Exploring the Use of Machine Learning to Parameterize Vertical Mixing in the Ocean Surface Boundary Layer

Jun-Hong Liang[1,2,3], Jianguo Yuan[1], Xiaoliang Wan[2,4], Jinliang Liu[1,*], Bingqing Liu[5], Hakun Jang[2], and Mayank Tyagi[2,6]

[1] Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana, USA

[2] Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana, USA

[3] Coastal Studies Institute, Louisiana State University, Baton Rouge, Louisiana, USA

[4] Department of Mathematics, Louisiana State University, Baton Rouge, Louisiana, USA

[5] the Water Institute of the Gulf, Baton Rouge, Louisiana, USA

[6] Department of Petroleum Engineering, Louisiana State University, Baton Rouge, Louisiana, USA

* Now at Pacific Northwest National Laboratory, Seattle, WA.

Corresponding author: Jun-Hong Liang (jliang@lsu.edu), Louisiana State University, Baton Rouge, LA, 70803

Abstract

In ocean and climate models, the simulation of upper-ocean temperature and salinity depends on mixing parameterization for ocean surface boundary layer turbulence. Existing mixing parameterizations are based on physical principles with empirical parameters. However, they are still imperfect, leading to biases in the simulation of physical states in the upper ocean. In this study, we explore the use of the data-based machine learning technique, specifically, a deep neural network model, for the effects of vertical mixing in the ocean surface boundary layer. The model is trained using process-oriented simulations of the upper-ocean turbulence driven by realistic forcing conditions at a mid-latitude ocean climate station, viz. the Ocean Station Papa. The deep neural network model outperforms traditional physics-based parameterizations that relate the mixing effects to surface forcing using deterministic formulas. The deep neural network model is also used to explore two currently debated issues in the development of physics-based mixing parameterizations, viz. the representation of wave forcing and the history of forcing conditions.

1. Introduction

Ocean surface boundary layer (OSBL) turbulence plays an important role in the ocean environment and global climate. It mediates the rate of exchange of heat and materials between the atmosphere and the interior ocean [e.g., Sallee et al. 2013; Liang et al. 2013], controls the effect of the ocean on the atmosphere by determining the temperature of the sea surface [e.g., Chen et al. 1994; Richards et al. 2009], modulates ocean ecosystem by setting the physical and chemical environment of the euphotic zone [e.g., Taylor and Ferrari 2011], and alter the dispersion and transport of pollutants in the near-surface ocean [e.g., Liang et al. 2018 and 2021; Kukulka 2020]. With continued efforts using advanced observational techniques and high-fidelity computer simulations, it is now understood that OSBL turbulence is primarily driven by three processes at the sea surface, viz. wind [e.g., Skyllingstad et al. 1999], heating/cooling [e.g., Li et al. 2005; Pearson et al. 2015], and ocean surface gravity waves [e.g., D'Asaro et al. 2014; Qiao et al. 2004]. It is also altered by other factors including density stratification [Price and Sundermeyer 1999], the earth's rotation [Liu et al. 2018], the depth of the water column [e.g., Tejada-Martinez and Grosch 2007; Yan et al. 2022], larger-scale horizontal density gradient [Fan et al. 2018 and 2020] and lateral currents such as tidal and submesoscale currents [e.g., Hamlington et al. 2014; Sullivan and McWilliams 2019; Yuan and Liang 2021].

In hindcast/forecast ocean and climate models, OSBL turbulence cannot be explicitly simulated currently and likely in the near future [Fox-Kemper et al. 2014 and 2019] for the following reasons: Most models for the realistic ocean have to be configured on grids that are coarser than the scale of OSBL turbulence (tens of centimeters to meters). Even in a small-domain setting where model grids could be as fine as at a scale of meters, those models are based on hydrostatic approximation that excludes OSBL turbulence. In addition to the abovementioned computational and physical limitations, numerically, hindcast/forecast models utilize finite volume or finite difference schemes that are flexible and efficient, but less accurate than spectral methods used in models to compute turbulence. In an idealized version of hindcast/forecast model that neglects the effect of horizontal processes, the effect of ocean surface boundary layer on prognostic variables are calculated as follows,

$$\left.\frac{\partial \bar{C}}{\partial t}\right|_{mixing} = \frac{\partial \overline{w'C'}}{\partial z} \tag{1a}$$

where *C* represents a tracer such as temperature and salinity, and *w* is vertical velocity. Prime indicates fluctuation associated with turbulence, and overbar means ensemble-averaged quantity without signals of turbulence. Subscript mixing denotes that this is the time tendency due only to vertical turbulent mixing by OSBL turbulence. Other terms and their effect on the time tendency are neglected for simplicity. Turbulent fluctuations, i.e., *w*' and *C*' in the above equation, cannot be calculated in hindcast and forecast models. They must be approximated using averaged variables, i.e., $\bar{C}$, and forcing conditions, through a parameterization as follows,

$$\frac{\partial \overline{w'C'}}{\partial z} = \frac{\partial}{\partial z}\left[K_T\left(\frac{\partial \bar{C}}{\partial z} + \gamma_C\right)\right] \tag{1b}$$

where K_T are the vertical diffusivity; and Υ_C is the counter-gradient (or non-local) term [e.g., Deardorff 1966] in some parameterizations to account for tracer transport that is not inversely related to the spatial gradient of the tracer. The counter-gradient term is important when coherent structures, such as buoyancy-driven convective cells and wave-driven Langmuir circulations, are dominant. Those coherent structures fill the whole ocean surface boundary layer, and tracer transport is no longer local and unrelated to the local tracer gradient, i.e., $\frac{\partial \bar{c}}{\partial z}$ in equation (1).

Parameters, i.e., K_T, and Υ_C, are deterministic functions of variables including surface forcing conditions (wind, wave, and buoyancy flux), and water column conditions. Their treatment is different in the two classes of commonly used physics-based mixing parameterization: In the first-order parameterizations, such as the K-Profile Parameterization [e.g., Large et al. 1994], the dependence of the parameters on surface forcing and water column conditions is direct. In the second-moment parameterizations, K_T, and Υ_C are diagnosed from turbulence statistics such as the kinetic energy, the length scales and the dissipation rate of turbulence that are prognostically calculated in the model [e.g., Kantha and Clayson 1994; Umlauf and Burchard 2003; Reichl and Hallberg 2018]. Those parameterizations usually work well for flows under one of the three forcing conditions (wind, wave and buoyancy forcing) for the following two reasons: (1) theories exist for the velocity scale of turbulence driven by each of those forcing conditions [e.g., Belcher et al. 2012; D'Asaro et al. 2014], and (2) the limited number of parameters in those parameterizations are tuned using data, from either field observations or high-fidelity process-oriented computer simulations, when one of the three forcing conditions dominates or exists [e.g., Harcourt 2015; Reichl et al. 2016]. They are less accurate when the three forcing conditions are similarly important or when the surface buoyancy flux is stabilizing [e.g., Li et al. 2019]. However, the realistic ocean is usually under the comparable influence of all three forcing conditions (Fig. 1a) or with surface buoyancy flux stabilizing the upper ocean (Fig. 1b). Biases in the simulated upper-ocean states using those parameterizations remain, degrading our hindcasting/forecasting ability of the upper-ocean states, the marine ecosystem, and the coupling of the ocean with the atmosphere [e.g., Belcher et al. 2012].

Given the aforementioned challenges in traditional physics-based OSBL mixing parameterizations, this study explores the use of machine learning, specifically, a deep neural network (DNN) model, to parameterize the vertical mixing effects of OSBL turbulence. In recent years, machine learning techniques, that are data-based, have attracted attention in the community of atmospheric and oceanic sciences and have been investigated for different applications, such as ensemble weather forecasting [e.g., Rasp and Lerch 2018], the parameterization of convection in atmospheric models [e.g., Brenowitz and Bretherton 2018;

Gentine et al. 2018], the coupling between the ocean and the atmosphere under hurricanes [Jiang et al. 2018], the parameterization for ocean mesoscale eddies [e.g., Bolton and Zanna 2019; Zanna and Bolton 2020], and the prediction of ocean currents from observations [e.g., Liu and Weisberg 2005] and satellite images [e.g., Zeng et al. 2015].

The specific objectives of the study are (1) to train a DNN model to parameterize the mixing effects of OSBL turbulence; (2) to compare the DNN model with a few traditional physics-based parameterizations using a single-column model; and (3) to explore the significance of different forcing parameters using the DNN model. The rest of the paper is organized as follows: Section 2 describes the DNN model and a few traditional deterministic parameterizations used for comparison; Section 3 describes the data used to develop and evaluate the DNN model; Section 4 presents and discusses model results; and Section 5 summarizes the important conclusions and discusses possible future research directions.

2. Model Description

2.1 A Physics-informed Deep Neural Network model for the Vertical Mixing by Ocean Surface Boundary Layer Turbulence

A feedforward Deep Neural Network (DNN) model [e.g., Goodfellow et al. 2016] is used in this study. As opposed to representing the vertical mixing effect by OSBL turbulence using deterministic functions such as equation (1), a DNN model provides an effective and flexible approximation of non-linear mapping between an input layer containing profiles of temperature and salinity and forcing conditions and an output layer containing profiles of the time derivatives of the variables, i.e., the left-hand-side term in equation (1a). The profiles of the time derivative of temperature and salinity are then used to prognostically calculate the profiles of temperature and salinity like using a traditional physics-based parameterization. We also experimented using the profiles of turbulent fluxes, i.e., the right-hand-side term in equation 1(a), as output, and the prognostic temperature and salinity profiles are the same. The input and output layers are connected by one or multiple hidden layers (Fig.3). Assume there are a total of *M* layers including the input (*j* = 1), hidden (*j* = 2 to *M*-1) and output (*j* = *M*) layers. The output of neuron *i* in layer *j* with *N*(*j*) neurons, i.e., $X_{i,j}$, is calculated using outputs from the previous layer, i.e., $X_{i,j-1}$ as [Goodfellow et al. 2016],

$$X_{i,j} = f\left(\sum_{k=1}^{N(j-1)} w_{k,i,j-1} X_{k,i,j-1} + b_{k,i,j-1}\right) \tag{2}$$

where $w_{k,i,j-1}$ and $b_{k,i,j-1}$ are the weight and bias directing from neuron *k* in layer *j*-1 to neuron *i* in layer *j*, respectively; and *f* is the activation function. The optimized set of weights and biases are determined through a learning process. In other words, each neuron represents a simple operation defined by an activation function. The output of each neuron is used as input to neurons in the next layer in a feedforward DNN model. Operations in each neuron collectively contribute to the mapping from the input layer to the output layer.

In addition to the set of weights and bias, there are a few hyperparameters in the model, i.e., those predefined and not optimized during each learning including the activation function, the loss function and the numbers of hidden layers and neurons. These hyperparameters are selected after then learning stage using data not used in the learning. The Leaky Rectified Linear Unit (LeakyReLU, $a(x) = \max(0.1x, x)$) [e.g., Mass et al. 2014], one of the popular ones in DNN models, is used as the nonlinear activation function in this study. It was reported by Gentine et al. [2018] that the LeakyReLU show the best performance among a few activation functions. Our own sensitivity experiments also showed that the LeakyReLU results in the fastest decrease in both the training and validation losses over epochs and best final scores among a few activation functions, including Rectified Linear Unit (ReLU), tangent hyperbolic (tanh) and the Sigmoid ($S(x) = (1 + e^{-x})^{-1}$). The loss function that is used to gauge the performance of the trained model is defined as,

$$Loss = \sum (X_{prediction} - X_{truth})^2 + \alpha\left(\left|\frac{dHC}{dt} - Q_{heat}\right| + \left|\frac{dSC}{dt} - Q_{salt}\right|\right) \tag{3}$$

Where $X_{prediction}$ and X_{truth} are the variables in the output layer and from the data (the LES solutions in this study), respectively; *HC* and *SC* stand for water-column heat and salt content, respectively; and α is a constant set to 0.1 in this study. The first term on the right-hand-side of equation (3) measures the deviation of the predicted results from the data. The second right-hand-side term is a penalty when the prediction violates heat and salt conservation. Penalizing the loss when physical principles are violated is one of the popular approaches to add physical constraint to a DNN model. With the loss function, the trained DNN model not only best re-produces the data, but also best abides by conservation laws.

To seek the optimized set of $w_{k,i,j-1}$ and $b_{k,i,j-1}$ that best map variables in the input layer to those in the output layer, available data are first separated into three independent sets, a training dataset, a validation set and a prediction (testing) set. The determination of the optimal set of $w_{k,i,j-1}$ and $b_{k,i,j-1}$ is formulated as a supervised learning problem, where the main goal is to minimize the loss function based on the training data set. The optimization of $w_{k,i,j-1}$ and $b_{k,i,j-1}$ is achieved by the stochastic gradient descent methods, which iteratively choose new values of $w_{k,i,j-1}$ and $b_{k,i,j-1}$ to reduce the loss function. The learning process is repeated for 2,000 iterations, the parameters that result in the smallest generalization error of the model, which is the loss estimated on a validation dataset that is independent of the training set, are selected. Sensitivity experiments with 4,000 iterations were conducted to confirm that the generalization error will not further decrease after 2,000 iterations.

The training of the DNN model is conducted using the Python library TensorFlow (http://www.tensorflow.org) and the Python library Keras (https://keras.io) (the computer code can be accessed through the link in the acknowledgement section). Using Nvidia GeForce RTX 2080 Ti GPU, each training and validation process takes about 10 minutes to complete.

A number of tests using the DNN model, with different model hyperparameters and inputs, are reported in this paper (Table 1). In the control simulation, the DNN model has 1 hidden layer and 256 neurons in the hidden layer. The numbers of hidden layer and neuron result in the smallest error and are selected by sensitivity experiments that systematically vary the two numbers. The input layer includes profiles of temperature and salinity as well as forcing including wind vector, surface buoyancy flux, the vertical profiles of the Stokes drift at the current time step (t) and the previous time step ($t - \Delta t$, with Δt = 30 minutes). Here, The Stokes drift is wave-averaged current associated with wave and drives wave-driven Langmuir turbulence when it interacts with ocean currents [e.g., Craik and Leibovich 1976]. In addition to the control simulation, two groups of sensitivity experiments are conducted to decide the input that constructs the best model. The first group of sensitivity experiments is devised to explore model performance under different representations of surface wave forcing. While the importance of wave forcing is realized, how it is represented differs in different parameterizations [e.g., Li et al. 2019]. In addition to the control simulation where the vertical profile of the Stokes drift vector is used to represent wave effect, three sensitivity experiments, using the surface Stokes drift magnitude, and the surface Stokes drift vector, respectively, are conducted. The second group of sensitivity experiments, in which the input layer includes forcing without history (t) and that with a 1-hour history (t, $t - \Delta t$, and t - $2\Delta t$), respectively, is designed to test the effect of forcing history on the performance of mixing parameterizations.

2.2 The General Ocean Turbulence Model (GOTM)

The performance of the DNN models is compared with conventional OSBL mixing parameterizations with deterministic formulas. For this purpose, the General Ocean Turbulence Model (GOTM) [Burchard et al. 1999; Umlauf and Burchard 2005; Umlauf et al. 2014] is used. The GOTM is a library and testbed of parameterizations for vertical mixing by OSBL turbulence. It belongs to the single-column model that excludes the impacts of horizontal processes. The GOTM includes several commonly used OSBL mixing parameterizations. In this study, two variants of the K-profile parameterizations [KPP, Large et al. 1994] that is a popular first-order model, including the KPP-CVmix [e.g., Large et al. 1994; Li et al. 2021; Van Roekel et al. 2018], and the KPPLT-LF17 [Li and Fox-Kemper 2017], are included in the comparison. It is shown in Li et al. [2019] that the performance of KPPLT-LF17 is similar to some other variants of the KPP including wave effects [e.g., van Roekel et al. 2012; Reichl et al. 2016]. The KPP-CVMix is the KPP in the Community Ocean Vertical Mixing (CVMix) project [Griffies et al. 2015] and is used in a few global ocean models, such as the Parallel Ocean Program Version 2 [POP2; Smith et al. 2010]. It includes the mixing effects of wind- and cooling-driven turbulence, but not those of wave-driven Langmuir circulations. The other KPP variant, i.e., KPPLT-LF17, includes the effect of Langmuir turbulence (LT) in the framework of KPP, by enhancing both the magnitude of diffusivity and the entrainment at the base of the OSBL. They differ in the forms of the enhancement factors for diffusivity and entrainment (See Li et al 2019 for a detailed comparison of the three parameterizations). In addition to the KPP, another type of commonly used parameterization is second-moment closure schemes [e.g., Harcourt 2015; Kantha and Clayson 2004]. It was concluded in Li et al. [2019] that those schemes perform similarly to the KPP, and therefore the comparison between the DNN model and different variants of the KPP is representative of that between the data-based DNN model and traditional physics-based parameterizations.

3. Data Description

The data for training, validating, and testing the DNN model and for evaluating traditional physics-based parameterization in the GOTM are turbulence-resolving solutions for Ocean Station Papa (50ºN 145ºW) calculated using the National Center for Atmospheric Research Large Eddy Simulation (NCAR-LES) model [e.g., Sullivan et al. 1996]. The NCAR-LES model has been extensively used to study OSBL turbulence driven by one or a combination of wind, wave, and heating/cooling [e.g., Sullivan and McWilliams 2010] and has been shown to accurately reproduce in situ observations when the effect of OSBL turbulence dominates [e.g., Liang et al. 2020]. Solutions from LES models are traditionally used to derive physics-based parameterizations with deterministic formulas for the effects of OSBL turbulence [e.g., Chor et al. 2021; Yang et al. 2015; Sinha et al. 2015; van Roekel et al. 2018] as they fully resolve OSBL turbulence yet exclude all other larger-scale processes. Those solutions are commonly used to tune and evaluate parameterizations in a 1-D model setting, such as within the framework of the GOTM [e.g., van Roekel et al. 2018; Li et al. 2019]. The Ocean Station Papa, located at the North Pacific subpolar gyre ~ (50ºN 145ºW, Fig. 1c), is selected since continuous high-resolution measurements of physical and chemical states and fluxes at and near both sides of the air-sea interfaces. Insights into physical and biogeochemical processes in the upper ocean have been gained through the analysis of observation at the station and accompanying computer simulations [e.g., Alford et al. 2012; Cronin et al. 2015; Kaminski et al. 2021].

For this study, the NCAR-LES model was run for about nine years, from September 2011 to June 2019, during which high-resolution observations of wind, wave, surface heat flux, and the profiles of temperature and salinity [e.g., Cronin et al. 2015; Thomson et al. 2013] are available. There is a period between fall 2017 and spring 2018 when the directional wave spectrum is not available, and that period is excluded in the simulation. The forcing conditions, including wind, wave and surface buoyancy forcing were applied uniformly across horizontal locations of the domain. The model was restarted every ten days. During each restart, in-situ profiles of temperature and salinity were used as initial conditions across horizontal locations of the domain.

A 6-hour simulation with constant forcing at the start of the period is used to spin up the turbulence field. Observed surface wind, wave, and surface heat flux were used as surface forcing conditions. Salinity flux is set to zero, although it was used in the 15-day simulation in Liang et al. [2017], as the observation of precipitation at the station was sporadic. The forcing conditions at this station during the multi-year simulation cover a wide range of meteorological conditions that are representative of mid-latitude oceans. Turbulence is not predominantly governed by one of three types, although the dominance of Langmuir turbulence is more common than that of the other two types of turbulence. Compared to the OSBL turbulence of the global ocean, OSBL turbulence at the station is more influenced by wave-driven Langmuir turbulence and is much less influenced by buoyancy-driven convective turbulence. Figures 2(a) to 2(c) show the comparison between the LES solutions and observation. The model agrees generally with the observations. Slight deviation of model solutions from the observation is expected as the LES model includes only OSBL turbulence and the effects from larger-scale circulations, i.e., submesoscale, mesoscale, and basin-scale currents, are not excluded. It should be noted that Liang et al. [2017] showed that the same LES model and configuration accurately reproduce the physical and chemical environment in the OSBL during a 15-day period without significant influence from processes other than OSBL turbulence. The exclusion of other processes makes the LES solutions superior over *in situ* data for the purpose of developing OSBL mixing parameterization. For the study, horizontally and temporally averaged temperature, and salinity profiles were archived every half hour. Excluding the first 12 hour of each simulation, there are approximately 80,000 vertical profiles of temperature and salinity, respectively. Around 72%, 14.5% and 13.5% of the remaining solutions and the corresponding forcing conditions were used for the training, validation and testing of the DNN model, respectively.

4. Results

4.1 Performance of the DNN model

The skill of the DNN model is first demonstrated by comparison of a 9-day run with the LES solutions, which are considered the truth, and with solutions using the conventional physics-based parameterizations. Note that the performance assessment in this section is based on the testing dataset from the LES solutions that is independent of the training and validating datasets and is not used for the learning of the DNN model. Figure 4 compares the prediction by the DNN model and two different traditional parameterizations (KPP-CVMix and KPPLT-VR12) for a 9-day period in December 2010. The wind is moderate at the beginning of the period, with a speed 10-m above the sea level (U_{10}) around 10 m/s (Fig. 4a). The wind weakens to about 5 m/s at around day 2, and slowly strengthens to more than 18 m/s at around day 5. Two different forms of Langmuir numbers, i.e., MSM97 [McWilliams et al. 1997] and VR12 [van Roekel et al. 2012], are plotted in Fig. 4b. There is no substantial difference between MSM97 and another popular Langmuir number proposed by Harcourt and D'Asaro [2008] (not shown). The difference between VR12 and MSM97 is evident as VR12 includes the effect of wind-wave mis-alignment, shows more variability than MSM97. Langmuir number is mostly close to 0.3 during most of the period, implying the existence of wave-driven Langmuir turbulence during the period. It briefly goes above (below) 0.3 when the wind strengthens (weakens), implying the dominance of wind- (wave-) driven turbulence during those moments. The net surface heat flux cools the OSBL and there are a few episodic rain events.

The OSBL continuously cools and deepens during the period. It also gets saltier when high-salinity water in the thermocline is entrained into the OSBL (Figs. 4c and 4d). All three models, including the DNN model, the KPP-CVMix and the KPP-LF17, capture the trends for mixed layer depth, temperature and salinity in the OSBL (Figs 4e to 4j). The predicted mixed layer depth by the DNN model closely follows that by the LES model. The predicted temperature in the OSBL by the DNN model is slightly cooler than the truth (the LES solutions), on the order of 0.1 ºC (Fig. 4e). Traditional deterministic parameterizations, i.e., the two variants of the KPP model, predict a substantially warmer mixed layer, by more than 1ºC (Figs. 4g and 4i). Similar to that for temperature, the error for salinity is smaller for the DNN model than for the two traditional parameterizations (Figs. 4f, 4h and 4j). By comparing the mixed layer depth, it is obvious that the two KPP models predict a mixed layer shallower than the truth while the mixed layer depth diagnosed from the DNN solutions closely follows the truth. It should be noted that the KPP-LF17, i.e., the KPP that includes wave-induced mixing, does predict a slightly deeper, cooler and saltier OSBL than the KPP-CVMix does. The difference between KPP-LF17 and KPP-CVMix in our simulations qualitatively agrees with Fig. 2 in Li et al. [2019]. However, the difference between KPP-LF17 and KPP-CVMix is much smaller than that between the two and the truth, therefore is not evident in Fig. 4, implying that including wave effect in a traditional physics-based model still cannot match the truth as well as the DNN model does. Like over the global ocean, the most common meteorological condition at the OSP is also when the three types of turbulence are similarly important or when the surface buoyancy flux is stabilizing (Fig. 1). Those are the conditions when traditional physics-based parameterizations struggle [Li et al. 2019].

The skill of the DNN model for all prediction periods is evaluated using the statistics of errors (Fig. 5). Both the mean and the standard deviation of the errors for both temperature and salinity are significantly smaller for the solutions using the DNN model than those using the two variants of the KPP model. Both KPP-CVMix and KPP-LF17 systematically predict warmer and fresher OSBL while the mean error for both temperature and salinity using the DNN model is less obvious. The comparison of model error statistics confirms that the data-based DNN model on average outperforms the two traditional physics-based parameterizations. Although the learning of DNN model is not based on the testing dataset, the DNN model performs well as the forcing conditions in the testing dataset, i.e., the input, overlaps with those in the training and validation datasets in the parameter space shown in Fig. 1.

4.2 Discussion Based on Sensitivity Experiments

In this subsection, the sensitivity of the DNN model to model inputs, including the forms of wave forcing and the history of forcing is evaluated using the Taylor diagrams (Fig. 6), respectively. In the Taylor diagram, three matrices, viz. the mean absolute error (solid grey lines), the root mean square error (solid black lines) and correlation (dashed black lines) representing the error, the scattering of the prediction, and the similarity in pattern, respectively, are presented in the same figure. Results from the two traditional parameterizations are not included in the comparison as the errors from those two models are much larger than those using the DNN framework. It should be noted that the optimal numbers of hidden layers and neurons in each layer are different when input variables for DNN model are different (Table 1). Sensitivity experiments by altering those hyperparameters (not shown) were conducted to select those optimal numbers.

4.2.1 The Importance of Stokes Drift Profile

The first group of sensitivity experiments examine the choice of wave forcing in model performance. Those two tests include one using surface Stokes drift without directional information (green cross) and one using surface Stokes drift vector (blue cross). The DNN models with surface Stokes as input are used to mimic the input of commonly used first-order parameterization for Langmuir circulations, such as the KPP-LF17 [Li and Fox-Kemper 2017]. In those parameterizations, a turbulent Langmuir number that is a function of surface or near-surface Stokes drift is included as a parameter to quantify the effects of wave-driven Langmuir turbulence. There are a few variants of turbulent Langmuir number (La_t) used in different parameterization. Here, the DNN model using Stokes drift magnitude as an input corresponds to parameterizations [e.g., McWilliams and Sullivan 2000] using the La_t originally defined in McWilliams et al. [1997]. The DNN model using Stokes drift vector mimics parameterizations [e.g., van Roekel et al. 2012; Li and Fox-Kemper 2017] using La_t that considers wind-wave misalignment [e.g., van Roekel et al. 2012]. Note that there is a third popular variant of La_t that uses near-surface averaged Stokes drift instead of surface Stokes drift [e.g., Harcourt and D'Asaro 2008].

The solutions of the two sensitivity experiments have larger errors than those of the control simulation (Figs. 6a and 6b). The control simulation (red dots) out-performs the two sensitivity experiments (blue pluses and green crosses) in terms of root mean square error (RMSE) by more than 15%, implying that the detailed Stokes drift profile is better than surface Stokes drift in representing the effect of waves. This is expected, as the subsurface profile of Stokes drift is complicated when both swell and wind wave are present [e.g., McWilliams et al. 2014; Breivik and Christensen 2020] and the detailed profile is important in determining the production of turbulent kinetic energy in the OSBL.

4.2.2 The Importance of Forcing History

The second group of sensitivity experiments are designed to test an important assumption in first-order OSBL turbulence parameterizations such as the KPP. In those parameterizations, forcing conditions at the current time step are used and the underlying assumption is that OSBL turbulence and its mixing effect are always in equilibrium with surface forcing conditions (wind, wave and surface buoyancy flux). Over much of the global ocean, however, surface forcing conditions are always changing associated with atmospheric variability at the weather and climatic scales. Consequently, OSBL turbulence and its mixing effect are seldom in equilibrium with forcing conditions. It has never been evaluated how significant the effect of forcing history is.

The length of the forcing history could be decided by considering the following scaling: The velocity scale of ocean surface boundary layer turbulence at mild to moderate forcing is ~10^{-2} m/s (1 cm/s), and the vertical scale of ocean surface boundary layer in most of the global ocean is ~10^{1}m, so the eddy turnover time and the turbulence response time scale are on the order of 10^{3} s. In a recent study by Wang and Kukulka [2021] where the LES model is driven by an abruptly changing wind, it is shown that turbulence response time is at a scale of 10^{3} seconds since the wind abruptly changes direction. With that consideration, a 30-minute history of forcing conditions (both t and t - Δt with Δt = 30 minutes) is used as input in the control simulation and in the first group of sensitivity experiments.

Comparisons of errors in the control simulation driven by a 30-minute history of forcing (red dots) and the sensitivity experiment driven by forcing conditions without history (yellow triangles) mimicking first-order parameterizations (Figs. 6a and 6b) show that the control simulations perform better by more than 15% in terms of RMSE. Another sensitivity experiment driven by a one-hour history of forcing conditions (purple squares) yields similar results as the control simulation and out-performs the sensitivity experiment driven by forcing conditions without history. Therefore, including the history forcing in OSBL mixing parameterizations improve the prediction of upper-ocean states.

5. Summary and Future Research Directions

In this study, a deep neural network (DNN) model that is a type of machine learning model for the effect of ocean surface boundary layer (OSBL) turbulence is trained using 9-year process-oriented numerical solutions for the Ocean Station Papa (OSP) that is at the subpolar Pacific Ocean. Computer code for training the DNN model is available through a link in the acknowledgement section. The DNN model is evaluated and compared against two popular traditional physics-based parameterizations using deterministic formulas, i.e., two variants of the K-Profile Parameterizations (KPP) including the KPP-CVMix [e.g., Large et al. 1994; Griffies et al. 2015] and the KPP-LF17 [Li and Fox-Kemper 2017]. It is also used to investigate the choice of forcing conditions in parameterizations for OSBL turbulence. Important conclusions from the results are:

(1) The data-based machine learning model, viz. the physics-informed deep neural network (DNN) model outperforms two popular traditional physics-based parameterizations, namely, the KPP-CVMix and the KPP-LF17.

(2) Including wave forcing improves the performance of the DNN model. The use of a Stokes drift profile is superior to the use of surface Stokes drift.

(3) Including a 30-minute or 1-hour history of forcing conditions as input for the DNN model improves the prediction over the use of forcing conditions without any history.

While the profiles of Stokes drift and the history of forcing are not in first-order parameterizations like the KPP, they are inherently in second-order closures [e.g., Umlauf and Burchard 2003 and 2005; Kantha and Clayson 2004; Harcourt 2015], In second-order closures, turbulent characteristics, such as its intensity, dissipation rate and length scale, is prognostically calculated using equations that includes the vertical profile of the Stokes drift and the history of forcing is retained during time integration.

Our study shows the promise of a DNN model for the parameterization of vertical mixing in the OSBL. In this study, the DNN model was trained and tested for conditions at Ocean Station Papa that is representative of conditions at the mid latitude. The model should be applicable to regions under forcing conditions within the parameter space shown in Figs. 1(a) and 1(b). Before the application of the trained DNN model to a region, forcing conditions in that region need to be examined to ensure that they are within the parameter space of the training data. Figures 1(a) and 1(b) also show that the training data still miss some turbulence regimes, most notably convective turbulence in deep mixed layers that is typical at high latitudes and strongly heated boundary layer that is at the tropical regions. Process-oriented solutions from the Large Eddy Simulation model at a variety of geographic locations, such as those at the Southern Ocean [e.g., Large et al. 2019], in the tropics and in other ocean regions, need to be included in the training and validation datasets to expand the parameter space that the trained DNN model can tackle. Recently, LES simulations for the ocean under realistic hurricane conditions are also available [e.g., Rabe et al. 2015; Liang et al. 2020]. If those data are added to the training and validation datasets, the trained DNN model will also be used for the ocean under those extreme conditions. Finally, while the current study tests the DNN model in a single-column model which is an idealized version of a hindcasting/forecasting ocean model, future research will implement the DNN model in hindcast/forecast ocean models and test it in regional and global oceans. Our ongoing efforts to implement the DNN model in the Coupled-Ocean-Atmosphere-Wave-Sediment transport (COAWST) model [Warner et al. 2010] and test it in a coupled-ocean-wave configuration for the Gulf of Mexico [Abolfazli et al. 2020] will be reported in a future manuscript.

Acknowledgement

We thank two anonymous reviewers for their insightful comments and suggestions. JHL and JY were supported by the National Science Foundation (NSF) through grant OCE-1945502. XW was supported by the NSF through grant DMS-1913163. The DNN model and data used to train and test the model are attached for review and will be made available to the community through a GitHub page here. Physical measurements from the NOAA surface mooring are available from http:// www.pmel.noaa.gov/OCS/Papa. Gas measurements from the NOAA surface mooring are available at http://www. pmel.noaa.gov/co2/story/Papa. Wave data are available at http://www.apl.washington.edu/projects/station_ papa/summary.html. Computations were performed on supercomputing facilities at Louisiana State University, and through the Louisiana Optical Network Infrastructure (LONI).

References

Abolfazli, E., J. -H. Liang, Y. Fan, Q. J. Chen, N. D. Walker, J. and Liu, 2020: Surface gravity waves and their role in ocean-atmosphere coupling in the Gulf of Mexico. *J. Geophys. Res. Oceans*, **125**, e2018JC014820.

Alford, M. H., M. F. Cronin, and J. M. Klymak, 2012: Annual cycle and depth penetration of wind-generated near-inertial internal waves at Ocean Station Papa in the northeast Pacific. *J. Phys. Oceanogr.*, **42,** 889-909.

Belcher, S. E., and Coauthors, 2012: A global perspective on Langmuir turbulence in the ocean surface boundary layer. *Geophys. Res. Lett.*, **39,** L18605.

Bolton, T., & L. Zanna, 2019: Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, **11**, 376– 399.

Brenowitz, N. D., and C. S. Bretherton, 2018: Prognostic validation of a neural network unified physics parameterization. *Geophys. Res. Lett.*, **45,** 6289-6298.

Burchard, H., K. Bolding, and M. R. Villarreal, 1999: *GOTM, a general ocean turbulence model: theory, implementation and test cases.* Space Applications Institute.

Breivik, Ø. and K. H. Christensen, 2020: A combined stokes drift profile under swell and wind sea. *J. Phys. Oceanogr.*, **50**, 2819-2833.

Chen, D., A. J. Busalacchi, and L. M. Rothstein, 1994: The roles of vertical mixing, solar radiation, and wind stress in a model simulation of the sea surface temperature seasonal cycle in the tropical Pacific Ocean. *J. Geophys. Res. Oceans*, **99,** 20345-20359.

Chor, T. J. C. McWilliams, and M. Chamecki, 2021: Modifications to the K-Profile parameterization with nondiffusive fluxes for Langmuir turbulence. *J. Phys. Oceanogr.*, **51(5)**, 1503-1521.

Craik, A. D. and S. Leibovich, 1976: A rational model for Langmuir circulations. *J. Fluid Mech.*, **73**, 401-426.

Cronin, M. F., N. A. Pelland, S. R. Emerson, and W. R. Crawford, 2015: Estimating diffusivity from the mixed layer heat and salt balances in the N orth P acific. *J. Geophys. Res. Oceans*, **120,** 7346-7362.

D'Asaro, E. A., J. Thomson, A. Shcherbina, R. Harcourt, M. Cronin, M. Hemer, and B. Fox-Kemper, 2014: Quantifying upper ocean turbulence driven by surface waves. *Geophys. Res. Lett.*, **41,** 102-107.

Deardorff, J.W., 1966. The counter-gradient heat flux in the lower atmosphere and in the laboratory. *J. Atmos. Sci*, **23**, 503-506.

Fan, Y. L., E. Jarosz, Z. T. Yu, W. E. Rogers, T. G. Jensen, and J. H. Liang, 2018: Langmuir turbulence in horizontal salinity gradient. *Ocean Model.*, **129,** 93-103.

Fan, Y., Z. Yu, I. Savelyev, P. P. Sullivan, J.-H. Liang, T. Haack, E. Terrill, T. De Paolo, K. Shearman (2020). The effect of Langmuir turbulence under complex real oceanic and meteorological forcing. *Ocean Model.*, **149,** 101601.

Fox-Kemper, B., S. Bachman, B. Pearson, and S. Reckinger, 2014: Principles and advances in subgrid modeling for eddy-rich simulations. CLIVAR Exchanges, 19(2), 42-46.

Fox-Kemper, B., and Coauthors, 2019: Challenges and prospects in ocean circulation models. *Front. Mar. Sci.*, **6,** 65.

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.

Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep learning.* MIT press.

Griffies, S. M., and Coauthors, 2015: Theory and numerics of the community ocean vertical mixing (CVMix) project. *Tech. Rep.*

Hamlington, P. E., L. P. Van Roekel, B. Fox-Kemper, K. Julien, & G. P. Chini, (2014). Langmuir-submesoscale interactions: Descriptive analysis of multiscale frontal spindown simulations. *J. Phys. Oceanogr.*, **44(9)**, 2249-2272.

Harcourt, R. R., 2015: An improved second-moment closure model of Langmuir turbulence. *J. Phys. Oceanogr.*, **45,** 84-103.

Harcourt, R. R., and E. A. D'Asaro, 2008: Large-eddy simulation of Langmuir turbulence in pure wind seas. *J. Phys. Oceanogr.*, **38,** 1542-1562.

Jiang, G. Q., J. Xu, and J. Wei, 2018: A deep learning algorithm of neural network for the parameterization of typhoon-ocean feedback in typhoon forecast models. *Geophys. Res. Lett.*, **45,** 3706-3716.

Kaminski, A. K., E. A. D'Asaro, A. Y. Shcherbina, and R. R. Harcourt, 2021: High-resolution observations of the North Pacific transition layer from a Lagrangian float. *J. Phys. Oceanogr.*, **51,** 3163-3181.

Kantha, L. H. and C. A. Clayson, 1994: An improved mixed layer model for geophysical applications. *J. Geophys. Res. Oceans*, **99,** 25235-25266.

Kantha, L. H., and C. A. Clayson, 2004: On the effect of surface gravity waves on mixing in the oceanic mixed layer. *Ocean Model.*, **6,** 101-124.

Kukulka, T., 2020: Horizontal Transport of Buoyant Material by Turbulent Jets in the Upper Ocean. *J. Phys. Oceanogr.*, **50,** 827-843.

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. *Rev. Geophys.*, **32,** 363-403.

Large, W. G., P. G. Patton, A. K. DuVivier, P. P. Sullivan, L. and Romero, 2019: Similarity theory in the surface layer of large-eddy simulations of the wind-, wave-, and buoyancy-forced southern ocean. *J. Phys. Oceanogr.*, **49,** 2165-2187.

Li, M., C. Garrett, and E. D. Skyllingstad, 2005: A regime diagram for classifying turbulent large eddies in the upper ocean. *Deep Sea Res. Part I Oceanogr.*, **52,** 259-278.

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.

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. *Geosci. Model Dev. Discuss.***,** 1-30.

Li, Q., and Coauthors, 2019: Comparing ocean surface boundary vertical mixing schemes including Langmuir turbulence. *J. Adv. Model. Earth Syst.*, **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. *J. Phys. Oceanogr.*, **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. *Glob. Biogeochem. Cycles*, **27,** 894-905.

Liang, J.-H., and Coauthors, 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.

Liang, J.-H., and Coauthors, 2020: Suppression of CO2 outgassing by gas bubbles under a hurricane. *Geophys. Res. Lett.*, **47,** e2020GL090249.

Liang, J.-H., and Coauthors, 2021: Including the effects of subsurface currents on buoyant particles in Lagrangian particle tracking models: Model development and its application to the study of riverborne plastics over the Louisiana/Texas shelf. *Ocean Modell.*, **167,** 101879.

Liu, J., J. -H. Liang, J. C. McWilliams, P. P. Sullivan, Y. Fan, and Q. Chen, 2018: Effect of planetary rotation on oceanic surface boundary layer turbulence. *J. Phys. Oceanogr.*, **48,** 2057-2080.

Liu, Y., and R. H. Weisberg, 2005: Patterns of ocean current variability on the West Florida Shelf using the self-organizing map. *J. Geophys. Res. Oceans*, **110,** C06003.

Maas, A. L., A. Y. Hannun, and A. Y. Ng, 2013: Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*, Citeseer, 3.

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, & P. P. Sullivan, 2014: Langmuir turbulence in swell. *J. Phys. Oceanogr.*, **44**, 870-890.

Pearson, B. C., A. L. 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.

Price, J. F., and M. A. Sundermeyer, 1999: Stratified Ekman layers. *J. Geophys. Res. Oceans*, **104,** 20467-20494.

Qiao, F., Yuan, Y., Yang, Y., Zheng, Q., Xia, C., & Ma, J. (2004). Wave-induced mixing in the upper ocean: Distribution and application to a global ocean circulation model. *Geophys. Res. Lett.*, 31(11).

Rabe, T. J., T. Kukulka, I. Ginis, T. Hara, B. G. Reichl, E. A. D'Asaro, R. R. Harcourt, P. P. and Sullivan, 2015: Langmuir turbulence under hurricane gustav (2008). *J. Phys. Oceanogr.*, **45**, 657-677.

Rasp, S., and S. Lerch, 2018: Neural networks for postprocessing ensemble weather forecasts. *Mon. Weather Rev.*, **146,** 3885-3900.

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.

Reichl, B. G., D. Wang, T. Hara, I. Ginis, and T. Kukulka, 2016: Langmuir turbulence parameterization in tropical cyclone conditions. *J. Phys. Oceanogr.*, **46,** 863-886.

Richards, K. J., S.-P. Xie, and T. Miyama, 2009: Vertical mixing in the ocean and its impact on the coupled ocean–atmosphere system in the eastern tropical Pacific. *J. Clim.*, **22,** 3703-3719.

Sallée, J.-B., E. Shuckburgh, N. Bruneau, A. J. S. Meijers, T. J. Bracegirdle, Z. Wang, & T. Roy, 2013: Assessment of Southern Ocean Water Mass Circulation and Characteristics in CMIP5 Models: Historical Bias and Forcing Response. *J. Geophys. Res.-Oceans*, **118**, 1830-1844.

Sinha, N., A. E. Tejada-Martínez, C. Akan, & C. E. Grosch, 2015: Toward a K-profile parameterization of Langmuir turbulence in shallow coastal shelves. *J. Phys. Oceanogr.*, **45(12)**, 2869-2895.

Skyllingstad, E. D., W. D. Smyth, J. N. Moum, and H. Wijesekera, 1999: Upper-ocean turbulence during a westerly wind burst: A comparison of large-eddy simulation results and microstructure measurements. *J. Phys. Oceanogr.*, **29,** 5-28.

Smith, R., and Coauthors, 2010: The parallel ocean program (POP) reference manual ocean component of the community climate system model (CCSM) and community earth system model (CESM). *LAUR-01853*, **141,** 1-140.

Sullivan, P.P. and J.C. McWilliams, 2010: Dynamics of winds and currents coupled to surface waves. *Annu. Rev. Fluid Mech.*, **42**, 19-42.

Sullivan, P. P., and J. C. McWilliams, 2019: Langmuir turbulence and filament frontogenesis in the oceanic surface boundary layer. *J. Fluid Mech.*, **879,** 512-553.

Sullivan, P. P., J. C. McWilliams, and C.-H. Moeng, 1996: A grid nesting method for large-eddy simulation of planetary boundary layer flows. *Bound.-Layer Meteor.*, **80**, 167-202.

Taylor, J. R., and R. Ferrari, 2011: Ocean fronts trigger high latitude phytoplankton blooms. *Geophys. Res. Lett.*, **38,** L23601.

Tejada-Martinez, A. E., and C. E. Grosch, 2007: Langmuir turbulence in shallow water. Part 2. Large-eddy simulation. *J. Fluid Mech.*, **576**, 63-108.

Thomson, J., E. A. D'Asaro, M. F. Cronin, W. E. Rogers, R. R. Harcourt, and A. Shcherbina, 2013: Waves and the equilibrium range at ocean weather station p, *J. Geophys. Res. Oceans*, **118**, 5951-5962.

Umlauf, L. and H. Burchard, 2003: A generic length-scale equation for geophysical turbulence models. *J. Mar. Res.*, **61,** 235-265.

Umlauf, L., and H. Burchard, 2005: Second-order turbulence closure models for geophysical boundary layers. A review of recent work. *Cont. Shelf Res.*, **25,** 795-827.

Umlauf, L., H. Burchard, H., & K. Bolding, (2014). GOTM Sourcecode and Test Case Documentation.

Van Roekel, L., B. Fox-Kemper, P. Sullivan, P. Hamlington, and S. Haney, 2012: The form and orientation of Langmuir cells for misaligned winds and waves. *J. Geophys. Res. Oceans*, **117,** C05001.

Van Roekel, L., and Coauthors, 2018: The KPP boundary layer scheme for the ocean: Revisiting its formulation and benchmarking one-dimensional simulations relative to LES. *J. adv. model. earth syst.*, **10,** 2647-2685.

Wang, X., and T. Kukulka, 2021: Ocean Surface Boundary Layer Response to Abruptly Turning Winds. *J. Phys. Oceanogr.*, **51,** 1779-1794.

Warner, J. C., B. Armstrong, R. He, J. B. and Zambon, 2010: Development of a coupled ocean–atmosphere–wave–sediment transport (COAWST) modeling system. *Ocean model.*, **35**, 230-244.

Yang, D., B. Chen, B., M. Chamecki, & C. Meneveau, 2015: Oil plumes and dispersion in Langmuir, upper-ocean turbulence: Large-eddy simulations and K-Profile parameterization. *J. Geophys. Res.-Oceans*, **120**, 4729-4759.

Yan, C., J. C. McWilliams, and M. Chamecki (2022). Overlapping boundary Layers in Coastal Oceans, *J. Phys. Oceanogr.*, 52(4), 627-646.

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.

Zanna, L., and T. Bolton, 2020: Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.*, **47,** e2020GL088376.

Zeng, X., Y. Li, and R. He, 2015: Predictability of the loop current variation and eddy shedding process in the Gulf of Mexico using an artificial neural network approach. *J. Atmos. Ocean. Technol.*, **32**, 1098-1111.

Table 1. Sensitivity experiments using the DNN model described in Section 3. Under the third column "forcing conditions": U_{10} is the wind at 10 meter above the sea level. $B(z = 0)$ is the surface buoyancy flux, $\overrightarrow{U_{st}(z=0)}$ is the surface Stokes drift vector, and $U_{st}(z=0)$ is the magnitude of the surface Stokes drift.

Figure 1. (a) Regime diagram for mixing in the ocean surface boundary layer when surface buoyancy flux is destabilizing [Belcher et al. 2012 and Li et al. 2019]. The thick black lines encompass parameter space where one of the three types of turbulence dominates. The thin black lines are contours for the probability (30%, 60%, 90% and 99%) of a certain parameter combination in the global ocean. The thin red lines are contours for the probability (30%, 60%, 90% and 99%) of a certain parameter combination at ocean station Papa (50ºN 145ºW). The grey scattered circles are individual data points from observations at OSP; (b) same as (a) except when surface buoyancy flux is stabilizing. (c) Geographic location of Ocean Station Papa (OSP).

Figure 2. The comparison of mixed layer depth (a), mixed layer temperature (b), and mixed layer salinity (c) between the LES solutions and *in situ* observation, respectively.

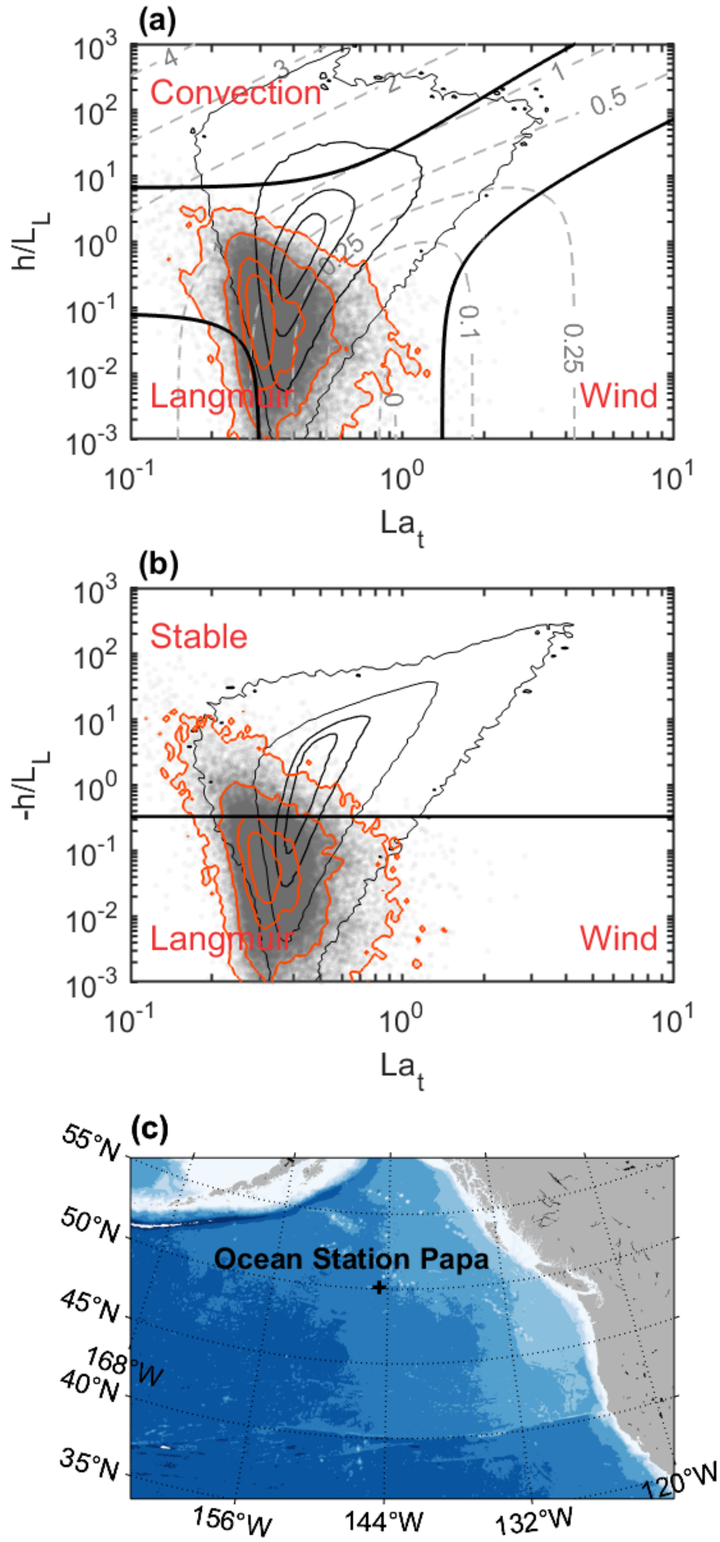
Figure 3. The architect of the Deep Neural Network (DNN) model.

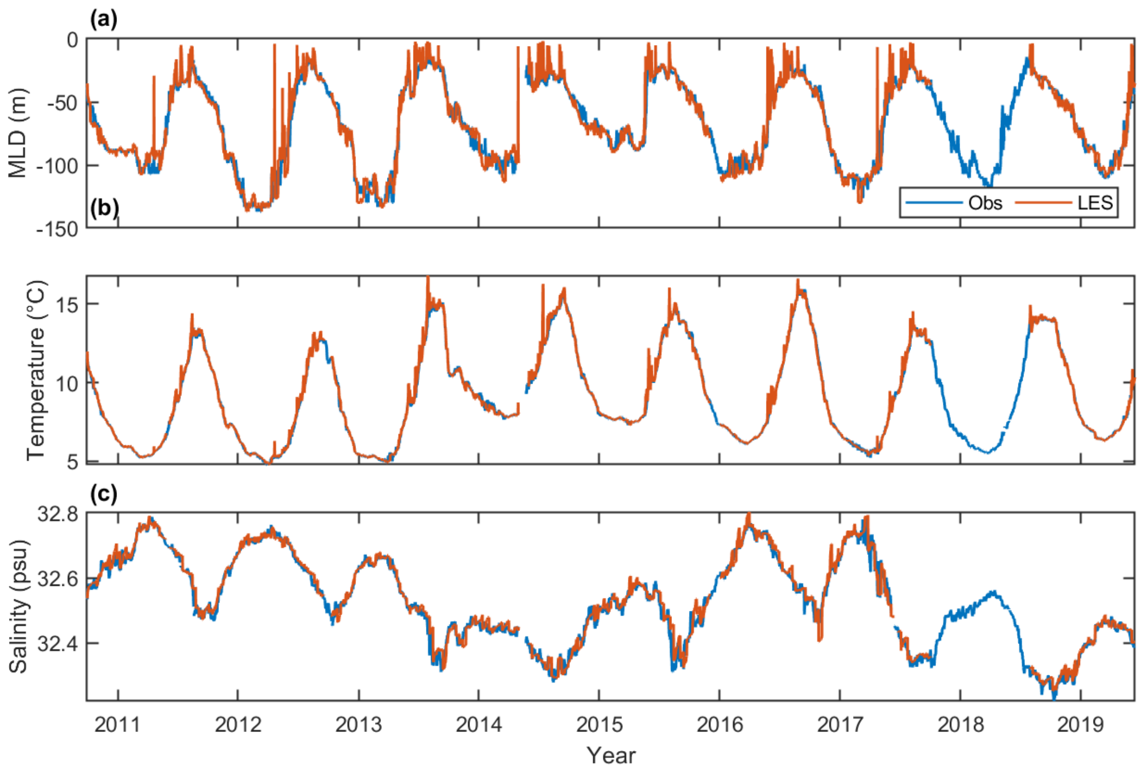
Figure 4. Meteorological conditions and solutions during a 9 day period starting from Dec. 19th 2010. (a) Wind speed at 10-m above sea level (U_{10}) and turbulent Langmuir number (La_t) based on formulas by McWilliams et al. [1997] (MSM97) and van Roekel et al. [2012] (VR12). Dashed line in the panel indicates $La_t = 0.3$, below which turbulence is dominated by wave-driven Langmuir turbulence. (b) surface heat flux and evaporation minus precipitation; (c) and (d) the simulated profiles of temperature and salinity using the LES model (truth), respectively. (e) and (f) the error of simulated temperature and salinity by the Deep Neural Network (DNN) model, respectively (here, error is defined as difference from the LES solutions, i.e., $E(x) = x_{DNN} - x_{truth}$). (g) and (h) the error of the simulated temperature and salinity profiles using the KPP-CVmix model [Large et al. 1994; Li et al. 2021]. (i) and (j) the error of the simulated temperature and salinity profiles using the KPP-LF17 model [Li and Fox-Kemper 2017].

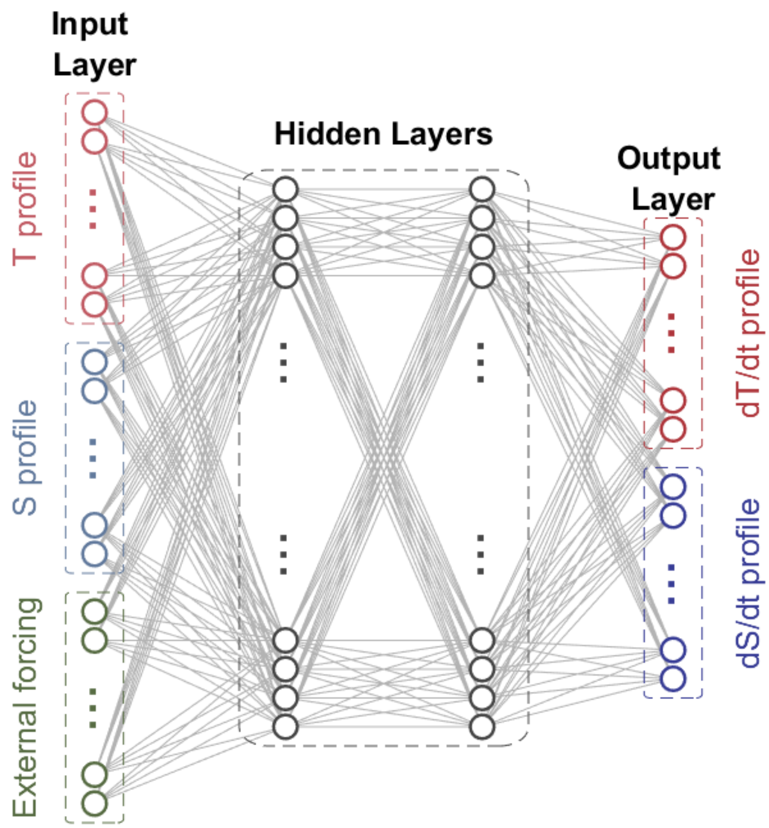
Figure 5. (a) to (c): the mean of modeled temperature errors by the DNN, the KPP-CVMix [e.g., Large et al. 1994] and the KPP-LF17 [Li and Fox-Kemper 2017], respectively ($\overline{E[T]} = \frac{1}{N}\sum_{n=1}^{N} E[T]$, where $E(x) = x_{model} - x_{truth}$ with the LES solutions considered the truth and N is the number of records); (d) to (f) the standard deviation of model errors by the DNN, the KPP-CVMix and the KPP-LF17, respectively ($\sigma_{E[T]} = \frac{1}{N}\sqrt{\sum_{n=1}^{N}\left(E[T] - \overline{E[T]}\right)^2}$). (g) to (i): same as panels (a) to (c), but for salinity. (j) to (l): same as panels (d) to (f), but for salinity.

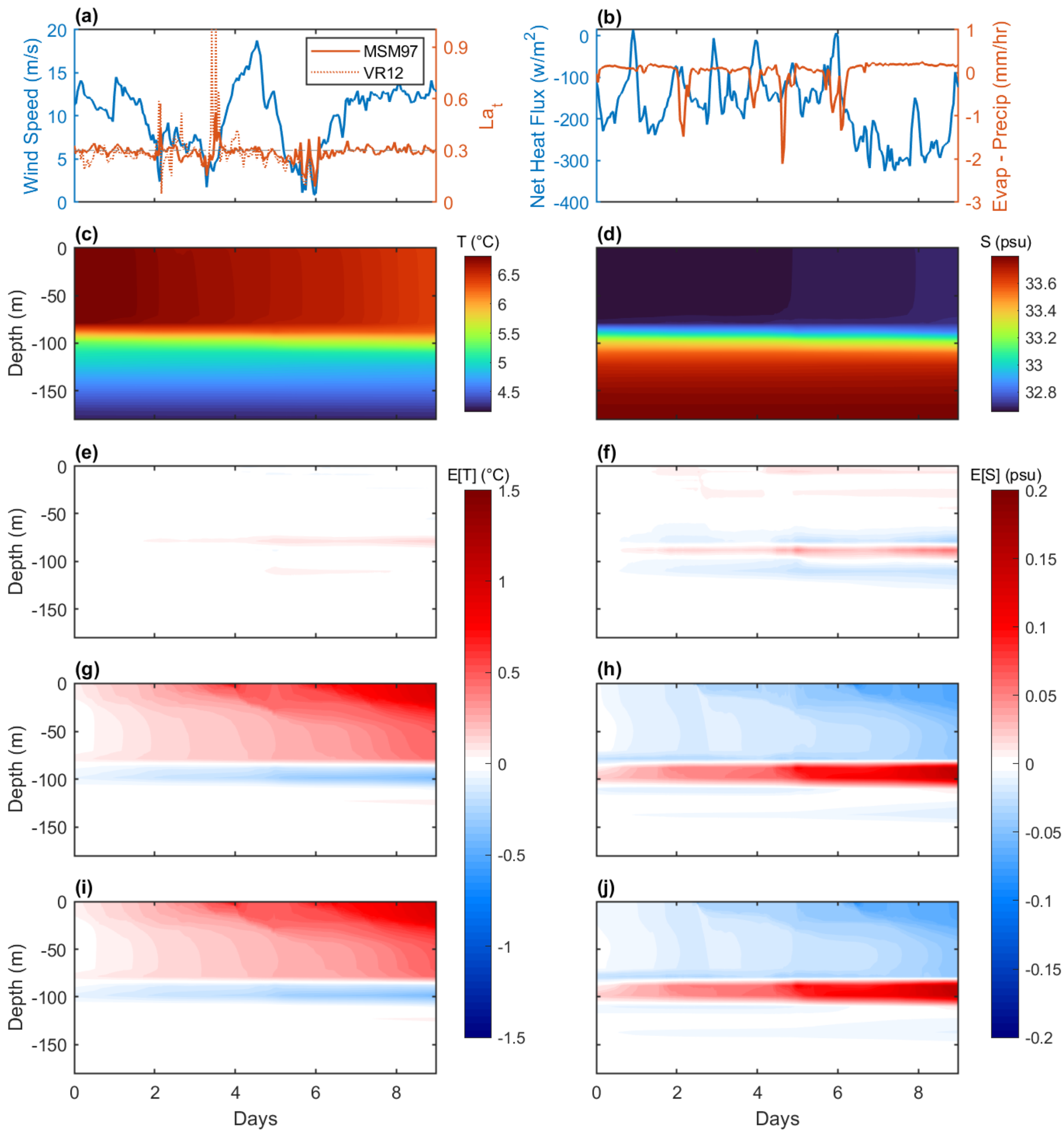
Figure 6. Taylor diagram for the rate of change of temperature (panel a) and salinity (panel b) in the OSBL, respectively. The solid grey lines, solid black lines, and dashed black lines are contours of normalized root mean square error $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(F_i - O_i)^2} \Big/ \left[\frac{1}{N}\sqrt{\sum_{n=1}^{N}\left(O_i - \overline{O_i}\right)^2}\right]$ with F the prediction by the DNN model and O the truth (LES solutions), normalized standard deviation $\sigma_{F_i} = \sqrt{\sum_{n=1}^{N}\left(F_i - \overline{F_i}\right)^2} \Big/ \sqrt{\sum_{n=1}^{N}\left(O_i - \overline{O_i}\right)^2}$, and correlation $Cor = \sum\left(F_i - \overline{F}_i\right)\left(O_i - \overline{O}_i\right) / \sqrt{\sum\left(F_i - \overline{F}_i\right)^2 \sum\left(O_i - \overline{O}_i\right)^2}$, respectively.

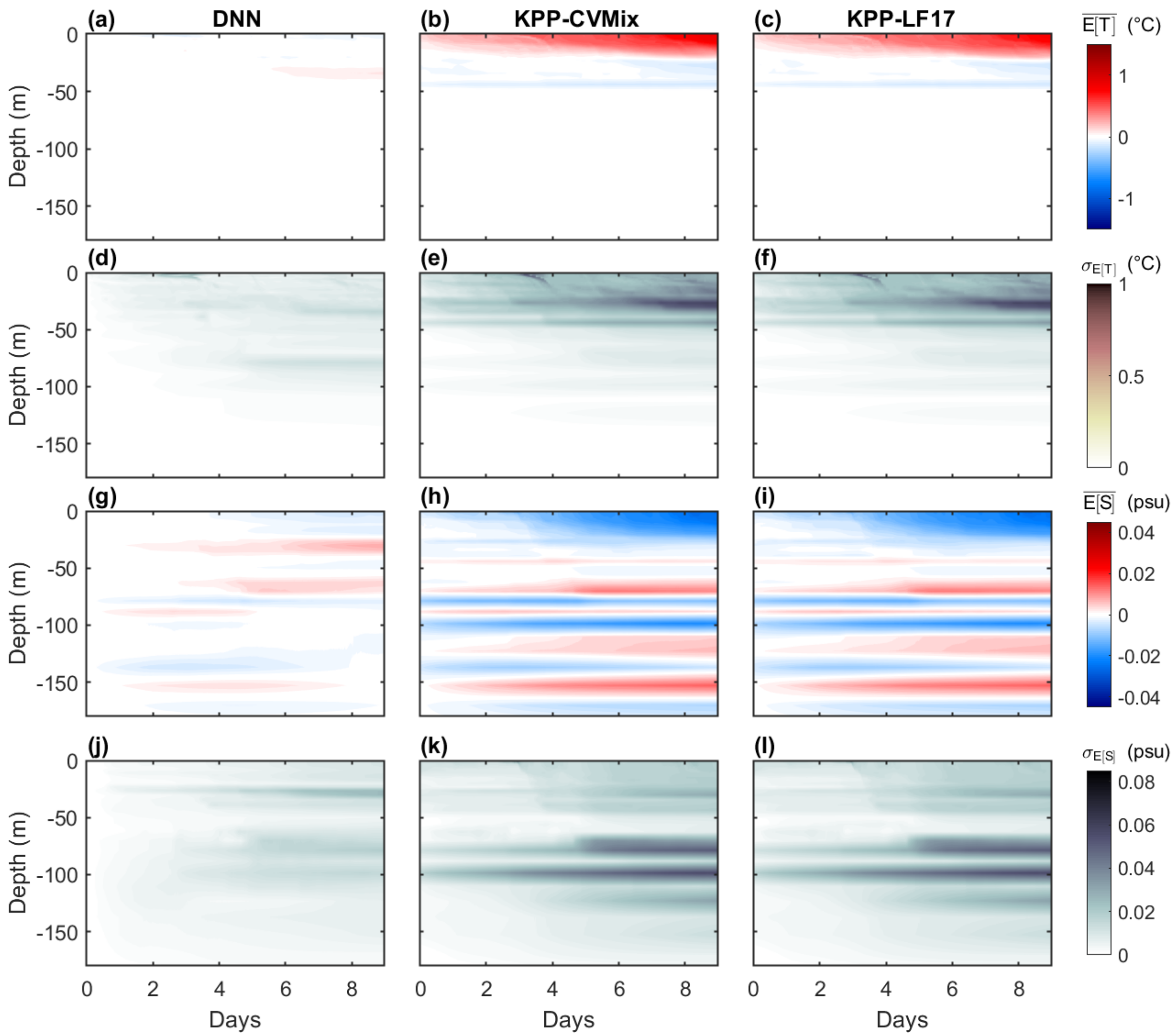
Simulation name	The combination of Neuron and hidden layer	Forcing conditions	Length of forcing condition
Control simulation	256×1	$\overrightarrow{U_{10}}$, $B(z=0)$, $\overrightarrow{U_{st}(z)}$	30-min (t and t - Δt)
Sensitivity Experiment 1	[128 × 1, 256 × 1]	$\overrightarrow{U_{10}}$, $B(z=0)$, [$\overrightarrow{U_{st}(z=0)}$, $U_{st}(z=0)$,]	30-min (t and t - Δt)
Sensitivity Experiment 2	128 × 1 (no history) 256 × 1 (1-hour history)	$\overrightarrow{U_{10}}$, $B(z=0)$, $\overrightarrow{U_{st}(z)}$	[no history (t), 1-hour (t, $t-\Delta t$, t - $2\Delta t$)]


(a)
Convection
Langmuir
Wind
h/L_L
La_t
(b)
Stable
Langmuir
Wind
-h/L_L
La_t
(c)
Ocean Station Papa
55°N
50°N
45°N
40°N
35°N
168°W
156°W
144°W
132°W
120°W

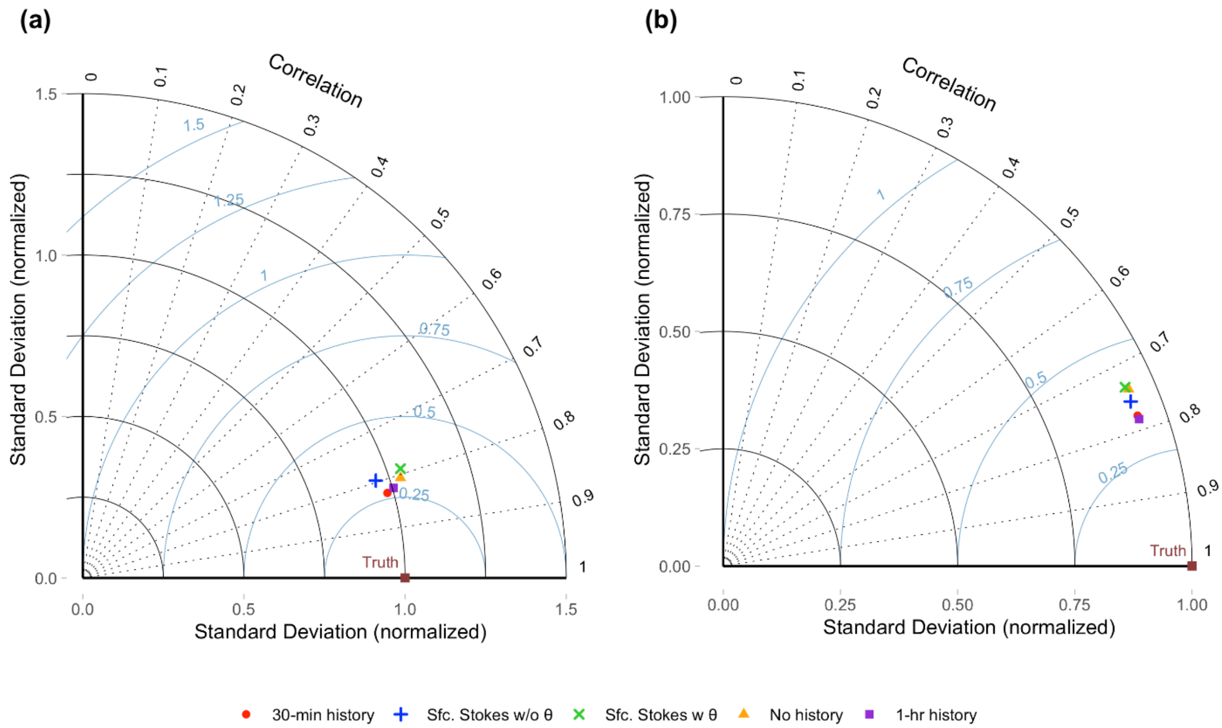

(a)
(b)
(c)
MLD (m)
Temperature (°C)
Salinity (psu)
Obs
LES
Year
0
-50
-100
-150
15
10
5
32.8
32.6
32.4
2011
2012
2013
2014
2015
2016
2017
2018
2019


Input Layer
Hidden Layers
Output Layer
T profile
S profile
External forcing
dT/dt profile
dS/dt profile


(a)
Wind Speed (m/s)
La$_t$
MSM97
VR12
(b)
Net Heat Flux (w/m^2)
Evap - Precip (mm/hr)
(c)
T (°C)
(d)
S (psu)
(e)
E[T] (°C)
(f)
E[S] (psu)
(g)
(h)
(i)
(j)
Depth (m)
Days



(a) DNN
(b) KPP-CVMix
(c) KPP-LF17
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
Depth (m)
0
-50
-100
-150
0
2
4
6
8
Days
$\overline{E[T]}$ (°C)
1
0
-1
$\sigma_{E[T]}$ (°C)
1
0.5
0
$\overline{E[S]}$ (psu)
0.04
0.02
0
-0.02
-0.04
$\sigma_{E[S]}$ (psu)
0.08
0.06
0.04
0.02
0


(a)
Correlation
Standard Deviation (normalized)
Standard Deviation (normalized)
Truth
(b)
Correlation
Standard Deviation (normalized)
Standard Deviation (normalized)
Truth
30-min history
Sfc. Stokes w/o θ
Sfc. Stokes w θ
No history
1-hr history