

1 **Enhancing Ocean Model Accuracy with Neural Network-Informed K-Profile**
2 **Parameterization for Vertical Mixing: A Test Case in the Gulf of Mexico**

3
4 Jianguo Yuan,^{a,b} Jun-Hong Liang,^{a,b} Eric P. Chassignet,^c Alexandra Bozec,^c and Alan J. Wallcraft^c

5
6 *^a Department of Oceanography & Coastal Sciences, Louisiana State University, Baton Rouge, LA, USA*

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

8 *^c Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA*

9
10 *Corresponding author: Jun-Hong Liang, jliang@lsu.edu*

ABSTRACT

Turbulent mixing in the ocean surface boundary layer is unresolved in ocean models and is parameterized using formulas based on physics 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 is 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 DNN-informed KPP improves sea surface temperature (SST) simulations, particularly in winter, and effectively mitigates the warm SST biases in the GOM simulated using the original KPP. Additionally, it enhances the accuracy of the simulated mixed layer depth (MLD), better capturing its magnitude, distribution, and variability. Sensitivity experiments demonstrate that the inclusion of ocean surface waves in the DNN input is essential for optimal model performance.

Keywords: Turbulent mixing; K-profile parameterization; Neural network; Sea surface temperature; Mixed layer depth; HYCOM

1. Introduction

The Ocean Surface Boundary Layer (OSBL), typically extending from a few tens to hundreds of meters below the ocean surface, is a critical zone dominated by various turbulent mixing processes such as wind-driven shear, buoyancy forcing, and wave-induced Langmuir turbulence (Belcher et al. 2012; Liang et al. 2013; 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; 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 has a profound impact on 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 produced by turbulence-resolving models such as Large Eddy Simulations (LES). While observational data were directly utilized to develop OSBL mixing parameterization in early studies (e.g., Large et al. 1994), they are impacted by submesoscale, mesoscale and other larger-scale processes (Thompson et al. 2016), making them less suitable than turbulence-resolving simulations for this purpose. Properties extracted from LES solutions have been widely used to derive or calibrate structure functions in parameterization schemes (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 General Ocean Turbulence Model (GOTM), which serve as efficient testbeds for structural comparisons and sensitivity analysis (e.g., He and Chen 2011; Li et al. 2019; George et al. 2025; Wang et al. 2025). Ultimately, parameterization schemes must be implemented into three-dimensional ocean models, where their impacts on key state variables, such as SST and MLD, can be rigorously evaluated against observational datasets (e.g., Li et al. 2016; Ali et al. 2019; Sane et al. 2023; Sane et al. 2025).

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), yet they remain inaccurate under certain conditions, as some physical processes modulating turbulent mixing were not originally considered and included in the parameterizations. One of such examples is the effect of non-breaking surface waves and the associated Langmuir turbulence, which LES simulations (e.g., Skillingstad et al. 1996; McWilliams et al. 1997; McWilliams and Sullivan 2000) and observations (e.g., Weller and Price 1988; D'Asaro and Dairiki 1997) show can

elevate turbulent mixing intensity, deepening MLDs and enhancing entrainment fluxes. However, when wind and wave are misaligned, turbulent mixing can be suppressed (e.g., Van Roekel et al. 2012).

Attempts have been made by various studies to incorporate non-breaking 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 wave effects (e.g., Li et al. 2019; Li et al. 2021; Yuan et al. 2024). In addition, no consensus has been reached on the optimal approach (Li et al. 2019).

Wave-informed KPP schemes have been incorporated into three-dimensional (3D) ocean models and evaluated against both baseline models without wave effects and observational data. The outcomes using specific schemes vary depending on the specific 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 also shown that some of the schemes tends 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 mid-latitudes in northern hemisphere in winter, simulated MLDs may have already been overestimated in baseline models (Fan and Griffies 2014; Li et al. 2016; Ali et al. 2019). In such cases, further adding wave enhancement can degrades 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 concern

(Brenowitz et al. 2020; Rasp 2020; Chattopadhyay and Hassanzadeh 2023). Other approaches have used DNNs to predict key parameters within well-established schemes while keeping the remainder of the parameterization structure unchanged (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 factors in KPP scheme: the enhancement factors for the velocity scale and unresolved shear. The trained DNNs were implemented into GOTM model and compared against traditional deterministic models. Their results show that the GOTM model with DNN informed KPP not only computational efficient but also yields reduced errors in both sea surface temperature (SST) and mixed layer depth (MLD). Furthermore, their results also revealed more complex, yet structured relationships between enhancement factors and physical variables such as MLD and turbulent Langmuir number, relationships that are not captured by traditional wave-informed deterministic schemes.

This study builds upon and further extends Yuan et al. (2024)'s work by adopting their approach of using deep neural networks (DNNs) to predict the two enhancement factors in the KPP scheme. However, instead of testing the method in a one-dimensional model, the well-tuned DNNs are integrated into the three-dimensional HYbrid Coordinate Ocean Model (HYCOM) to simulate the evolution of temperature, salinity, MLD and large-scale circulation in the Gulf of Mexico (GOM). The model outputs are compared against both observational data and HYCOM simulations using the baseline KPP scheme without wave effects, to evaluate whether the DNN-informed KPP reduces the biases observed in the traditional parameterization and better aligns with observations. In addition, the DNN inputs are revised to be more representative of the broader GOM region, rather than being limited to a single fixed location as in Yuan et al. (2024).

The remainder of this paper is organized as follows: Section 2 describes the generation of the training data for the DNNs, the configuration of the DNN models, and the implementation of the trained DNNs into the HYCOM framework. Section 3 presents the simulation results from HYCOM, 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 key coefficients introduced to the baseline model to account for wave effect. 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 the vertical velocity. Overbars indicate ensemble averages, while primes denote turbulent fluctuations. K_x is the diffusivity or viscosity. γ_x accounts for the contribution due to non-local effects that are not proportional to local gradients only applied to scalar fluxes (e.g., temperature and salinity) and equals to zero for momentum. In this study, only K_x is modified.

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

$$K_x(\sigma) = w_x(\sigma) h G_x(\sigma) \quad (2)$$

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

The OSBL depth is diagnosed using the Bulk Richardson number:

$$Ri_b(z) = \frac{z \left(b_r - \overline{b(z)} \right)}{\left(\mathbf{u}_r - \overline{\mathbf{u}(z)} \right)^2 + U_t^2(z)} \quad (3)$$

Where b is the buoyancy, $\mathbf{u}$ is the water current vector. The subscript r represents the reference value. The effect of unresolved shear is parameterized as:

$$U_t^2(z) = \frac{C_v N(z) w_x(z) |z|}{Ri_c} \quad (4)$$

Where C_v is a dimensionless constant and $N(z)$ is the Brunt-Väsälä frequency.

To account for additional physical processes that are absent (e.g., non-breaking 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, equations (1) and (4) become:

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

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

Here, the subscript “LMD” (Large, McWilliams and Doney) 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 semi-empirical equations tuned with LES solutions of a limited set of forcing conditions and only used to quantify effects due to non-breaking 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, 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 model for the OSBL (NCAR-LES, Sullivan and McWilliams 2010). The NCAR-LES model has been applied to study mixing processes under various external atmospheric forcing, turbulent mixing and entrainment enhancement due to breaking and non-

breaking waves, mixing and entrainments associated with a frontal zone and submesoscale instabilities, tracer dispersion driven by turbulence, and ocean response to extreme forcing such as hurricanes (Kukulka et al. 2009; Hamlington et al. 2014; Liang et al. 2018; Yuan and Liang 2021). Turbulence-resolving models such as the NCAR-LES model offers an economical approach to generate extensive training data sets for machine learning based parameterizations. In this study, a suite of NCAR-LES simulations are conducted at various locations in the GOM (Fig. 1) to generate turbulence-resolving solutions for DNN training.

Over 700 locations within the domain (Fig. 1) over water points are randomly selected to ensure that our LES ensemble captures turbulent mixing under a wide range of meteorological, water-column, and geographic conditions across the GOM. A 5-day LES simulation is performed at each location with the first day of each simulation treated as spin-ups and thus discarded, while solutions from days 2 to 5 are processed for DNN training. Each LES run corresponds to a randomly selected distinct 5-day period. For example, the simulation at location $93.24^{\circ}W, 28.78^{\circ}N$ spans from Dec.26 2010 to Dec.30 2010, while the simulation at location $82.2^{\circ}W, 18.73^{\circ}N$ covers from Sep.22 2011 to Sep.26 2011. Altogether, these simulations span the period of 10 years from December 26, 2010, to December 28, 2020.

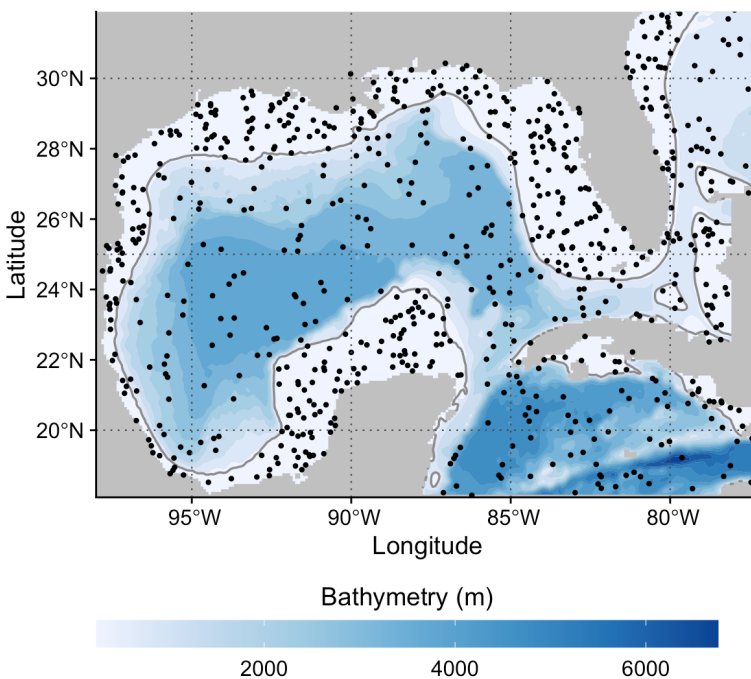


Fig. 1. Map of Gulf of Mexico study region. The blue shading indicates bathymetry, with grey contour lines at 300 meters. Black dots are the selected locations where large eddy simulations are conducted. Land mask is shown in grey.

All LES simulations are forced by atmospheric and oceanic conditions derived from the ERA5 reanalysis (Hersbach et al. 2020), including zonal and meridional winds, net short-wave radiation, net heat flux excluding short wave 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 and forcing fields are interpolated to the specific spatial 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, the shortwave radiation penetrates the water column. Here, the Jerlov water type 1 (Jerlov 1976) is used to calculate the shortwave radiation at depth z , given by:

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

$q_{sw}(0)$ is the net shortwave radiation at the surface, and the constants $r = 0.58, \mu_1 = 0.35, \mu_2 = 23$ correspond to Jerlov water type 1. In HYCOM model, 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 bottom-reaching before adjustment. 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.

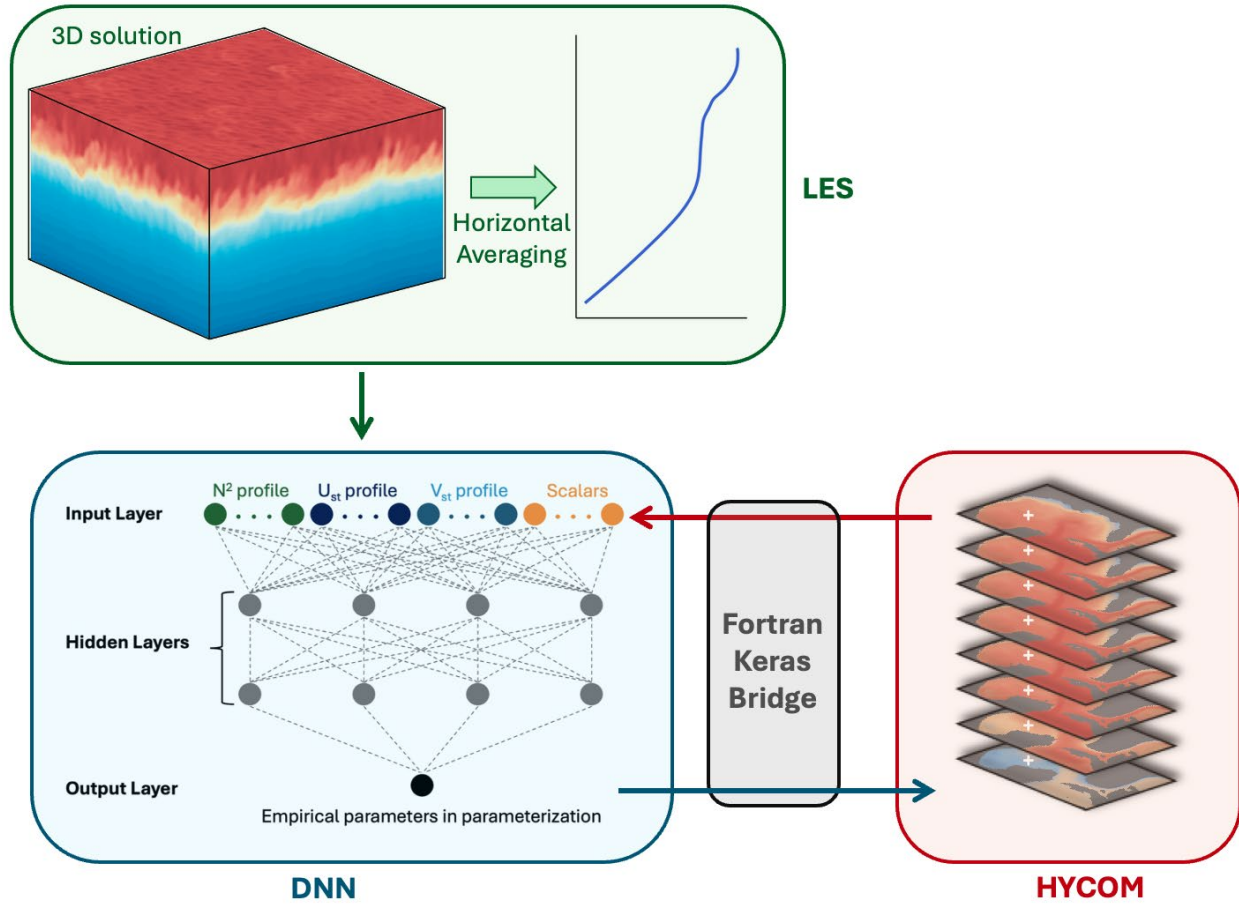
217

218 *c. Data standardization and DNN training*

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

229 Compared with Yuan et al (2024), the DNN models in this study introduce three additional
230 variables into the input array: the latitude to indicate effects due to Coriolis effects, the OSBL
231 depth to indicate the absolute depth range of input profile variables, and the bathymetry depth
232 indicating the data sparsity passed from HYCOM solutions to DNN models. It also replaces
233 temperature and salinity profiles with profiles of the Brunt-Vaisala frequency squared, to avoid
234 DNNs making unplausible predictions when the simulated temperature and salinity are outside
235 the ranges of the training data, a possibility in multi-year or multi-decadal simulations.

236 Three DNN models are trained to predict two scalar variables: models $D_{\eta 0}$ and D_{η} are used
237 to predict the enhancement to the unresolved shear contribution (η), while model D_{ϵ} is used to
238 predict the velocity scale coefficient (ϵ) in KPP formula. In shallow-water regions, h often
239 reaches the seafloor even when $\eta = 0$, resulting in a high occurrence of $\eta = 0$ in the data. To
240 better capture this, we adopt a two-step prediction strategy: model $D_{\eta 0}$ first predicts whether $\eta =$
241 0, and if not, D_{η} predicts the actual non-zero value of η . For predicting η , h from the previous
242 time step is included as an input. When predicting ϵ in model D_{ϵ} , h at current time step is used,
243 as ϵ depends directly on h at the same time step.



244

245 Fig. 2. Conceptual framework of the modeling system. LES generates turbulence-resolving 3D
 246 fields that are horizontally averaged and employed to train DNNs. In HYCOM runs, the model
 247 provides profile and scalar variables as inputs to the trained DNNs, and the predicted empirical
 248 parameters are passed back to HYCOM. The information exchange between HYCOM and the
 249 DNN is implemented through the Fortran Keras Bridge (FKB).

250 All input variables are normalized before being fed into DNNs. Each input variable is
 251 standardized by subtracting its mean, then divided by its standard deviation. For scalar input
 252 variables, an additional sigmoid transformation $f(x) = 1/(1 + e^{-x})$ is then applied to the
 253 normalized value. This use of the sigmoid function is heuristic, introduced after we observed that
 254 it improves prediction accuracy during model training. For the outputs, η in D_η and ϵ in D_ϵ are
 255 transformed using a base-10 logarithm, i.e., $\eta_{norm} = \log_{10} \eta$ and $\epsilon_{norm} = \log_{10} \epsilon$. Data are
 256 randomly separated into training, validation and testing datasets at a ratio of approximately 6:2:2.
 257 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. The binary cross-entropy is used as loss function in $D_{\eta 0}$, while the mean absolute error is used in D_{η} and D_{ϵ} . A leaky ReLU activation, $f(x) = \max(x, \alpha x)$ with $\alpha = 0.1$, is applied in all hidden layers, and the Adam optimizer is used for training. The DNN models are trained and tested using TensorFlow and Keras in R.

To identify the best-performing models, we evaluate architectures ranging from 1 to 10 hidden layers and from 4 up to 128 neurons in each hidden layer. Each model is trained for 1000 epochs, with the learning rate reduced by a factor of 0.1 if no improvement in validation loss is observed over 5 consecutive epochs. Models with the lowest validation losses (for D_{η} and D_{ϵ}) and the highest classification accuracy (for $D_{\eta 0}$) are selected for implementation in the HYCOM model. Configurations of the best trained DNN models are shown in Table 1. Model performance, including training and validation losses over epochs, and the comparison between predicted and target value distributions, is shown in Figs. S1 and S2. Offline tests confirm that the trained models reliably capture the essential data distributions, justifying their implementation into ocean models.

Scheme	Model	Hidden layers	Neurons in each hidden layer
$KPP_{DNN,NW}$	$D_{\eta 0}$	2	16
	D_{η}	5	8
	D_{ϵ}	5	24
$KPP_{DNN,WV}$	$D_{\eta 0}$	5	4
	D_{η}	3	16
	D_{ϵ}	3	12

Table 1. Configurations of the best trained DNNs used in KPP_{DNN} schemes. $KPP_{DNN,NW}$ excludes Stokes drift, while $KPP_{DNN,WV}$ includes it. Model $D_{\eta 0}$ 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 .

d. The implementation of DNN informed KPP into HYCOM

The Hybrid Coordinate Ocean Model (HYCOM, Bleck 2002; Chassignet et al. 2003; Chassignet et al. 2009) is a three-dimensional 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 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 large-scale circulation.

This study modifies the KPP related subroutines in HYCOM to incorporate the use of well-tuned DNNs for predicting η and ϵ . Same as in Yuan et al. (2024), 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 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 de-normalized 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 model configuration

The HYCOM model domain is shown in Fig. 1. It spans from 18°N to 32°N in latitude and from 77°W to 98°W in longitude. As shown in Fig. 1, 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 grids in the zonal and meridional directions, respectively, and a horizontal resolution of the model is $1/12^\circ$. Vertically, the model

is configured with 41 hybrid layers. The upper 83 meters of the water column consists of 14 sigma-Z layers, which are terrain-following where the seafloor depth is less than 83 meters. In water deeper than 83 meters, layers 15 to 41 are isopycnal unless they outcrop into the near-surface Z (fixed depth) regime. The top layer is 1 meter thick everywhere, so the model's SST is actually the average over the top 1 m.

Three sets of HYCOM simulations are conducted in this study. HY_{LMD} employs the KPP_{LMD} scheme to parameterize effect due to vertical mixing in the OSBL. $HY_{DNN,WV}$ uses the $KPP_{DNN,WV}$ scheme, with Stokes drift profiles included in the input features are incorporated into HYCOM to predict coefficients, η and ϵ . $HY_{DNN,NW}$ uses $KPP_{DNN,NW}$ scheme, excluding Stokes drift information from the input arrays. 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° Reanalysis on Jan. 1, 2001 (GLBu0.08, https://tds.hycom.org/thredds/catalogs/GLBu0.08/expt_19.1.html) and integrated for 22 years from year 2001 to year 2022, with the first year treated as model spin-up. Atmospheric forcing required by HYCOM is provided by the hourly Climate Forecast System Reanalysis (CFSR) for year 2001~2010, and from the hourly Climate Forecast System Version 2 (CFSv2) for year 2011~2022. Stokes drift is one of the inputs in $HY_{DNN,WV}$. To avoid the high computational cost of coupled ocean-wave modeling (Li et al. 2016), Stokes drift profiles in both zonal and meridional directions are computed offline from ERA5 wave spectra (Hersbach et al. 2020). These wave forcing fields are treated as forcing by HYCOM, 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. Although our primary goal is to demonstrate how DNNs can improve the KPP scheme's representation of vertical mixing, it is important to note that biases may stem from other processes as well, such as the misrepresentation of large-scale horizontal circulations, 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 is the Global Ocean Sea Surface Temperature and Sea Ice Reprocessed produced by the UK 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 2002–2021, separately for summer months (June, July, and August) and winter months (December, January, and February).

Fig. 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, with SSTs there 3 to 4 lower than offshore waters, and strong horizontal temperature gradients, are results of summer-time coastal upwelling (e.g., Merino 1997; Zavala-Hidalgo et al. 2003; Zavala-Hidalgo et al. 2006; Jouanno et al. 2018).

Figs. 3c, 3e and 3g display the SST differences between model simulations and observations in summer, for HY_{LMD} , $HY_{DNN,NW}$, and $HY_{DNN,WV}$, respectively. In the deep open GOM, SST deviations are small, within $\pm 0.25^\circ\text{C}$ from observations. Positive deviations dominate in HY_{LMD} , while negative deviations prevail in $HY_{DNN,NW}$ and $HY_{DNN,WV}$. However, both models exhibit larger errors along the coasts, particularly in regions with sharp SST gradients, such as the upwelling zones, highlighting the persistent challenge of resolving coastal dynamics.

Fig. 3b 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, 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.

Figs. 3d, 3f and 3h show the winter SST deviations from observations for HY_{LMD} , $HY_{DNN,NW}$, and $HY_{DNN,WV}$, respectively. Except in 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 smaller. $HY_{DNN,NW}$ does not mitigate this domain-wide warm bias. In fact, significant positive anomalies 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 anomaly belt evident in HY_{LMD} and $HY_{DNN,NW}$ becomes less distinct. However, near the LC core region, a bias dipole pattern emerges, characterized by a strong positive anomaly ($> 0.75^{\circ}\text{C}$) to the west (north of Yucatan Peninsula) and negative anomaly ($< -0.75^{\circ}\text{C}$) to the east (northwest of Cuba).

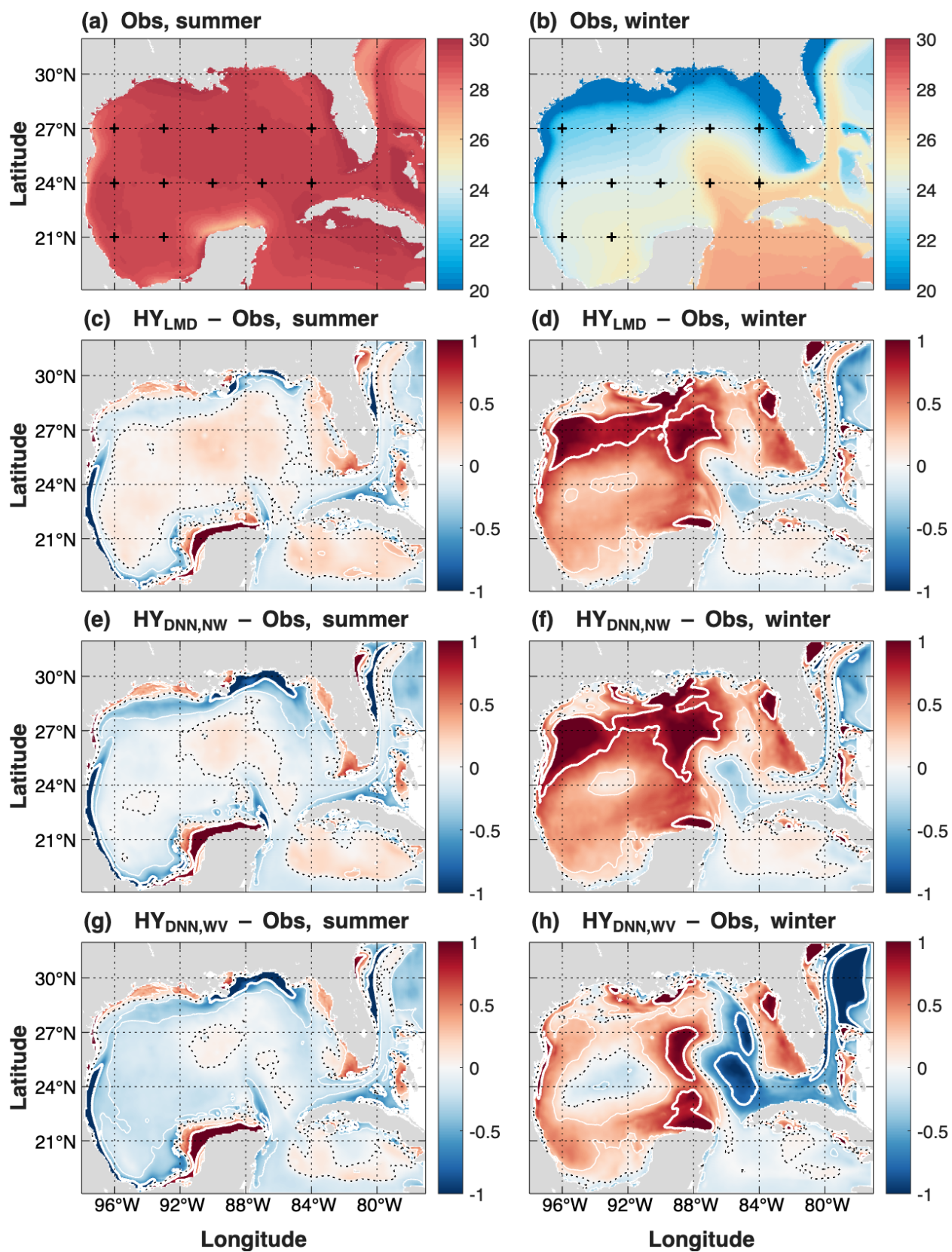


Fig. 3. Climatological mean sea surface temperature (SST) in the Gulf of Mexico (GOM) from 2002 to 2021. The left column (panels a, c, e and g) shows summer (June–August) SST, and the right column (panels b, d, f and h) shows winter (December–February) SST. Panels (a) and (b): observed mean SST. Panels (c) and (d): SST bias between the HY_{LMD} and observations. Panels (e) and (f): SST bias between $HY_{DNN,NW}$ and observations. Panels (g) and (h): SST bias between $HY_{DNN,WV}$ and observations. The black dashed lines in panels (b–h) indicate the 0 °C bias contour, thin black solid lines indicate ± 0.25 °C bias, and thick white lines indicate ± 0.75 °C bias. The plus signs in panels (a) and (b) mark the locations of selected sites used for the bias distribution analysis shown in Fig. 5.

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 observed standard deviations. In summer, $HY_{DNN,WV}$ outperforms HY_{LMD} , showing a stronger correlation and a slightly lower RMSE. In contrast, $HY_{DNN,NW}$ performs the worst among the three, implying that omitting Stokes drift information limits the predictive capability of the trained DNNs. In winter, the differences between HY_{LMD} and $HY_{DNN,WV}$ become subtler. HY_{LMD} achieves higher correlation whereas standard deviation by $HY_{DNN,WV}$ is closer to unity. Their RMSEs are comparable. Once again, $HY_{DNN,NW}$ shows degraded performance, highlighting the consistent importance of including Stokes drift profiles as input features for improved model skill.

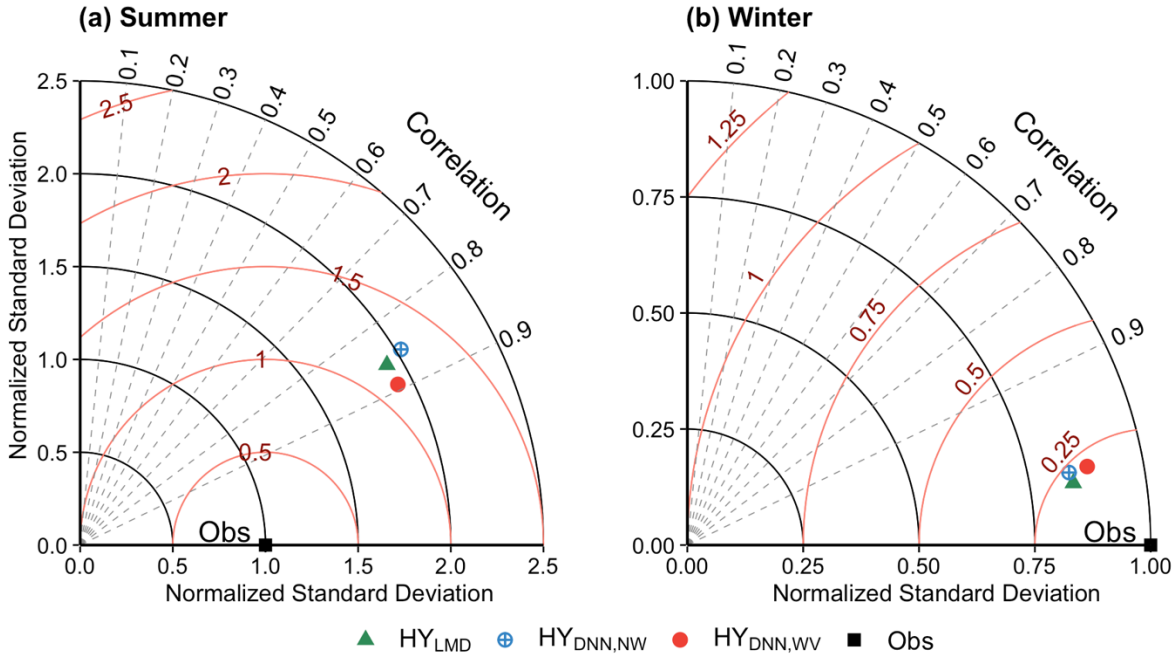


Fig. 4. Taylor diagrams comparing the mean sea surface temperature (SST) in the Gulf of Mexico (GOM) from 2002 to 2021. Panel (a) shows results for summer (June–August), and panel (b) for winter (December–February). In each diagram, the azimuthal position indicates the correlation coefficient between model simulations and observations, the radial distance indicates the normalized standard deviation (relative to the observed SST), and the red contours represent the normalized root mean square error (RMSE).

To assess the localized performance of SST simulations, 12 representative sites, marked by plus signs in Figs. 3a and 3b, are selected. Fig. 5 compares the density distributions of daily SST biases at these locations for the three simulations. At most sites, $HY_{DNN,WV}$ exhibits the best agreement with observations, characterized by bias distributions that are sharply peaked near zero and more symmetric about the zero-bias axis. In contrast, HY_{LMD} and $HY_{DNN,NW}$ often show rightward-shifted density peaks, particularly at locations (c–g, i–k), indicating a systematic warm bias relative to observations.

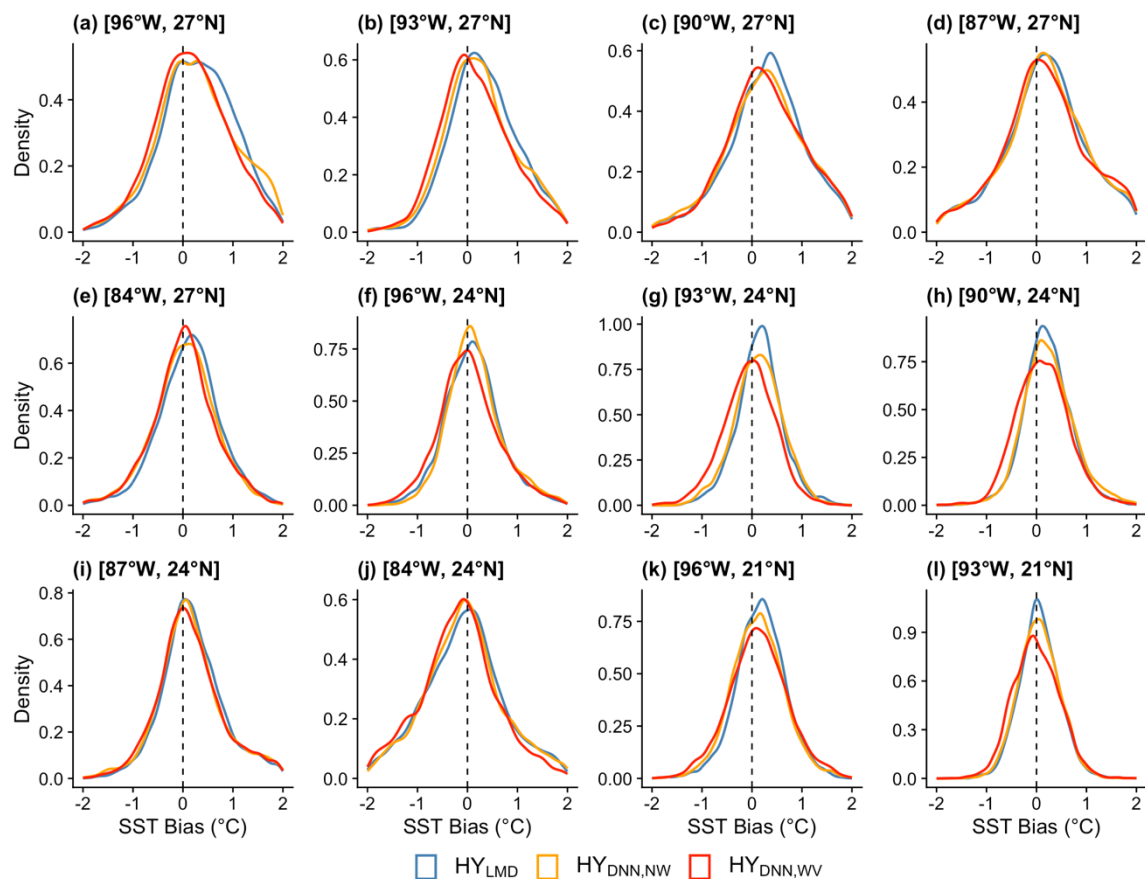


Fig. 5. Probability density distributions of daily SST biases (model minus observation) at 12 selected sites in the GOM (shown as plus signs in Fig. 3) during 2002–2021. Each panel corresponds to a specific site, with the longitude and latitude indicated in the panel title.

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 a reference depth of 10 m. Observed MLDs are calculated using vertical profiles of temperature, salinity, and pressure, obtained from Argo floats (Wong et al. 2020) in the GOM. Although 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 deep open ocean before 2009 (Roemmich and Team 2009), increased deployment in marginal seas like GOM started from 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 measured the open ocean where the water exceeds 1000 m,

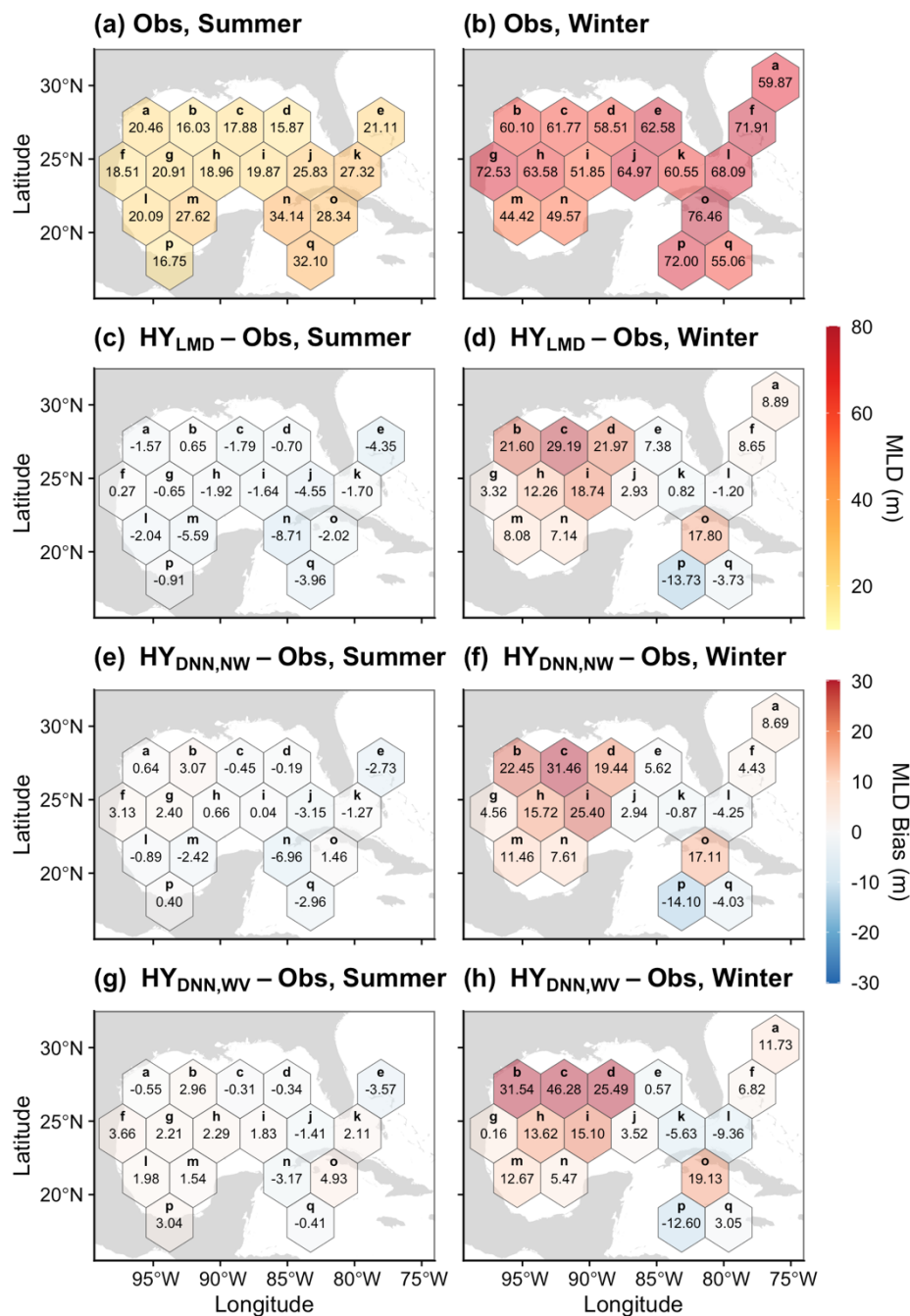
while observations in shallower regions (<1000 m) are scarce (Fig. S3b). Even within the deep ocean regions of the GOM, Argo profiles are not uniformly distributed, as data is more abundant in the northern and eastern portions, whereas the southwestern area, including the Bay of Campeche, has significantly fewer observations.

Fig. 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 (green) and the three 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 enable a fair and consistent comparison with the unevenly distributed Argo observations in both space and time, 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 meters. Spatial variations in summer MLDs are relatively small, with the deepest MLD near Yucatan Channel (~ 35 m), the core region of LC, while the shallowest MLD offshore of Louisiana (~ 15 m), where river freshwater flux from Mississippi river helps stabilizing the upper ocean. Mean MLD biases from all three HYCOM simulations are small. HY_{LMD} (Fig. 6c) underestimates MLD over most of the gulf, with the largest negative biases (~ 10 m) in the Yucatan Channel. With wave information incorporated, both $HY_{DNN,NW}$ and $HY_{DNN,WV}$ (Figs. 6e and 6g) deepens the MLDs, with positive biases prevailing in the west GOM and shallow MLD biases reduced in the Yucatan Channel and west of Florida Peninsula.

In contrast, the observed mixed layer is substantially deeper, and its depth is also more spatially variable in winter over the GOM (Fig. 6b). Two distinct regions exhibit particularly deep winter MLDs: one near the Loop Current system, and the other in the west-central GOM (hexagons g and h). All three HYCOM simulations tend to overestimate winter MLDs across the GOM, especially in the north-central region. Among the evaluated regions, hexagon c exhibits the largest positive MLD bias, with mean winter MLD overestimated by ~ 25 - 30 m in HY_{LMD} (Fig. 6d). The inclusion of wave effects in $HY_{DNN,NW}$ and $HY_{DNN,WV}$ further increases the bias, reaching ~ 30 m (Fig. 6f) and ~ 40 m (Fig. 6h), respectively. This suggests that while wave-

458 induced mixing enhances vertical mixing in certain areas, it may also lead to excessive
 459 deepening of the MLD in regions where background stratification is already weak.



460
 461 Fig. 6. Comparison of seasonal mean mixed layer depth (MLD) between HYCOM
 462 simulations and observations. The left column presents MLD climatology for summer (June–
 463 August), and the right column for winter (December–February). Panel (a) and (b) show the
 464 observed mean MLDs, while panels (c)–(h) display mean MLD biases (simulation minus

observation) for HY_{LMD} , $HY_{DNN,NW}$ and $HY_{DNN,WV}$. 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.

However, the larger overestimation of winter MLDs in $HY_{DNN,NW}$ and $HY_{DNN,WV}$ compared to HY_{LMD} does not necessarily indicate poorer performance. To gain a more complete picture of model behavior, boxplots of winter MLD distributions are constructed for five representative winter hexagonal regions (c, d, e, h, and i in the right column of Fig. 6), as shown in Fig. 7. Even though the mean MLDs biases in $HY_{DNN,WV}$ are similar or even larger than in HY_{LMD} , its median MLDs more closely match the observed medians in most regions. In addition, there are large portion of observed winter MLDs falls shallower than 50m, representing shallow mixing events in winter. These shallow MLD events are rarely captured by HY_{LMD} and $HY_{DNN,NW}$, whose interquartile ranges are consistently deeper. In contrast, $HY_{DNN,WV}$ not only better reproduces the observed median MLD values but also produces more consistent interquartile ranges that have a stronger ability to capture the shallow MLD events. An exception is found in hexagon c, where none of the simulations adequately capture the observed MLD distribution. The interquartile ranges from all three models are 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 stratification in this region.

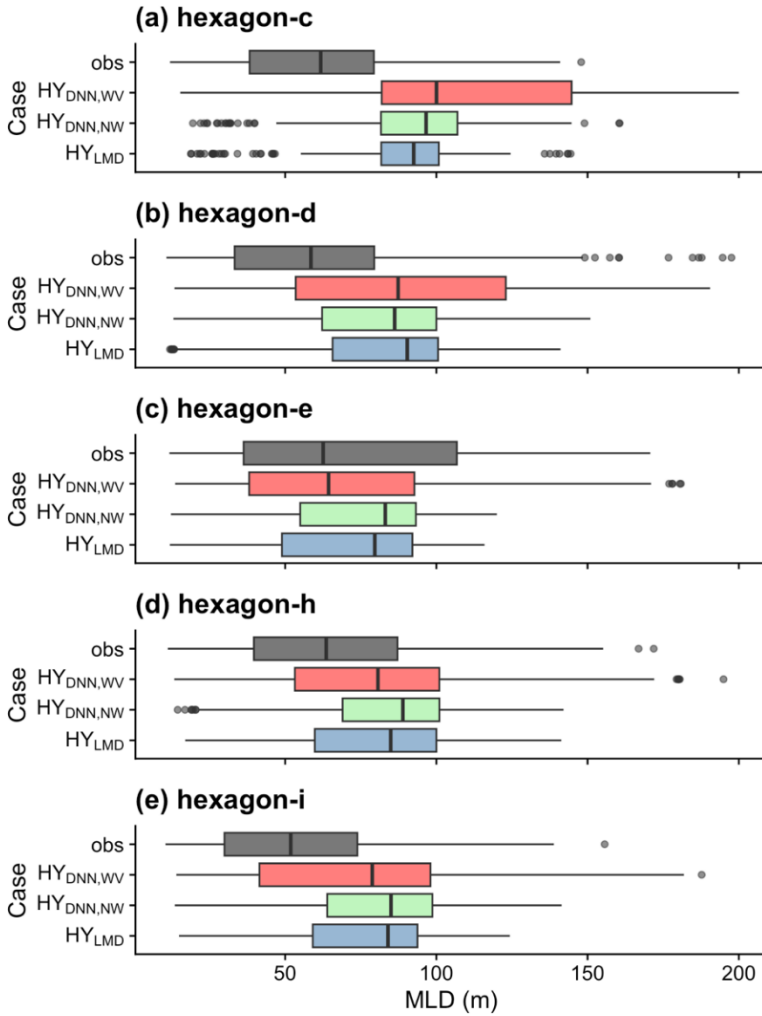


Fig. 7. Boxplot of winter MLD climatology distribution from HYCOM simulations and observations in five selected hexagonal regions (c, d, e, h, and i), as indicated in Fig. 6b. Each panel compares the distribution of MLDs during winter (December–February) among the observation (Obs) and HYCOM simulation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

c. Response of the DNN-informed KPP to Different Forcing Regimes

Surface forcing in the GOM varies markedly in space and time. Key drivers include wind conditions, wind-induced Langmuir waves, as well as 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). Fig. 8 replicates this two-panel diagram: the top row represents destabilizing buoyancy conditions, while the bottom row stabilizing buoyancy conditions. The turbulent Langmuir number La_t along the x axis, the square root of the ratio of friction velocity divided by the surface Stokes drift magnitude as defined in McWilliams et al. (1997), quantifies the relative importance of wind/wave-driven Langmuir turbulence over wind-induced shear turbulence, while the ratio between mixed layer depth h and Langmuir stability depth L_L along y-axis quantifies the relative importance of buoyancy over wind and waves.

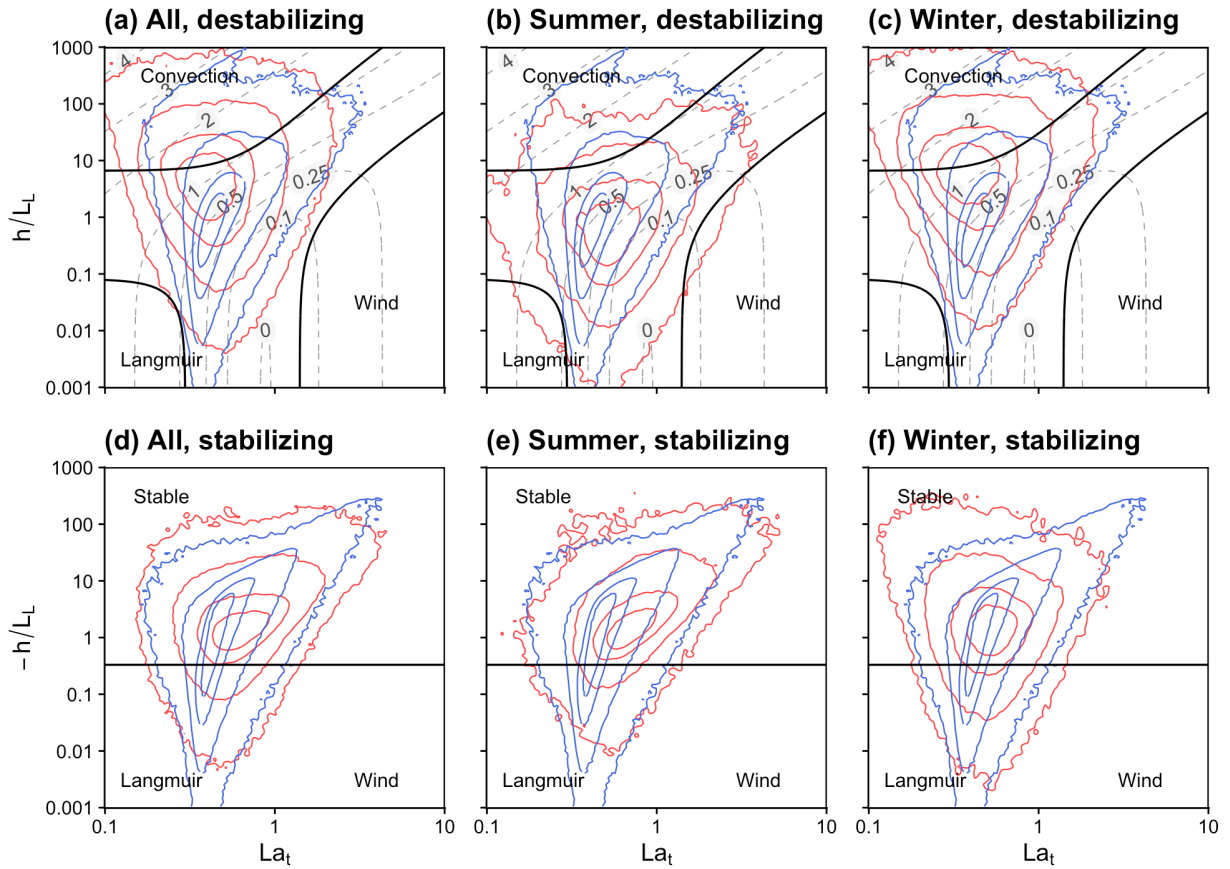


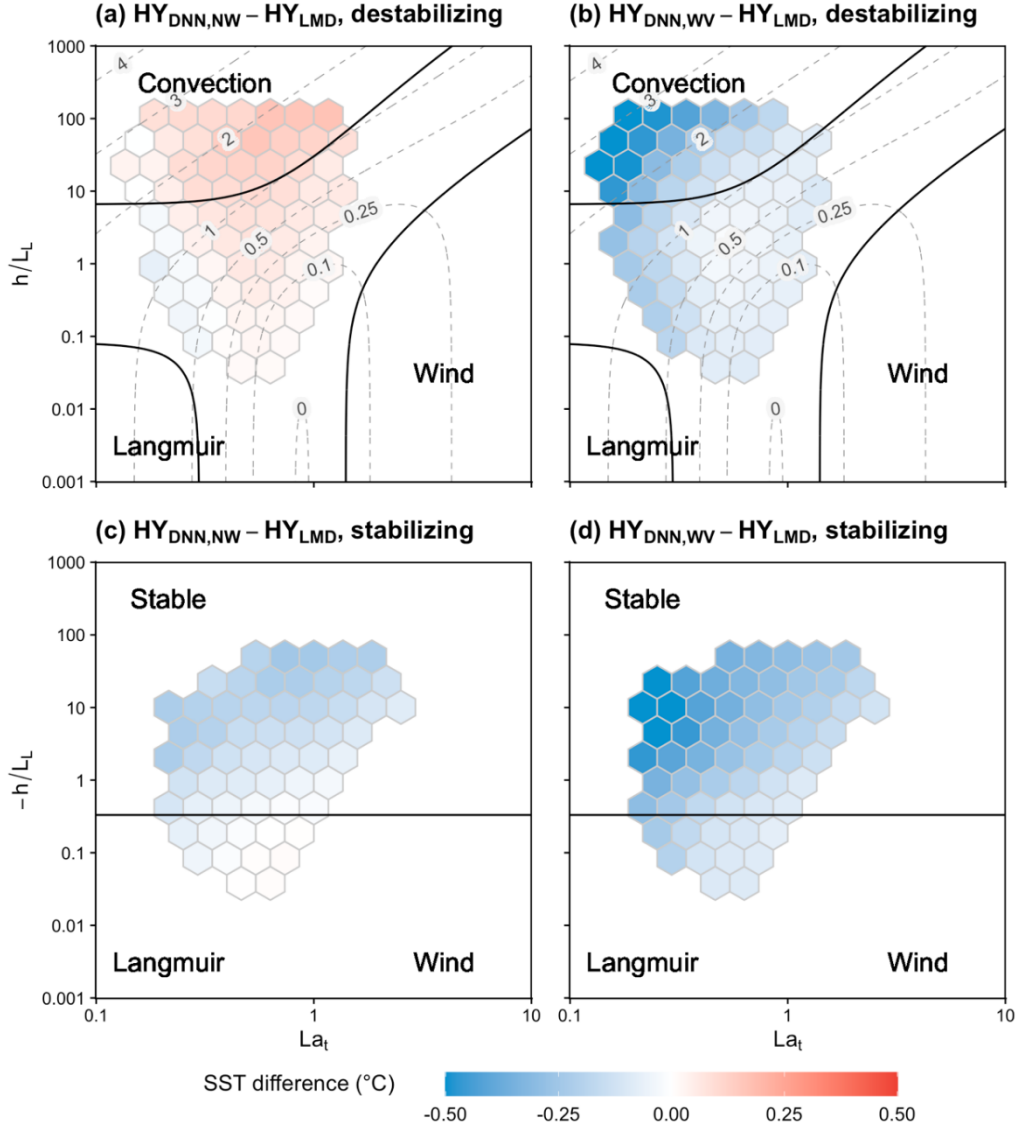
Fig. 8. Regime diagrams of forcing conditions in the GOM from 2002 to 2021. The top row (Panels a-c) depicts conditions under destabilizing buoyancy forcing, while the bottom row (panels d-f) shows conditions under stabilizing buoyancy forcing. In the destabilizing cases, thin dashed contours indicate 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 blackline with $-h/L_L = 1/3$ represents the maximum equilibrium value based on Pearson et al. (2015). Blue contours

represent the probability distribution (30%, 60%, 90% and 99%) of global ocean conditions from Li et al. (2019), while red panels indicate the corresponding distribution in the GOM over the full 20-year period (panels a and d), during summer months (panels b and e) and during winter months (panels c and f).

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. While unstable buoyancy forcing can be the primary driver of OSBL turbulence, wind and wave are rarely the dominant driver. The forcing conditions in the GOM over 20 years are overlayed in Fig. 8 as red contours, with the left, middle, 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, 90% of the winter daily buoyancy forcing is destabilizing, and the winter forcing regime distributions under destabilizing forcing resemble the annual pattern, while 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 DNN-informed KPP schemes affect SST and MLD under different regimes, we calculated daily differences between $HY_{DNN,NW}$ and $HY_{DNN,WV}$ relative to HY_{LMD} . These differences are aggregated into hexagonal bins within the regime diagram, and the mean values within each bin are mapped to visualize systematic patterns.

Fig. 9 shows the hexagonal regime diagrams for SST differences. A consistent pattern across both $HY_{DNN,NW}$ and $HY_{DNN,WV}$ is that as wave effects become more dominant, there is a negative trend of SST, indicating stronger wave induced turbulent mixing efficiently reduces SSTs. Under destabilizing buoyancy forcing condition, $HY_{DNN,NW}$ shows slightly warmer SSTs than the HY_{LMD} (Fig. 8a), further amplifying KPP's warm bias in the GOM during winter. In contrast, $HY_{DNN,WV}$ consistently produces lower SSTs across all regimes (Figs. 8b and 8d), effectively correcting the warm bias.



540

541 Fig. 9. Regime diagrams illustrating the SST differences of HYCOM simulations,
 542 $HY_{DNN,NW}$ (panels a and c) and $HY_{DNN,WV}$ (panels b and d), relative to HY_{LMD} , under different
 543 turbulence forcing regimes. The top row (panels a and b) corresponds to the destabilizing
 544 buoyancy forcing, while the bottom row (panels c and d) corresponds to stabilizing buoyancy
 545 forcing. The color of each hexagon represents the mean SST difference, averaged over all
 546 samples whose forcing conditions fall within that hexagonal bin. Only hexagons with over
 547 100,000 samples are shown. Definition of the gray dashed contours and black solid lines are
 548 consistent with those in Fig. 8.

549 Fig. 10 shows corresponding MLD differences. Unlike SST, MLD difference changes
 550 exhibit a stronger dependence on buoyancy forcing than on wave effects. Under destabilizing
 551 buoyancy conditions (Figs. 10a and 10b), deeper MLDs are found in regions dominated by
 552 convective mixing, with $HY_{DNN,WV}$ exhibiting more pronounced deepening, potentially

contributing to the higher chance of extreme deep MLDs (over 150m) in winter (shown in Fig. 7). Under stabilizing buoyancy condition (Figs. 10c and 10d), MLD deepens with increasing stability. As reported by Yuan et al. (2024), the enhancement of η is stronger when MLD is shallower, corresponding to stronger stabilizing forcing. This may be linked to that the strength of Langmuir circulation is intense near surface (Weller and Price 1988; McWilliams et al. 1997; Li and Fox-Kemper 2020), where wave effects become more effective in enhancing mixing and thus deepening the MLD.

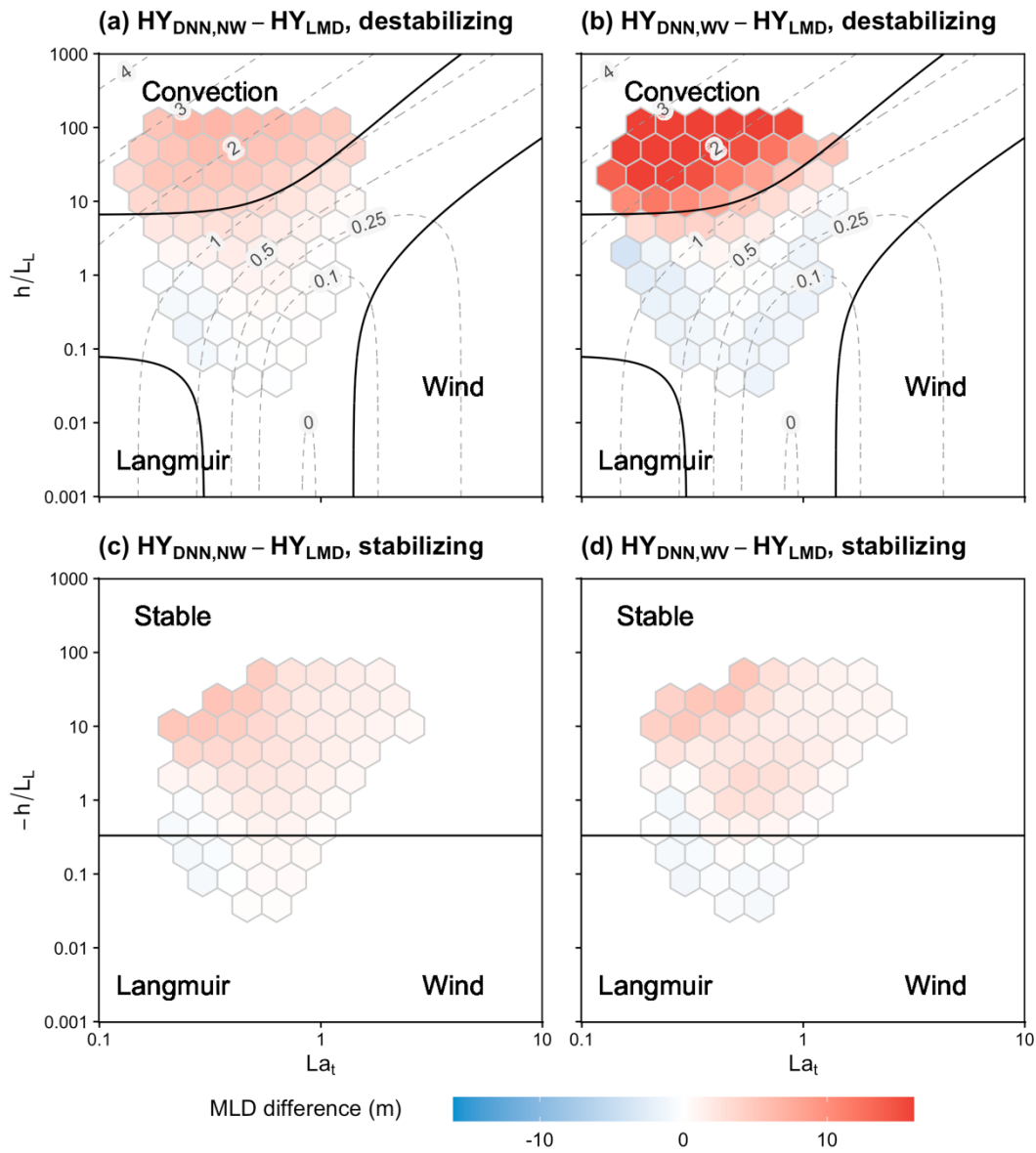


Fig. 10. Same as Fig. 9, but for MLD differences.

Another notable pattern from Fig. 10 is the shoaling of MLDs when wind and wave forcing are more dominating than the buoyancy forcing, especially under destabilizing buoyancy forcing in $HY_{DNN,WV}$. This shoaling results in large portion of shallower MLDs than in HY_{LMD} , thus better agreeing with the distribution of observed MLDs. It is important to note that, this shoaling does not suggest that wave effects potentially suppress turbulent mixing in the real ocean. Instead, it is correcting the overestimation of MLDs by simulations using KPP scheme, whose parameters were derived from datasets outside the GOM.

4. Discussion

In ocean models, biases of simulated SSTs and MLDs arise from inaccurate representation of several interacting processes, including small-scale turbulence parameterization, submesoscale, mesoscale dynamics and large-scale circulation. In this study, only the parameterization for small-scale turbulent mixing is improved. As shown in Fig. S4 in the supplementary materials, KPP_{DNN} schemes have limited impact on sea surface heights (SSH) and high-frequency sea level anomaly (HSLA) fields, and similar bias patterns persist in all three HYCOM simulations when compared with observations, indicating that unmodified large-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 bias, implying stronger anticyclonic circulation, exists in all three simulations in the northwest GOM. Within this dome lies a saddle region near 93°W , characterized by sharp SSH anomaly gradients on both its eastern and western flanks (Figs. S4c, S4e and S4g). This saddle region corresponds to hexagon c during winter months (Fig. 6), where the simulated MLDs from 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 three simulations exhibit similar climatological large-scale circulation, the differences in SSTs and MLDs among simulations can be primarily attributed to the distinct turbulent mixing parameterizations.

Although strong misalignment between wind and wave, when they are opposing each other, suppresses 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. This is also found in $HY_{DNN,WV}$, where mean MLDs are deepened over climatological scales. It indicates enhanced mixing due to Langmuir turbulence is more prevalent than suppressed mixing when waves oppose wind and underscore the importance of incorporating wave effects into vertical mixing parameterization schemes.

Unlike previous works on enhancing KPP_{LMD} scheme mainly only considering wave enhancements, our KPP_{DNN} schemes also incorporate additional relationships into enhancements, such as dependencies on buoyancy stability that are not represented in wave enhancements in deterministic formulations. In the GOM, this results in MLD shoaling 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 superior performance of $HY_{DNN,WV}$ in terms of simulated SST and MLD relative to $HY_{DNN,NW}$ highlights the extensive range of wind and wave conditions even over a regional ocean such as the GOM and underscores the significance of incorporating wave conditions in mixing parameterizations.

5. Conclusion

In this study, deep neural networks (DNNs) are trained on over 700 turbulence-resolving large eddy simulation (LES) solutions, driven by a wide range of different forcing conditions in the Gulf of Mexico region 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. Three HYCOM simulations are conducted to evaluate the impact of non-breaking surface waves on vertical mixing. HY_{LMD} uses the baseline KPP_{LMD} scheme without any wave effects, $HY_{DNN,NW}$ excludes Stokes drift information from DNN inputs, and $HY_{DNN,WV}$ includes it.

Key conclusions are summarized as follows:

- $HY_{DNN,WV}$ improves the representation of sea surface temperature (SST) and mixed layer depth (MLD) distribution, especially during winter when traditional schemes often suffer from systematic biases.
- Under conditions where wave-induced turbulence is more dominant, $HY_{DNN,WV}$ tends to produce a stronger cooling effect in SST compared to the 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} .
- However, improvements in vertical mixing alone are not sufficient to resolve SST and MLD biases, since some biases are caused by large-scale circulation in the GOM.

This study demonstrates the promise of using machine learning to enhance traditional vertical-mixing schemes, yielding improved SST and MLD using three-dimensional realistic ocean models. However, it also indicates that SST and MLD variability depend not only on vertical mixing, but the impact of large-scale circulation is also critical. Future model development should consider combining DNN-based schemes with targeted approaches to capture the structure and variability of the LC, potentially through additional machine learning tools or deterministic corrections grounded in physical diagnostics. Moreover, expanding the coverage and diversity of LES training datasets, especially in dynamically complex regions such as high latitudes, equatorial areas and coastal regions, 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.

Acknowledgements.

JY and JHL were supported by the Office of Naval Research under Grant No. N00014-23-1-2553. JHL was also supported by the National Science Foundation under Grant No. OCE1945502. 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 models is attached to the submission and will be made available through a GitHub page in the final version of the paper.

REFERENCES

- Ali, A., and Coauthors, 2019: A comparison of Langmuir turbulence parameterizations and key wave effects in a numerical model of the North Atlantic and Arctic Oceans. *Ocean Modelling*, **137**, 76-97.
- Belcher, S. E., and Coauthors, 2012: A global perspective on Langmuir turbulence in the ocean surface boundary layer. *Geophysical Research Letters*, **39**, L18605.
- Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean modelling*, **4**, 55-88.
- Brenowitz, N. D., T. Beucler, M. Pritchard, and C. S. Bretherton, 2020: Interpreting and stabilizing machine-learning parametrizations of convection. *Journal of the Atmospheric Sciences*, **77**, 4357-4375.

- 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. *Journal of Physical Oceanography*, **33**, 2504-2526.
- Chassignet, E. P., and Coauthors, 2009: US GODAE Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography*, **22**, 64-75.
- 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 preprint arXiv:2304.07029*.
- D'Asaro, E. A., 2014: Turbulence in the upper-ocean mixed layer. *Ann Rev Mar Sci*, **6**, 101-115.
- D'Asaro, E. A., and G. T. Dairiki, 1997: Turbulence intensity measurements in a wind-driven mixed layer. *Journal of physical oceanography*, **27**, 2009-2022.
- Fan, Y., and S. M. Griffies, 2014: Impacts of parameterized Langmuir turbulence and nonbreaking wave mixing in global climate simulations. *Journal of Climate*, **27**, 4752-4775.
- Fox-Kemper, B., L. Johnson, and F. Qiao, 2022: Ocean near-surface layers. *Ocean mixing*, Elsevier, 65-94.
- Fox-Kemper, B., and Coauthors, 2019: Challenges and Prospects in Ocean Circulation Models. *Frontiers in Marine Science*, **6**, 65.
- Gargett, A. E., and J. Marra, 2002: Effects of upper ocean physical processes (turbulence, advection and air-sea interaction) on oceanic primary production. *The sea*, Wiley & Sons, 19-49.
- Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, **45**, 5742-5751.
- George, J., M. 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. *Journal of Physical Oceanography*, **55**, 1067-1081.
- 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 Sensing*, **12**, 720.
- Group, G. B. C., 2019: The GEBCO_2019 Grid - a continuous terrain model of the global oceans and land. NERC British Oceanographic Data Centre.
- 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. *Journal of Physical Oceanography*, **44**, 2249-2272.
- Harcourt, R. R., 2015: An Improved Second-Moment Closure Model of Langmuir Turbulence. *Journal of Physical Oceanography*, **45**, 84-103.
- Harcourt, R. R., and E. A. D'Asaro, 2008: Large-eddy simulation of Langmuir turbulence in pure wind seas. *Journal of Physical Oceanography*, **38**, 1542-1562.
- He, H., and D. Chen, 2011: Effects of surface wave breaking on the oceanic boundary layer. *Geophysical Research Letters*, **38**, L07604.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999-2049.
- Holte, J., L. D. Talley, J. Gilson, and D. Roemmich, 2017: An Argo mixed layer climatology and database. *Geophysical Research Letters*, **44**, 5618-5626.

- Iyer, S., K. G. Hughes, and J. N. Moum, 2025: Predicting ocean turbulence across orders of magnitude using neural networks trained on multiyear observations. *Artificial Intelligence for the Earth Systems*, **4**, 240093.
- 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**, 18-28.
- Jerlov, N. G., 1976: *Marine optics*. Elsevier.
- Jouanno, J., E. Pallàs-Sanz, and J. Sheinbaum, 2018: Variability and dynamics of the Yucatan upwelling: high-resolution simulations. *Journal of Geophysical Research: Oceans*, **123**, 1251-1262.
- Kantha, L. H., and C. A. Clayson, 1994: An Improved Mixed-Layer Model for Geophysical Applications. *Journal of Geophysical Research-Oceans*, **99**, 25235-25266.
- Kukulka, T., A. J. Plueddemann, J. H. Trowbridge, and P. P. Sullivan, 2009: Significance of Langmuir circulation in upper ocean mixing: Comparison of observations and simulations. *Geophysical Research Letters*, **36**, L10603.
- Large, W. G., and P. R. Gent, 1999: Validation of vertical mixing in an equatorial ocean model using large eddy simulations and observations. *Journal of Physical Oceanography*, **29**, 449-464.
- Large, W. G., J. C. McWilliams, and S. C. Doney, 1994: Oceanic Vertical Mixing - a Review and a Model with a Nonlocal Boundary-Layer Parameterization. *Reviews of Geophysics*, **32**, 363-403.
- 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 Biogeochemical Cycles*, **27**, 1001-1012.
- Li, Q., and B. Fox-Kemper, 2017: Assessing the Effects of Langmuir Turbulence on the Entrainment Buoyancy Flux in the Ocean Surface Boundary Layer. *Journal of Physical Oceanography*, **47**, 2863-2886.
- , 2020: Anisotropy of Langmuir turbulence and the Langmuir-enhanced mixed layer entrainment. *Physical Review Fluids*, **5**, 013803.
- Li, Q., 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. *Geoscientific Model Development*, **14**, 4261-4282.
- Li, Q., 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 Modelling*, **103**, 145-160.
- Li, Q., and Coauthors, 2019: Comparing Ocean Surface Boundary Vertical Mixing Schemes Including Langmuir Turbulence. *Journal of Advances in Modeling Earth Systems*, **11**, 3545-3592.
- Liang, J. H., X. Wan, K. A. Rose, P. P. Sullivan, and J. C. McWilliams, 2018: Horizontal Dispersion of Buoyant Materials in the Ocean Surface Boundary Layer. *Journal of Physical Oceanography*, **48**, 2103-2125.
- 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 Biogeochemical Cycles*, **27**, 894-905.
- Liang, J. H., 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 Modelling*, **176**, 102059.

- McPhee, M. G., 1999: Parameterization of mixing in the ocean boundary layer. *Journal of Marine Systems*, **21**, 55-65.
- , 2008: *Air-ice-ocean interaction: Turbulent ocean boundary layer exchange processes*. Springer.
- McWilliams, J. C., and P. P. Sullivan, 2000: Vertical mixing by Langmuir circulations. *Spill Science & Technology Bulletin*, **6**, 225-237.
- McWilliams, J. C., P. P. Sullivan, and C. H. Moeng, 1997: Langmuir turbulence in the ocean. *J Fluid Mech*, **334**, 1-30.
- McWilliams, J. C., E. Huckle, J. H. Liang, and P. P. Sullivan, 2014: Langmuir Turbulence in Swell. *Journal of Physical Oceanography*, **44**, 870-890.
- Merino, M., 1997: Upwelling on the Yucatan Shelf: hydrographic evidence. *Journal of Marine systems*, **13**, 101-121.
- 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. *Journal of Physical Oceanography*, **46**, 57-78.
- Ott, J., M. Pritchard, N. Best, E. Linstead, M. Curcic, and P. Baldi, 2020: A Fortran-Keras Deep Learning Bridge for Scientific Computing. *Scientific Programming*, **2020**, 1-13.
- 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. *Journal of Physical Oceanography*, **45**, 2897-2911.
- Qiao, F. L., Y. L. Yuan, Y. Z. Yang, Q. N. Zheng, C. S. Xia, and J. A. Ma, 2004: Wave-induced mixing in the upper ocean: Distribution and application to a global ocean circulation model. *Geophysical Research Letters*, **31**, L11303.
- 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). *Geoscientific Model Development*, **13**, 2185-2196.
- Rasp, S., M. S. Pritchard, and P. Gentile, 2018: Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, **115**, 9684-9689.
- Reichl, B. G., and R. Hallberg, 2018: A simplified energetics based planetary boundary layer (ePBL) approach for ocean climate simulations. *Ocean Modelling*, **132**, 112-129.
- Reichl, B. G., and Q. Li, 2019: A parameterization with a constrained potential energy conversion rate of vertical mixing due to Langmuir turbulence. *Journal of Physical Oceanography*, **49**, 2935-2959.
- Reichl, B. G., D. Wang, T. Hara, I. Ginis, and T. Kukulka, 2016: Langmuir Turbulence Parameterization in Tropical Cyclone Conditions. *Journal of Physical Oceanography*, **46**, 863-886.
- Roemmich, D., and A. S. Team, 2009: Argo: the challenge of continuing 10 years of progress. *Oceanography*, **22**, 46-55.
- Sane, A., B. G. Reichl, A. Adcroft, and L. Zanna, 2023: Parameterizing Vertical Mixing Coefficients in the Ocean Surface Boundary Layer Using Neural Networks. *Journal of Advances in Modeling Earth Systems*, **15**, e2023MS003890.
- Sane, A., B. Reichl, A. Adcroft, and L. Zanna, 2025: Machine Learned Equations for Vertical Mixing in the Ocean Surface Boundary Layer.
- 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 Biogeochemical Cycles*, **19**, GB3004-GB3004.

Skillingstad, E. D., T. Paluszkiwicz, D. W. Denbo, and W. D. Smyth, 1996: Nonlinear vertical mixing processes in the ocean: Modeling and parameterization. *Physica D: Nonlinear Phenomena*, **98**, 574-593.

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. *Journal of Advances in Modeling Earth Systems*, **10**, 3030-3048.

Sullivan, P. P., and J. C. McWilliams, 2010: Dynamics of Winds and Currents Coupled to Surface Waves. *Annual Review of Fluid Mechanics*, **42**, 19-42.

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. *Journal of Physical Oceanography*, **46**, 1285-1307.

Toole, J. M., 1998: Turbulent mixing in the ocean: Intensity, causes, and consequences. *Ocean modeling and parameterization*, Springer, 171-190.

Umlauf, L., and H. Burchard, 2003: A generic length-scale equation for geophysical turbulence models. *Journal of Marine Research*, **61**, 235-265.

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. *Journal of Geophysical Research: Oceans*, **117**, C05001.

Wang, P., J. C. McWilliams, J. Yuan, and J. H. Liang, 2025: Langmuir mixing schemes based on a modified K-profile parameterization. *Journal of Advances in Modeling Earth Systems*, **17**, e2024MS004729.

Weller, R. A., and J. F. Price, 1988: Langmuir Circulation within the Oceanic Mixed Layer. *Deep-Sea Research Part a-Oceanographic Research Papers*, **35**, 711-747.

Wong, A. P., and Coauthors, 2020: Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Frontiers in Marine Science*, **7**, 700.

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. *Journal of Physical Oceanography*, **51**, 2655-2680.

Yuan, J., J. H. Liang, 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). *Journal of Advances in Modeling Earth Systems*, **16**, e2024MS004405.

Zavala-Hidalgo, J., 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 dynamics*, **56**, 333-338.

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. *Journal of Geophysical Research: Oceans*, **108**, 3389.

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. *National Science Review*, **9**, nwac044.