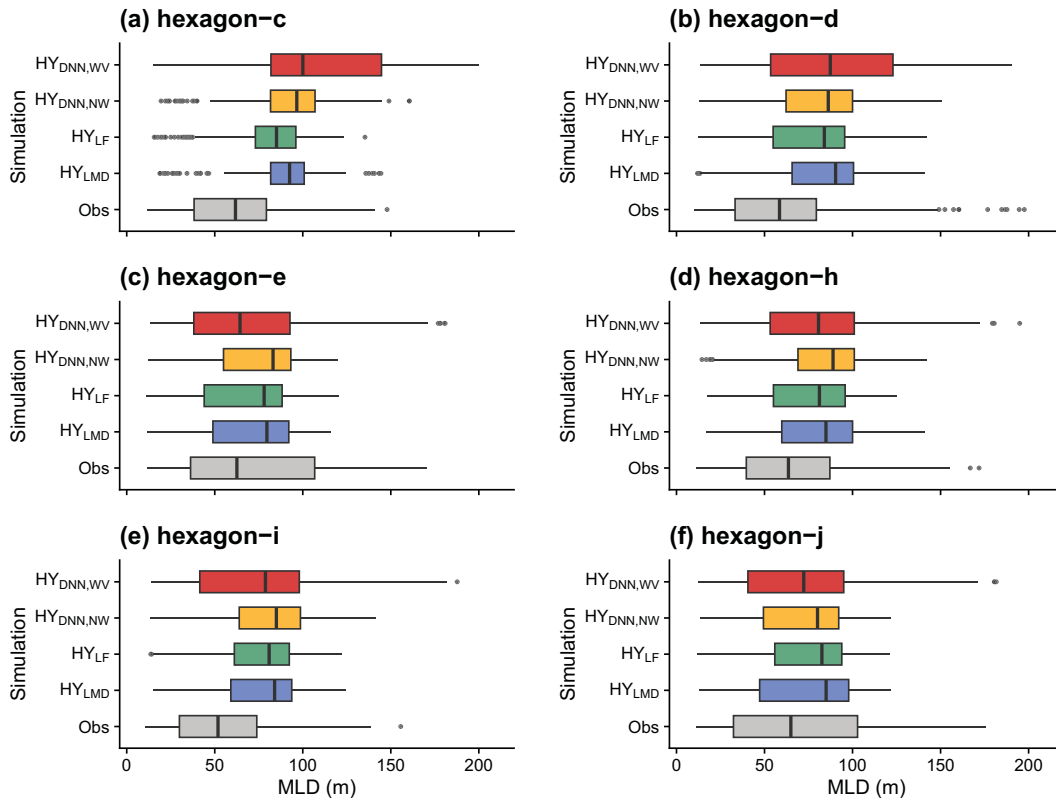


FIG. 7. Boxplot of winter MLD climatology distribution from HYCOM simulations and observations in five selected hexagonal regions (boxes c, d, e, h, i, and j indicated in Fig. 6). Each panel compares the distribution of MLDs during winter (December–February) among the observations (Obs) and HYCOM simulations HY_{LMD} , HY_{LF} , $HY_{DNN,NW}$, and $HY_{DNN,WV}$.

stratification in this region. These processes will be further discussed in section 4.

c. Response of the KPP_{DNN} to different forcing regimes

Surface forcing in the GOM varies markedly in space and time. Key drivers include wind conditions, wind-induced Langmuir turbulence, and buoyancy conditions. These processes collectively regulate near-surface turbulent mixing and therefore influence how the KPP_{DNN} schemes depart from the KPP_{LMD} scheme, ultimately shaping the climatology of SSTs and MLDs in HYCOM.

To visualize the relative importance of these drivers, we adopt the regime diagram introduced by Van Roekel et al. (2012) and later extended by Li et al. (2019). Figure 8 replicates this two-panel diagram: The top row represents destabilizing buoyancy conditions, while the bottom row represents stabilizing buoyancy conditions. The turbulent Langmuir number La_t , shown along the x axis, quantifies the relative importance of wind/wave-driven Langmuir turbulence over wind-induced shear turbulence. It is defined as the square root of the ratio of friction velocity divided by the surface Stokes drift magnitude as defined in McWilliams et al. (1997). The ratio between mixed layer depth h and Langmuir stability depth L_L , shown along the y axis, quantifies the relative importance of buoyancy over wind and waves.

Blue contours in Fig. 8 represent the global distribution of forcing regimes from Li et al. (2019), revealing that wind, wave, and buoyancy all play equally essential roles in modulating oceanic turbulence globally. Due to generally weak wind and wave conditions over the Gulf, unstable buoyancy forcing can be the primary driver of OSBL turbulence, while wind and wave are rarely the dominant drivers. The forcing conditions in the GOM over 20 years are overlaid in Fig. 8 as red contours, with the left, middle, and right columns representing distributions over all seasons, summer months, and winter months, respectively. Compared to the global ocean, the GOM exhibits slightly greater dominance of convective turbulence under destabilizing buoyancy forcing. This aligns with the finding by Li et al. (2019) that convection is more prominent at low latitudes. Seasonally, approximately 90% of the winter daily buoyancy forcing is destabilizing, and winter forcing regime distributions under destabilizing forcing resemble the annual pattern. In contrast, about 80% of the summer daily buoyancy forcing is stabilizing, and summer forcing regime distributions under stabilizing forcing are more representative of the overall stabilizing regime.

To investigate how the KPP_{DNN} schemes affect SST and MLD under different regimes, we calculated daily differences of HY_{LF} , $HY_{DNN,NW}$, and $HY_{DNN,WV}$ relative to HY_{LMD} . These differences are aggregated into hexagonal bins within

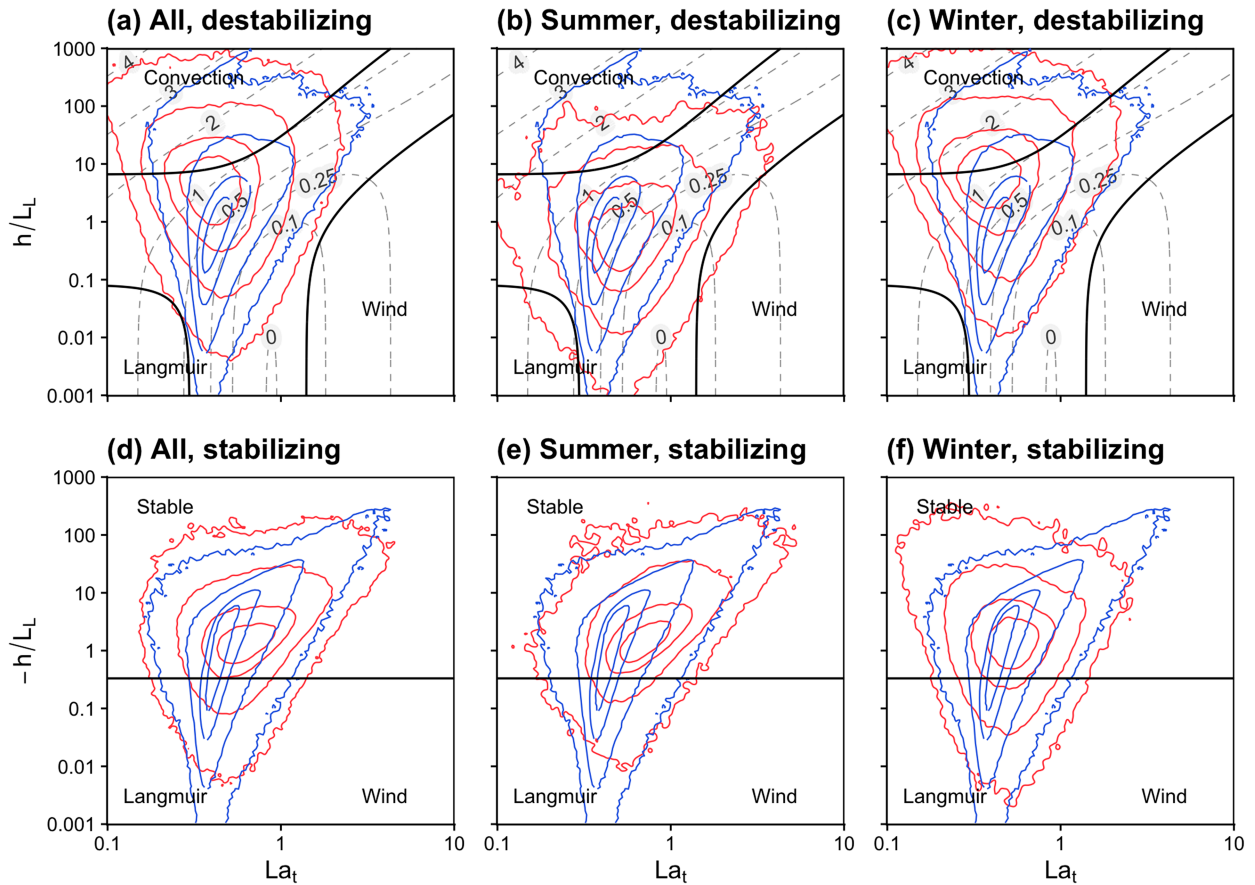


FIG. 8. Regime diagrams of forcing conditions in the GOM from 2002 to 2021. (a)–(c) Conditions under destabilizing buoyancy forcing; (d)–(f) conditions under stabilizing buoyancy forcing. In the destabilizing cases, thin dashed contours indicate the turbulent dissipation rate, and thick black lines enclose regimes where a single turbulence mechanism (wind-driven, Langmuir, or convective) contributes more than 90% of total dissipation. In the stabilizing cases, the horizontal black line with $-h/L_L = 1/3$ represents the maximum equilibrium value based on Pearson et al. (2015). Red contours indicate the probability distribution (30%, 60%, 90%, and 99%) of the seasonal ocean conditions in the GOM (a),(d) over the full 20-yr period, (b),(e) during summer months, and (c),(f) during winter months. Blue contours represent the probability distribution of all-season global ocean conditions from Li et al. (2019).

the regime diagram, and the mean values within each bin are mapped to visualize systematic patterns.

Figure 9 shows the hexagonal regime diagrams for SST differences. A consistent pattern across HY_{LF} , $HY_{DNN,NW}$, and $HY_{DNN,WV}$ is that SST differences are getting more negative as wave effects become more dominant, indicating that enhanced wave-induced turbulent mixing is associated with reduced SSTs. When La_t is greater than 1, SSTs in HY_{LF} are generally warmer than those in HY_{LMD} , thereby exacerbating the positive SST biases compared to observations (Figs. 9a,d). Under destabilizing buoyancy forcing conditions, $HY_{DNN,NW}$ predicts slightly warmer SSTs than HY_{LMD} (Fig. 9b), amplifying the KPP warm biases in the GOM during winter. In contrast, $HY_{DNN,WV}$ consistently predicts lower SSTs across all turbulent regimes (Figs. 9c,f), effectively mitigating the warm SST biases in winter.

Figure 10 shows the corresponding MLD differences. Unlike SST, MLD differences exhibit a stronger dependence on buoyancy forcing than on wave effects. Under destabilizing

buoyancy conditions (Figs. 10a–c), large positive MLD differences emerge when the dominant turbulence is convection. In these regimes, $HY_{DNN,WV}$ predicts more pronounced deeper MLDs, which may contribute to the higher chance of extreme deep MLDs (over 150 m) in winter (shown in Fig. 7). Under stabilizing buoyancy conditions (Figs. 10d–f), MLD increases with increasing buoyancy stability. As reported by Yuan et al. (2024), η is stronger when MLD is shallower, corresponding to stronger stabilizing forcing. This may be linked to the fact that the strength of Langmuir circulation is intense near the surface (Weller and Price 1988; McWilliams et al. 1997; Li and Fox-Kemper 2020), where wave effects are more effective at enhancing mixing and thus deepening the MLD.

Another notable pattern from Fig. 10 is that MLDs are shallower in $HY_{DNN,WV}$ than in HY_{LMD} when wind and wave forcing are dominant over buoyancy forcing, especially under destabilizing buoyancy conditions in $HY_{DNN,WV}$. This results in a large portion of shallower MLDs compared to HY_{LMD} , consistent with observed MLD ranges in Fig. 7. It is important

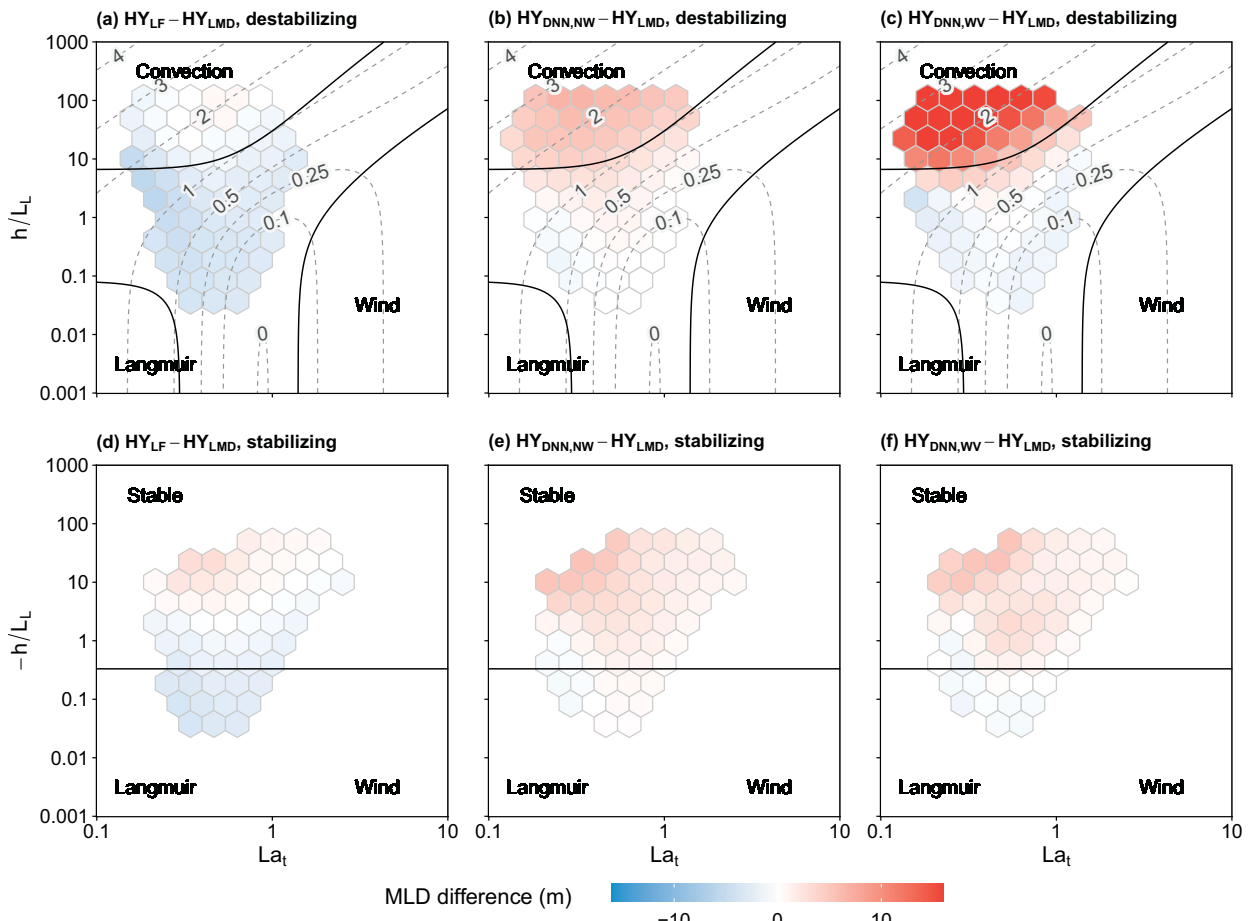


FIG. 9. Regime diagrams illustrating the SST differences of HYCOM simulations, (a),(d) HY_{LF} (b),(e) $HY_{DNN,NW}$, and (c),(f) $HY_{DNN,WV}$, relative to HY_{LMD} , under different turbulence forcing regimes. (a)–(c) Destabilizing buoyancy forcing; (d)–(f) stabilizing buoyancy forcing. The color of each hexagon represents the mean SST difference, averaged over all samples whose forcing conditions fall within that hexagonal bin. Only hexagons with over 100 000 samples are shown. Definitions of the gray dashed contours and black solid lines are consistent with those in Fig. 8.

to note that this does not suggest that wave effects suppress turbulent mixing in the real ocean. Instead, it suggests that the inclusion of wave-related information in $HY_{DNN,WV}$ can, under certain forcing regimes, reduce the tendency of the baseline KPP scheme to produce overly deep MLDs. Nevertheless, unrealistically deep MLDs still occur in some regions and conditions, indicating that further improvements are required.

4. Discussion

In ocean models, biases in simulated SSTs and MLDs arise from inaccurate representation of several interacting processes, including small-scale turbulence parameterization, submesoscale dynamics, mesoscale dynamics, and large-scale circulation. In this study, we focus exclusively on the parameterization of small-scale turbulent mixing. As shown in Fig. S4, KPP_{DNN} schemes have a limited impact on sea surface heights (SSHs) and high-frequency sea level anomalies (HSLAs), and similar bias patterns persist in all four HYCOM simulations when compared

with observations, indicating that unmodified regional-scale and mesoscale processes exert a dominant control on these features.

The persistent biases in SSH and HSLA fields also contribute to SST and MLD biases in the simulations. For example, compared to observations, a dome of positive biases, implying stronger anticyclonic circulation, exists in all four HYCOM simulations in the northwest GOM. Within this dome lies a saddle region near $93^\circ W$, characterized by sharp SSH anomaly gradients on both its eastern and western flanks (Figs. S4c,e,g,i). This saddle region corresponds to hexagon c during winter months (Fig. 6), where the simulated MLDs in all HYCOM simulations are significantly deeper than observed, with minimal overlap in interquartile ranges (Fig. 7a). This highlights that improving vertical mixing alone is insufficient to eliminate SST and MLD biases, as the background circulation errors persist. To further reduce uncertainties in simulated SST and MLD, additional efforts are necessary to better represent larger-scale processes in the models. Since the four simulations exhibit similar climatological regional-scale circulation, the differences in SSTs

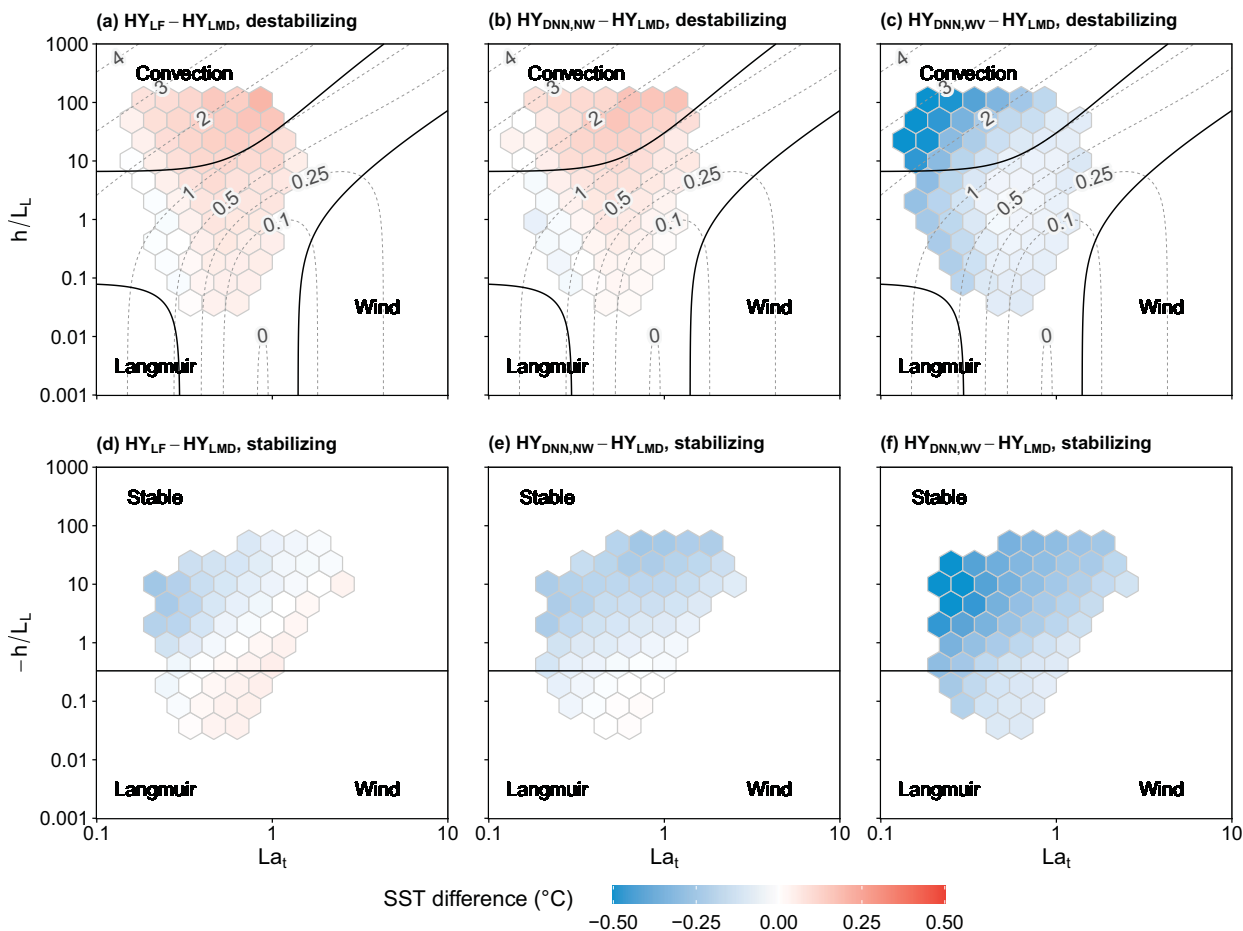


FIG. 10. As in Fig. 9, but for MLD differences.

and MLDs among simulations can be primarily attributed to the distinct turbulent mixing parameterizations.

Although strong misalignment between wind and wave can suppress mixing (e.g., Van Roekel et al. 2012; McWilliams et al. 2014), the overall effect of wave-enhanced KPP_{LMD} schemes is to increase turbulent mixing over interannual time scales in ocean models (e.g., Li et al. 2016; Ali et al. 2019), resulting in deeper mean MLDs than using the KPP_{LMD} scheme. A similar behavior is found in $HY_{DNN,WV}$, where mean MLDs are deepened over climatological scales. It indicates that enhanced mixing due to Langmuir turbulence is more prevalent than suppressed mixing when waves oppose wind and underscores the importance of incorporating wave effects into vertical mixing parameterization schemes.

Unlike previous works on enhancing the KPP_{LMD} scheme primarily through prescribed wave enhancements, our KPP_{DNN} schemes also incorporate additional relationships into enhancements, such as dependencies on buoyancy stability that are not represented in deterministic wave enhancement formulations. In the GOM, this results in shallower MLDs under specific wind-wave-buoyancy forcing combinations (Fig. 10), bringing the simulated MLD distributions closer to observations in $HY_{DNN,WV}$ (Fig. 7). This suggests that KPP_{LMD} , developed outside of the

GOM, can overestimate turbulent mixing intensity under certain forcing regimes in this basin and that simply adding wave effects without accounting for buoyancy stability correction could exacerbate existing biases. This also demonstrates that data-driven schemes like KPP_{DNN} offer the potential to adaptively balance competing mixing processes, providing a more flexible alternative to fixed empirical enhancements.

The reduced winter SST warm biases in the west GOM and the improved ability to capture the distribution of observed MLDs in $HY_{DNN,WV}$ relative to HY_{LMD} , HY_{LF} , and $HY_{DNN,NW}$ demonstrate that the $KPP_{DNN,WV}$ scheme can operate effectively across a range of wind and wave conditions within a regional ocean such as the GOM. These results suggest that the scheme is suitable for application to the entire ocean basin or the global ocean. These results also underscore the importance of incorporating nonbreaking wave effects, especially the Stokes drift profiles, in vertical mixing parameterizations.

5. Conclusions

In this study, deep neural networks (DNNs) are trained on over 700 turbulence-resolving large-eddy simulation (LES) solutions, driven by different forcing conditions in the Gulf of

Mexico (GOM) derived from reanalysis products. The DNN models are used to predict two key coefficients in the K -profile parameterization (KPP) scheme: the unresolved shear coefficient η and the velocity scale coefficient ϵ , as defined in equations (5) and (6). The trained DNN models are implemented into HYCOM to incorporate knowledge obtained from LES solutions into KPP. Four HYCOM simulations are conducted to evaluate the impact of nonbreaking surface waves on vertical mixing. The HY_{LMD} uses the baseline KPP_{LMD} scheme without any wave effects, HY_{LF} uses the deterministic non-breaking wave-enhanced KPP scheme developed by Li and Fox-Kemper (2017), HY_{DNN,NW} excludes Stokes drift information from DNN inputs, and HY_{DNN,WV} includes it.

Key conclusions are summarized as follows:

- HY_{DNN,WV} reduces the winter sea surface temperature (SST) warm biases in the west GOM and better captures the distribution of mixed layer depths (MLDs).
- Under conditions where wave-induced turbulence is more dominant, HY_{DNN,WV} tends to produce a stronger cooling effect in SST compared to HY_{LMD}. When destabilizing buoyancy forcing becomes more intense or stabilizing buoyancy becomes stronger, HY_{DNN,WV} generally results in a deepening of the mixed layer relative to HY_{LMD}.
- Improvements in vertical mixing alone are not sufficient to resolve SST and MLD biases, since some biases are caused by regional-scale circulation and mesoscale systems in the GOM.

By generalizing the KPP_DNN scheme in Yuan et al. (2024) from the 1D vertical column GOTM to the 3D HYCOM, the present study advances the capability of data-driven turbulence parameterizations to operate robustly across spatially heterogeneous and dynamically diverse ocean environments. Specifically, the proposed framework incorporates geographic-location-specific conditions into inputs and enables LES-informed wave-turbulence effects to interact consistently with regional-scale circulation, mesoscale variability, and regional forcing in HYCOM, which was not achievable in previous single-column implementations. This study also advances our understanding of how vertical mixing parameterization schemes impact 3D ocean model outputs by demonstrating that the use of machine learning to enhance traditional vertical mixing schemes improves SSTs and MLDs using 3D realistic ocean models, more than the use of physics-based wave-informed parameterization could do. During the course, it also shows that SST and MLD variabilities depend not only on vertical mixing but also on the critical impact of regional-scale circulation. Future model development will focus on testing DNN-based parameterizations in data assimilative models (Chassignet et al. 2007; Zavala-Romero et al. 2025) that better capture the structure and variability of the large and mesoscale variabilities such as the LC. Moreover, expanding the coverage and diversity of LES training datasets, for example, in dynamically complex regions such as high latitudes, equatorial areas, and coastal regions, as well as the interplay between OSBL turbulence and submesoscale currents, could further generalize the applicability of DNN-based parameterizations to the global ocean.

Overall, this study highlights the potential of machine learning to improve ocean model fidelity at critical air-sea interaction layers while pointing to the need for strategies that couple vertical and horizontal processes to achieve more realistic, physically consistent ocean simulations.

Acknowledgments. The authors thank two anonymous reviewers for their many comments and suggestions that improved the manuscript. J. Y. and J. H. L. were supported by the Office of Naval Research under Grant N00014-23-1-2553. J. H. L. was also supported by the National Science Foundation under Grant OCE1945502. E. P. C., A. B., and A. J. W. were supported by the Office of Naval Research through Grant N00014-23-1-2547. Computations were performed using the high-performance computing resources provided by Louisiana State University and the Louisiana Optical Network Initiative (LONI).

Data availability statement. The CFSR and CFSv2 atmospheric reanalysis forcing data used to drive both LES and HYCOM simulations are available from the UCAR Research Data Archive (<https://rda.ucar.edu/datasets/d093002/> for CFSR and <https://rda.ucar.edu/datasets/d094002/> for CFSv2). The ERA5 wave spectra data are available from the Climate Data Store (<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-complete>). The HYCOM source code with the KPP_DNN model is available at https://github.com/lsocean/KPPDNN_in_HYCOM.

REFERENCES

- Ali, A., K. H. Christensen, Ø. Breivik, M. Malila, R. P. Raj, L. Bertino, E. P. Chassignet, and M. Bakhoday-Paskyabi, 2019: A comparison of Langmuir turbulence parameterizations and key wave effects in a numerical model of the North Atlantic and Arctic Oceans. *Ocean Modell.*, **137**, 76–97, <https://doi.org/10.1016/j.ocemod.2019.02.005>.
- Belcher, S. E., and Coauthors, 2012: A global perspective on Langmuir turbulence in the ocean surface boundary layer. *Geophys. Res. Lett.*, **39**, L18605, <https://doi.org/10.1029/2012GL052932>.
- Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modell.*, **4**, 55–88, [https://doi.org/10.1016/S1463-5003\(01\)00012-9](https://doi.org/10.1016/S1463-5003(01)00012-9).
- Brenowitz, N. D., T. Beucler, M. Pritchard, and C. S. Bretherton, 2020: Interpreting and stabilizing machine-learning parameterizations of convection. *J. Atmos. Sci.*, **77**, 4357–4375, <https://doi.org/10.1175/JAS-D-20-0082.1>.
- Chassignet, E. P., L. T. Smith, G. R. Halliwell, and R. Bleck, 2003: North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *J. Phys. Oceanogr.*, **33**, 2504–2526, [https://doi.org/10.1175/1520-0485\(2003\)033<2504:NASWTH>2.0.CO;2](https://doi.org/10.1175/1520-0485(2003)033<2504:NASWTH>2.0.CO;2).
- , H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, R. Baraille, and R. Bleck, 2007: The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *J. Mar. Syst.*, **65**, 60–83, <https://doi.org/10.1016/j.jmarsys.2005.09.016>.
- , and Coauthors, 2009: US GODAE: Global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography*, **22** (2), 64–75, <https://doi.org/10.5670/oceanog.2009.39>.

- Chattopadhyay, A., and P. Hassanzadeh, 2023: Long-term instabilities of deep learning-based digital twins of the climate system: The cause and a solution. arXiv, 2304.07029v1, <https://doi.org/10.48550/arXiv.2304.07029>.
- D'Asaro, E. A., 2014: Turbulence in the upper-ocean mixed layer. *Annu. Rev. Mar. Sci.*, **6**, 101–115, <https://doi.org/10.1146/annurev-marine-010213-135138>.
- , and G. T. Dairiki, 1997: Turbulence intensity measurements in a wind-driven mixed layer. *J. Phys. Oceanogr.*, **27**, 2009–2022, [https://doi.org/10.1175/1520-0485\(1997\)027<2009:TIMIAW>2.0.CO;2](https://doi.org/10.1175/1520-0485(1997)027<2009:TIMIAW>2.0.CO;2).
- de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone, 2004: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J. Geophys. Res.*, **109**, C12003, <https://doi.org/10.1029/2004JC002378>.
- Fan, Y., and S. M. Griffies, 2014: Impacts of parameterized Langmuir turbulence and nonbreaking wave mixing in global climate simulations. *J. Climate*, **27**, 4752–4775, <https://doi.org/10.1175/JCLI-D-13-00583.1>.
- Fox-Kemper, B., and Coauthors, 2019: Challenges and prospects in ocean circulation models. *Front. Mar. Sci.*, **6**, 65, <https://doi.org/10.3389/fmars.2019.00065>.
- , L. Johnson, and F. Qiao, 2022: Ocean near-surface layers. *Ocean Mixing*, Elsevier, 65–94.
- Gargett, A. E., and J. Marra, 2002: Effects of upper ocean physical processes (turbulence, advection and air-sea interaction) on oceanic primary production. *Biological–Physical Interactions in the Ocean*, A. R. Robinson, J. J. McCarthy and B. J. Rothschild, Eds., The Sea—Ideas and Observations on Progress in the Study of the Seas, Vol. 12, Wiley and Sons, 19–49.
- GEBCO Bathymetric Compilation Group, 2019: The GEBCO_2019 Grid—A continuous terrain model of the global oceans and land. NERC British Oceanographic Data Centre, accessed 5 June 2024, https://kiyuu.bosai.go.jp/ExData/GEBCO/GEBCO_2019.htm.
- Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**, 5742–5751, <https://doi.org/10.1029/2018GL078202>.
- George, J., M. A. A. Teixeira, and I. Stiperski, 2025: A physically based modification of the KPP turbulence closure scheme for the upper ocean that accounts for the effects of Langmuir turbulence. *J. Phys. Oceanogr.*, **55**, 1067–1081, <https://doi.org/10.1175/JPO-D-24-0175.1>.
- Good, S., and Coauthors, 2020: The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sens.*, **12**, 720, <https://doi.org/10.3390/rs12040720>.
- Hamlington, P. E., L. P. Van Roekel, B. Fox-Kemper, K. Julien, and G. P. Chini, 2014: Langmuir–submesoscale interactions: Descriptive analysis of multiscale frontal spindown simulations. *J. Phys. Oceanogr.*, **44**, 2249–2272, <https://doi.org/10.1175/JPO-D-13-0139.1>.
- Harcourt, R. R., 2015: An improved second-moment closure model of Langmuir turbulence. *J. Phys. Oceanogr.*, **45**, 84–103, <https://doi.org/10.1175/JPO-D-14-0046.1>.
- , and E. A. D'Asaro, 2008: Large-eddy simulation of Langmuir turbulence in pure wind seas. *J. Phys. Oceanogr.*, **38**, 1542–1562, <https://doi.org/10.1175/2007JPO3842.1>.
- He, H., and D. Chen, 2011: Effects of surface wave breaking on the oceanic boundary layer. *Geophys. Res. Lett.*, **38**, L07604, <https://doi.org/10.1029/2011GL046665>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Holte, J., L. D. Talley, J. Gilson, and D. Roemmich, 2017: An Argo mixed layer climatology and database. *Geophys. Res. Lett.*, **44**, 5618–5626, <https://doi.org/10.1002/2017GL073426>.
- Iyer, S., K. G. Hughes, and J. N. Moum, 2025: Predicting ocean turbulence across orders of magnitude using neural networks trained on multiyear observations. *Artif. Intell. Earth Syst.*, **4**, 240093, <https://doi.org/10.1175/AIES-D-24-0093.1>.
- Jayne, S. R., D. Roemmich, N. Zilberman, S. C. Riser, K. S. Johnson, G. C. Johnson, and S. R. Piotrowicz, 2017: The Argo Program: Present and future. *Oceanography*, **30** (2), 18–28, <https://doi.org/10.5670/oceanog.2017.213>.
- Jerlov, N. G., 1976: *Marine Optics*. Elsevier, 230 pp.
- Jouanno, J., E. Pallàs-Sanz, and J. Sheinbaum, 2018: Variability and dynamics of the Yucatan upwelling: High-resolution simulations. *J. Geophys. Res. Oceans*, **123**, 1251–1262, <https://doi.org/10.1002/2017JC013535>.
- Kantha, L. H., and C. A. Clayson, 1994: An improved mixed layer model for geophysical applications. *J. Geophys. Res.*, **99**, 25 235–25 266, <https://doi.org/10.1029/94JC02257>.
- Kukulka, T., 2020: Horizontal transport of buoyant material by turbulent jets in the upper ocean. *J. Phys. Oceanogr.*, **50**, 827–843, <https://doi.org/10.1175/JPO-D-19-0276.1>.
- , A. J. Plueddemann, J. H. Trowbridge, and P. P. Sullivan, 2009: Significance of Langmuir circulation in upper ocean mixing: Comparison of observations and simulations. *Geophys. Res. Lett.*, **36**, L10603, <https://doi.org/10.1029/2009GL037620>.
- Large, W. G., and P. R. Gent, 1999: Validation of vertical mixing in an equatorial ocean model using large eddy simulations and observations. *J. Phys. Oceanogr.*, **29**, 449–464, [https://doi.org/10.1175/1520-0485\(1999\)029<0449:VOVMIA>2.0.CO;2](https://doi.org/10.1175/1520-0485(1999)029<0449:VOVMIA>2.0.CO;2).
- , J. C. McWilliams, and S. C. Doney, 1994: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, **32**, 363–403, <https://doi.org/10.1029/94RG01872>.
- Lévy, M., L. Bopp, P. Karleskind, L. Resplandy, C. Éthé, and F. Pinsard, 2013: Physical pathways for carbon transfers between the surface mixed layer and the ocean interior. *Global Biogeochem. Cycles*, **27**, 1001–1012, <https://doi.org/10.1002/gbc.20092>.
- Li, Q., and B. Fox-Kemper, 2017: Assessing the effects of Langmuir turbulence on the entrainment buoyancy flux in the ocean surface boundary layer. *J. Phys. Oceanogr.*, **47**, 2863–2886, <https://doi.org/10.1175/JPO-D-17-0085.1>.
- , and —, 2020: Anisotropy of Langmuir turbulence and the Langmuir-enhanced mixed layer entrainment. *Phys. Rev. Fluids*, **5**, 013803, <https://doi.org/10.1103/PhysRevFluids.5.013803>.
- , A. Webb, B. Fox-Kemper, A. Craig, G. Danabasoglu, W. G. Large, and M. Vertenstein, 2016: Langmuir mixing effects on global climate: WAVEWATCH III in CESM. *Ocean Modell.*, **103**, 145–160, <https://doi.org/10.1016/j.ocemod.2015.07.020>.
- , and Coauthors, 2019: Comparing ocean surface boundary vertical mixing schemes including Langmuir turbulence. *J. Adv. Model. Earth Syst.*, **11**, 3545–3592, <https://doi.org/10.1029/2019MS001810>.
- , J. Bruggeman, H. Burchard, K. Klingbeil, L. Umlauf, and K. Bolding, 2021: Integrating CVMix into GOTM (v6.0): A consistent framework for testing, comparing, and applying ocean mixing schemes. *Geosci. Model Dev.*, **14**, 4261–4282, <https://doi.org/10.5194/gmd-14-4261-2021>.

- Liang, J.-H., C. Deutsch, J. C. McWilliams, B. Baschek, P. P. Sullivan, and D. Chiba, 2013: Parameterizing bubble-mediated air-sea gas exchange and its effect on ocean ventilation. *Global Biogeochem. Cycles*, **27**, 894–905, <https://doi.org/10.1002/gbc.20080>.
- , S. R. Emerson, E. A. D’Asaro, C. L. McNeil, R. R. Harcourt, P. P. Sullivan, B. Yang, and M. F. Cronin, 2017: On the role of sea-state in bubble-mediated air-sea gas flux during a winter storm. *J. Geophys. Res. Oceans*, **122**, 2671–2685, <https://doi.org/10.1002/2016JC012408>.
- , X. Wan, K. A. Rose, P. P. Sullivan, and J. C. McWilliams, 2018: Horizontal dispersion of buoyant materials in the ocean surface boundary layer. *J. Phys. Oceanogr.*, **48**, 2103–2125, <https://doi.org/10.1175/JPO-D-18-0020.1>.
- , E. A. D’Asaro, C. L. McNeil, Y. Fan, R. R. Harcourt, S. R. Emerson, B. Yang, and P. P. Sullivan, 2020: Suppression of CO₂ outgassing by gas bubbles under a hurricane. *Geophys. Res. Lett.*, **47**, e2020GL090249, <https://doi.org/10.1029/2020GL090249>.
- , J. Yuan, X. Wan, J. Liu, B. Liu, H. Jang, and M. Tyagi, 2022: Exploring the use of machine learning to parameterize vertical mixing in the ocean surface boundary layer. *Ocean Modell.*, **176**, 102059, <https://doi.org/10.1016/j.ocemod.2022.102059>.
- McPhee, M. G., 1999: Parameterization of mixing in the ocean boundary layer. *J. Mar. Syst.*, **21**, 55–65, [https://doi.org/10.1016/S0924-7963\(99\)00005-6](https://doi.org/10.1016/S0924-7963(99)00005-6).
- , 2008: *Air-Ice-Ocean Interaction: Turbulent Ocean Boundary Layer Exchange Processes*. Springer, 216 pp.
- McWilliams, J. C., and P. P. Sullivan, 2000: Vertical mixing by Langmuir circulations. *Spill Sci. Technol. Bull.*, **6**, 225–237, [https://doi.org/10.1016/S1353-2561\(01\)00041-X](https://doi.org/10.1016/S1353-2561(01)00041-X).
- , —, and C.-H. Moeng, 1997: Langmuir turbulence in the ocean. *J. Fluid Mech.*, **334**, 1–30, <https://doi.org/10.1017/S0022112096004375>.
- , E. Huckle, J.-H. Liang, and P. P. Sullivan, 2012: The wavy Ekman layer: Langmuir circulations, breaking waves, and Reynolds stress. *J. Phys. Oceanogr.*, **42**, 1793–1816, <https://doi.org/10.1175/JPO-D-12-07.1>.
- , —, J. Liang, and P. P. Sullivan, 2014: Langmuir turbulence in swell. *J. Phys. Oceanogr.*, **44**, 870–890, <https://doi.org/10.1175/JPO-D-13-0122.1>.
- Merino, M., 1997: Upwelling on the Yucatan shelf: Hydrographic evidence. *J. Mar. Syst.*, **13**, 101–121, [https://doi.org/10.1016/S0924-7963\(96\)00123-6](https://doi.org/10.1016/S0924-7963(96)00123-6).
- Noh, Y., H. Ok, E. Lee, T. Toyoda, and N. Hirose, 2016: Parameterization of Langmuir circulation in the ocean mixed layer model using LES and its application to the OGCM. *J. Phys. Oceanogr.*, **46**, 57–78, <https://doi.org/10.1175/JPO-D-14-0137.1>.
- Ott, J., M. Pritchard, N. Best, E. Linstead, M. Curcic, and P. Baldi, 2020: A Fortran-Keras deep learning bridge for scientific computing. *Sci. Program.*, **2020**, 8888811, <https://doi.org/10.1155/2020/8888811>.
- Pearson, B. C., A. L. M. Grant, J. A. Polton, and S. E. Belcher, 2015: Langmuir turbulence and surface heating in the ocean surface boundary layer. *J. Phys. Oceanogr.*, **45**, 2897–2911, <https://doi.org/10.1175/JPO-D-15-0018.1>.
- Qiao, F., Y. Yuan, Y. Yang, Q. Zheng, C. Xia, and J. Ma, 2004: Wave-induced mixing in the upper ocean: Distribution and application to a global ocean circulation model. *Geophys. Res. Lett.*, **31**, L11303, <https://doi.org/10.1029/2004GL019824>.
- Rabe, T. J., T. Kukulka, I. Ginis, T. Hara, B. G. Reichl, E. A. D’Asaro, R. R. Harcourt, and P. P. Sullivan, 2015: Langmuir turbulence under Hurricane Gustav (2008). *J. Phys. Oceanogr.*, **45**, 657–677, <https://doi.org/10.1175/JPO-D-14-0030.1>.
- Rasp, S., 2020: Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: General algorithms and Lorenz 96 case study (v1.0). *Geosci. Model Dev.*, **13**, 2185–2196, <https://doi.org/10.5194/gmd-13-2185-2020>.
- , M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. USA*, **115**, 9684–9689, <https://doi.org/10.1073/pnas.1810286115>.
- Reichl, B. G., and R. Hallberg, 2018: A simplified energetics based Planetary Boundary Layer (ePBL) approach for ocean climate simulations. *Ocean Modell.*, **132**, 112–129, <https://doi.org/10.1016/j.ocemod.2018.10.004>.
- , and Q. Li, 2019: A parameterization with a constrained potential energy conversion rate of vertical mixing due to Langmuir turbulence. *J. Phys. Oceanogr.*, **49**, 2935–2959, <https://doi.org/10.1175/JPO-D-18-0258.1>.
- , D. Wang, T. Hara, I. Ginis, and T. Kukulka, 2016: Langmuir turbulence parameterization in tropical cyclone conditions. *J. Phys. Oceanogr.*, **46**, 863–886, <https://doi.org/10.1175/JPO-D-15-0106.1>.
- Roemmich, D., and The Argo Steering Team, 2009: Argo: The challenge of continuing 10 years of progress. *Oceanography*, **22** (3), 46–55, <https://doi.org/10.5670/oceanog.2009.65>.
- Sane, A., B. G. Reichl, A. Adcroft, and L. Zanna, 2023: Parameterizing vertical mixing coefficients in the ocean surface boundary layer using neural networks. *J. Adv. Model. Earth Syst.*, **15**, e2023MS003890, <https://doi.org/10.1029/2023MS003890>.
- , B. Reichl, A. Adcroft, and L. Zanna, 2026: Machine learned equations for vertical mixing in the ocean surface boundary layer. *Geophys. Res. Lett.*, in press.
- Schmittner, A., A. Oschlies, X. Giraud, M. Eby, and H. L. Simmons, 2005: A global model of the marine ecosystem for long-term simulations: Sensitivity to ocean mixing, buoyancy forcing, particle sinking, and dissolved organic matter cycling. *Global Biogeochem. Cycles*, **19**, GB3004, <https://doi.org/10.1029/2004GB002283>.
- Schultz, M. G., C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadler, 2021: Can deep learning beat numerical weather prediction? *Philos. Trans. Roy. Soc.*, **A379**, 20200097, <https://doi.org/10.1098/rsta.2020.0097>.
- Skyllingstad, E. D., T. Paluszkiwicz, D. W. Denbo, and W. D. Smyth, 1996: Nonlinear vertical mixing processes in the ocean: Modeling and parameterization. *Physica D*, **98**, 574–593, [https://doi.org/10.1016/0167-2789\(96\)00113-3](https://doi.org/10.1016/0167-2789(96)00113-3).
- Smith, K. M., P. E. Hamlington, K. E. Niemeyer, B. Fox-Kemper, and N. S. Lovenduski, 2018: Effects of Langmuir turbulence on upper ocean carbonate chemistry. *J. Adv. Model. Earth Syst.*, **10**, 3030–3048, <https://doi.org/10.1029/2018MS001486>.
- Sullivan, P. P., and J. C. McWilliams, 2010: Dynamics of winds and currents coupled to surface waves. *Annu. Rev. Fluid Mech.*, **42**, 19–42, <https://doi.org/10.1146/annurev-fluid-121108-145541>.
- Thompson, A. F., A. Lazar, C. Buckingham, A. C. Naveira Garabato, G. M. Damerell, and K. J. Heywood, 2016: Open-ocean submesoscale motions: A full seasonal cycle of mixed layer instabilities from gliders. *J. Phys. Oceanogr.*, **46**, 1285–1307, <https://doi.org/10.1175/JPO-D-15-0170.1>.
- Toole, J. M., 1998: Turbulent mixing in the ocean: Intensity, causes, and consequences. *Ocean Modeling and Parameterization*,

- E. P. Chassignet and J. Verron, Eds., NATO Science Series, Vol. 516, Springer, 171–190.
- Umlauf, L., and H. Burchard, 2003: A generic length-scale equation for geophysical turbulence models. *J. Mar. Res.*, **61**, 235–265, <https://doi.org/10.1357/002224003322005087>.
- Van Roekel, L. P., B. Fox-Kemper, P. P. Sullivan, P. E. Hamlington, and S. R. Haney, 2012: The form and orientation of Langmuir cells for misaligned winds and waves. *J. Geophys. Res.*, **117**, C05001, <https://doi.org/10.1029/2011JC007516>.
- Wang, P., J. C. McWilliams, J. Yuan, and J.-H. Liang, 2025: Langmuir mixing schemes based on a modified K-profile parameterization. *J. Adv. Model. Earth Syst.*, **17**, e2024MS004729, <https://doi.org/10.1029/2024MS004729>.
- Weller, R. A., and J. F. Price, 1988: Langmuir circulation within the oceanic mixed layer. *Deep-Sea Res.*, **35A**, 711–747, [https://doi.org/10.1016/0198-0149\(88\)90027-1](https://doi.org/10.1016/0198-0149(88)90027-1).
- Wong, A. P. S., and Coauthors, 2020: Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Front. Mar. Sci.*, **7**, 700, <https://doi.org/10.3389/fmars.2020.00700>.
- Yuan, J., and J.-H. Liang, 2021: Wind- and wave-driven ocean surface boundary layer in a frontal zone: Roles of submesoscale eddies and Ekman-stokes transport. *J. Phys. Oceanogr.*, **51**, 2655–2680, <https://doi.org/10.1175/JPO-D-20-0270.1>.
- , —, E. P. Chassignet, O. Zavala-Romero, X. Wan, and M. F. Cronin, 2024: The K-Profile Parameterization augmented by Deep Neural Networks (KPP_DNN) in the General Ocean Turbulence Model (GOTM). *J. Adv. Model. Earth Syst.*, **16**, e2024MS004405, <https://doi.org/10.1029/2024MS004405>.
- Zavala-Hidalgo, J., S. L. Morey, and J. J. O'Brien, 2003: Seasonal circulation on the western shelf of the Gulf of Mexico using a high-resolution numerical model. *J. Geophys. Res.*, **108**, 3389, <https://doi.org/10.1029/2003JC001879>.
- , A. Gallegos-García, B. Martínez-López, S. L. Morey, and J. J. O'Brien, 2006: Seasonal upwelling on the western and southern shelves of the Gulf of Mexico. *Ocean Dyn.*, **56**, 333–338, <https://doi.org/10.1007/s10236-006-0072-3>.
- Zavala-Romero, O., A. Bozec, E. P. Chassignet, and J. R. Miranda, 2025: Convolutional neural networks for sea surface data assimilation in operational ocean models: Test case in the Gulf of Mexico. *Ocean Sci.*, **21**, 113–132, <https://doi.org/10.5194/os-21-113-2025>.
- Zhu, Y., R.-H. Zhang, J. N. Moum, F. Wang, X. Li, and D. Li, 2022: Physics-informed deep-learning parameterization of ocean vertical mixing improves climate simulations. *Natl. Sci. Rev.*, **9**, nwac044, <https://doi.org/10.1093/nsr/nwac044>.