Geophysical Research Letters

RESEARCH LETTER
10.1029/2026GL122106

Machine Learned Equations for Vertical Mixing Coefficients in the Ocean Surface Boundary Layer

Aakash Sane[1], Brandon G. Reichl[2], Alistair Adcroft[1], and Laure Zanna[3]

[1]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA, [2]NOAA - Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [3]Courant Institute, New York University, New York, NY, USA

Key Points:

- We find interpretable equations for vertical diffusivity using machine learning-based equation discovery and least squares fitting
- Equations have similar skill as neural networks and reveal structural error in the baseline physics-based parameterization
- Equations elucidate shifting of the diffusivity peak toward surface under stabilizing and toward mid-boundary layer depth under convection

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Sane,
aakash.sane@princeton.edu

Citation:

Sane, A., Reichl, B. G., Adcroft, A., & Zanna, L. (2026). Machine learned equations for vertical mixing coefficients in the ocean surface boundary layer. *Geophysical Research Letters*, *53*, e2026GL122106. https://doi.org/10.1029/2026GL122106

Received 26 JAN 2026
Accepted 25 MAR 2026

Author Contributions:

Conceptualization: Aakash Sane
Funding acquisition: Alistair Adcroft, Laure Zanna
Methodology: Aakash Sane
Project administration: Brandon G. Reichl, Alistair Adcroft, Laure Zanna
Resources: Aakash Sane
Supervision: Brandon G. Reichl, Alistair Adcroft, Laure Zanna
Writing – review & editing: Aakash Sane, Brandon G. Reichl, Alistair Adcroft, Laure Zanna

Abstract Neural networks offer novel ways to parameterize unresolved ocean mixing but are challenging to interpret. Here, we derive compact equations that reproduce the behavior of neural networks trained on a second-moment closure data set. The resulting interpretable expressions employed in a physics-based first order closure scheme match neural-network performance in global forced simulations. They expose a structural error in the baseline physics-based scheme and describe how surface friction velocity, buoyancy flux, rotation, and boundary layer depth regulate diffusivity. The equations reveal a shift in the mixing peak toward the surface under stabilizing conditions and toward the mid-boundary layer depth under convective conditions. The diffusivity amplitude (set by the velocity scale) is controlled by surface shear and buoyancy flux. Equations yield a transparent, efficient, and physically grounded vertical diffusivity applicable for ocean models.

Plain Language Summary This study focuses on improving the representation of ocean turbulence in climate models. In many climate models, vertical mixing of the upper ocean is simplified using computationally efficient turbulence models. These models estimate the mixing process using physical rules and educated guesses, but they can miss important details, especially under changing wind forcing and solar heating. To address this, we previously developed neural networks that improved the accuracy of these estimates by mimicking more detailed models. However, these networks were difficult to interpret and computationally expensive to use in global simulations. In this work, we explore the use of equation discovery techniques that find simple, interpretable equations to replicate the effects of neural networks. We developed new equations that are easier to understand and less computationally demanding but still improve the accuracy of the model. The new equations were tested in a state-of-the-art ocean model to simulate the ocean. The results show that they can effectively capture the complex turbulent mixing processes occurring in the upper ocean. They do so with the same effectiveness as neural networks but without their computational burden.

1. Introduction

Turbulent mixing in the ocean surface boundary layer (OSBL) has significant implications for physical and biogeochemical processes in the upper ocean, and for large-scale circulation and climate. This mixing occurs on spatiotemporal scales that are smaller than those resolved in a typical ocean climate model, so it must be accurately parameterized. Various mixing parameterizations proposed for climate models disagree with each other and with observations, introducing uncertainty in ocean climate simulations (Li et al., 2019).

Deficiencies in such turbulence parameterizations may arise due to an incomplete understanding of physics, leading to incorrect equations formulated on the model grid. Uncertainty may also arise from ad-hoc choices that introduce poorly constrained parameters, contributing to both parametric and structural uncertainty (see Balaji et al. (2022), Christensen and Zanna (2022), and Schneider et al. (2024)). In OSBL schemes, *parametric* uncertainty reflects uncertain coefficients or tunable parameters within a chosen formulation, whereas *structural* uncertainty reflects missing or misrepresented physics or an inadequate functional form of the parameterization.

Recent progress in data-driven approaches allows us to alleviate these deficiencies. Machine learning is an attractive tool for the turbulence closure problem, as it can learn the mapping between inputs and outputs. Considerable progress has been made in the application of neural networks in the domain of earth sciences. Artificial neural networks (ANN) have been applied to model sub-grid momentum transport (Yuval & O'Gorman, 2023). Convolutional neural networks were able to learn data-assimilation increments, enabling state-dependent bias corrections (Chapman & Berner, 2024; Gregory et al., 2023, 2024; Watt-Meyer et al., 2021). For oceanic submeso and mesoscale turbulence, deep ANN, and convolutional neural networks have shown

promise in parameterizing sub-grid fluxes (Balwada et al., 2025; Bodner et al., 2025; Partee et al., 2022; Perezhogin et al., 2024, 2025; Zanna & Bolton, 2020; Zhang et al., 2025; Zhou et al., 2024). For oceanic vertical mixing in the boundary layer, Liang et al. (2022) modeled the temporal rate of change of temperature and salinity profiles while Yuan et al. (2024) modeled parameters in the KPP scheme (Large et al., 1994) by employing ANNs. Iyer et al. (2025) and Mashayek et al. (2022) use supervised learning to obtain turbulence properties from observations, while Chen et al. (2025) uses unsupervised learning to understand frontal features. The above list is not exhaustive, but it points toward the capability of machine learning to model sub-grid physics, often with better skill than traditional methods. Black-box neural networks are skillful, but their interpretation is challenging and often speculative. Here, we use equation discovery to circumvent this challenge and provide us with an interpretation of the learned networks.

1.1. Machine Learning for Parameterizations

We focus on improving the parameterization of OSBL diffusivity in a "first-order" scheme. First-order parameterizations impose a vertical profile of diffusivity. These approximations, although physically motivated, may not capture the complex physics observed in higher order closures. Here, we modify one such closure, called the energetics-based Planetary Boundary Layer scheme (ePBL, Reichl & Hallberg, 2018). This scheme provides the OSBL mixing in NOAA-GFDL's fourth generation climate model suite (see Adcroft et al. (2019)). ePBL diagnoses the diffusivity (κ) as the product of a turbulent velocity scale (v_0), surface boundary layer depth (h), and a structural shape function $g(\sigma)$. σ is the non-dimensional depth coordinate, $\sigma = -z/h$, where z is the depth:

$$\kappa(\sigma) = v_0 \cdot h \cdot g(\sigma). \tag{1}$$

In ePBL, the net mixing is constrained by the buoyancy flux through the energy available to homogenize stable buoyancy gradients (see Reichl & Hallberg, 2018 for details). The energetic calculation is used to estimate h, which is one component of diffusivity. The other two components, $g(\sigma)$ and v_0, are approximate and contain uncertain parameters. To reduce the structural error in $g(\sigma)$ and v_0, our earlier work used neural networks trained on second moment closure (SMC) data to predict the ePBL shape function $g(\sigma)$ and the velocity scale v_0 (Sane et al., 2023).

SMC scheme was used in a single column setting and forced at the surface. The forcing space spanned various wind stresses (0–20 N/m^2), surface heat (buoyancy) flux (−2,000–2,000 W/m^2), and different latitudes (0°–90°). For details related to generating data from SMC see Sane et al. (2023).

We introduce two functions to be optimized:

$$g(\sigma) = \mathcal{F}_1(|f|, B, u_*, h), \tag{2a}$$

$$v_0 = \mathcal{F}_2(|f|, B, u_*, h), \tag{2b}$$

where $|f|$ is the absolute value of the Coriolis parameter, B is the surface buoyancy flux, and u_* is the surface friction velocity. Here $\mathcal{F}_1$ and $\mathcal{F}_2$ are two target functions that need to be optimized. In the neural network enhanced version of ePBL (called ePBL_NN hereafter), $\mathcal{F}_1$ and $\mathcal{F}_2$ are neural networks; and in the equations-based enhancement (referred to as ePBL_EQ hereafter), they are expressions.

Although our previous work (Sane et al., 2023) has shown that machine learning can enhance ePBL based on SMC, there are drawbacks to employing black-box neural networks for routine ocean climate simulations. The opaque nature of neural networks can hinder a better understanding of the underlying physics and any structural deficiencies in the baseline scheme. In contrast, equations are easy to interpret and computationally less expensive than neural networks. Hence, it is preferable to apply machine learning for equation discovery on the SMC data to replace the networks. Such approaches have proven successful in mesoscale closure problems (Zanna & Bolton, 2020), yielding compact, transparent equations.

The choice of SMC for training data is made following Sane et al. (2023). We could have used OSBL large eddy simulations (LES) that directly simulate OSBL turbulence rather than relying on SMC parameterizations. However, LES data sets require significantly more computational resources to generate and therefore an LES data set that spans the entire parameter space of interest is not available at present. Considerable literature exists

signifying the importance of LES data sets and their role in elucidating OSBL physics (Large et al., 2019; Whitt et al., 2022). SMC schemes have been compared and constrained to LES data set in realistic ocean conditions (e.g., Reichl et al., 2024), suggesting they can be trusted to generate large training data sets. Hence, for this work we use SMC as the ground truth to develop equations for diffusivity.

1.2. Approach

We aim to capture the key behaviors between the inputs and outputs stated in Equations 2a and 2b. Following Sane et al. (2023): we train neural networks on the "Generic Length Scheme" (hereafter, GLS scheme; Umlauf & Burchard, 2003) instead of the $k-\epsilon$ scheme in Sane et al. (2023). The profile of κ in GLS is smoother than in $k-\epsilon$, making it easier to obtain a smooth function for the vertical structure of diffusivity (shape function). To generate the training data, single-column GLS runs were initialized with a stably stratified profile and then forced by constant surface wind stress and buoyancy flux for different latitudes (as in Sane et al. (2023)). The details of splitting the data set into training and testing for neural networks are exactly the same as given in our previous work (Sane et al., 2023). The loss curves in training the neural networks are given in Supporting Information S1. The trained networks predict $g(\sigma)$ and v_0 as a function of the four inputs $|f|, B, u_*, h$ shown in Equations 2a and 2b.

We obtain equations using the distilled data from the learned neural networks rather than from the raw data. Distillation involves using a neural network to predict certain properties and then applying equation discovery (machine learning or least squares fitting) on the network's predictions rather than on the original data set. This provides a regularized target for equation discovery, effectively constraining the search to the dominant structure captured by the neural network (Champion et al., 2019; Cranmer et al., 2020). As neural networks are function approximators, neural network inferred data should be coherent as compared to raw SMC data. This also smooths out, to some extent, any spurious or noisy behavior in the SMC scheme that could arise from ad-hoc choices of tunable parameters in the ground truth SMC scheme itself.

To aid the algorithm in finding dimensionally consistent equations, we non-dimensionalized the variables inspired by physics for generalization (e.g., Beucler et al., 2024). We use five parameters f, u_*, B, h, Ω (Earth's rate of rotation) to construct three non-dimensional numbers: boundary layer depth divided by Monin-Obukhov depth $L_h = (|B|h)^{1/3}/u_*$, boundary layer depth divided by Ekman depth $E_h = h|f|/u_*$, and the non-dimensional Coriolis parameter $f' = |f|/\Omega$.

To obtain closed-form equations, we applied genetic programming (GP) and least squares fitting to neural network inferred data. GP is a machine learning technique that gives transparent and interpretable expressions. It constructs expressions from prescribed variables and operators through evolution-inspired search algorithms (Koza, 1994; Ross et al., 2023). GP reduces dependency on heuristic assumptions and gives optimal expressions based on our choices of parameters and operators.

Following Ross et al. (2023), we opt for GP in this work since it can build an expression from variables without any a priori assumption about its structure. This is unlike regression techniques, where the structure of the equation is assumed. GP produced satisfactory, compact relations in several instances, but it did not consistently outperform least squares curve fitting.

Consequently, we prioritized parsimonious expressions consistent with expected behavior from data and physical understanding. When GP returned long expressions that were challenging to interpret, we used physically guided functional forms and fit coefficients by least squares. The resulting expressions are interpretable, involve only $\mathcal{O}(10)$ parameters, and are substantially less computationally expensive than the original networks. All discovered and least squares fitted equations are listed in Table 1.

2. Data-Driven Equations for Vertical Diffusivity

2.1. Modeling the Shape Function

From the training data, we diagnose the shape function as $g(\sigma) = \kappa(\sigma)/\max(\kappa(\sigma))$, following Sane et al. (2023). The neural network denoted as $g^{NN}(\sigma)$ predicts $g(\sigma)$ with high skill (see Figure 1a).

Table 1
Top: Equations and Definitions for Shape Function and Velocity Scales

Description	Equation		Figure
Non-dimensional numbers. sgn(B) < 0 is surface heating and sgn(B) > 0 is cooling	$L_h = \frac{(\lvert B\rvert h)^{1/3}}{u_*}, E_h = \frac{h\lvert f\rvert}{u_*}, \lambda = \sqrt{\frac{L_h^3}{E_h}}, f' = \frac{\lvert f\rvert}{\Omega}$	(4)	–
Shape function maximum	$\sigma_m = \frac{1}{c_1 + \frac{c_2}{\mathcal{F}\cdot E_h}}$	(5)	Figure 1c
Intermediate function for σ_m	$\mathcal{F} = \frac{1}{c_3 + c_4 \cdot e^{-\left(\mathrm{sgn}(B)\cdot c_5 \cdot {L_h}^3\right)}} + c_6$	(6)	–
v_0 Surface heating	$\frac{v_0}{u_*} = \frac{c_7}{\lambda + c_8 + \frac{c_9^2}{\lambda + c_9}}$	(7)	Figure 1e
v_0 Surface cooling	$\frac{v_0}{u_*} = \frac{c_{10}\cdot\lambda\cdot\sqrt{f'}}{1 + \frac{\left(c_{11}e^{\left(-c_{12}\cdot f'\right)} + c_{13}\right)}{\lambda^2}} + c_{14}$	(8)	Figure 1f
v_0^h Surface heating	$\frac{v_0^h}{u_*} = \frac{c_{14}}{c_{15}L_h^3 + c_{16}L_h^2 + 1}$	(9)	Figure 1g
v_0^h Surface cooling	$\frac{v_0^h}{u_*} = \frac{L_h}{c_{17} + \frac{c_{18}}{L_h^2}} + c_{14}$	(10)	Figure 1h

Coefficient	Value	Coefficient	Value	Coefficient	Value	Coefficient	Value
c_1	1.7908	c_6	1.5845	c_{11}	45.0	c_{16}	0.0944
c_2	0.6904	c_7	0.1550	c_{12}	2.8570	c_{17}	6.0277
c_3	0.0712	c_8	1.1120	c_{13}	3.290	c_{18}	15.7292
c_4	0.4380	c_9	0.8616	c_{14}	0.0785		
c_5	2.6821	c_{10}	0.0984	c_{15}	0.650		

Note. Neutral surface forcing ($B = 0$) is included in surface heating. The velocity scale v_0 depends on f, B, u_*, while v_0^h also includes h. Bottom: Values of coefficients used in the equations.

Using GP on profiles of $g(\sigma)$ yielded high-order polynomials that were challenging to constrain to strictly positive values. To overcome this, we transformed the problem into finding a relation between σ_m and the forcing parameters f, B, u_* and h. σ_m is the value of σ at maximum diffusivity and is defined by $g(\sigma_m) = 1$.

Even on the reduced problem (finding the expression of σ_m instead of $g(\sigma)$), GP provided an unsatisfactory outcome; hence, we used least squares fitting to obtain an expression for σ_m. The formulation, presented as Equation 5 in Table 1, depends only on the two non-dimensional numbers L_h and E_h (see Equation 4 in Table 1). There is an intermediate function $\mathcal{F}$ that depends only on L_h. $\mathcal{F}$ has been found using empirical fitting by observing data for σ_m. A smaller σ_m means the peak of $g(\sigma)$ lies nearer to the surface, indicating stronger near-surface mixing. This limit occurs during surface heating, when $L_h^3 \to -\infty$ and $E_h \to 0$, and more generally, whenever the boundary-layer depth h is less than both the Monin–Obukhov ${u_*}^3/|B|$ and the Ekman depth scale $u_*/|f|$. Under these conditions, as illustrated in Figure 1c, the diffusivity maximum shifts upward, concentrating mixing near the surface. In KPP, the mixing coefficient is similarly decreased as one moves away from the surface for the stable boundary layer by decreasing the velocity scale with $|z|/L_h$, which was justified based on available atmospheric observations of the stable boundary layer (see Figure 2 in Large et al. (1994)).

For surface cooling conditions, σ_m has shallower values for $L_h^3 \to 0$ and $E_h \to 0$, and as L_h and E_h increase, diffusivity intensification reaches the midpoint of the boundary layer at around $\sigma \approx 0.55$. This can be observed from Equation 5 in Table 1. Assume a large value of E_h, say 1.0, and assume $L_h^3 \to \infty$. Then Equation 5 in Table 1 gives: $\sigma_m = 0.55$. For both L_h and E_h approaching asymptotically large positive values (under surface cooling conditions), Equation 5 in Table 1 gives $\sigma_m = 1/c_1 = 0.558$. This shows that for strong convective conditions, the diffusivity has a maximum at ≈0.55, that is, slightly deeper than the midpoint of the boundary layer. In Figure 1c, the first quadrant shows values of σ_m tending toward 0.55. This may indicate an equilibration of the boundary layer vertical structure under strong convective conditions. Again, a related behavior can be noted in KPP, where it is accomplished through an increase in the velocity scale moving away from the surface.

We set $g^{EQ}(\sigma)$ with a quadratic profile for $0 \le \sigma \le \sigma_m$ and a cubic profile for $\sigma_m \le \sigma \le 1$:

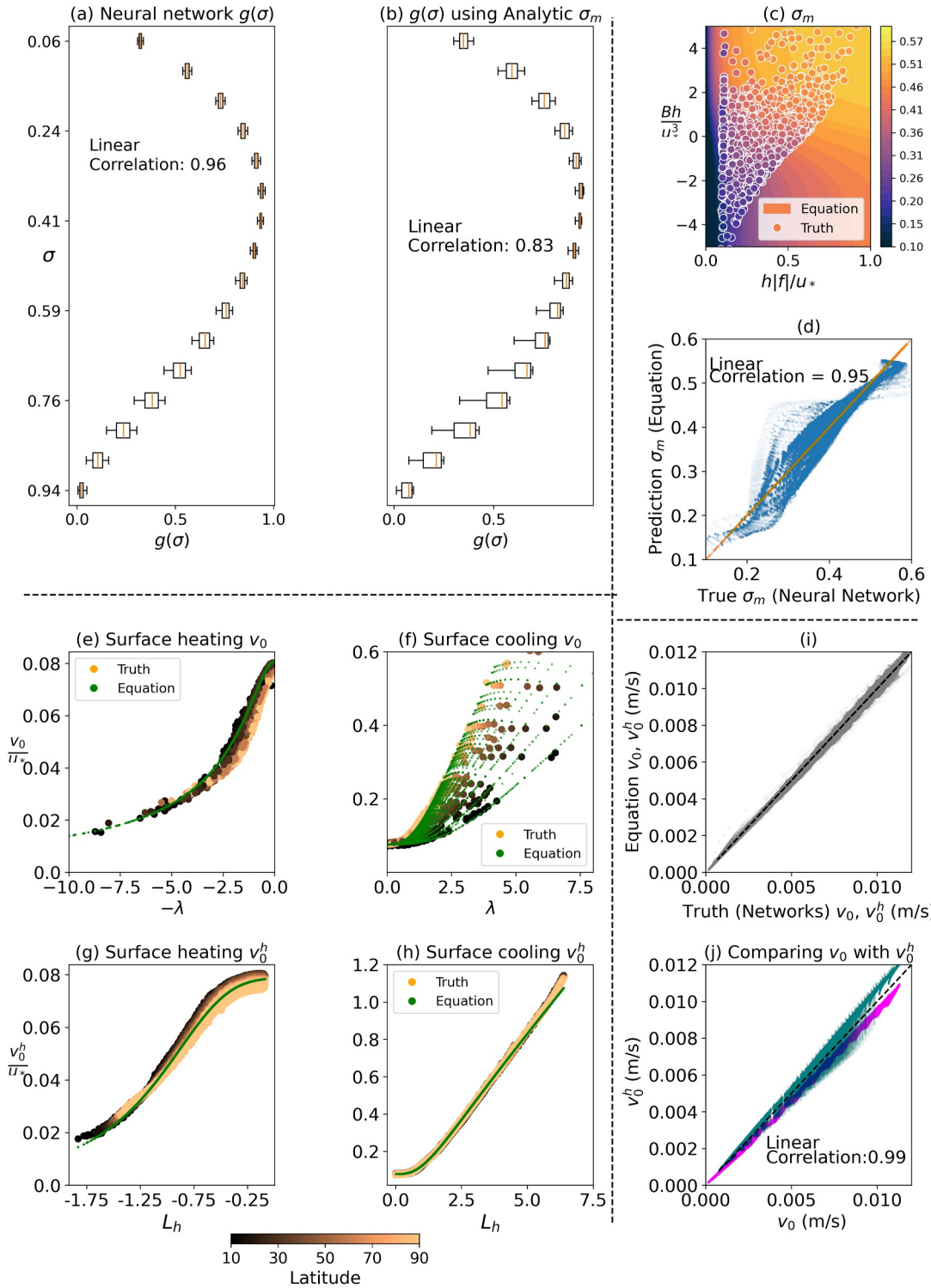

Figure 1. Skill of machine learning discovered and least squares fitted equations. (a, b) Skill of neural network and equation based predictions for the shape function, respectively. (b) Shape function uses analytical σ_m with quadratic-cubic vertical profile (Equation 3) plotted against second moment closure data. Box plots in panels (a, b) show the error statistics while high linear correlation shows agreement of the machine learned and least square fitted equations with the ground truth. (c) Equation for σ_m, uses two non-dimensional numbers: $L_h = (|B|h)^{1/3}/u_*$ and $E_h = h|f|/u_*$, $L_h E_h$ (d) shows parity plot. Plots (e–h): velocity scale v_0 depends on $\lambda = \sqrt{L_h^3/E_h}$, (e) shows surface heating conditions (Equation 7 in Table 1), and (f) shows cooling conditions (Equation 8 in Table 1). The second formulation for velocity scale v_0^h depends on L_h, plot (g) surface heating (Equation 9 in Table 1) and (h) cooling conditions (Equation 10 in Table 1). (i) Parity plot of v_0 and v_0^h against their ground truth. (j) Agreement of v_0 with v_0^h.

$$g^{EQ}(\sigma) = \begin{cases} \frac{2\sigma}{\sigma_m} - \frac{\sigma^2}{\sigma_m^2} & \text{if } 0 \le \sigma \le \sigma_m \\ 2\left(\frac{\sigma - \sigma_m}{1 - \sigma_m}\right)^3 - 3\left(\frac{\sigma - \sigma_m}{1 - \sigma_m}\right)^2 + 1 & \text{if } \sigma_m \le \sigma \le 1 \end{cases} \quad (3)$$

g^{EQ} is continuous and differentiable at $\sigma = \sigma_m$. The quadratic profile for $\sigma \le \sigma_m$ agrees with the true profiles (see Figure 1b). We use a cubic because it is the simplest polynomial that satisfies the boundary condition $g(\sigma = 1) = 0$. These choices of boundary conditions are inspired by the KPP scheme (Large et al., 1994), except that the KPP scheme does not include a model for σ_m. The diffusivity gradient at $\sigma = 1$ is assumed to be zero. We found that modeling a nonzero gradient near $\sigma = 1$ was challenging using equations (either by GP or least squares fitting); however, neural networks were able to capture the slope from the training data.

Figure 1b shows $g^{EQ}(\sigma)$ using σ_m as given by Equation 5 in Table 1. The error bars are plotted for all cases around the mean true profile at sixteen σ points within the boundary layer. The linear correlation at each σ was calculated and weighted by the mean values of the profile to obtain an overall correlation. The equation-based profiles show an accuracy comparable to that of neural networks. The accuracy decreases in the lower part as a result of imposing a cubic curve that may be uncertain in the entraining region of the boundary layer.

2.2. Modeling the Velocity Scale

The velocity scale v_0 is diagnosed as $v_0 = \langle \max(\kappa(\sigma))/h \rangle$ from the SMC data, as given in Sane et al. (2023). The angled brackets denote averaging over h. In a global simulation, if h goes out of the bounds of the training data, then the neural network can make a spurious prediction. We avoided this in our earlier work (Sane et al., 2023) by averaging each forcing case (f, B, u_*) over h to remove its dependence. Removing h has the disadvantage that, in the limit of pure convection, v_0 does not scale as $(Bh)^{1/3}$, which is the well-known Deardorff scaling (Deardorff, 1970). To overcome this, we have a second model for the velocity scale diagnosed without averaging over h, which we denote as v_0^h. It is diagnosed from SMC as $v_0^h = \max(\kappa(\sigma))/h$, and this variable depends on h. Equations for v_0 and v_0^h have been given in Table 1.

2.2.1. Surface Heating and Neutral Conditions

We found an expression for v_0 (Equation 7 in Table 1) under surface heating conditions as a function of λ using GP. Equation 7 in Table 1 approximates the neural network model used to predict v_0. v_0/u_* lies between two values 0 and 0.0785 (see Figure 1e). There exists a transition between two dynamical regimes: one dominated by shear and the other by surface buoyancy. The transition occurs at $\lambda \approx -1$. For $\lambda \ll -1$, $v_0/u_* \to 0.0$, indicating that strong surface heating prevents the OSBL from developing. For $-1 \le \lambda \le 0$, $v_0/u_* \to 0.0785$, implying shear-dominated dynamics in the presence of weak surface heating. As $\lambda \to 0$, v_0 tends to be linear with u_*. Near $\lambda = -1$, shear stress and surface heating contribute equally to v_0 for all latitudes.

For v_0^h under surface heating and neutral conditions GP gave Equation 9 in Table 1 which is a relation between v_0^h and L_h. Equation 9 in Table 1 is an alternative to Equation 7 in Table 1. The distinction is that it estimates the actual velocity scale instead of the velocity scale v_0 averaged over h. GP was tried with various non-dimensional numbers as predictors, and GP found that L_h was relevant to predict v_0^h under surface heating conditions. Similar to v_0, we can observe from Equation 9 in Table 1 that the transition between the shear and buoyancy dominated regimes occurs around $L_h \approx -1$. As $L_h \to 0$ for $L_h > -1$, $v \propto u_*$ and as $L_h \ll -1$, $v_0^h \to 0$ imply that the surface heating flux has restricted the evolution of the boundary layer (see Figure 1g).

Equations 7 and 9 in Table 1 differ from the control ePBL scheme. In the control ePBL scheme $v_0 \propto u_*$ for all values of stabilizing B, it does not depend on B under surface stabilizing conditions (see Equation 9 in Reichl and Li (2019) for more details). This restricts the velocity scale to depend only on wind shear, which disagrees with SMC.

2.2.2. Surface Cooling Conditions

For v_0 under surface cooling conditions, GP failed to yield a compact and interpretable solution. However, inspired by the data shown in Figure 1f, we found the relation to be Equation 8 in Table 1 using empirical fitting. We found the unknown constants c_{10} to c_{14} using Python Scipy's curvefit optimal fitting algorithm. For $\lambda < 1$ and $\lambda \to 0$, shear (u_*) is expected to dominate over convection, so v/u_* should reach a constant value, implying $v \propto u_*$ in this limit. As $\lambda \gg 0$, convection becomes dominant relative to shear in determining v_0. The data show that in this regime, v_0/u_* tends to increase linearly with λ for every f'. This motivates a function that transitions between a constant value as λ goes from 0 to a linear function as $\lambda \gg 1$ as seen in Figure 1f. Hence, a function of the form $\frac{v}{u_*} \approx \frac{A\lambda}{B+C/\lambda^n}$ is appropriate, where A, B, C and n are unknowns to be optimized. To satisfy linearity for strong convection $(\lambda \gg 1)$, which depends on latitude, A must be a function of f'. We chose a square root function as it matched the data well and was convenient to cancel out $\sqrt{1/f}$ in the numerator due to $\sqrt{f'}$. This reasoning led us to formulate Equation 8 in Table 1, and it also provides insight into how v_0/u_* behaves under convective forcing when the variables f, u_*, B are used as inputs.

Equation 8 in Table 1 is formulated to emulate the neural network trained on f, u_*, B as inputs. A caveat of not having h in the inputs is that, under the limit of pure convection, $\lambda \to \infty$, $v_0 \to \sqrt{B/\Omega}$. Dimensionally, v_0 is correct and numerically a viable alternative that mimics the use of neural networks in a production model by avoiding out-of-sample inputs. However, it does not match the well-known Deardorff scaling.

To agree with the known Deardorff scaling, the model for v_0^h as given by Equation 10 in Table 1 is better suited, as it includes h as one of the inputs. We used a simple equation with three parameters (c_{14}, c_{17}, c_{18}) to model v_0^h for surface cooling conditions. Only L_h was found to be sufficient to capture variations in v_0^h/u_* (see Figure 1h). As per Equation 10 in Table 1, for smaller values of L_h, that is, approximately between 0 and 2, $v_0^h/u_* = \left(L_h^3/c_{18}\right) + c_{14}$ that is, $v_0^h/u_* \approx L_h^3$. Near zero, that is, $L_h = 0$, shear dominates over buoyancy and $v_0^h/u_* = c_{14}$. For larger values of L_h, that is, $L_h \gg 1$, buoyancy dominates, and Equation 10 in Table 1 tends to $v_0^h/u_* = L_h/c_{17}$. Hence, $v_0^h \approx (Bh)^{1/3}$, allowing the velocity scale to agree with the known scaling under the limit of pure convection.

3. Impact of Improved Diffusivity in Global Runs: Mixed Layer Depths

We tested ePBL_EQ in a 1/4° global MOM6 configuration following Adcroft et al. (2019) that uses the JRA55-do atmospheric reanalysis forcing (Tsujino et al., 2018) for the duration from 1958 to 2017. Four simulations were run with: (a) the default OM4 ePBL, denoted as "ePBL"; (b) the ePBL with enhanced diffusivity using neural networks, denoted as ePBL_NN; (c) ePBL with enhanced diffusivity using the discovered equations with v_0 for velocity scale, denoted as ePBL_EQ (without h). EQ stands for equation discovery; and (d) ePBL_EQ (with h) uses v_0^h for velocity scale.

To demonstrate the improvements to OM4 relative to Sane et al. (2023), all runs are evaluated for mixed layer properties against ARGO data (Argo, 2024). We present a comparison for the summer MLD averaged over the years 2003–2017 in Figure 2. Summer MLD is defined as the minimum of the monthly averaged MLD at each grid point. We calculated MLDs using the potential anomaly criterion of 25 J/m^2 as defined in Reichl et al. (2022).

As in Sane et al. (2023), the summer MLD bias is reduced in ePBL_NN relative to ePBL, with the global mean RMSE decreasing from 4.38 to 3.83 m using the potential anomaly criterion (Figure 2) and from 7.22 to 6.07 m using the density threshold criterion (see Figure S6 in Supporting Information S1). We find a strong latitude dependence of RMSE between the model and the observations, as shown in Figure 2. There is significant bias reduction for latitudes ranging from about −40° to 40°. ePBL_EQ with and without h (Figures 2g and 2h) have a similar impact on summer MLD as ePBL_NN. ePBL_EQ shows slightly more error than ePBL_NN in terms of equatorial bias, but the RMS values indicate that ePBL_EQ is closer to ePBL_NN than to ePBL, as expected. Both ePBL_EQ with and without h perform well enough to be considered viable alternatives to ePBL_NN.

The winter MLD biases are similar for all cases and have not been worsened due to the new diffusivity or by considering h in ePBL_EQ (see Supporting Information S1). The winter biases could be due to missing effects of lateral processes in the model, hence unaffected by changing the OSBL vertical diffusivity (Treguier et al., 2023). The summer and winter MLD results are robust to using the density threshold criterion of 0.03 kg/m^3 (de Boyer

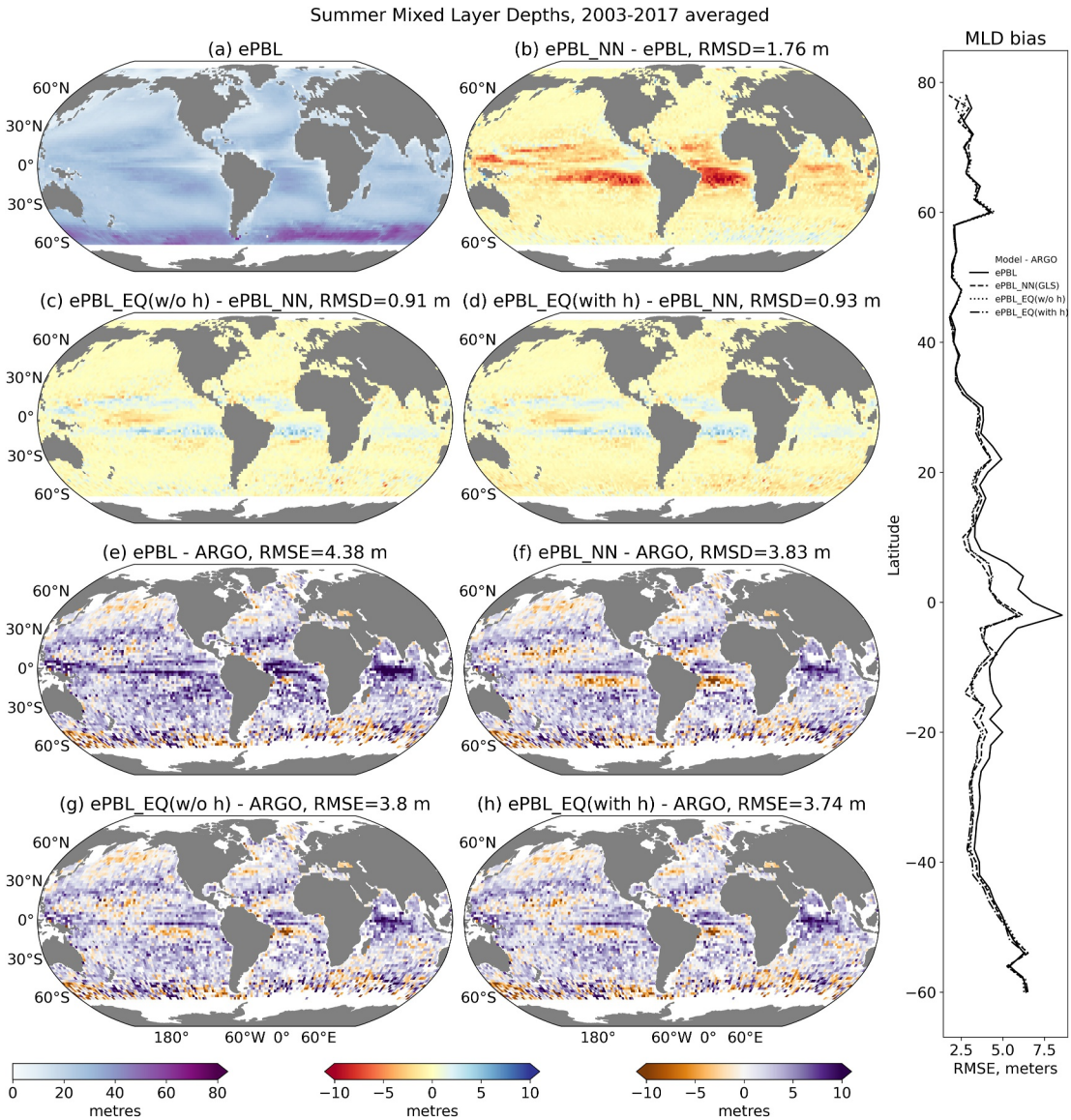

Figure 2. (a) Summer MLD in the control run ePBL found using the potential anomaly criterion (Reichl et al., 2022). Difference between model runs: (b) Difference between ePBL_NN and ePBL, (c) ePBL_EQ (without h) and ePBL_NN, (d) ePBL_EQ (with h) and ePBL_NN. Model–Observation comparisons: difference between (e) ePBL and ARGO, (f) ePBL_NN and ARGO, (g) ePBL_EQ (without h) and ARGO, and (h) ePBL_EQ (with h) and ARGO. Right image: Latitude-dependent RMSE for summer MLD for the four runs. ePBL_EQ (without h) uses a velocity scale that does not have h as input, while ePBL_EQ (with h) includes it. As compared to the control run that uses the baseline ePBL, the machine-learning-enhanced diffusivity reduces bias in the summer MLD. The model with the machine-learned equations shows similar response to neural networks and hence can replace the neural network. The response of OM4 with a velocity scale that uses h as input agrees closely with the velocity scale that does not use h as input.

Montégut et al., 2004) as shown in the Supporting Information S1. These results are consistent with our previous work (Sane et al., 2023).

4. Discussions

4.1. Dependence of Diffusivity on Forcing Parameters

The machine learned and least-squares-fitted equations reveal the dependence of the shape function and velocity scale on the following non-dimensional numbers $L_h = (|B|h)^{1/3}/u_*$, $E_h = h|f|/u_*$, and $\lambda = \sqrt{L_h^3/E_h}$ (see Table 1). The ePBL_EQ shape-function model sets the location of maximum diffusivity σ_m based on L_h and E_h.

For surface heating conditions, as $L_h \to -\infty$ and $E_h \to 0$, σ_m shifts toward the surface. As $L_h \to \infty$, that is, for surface cooling conditions, σ_m asymptotes to 0.55, implying that the maximum diffusivity is at the mid-point within the boundary layer. σ_m informs us about the intensity of mixing within the boundary layer. The dependence of $g(\sigma)$ on L_h and E_h allows a mechanistic pathway for setting the structure of diffusivity based on forcing parameters. In the baseline ePBL scheme, the effective shape function had its maximum fixed at 0.33, and this universal shape function did not capture the variation due to the surface forcing. The equations for the velocity scale show how it varies with respect to L_h and λ. Surface shear and rotation dominate v_0 for $\lambda \leq 1$, and for larger values, buoyancy is important. v_0^h is set by surface shear when $|L_h| < 1$, and buoyancy becomes relevant when $|L_h| > 1$. Under pure convective conditions, $v_0^h \approx (Bh)^{1/3}$, that is, it agrees with the well-known convective velocity scaling.

The equations for σ_m and the velocity scale are diagnostic and are constrained by the GLS SMC. They describe how a SMC sets diffusivity. The equations approximate a neural network and hence provide insight into the physics it captures.

4.2. Sensitivity of Including *h* as an Input in the Velocity Scale

Figure 1i shows that the equations of v_0, v_0^h have high skill compared to the ground truth (velocities inferred from neural networks). v_0 and v_0^h differ by definition, but they are highly correlated in the single column SMC training data set (see Figure 1j). Although agreement is acceptable in our training data set, v_0^h will disagree with v_0 for boundary layer depths deeper than those present in the training data under strong convective conditions. This is due to v_0^h agreeing with the known convective velocity scale. However, these differences were found to have minimal impact in global runs. This suggests that variability in v_0^h due to h has negligible impact on MLDs in the global runs.

4.3. Which Missing Physics Is Causing the Bias Reduction for Summer MLD in the Tropics?

The summer MLD improvements in ePBL_NN and ePBL_EQ (Figure 2) arise from a realistic treatment of diffusivity under surface heating conditions. The baseline ePBL scheme is over-diffusive when the surface buoyancy flux is stabilizing, producing excessive mixing and an overly deep summer mixed layer. ePBL sets the velocity scale solely as a function of u_* (Reichl & Li, 2019), while the velocity scale in ePBL_EQ includes surface buoyancy flux as an additional predictor (Equations 7 and 9 in Table 1) and thus captures the response of the GLS scheme to surface heating. The velocity scale in ePBL is insensitive to changes in B for surface stabilizing conditions, whereas in ePBL_NN and ePBL_EQ, the velocity scale decreases as B increases. This reduces κ, thereby reducing mixing, an effect noticeable in the tropics due to the presence of high surface heating values.

Consequently, ePBL_EQ yields a shallower summer mixed-layer depth consistent with known boundary layer physics: stabilizing heat flux increases stratification and lowers diffusivity, making ePBL_EQ κ physically more consistent with SMC than the control ePBL. We note a similar recent finding on biases associated with the energetics of ePBL mixing in the tropics (Reichl et al., 2024), which has complementary effects on tropical mixed layer processes to those examined here. These impacts will be tested together in future work.

4.4. Limitations

A limitation of our work is the inheritance of deficiencies and biases from GLS, our ground truth scheme. GLS employs a downgradient turbulent flux assumption, which is not valid under non-local flux conditions. Recent advances in incorporating non-local effects into SMC turbulent mixing schemes (Garanaik et al., 2024; Legay et al., 2025; Perrot & Lemarié, 2025) may address this issue. Additionally, diffusivity only accounts for shear- and buoyancy-driven turbulence, without incorporating the effects of waves that are relevant in climate simulations (Orenstein et al., 2022). The dominant wave effect, Langmuir turbulence, has been parameterized through m^*, which is the rate of change of kinetic energy to potential energy (see Reichl & Li, 2019 for more details), but the impact of Langmuir turbulence on diffusivity profiles requires further investigation and could contribute to remaining biases in Figure 2. The quadratic-cubic equation for $g(\sigma)$ assumes a cubic profile that does not align with the output of neural networks, which could affect the entrainment region at the base of the boundary layer.

4.5. Broader Applicability to Other First-Order Closures

The vertical diffusivity or mixing coefficients are applicable beyond the ePBL scheme. They can be applied to first-order closures that are constructed from a shape-function and an amplitude (velocity scale), such as KPP. The equations or neural networks only require four inputs: u_*, B, $|f|$, and h. These inputs are readily available in any scheme. B, f, and u_* are forcing parameters. h is estimated in a first-order closure schemes using a turbulence criterion (e.g., turbulent kinetic energy criteria (Gaspar et al., 1990) or bulk Richardson number (Large et al., 1994)).

Equation 3, Equations 5 and 6 in Table 1 give the shape function $g(\sigma)$. The set of Equations 7 and 8 or Equations 9 and 10 in Table 1 can be used to obtain the velocity scale. To agree with the known convective scaling the latter set of equations (Equations 9 and 10 in Table 1) that use h as an input are preferred.

5. Conclusions

We improve the vertical diffusivity used in a first-order closure employed in a CMIP-class ocean model in two stages. First, neural networks were trained against a GLS second-moment closure and shown to reproduce GLS's structure and velocity scale with high skill. Second, we replaced the black-box networks with compact analytical expressions obtained through a combination of GP and least squares fitting. GP performed well for relationships that were relatively simple, while least squares fitting proved necessary for the relatively complex data. In the OMIP runs, the discovered equations with O(10) parameters for OSBL diffusivity showed similar skill to neural networks that had O(1000) parameters. This is vital as global runs with neural networks with O(1000) are ≈5%–10% costlier than ePBL (Sane et al., 2023). ePBL and ePBL_EQ have the same cost in a global run. Equations in ePBL_EQ set the diffusivity in a σ grid similar to ePBL_NN and hence are insensitive to changing the vertical grid size (see Sane et al., 2023). The ML velocity scale modulates mixing under strong surface heating, rectifying the structural deficiency of the baseline ePBL scheme and yielding shallower summer mixed layers that reduce the tropical summer MLD bias.

The equations provide insight into the black-box nature of neural networks and offer a practical, interpretable alternative. Our methodology can be applied to enhance other first-order closures used in global simulations, such as the KPP scheme.

Acknowledgments
AS, AA, and LZ received M^{2}LInES research funding through Schmidt Sciences, LLC. LZ was also partially funded through NOAA NA19OAR4310364-T1-01. AA also received support from NA18OAR4320123, from the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce. Statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration or the U.S. Department of Commerce. We thank various members of the M^{2}LInES project for brainstorming about physics and machine learning. We used the Stellar computational resources provided by Princeton University and the National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL). We thank Dr. Wenda Zhang and Dr. Theresa Cordero for providing valuable feedback to improve the manuscript. We also thank Dr. Jun-Hong Liang and two anonymous reviewers for their constructive comments that helped us improve the manuscript. The authors thank the international Argo project and the various associated national programs for collecting and freely distributing the data set. AS also thanks Dr. Josiah Carberry, whose spectral presence was a constant, if slightly unnerving, source of inspiration.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The code and data required to reproduce the neural network and genetic programming training are available at https://doi.org/10.5281/zenodo.19616567 (Sane, 2026). This repository also contains the code used to generate all figures in the manuscript. In addition, we provide the source code of MOM6 used in this study, including the implementations of the neural-network-based and equation-based diffusivity schemes. The MOM6 source includes the `MOM_energetic_PBL.F90` module, with subroutines that perform inference using the trained neural networks. The module also contains subroutines for the equation-based diffusivity scheme. The weights and biases of the shape-function and velocity-scale neural networks are provided as NetCDF files.

The equation-based diffusivity has been incorporated into the current NOAA-GFDL MOM6 source code, available at https://github.com/NOAA-GFDL/MOM6.git.

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3167–3211. https://doi.org/10.1029/2019MS001726

Argo. (2024). Argo float data and metadata from Global Data Assembly Centre (ARGO GDAC) [Dataset]. *SEANOE*. Retrieved from https://www.seanoe.org/data/00311/42182/

Balaji, V., Couvreux, F., Deshayes, J., Gautrais, J., Hourdin, F., & Rio, C. (2022). Are general circulation models obsolete? *Proceedings of the National Academy of Sciences of the United States of America*, *119*(47), e2202075119. https://doi.org/10.1073/pnas.2202075119

Balwada, D., Perezhogin, P., Adcroft, A., & Zanna, L. (2025). *Design and implementation of a data-driven parameterization for mesoscale thickness fluxes*. Authorea Preprints.

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., et al. (2024). Climate-invariant machine learning. *Science Advances*, *10*(6), eadj7250. https://doi.org/10.1126/sciadv.adj7250

Bodner, A., Balwada, D., & Zanna, L. (2025). A data-driven approach for parameterizing ocean submesoscale buoyancy fluxes. *Journal of Advances in Modeling Earth Systems*, *17*(11), e2025MS004991. https://doi.org/10.1029/2025MS004991

Champion, K., Lusch, B., Kutz, J. N., & Brunton, S. L. (2019). Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22445–22451. https://doi.org/10.1073/pnas.1906995116

Chapman, W. E., & Berner, J. (2024). *A state-dependent model-error representation for online climate model bias correction*. Authorea Preprints.

Chen, J., Weisberg, R. H., Zheng, L., Hu, C., Zhang, Y., Nickerson, A. K., et al. (2025). Frontogenesis by material convergence in an estuary and its adjacent coastal ocean. *Limnology and Oceanography*, *71*(1), e70276. https://doi.org/10.1002/lno.70276

Christensen, H., & Zanna, L. (2022). *Parametrization in weather and climate models*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228620.013.826

Cranmer, M., Sanchez Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., & Ho, S. (2020). Discovering symbolic models from deep learning with inductive biases. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 17429–17442). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2020/file/c9f2f917078bd2db12f23c3b413d9cba-Paper.pdf

Deardorff, J. W. (1970). Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection. *Journal of the Atmospheric Sciences*, *27*(8), 1211–1213. https://doi.org/10.1175/1520-0469(1970)027<1211:cvatsf>2.0.co;2

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*(C12), C12003. https://doi.org/10.1029/2004JC002378

Garanaik, A., Pereira, F. S., Smith, K., Robey, R., Li, Q., Pearson, B., & Van Roekel, L. (2024). A new hybrid mass-flux/high-order turbulence closure for ocean vertical mixing. *Journal of Advances in Modeling Earth Systems*, *16*(1), e2023MS003846. https://doi.org/10.1029/2023ms003846

Gaspar, P., Grégoris, Y., & Lefevre, J.-M. (1990). A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station papa and long-term upper ocean study site. *Journal of Geophysical Research*, *95*(C9), 16179–16193. https://doi.org/10.1029/jc095ic09p16179

Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., & Zanna, L. (2023). Deep learning of systematic sea ice model errors from data assimilation increments. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003757. https://doi.org/10.1029/2023ms003757

Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., & Zanna, L. (2024). Machine learning for online sea ice bias correction within global ice-ocean simulations. *Geophysical Research Letters*, *51*(3), e2023GL106776. https://doi.org/10.1029/2023gl106776

Iyer, S., Hughes, K. G., & Moum, J. N. (2025). Predicting ocean turbulence across orders of magnitude using neural networks trained on multiyear observations. *Artificial Intelligence for the Earth Systems*, *4*(3), 240093. https://doi.org/10.1175/aies-d-24-0093.1

Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, *4*(2), 87–112. https://doi.org/10.1007/bf00175355

Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, *32*(4), 363–403. https://doi.org/10.1029/94rg01872

Large, W. G., Patton, E. G., DuVivier, A. K., Sullivan, P. P., & Romero, L. (2019). Similarity theory in the surface layer of large-eddy simulations of the wind-wave-and buoyancy-forced southern ocean. *Journal of Physical Oceanography*, *49*(8), 2165–2187. https://doi.org/10.1175/jpo-d-18-0066.1

Legay, A., Deremble, B., & Burchard, H. (2025). Derivation and implementation of a non-local term to improve the oceanic convection representation within the k–ε parameterization. *Journal of Advances in Modeling Earth Systems*, *17*(1), e2024MS004243. https://doi.org/10.1029/2024ms004243

Li, Q., Reichl, B. G., Fox-Kemper, B., Adcroft, A. J., Belcher, S. E., Danabasoglu, G., et al. (2019). Comparing ocean surface boundary vertical mixing schemes including Langmuir turbulence. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3545–3592. https://doi.org/10.1029/2019MS001810

Liang, J.-H., Yuan, J., Wan, X., Liu, J., Liu, B., Jang, H., & Tyagi, M. (2022). Exploring the use of machine learning to parameterize vertical mixing in the ocean surface boundary layer. *Ocean Modelling*, *176*, 102059. https://doi.org/10.1016/j.ocemod.2022.102059

Mashayek, A., Reynard, N., Zhai, F., Srinivasan, K., Jelley, A., Naveira Garabato, A., & Caulfield, C.-C. P. (2022). Deep ocean learning of small scale turbulence. *Geophysical Research Letters*, *49*(15), e2022GL098039. https://doi.org/10.1029/2022gl098039

Orenstein, P., Fox-Kemper, B., Johnson, L., Li, Q., & Sane, A. (2022). Evaluating coupled climate model parameterizations via skill at reproducing the monsoon intraseasonal oscillation. *Journal of Climate*, *35*(6), 1873–1884. https://doi.org/10.1175/JCLI-D-21-0337.1

Partee, S., Ellis, M., Rigazzi, A., Shao, A. E., Bachman, S., Marques, G., & Robbins, B. (2022). Using machine learning at scale in numerical simulations with SMARTSIM: An application to ocean climate modeling. *Journal of Computational Science*, *62*, 101707. https://doi.org/10.1016/j.jocs.2022.101707

Perezhogin, P., Adcroft, A., & Zanna, L. (2025). Generalizable neural-network parameterization of mesoscale eddies in idealized and global ocean models. *Geophysical Research Letters*, *52*, e2025GL117046. https://doi.org/10.1029/2025GL117046

Perezhogin, P., Zhang, C., Adcroft, A., Fernandez-Granda, C., & Zanna, L. (2024). A stable implementation of a data-driven scale-aware mesoscale parameterization. *Journal of Advances in Modeling Earth Systems*, *16*(10), e2023MS004104. https://doi.org/10.1029/2023ms004104

Perrot, M., & Lemarié, F. (2025). Energetically consistent eddy-diffusivity mass-flux convective schemes: 2. Implementation and evaluation in an oceanic context. *Journal of Advances in Modeling Earth Systems*, *17*(7), e2024MS004616. https://doi.org/10.1029/2024ms004616

Reichl, B. G., Adcroft, A., Griffies, S. M., & Hallberg, R. (2022). A potential energy analysis of ocean surface mixed layers. *Journal of Geophysical Research: Oceans*, *127*(7), e2021JC018140. https://doi.org/10.1029/2021JC018140

Reichl, B. G., & Hallberg, R. (2018). A simplified energetics based planetary boundary layer (ePBL) approach for ocean climate simulations. *Ocean Modelling*, *132*, 112–129. https://doi.org/10.1016/j.ocemod.2018.10.004

Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential energy conversion rate of vertical mixing due to Langmuir turbulence. *Journal of Physical Oceanography*, *49*(11), 2935–2959. https://doi.org/10.1175/JPO-D-18-0258.1

Reichl, B. G., Wittenberg, A. T., Griffies, S. M., & Adcroft, A. (2024). Improving equatorial upper ocean vertical mixing in the NOAA/GFDL OM4 model. *Earth and Space Science*, *11*(10), e2023EA003485. https://doi.org/10.1029/2023EA003485

Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Benchmarking of machine learning ocean subgrid parameterizations in an idealized model. *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003258. https://doi.org/10.1029/2022ms003258

Sane, A. (2026). Data and code for the manuscript "machine learned equations for vertical mixing coefficients in the ocean surface boundary layer" [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.19616567

Sane, A., Reichl, B. G., Adcroft, A., & Zanna, L. (2023). Parameterizing vertical mixing coefficients in the ocean surface boundary layer using neural networks. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003890. https://doi.org/10.1029/2023ms003890

Schneider, T., Leung, L. R., & Wills, R. C. J. (2024). Opinion: Optimizing climate models with process knowledge, resolution, and artificial intelligence. *Atmospheric Chemistry and Physics*, *24*(12), 7041–7062. https://doi.org/10.5194/acp-24-7041-2024

Treguier, A. M., de Boyer Montégut, C., Bozec, A., Chassignet, E. P., Fox-Kemper, B., McC. Hogg, A., et al. (2023). The mixed-layer depth in the ocean model intercomparison project (OMIP): Impact of resolving mesoscale eddies. *Geoscientific Model Development*, *16*(13), 3849–3872. https://doi.org/10.5194/gmd-16-3849-2023

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., et al. (2018). JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-DO) [Dataset]. *Ocean Modelling*, *130*, 79–139. https://doi.org/10.1016/j.ocemod.2018.07.002

Umlauf, L., & Burchard, H. (2003). A generic length-scale equation for geophysical turbulence models. *Journal of Marine Research*, *61*(2), 235–265. https://doi.org/10.1357/002224003322005087

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., et al. (2021). Correcting weather and climate models by machine learning nudged historical simulations. *Geophysical Research Letters*, *48*(15), e2021GL092555. https://doi.org/10.1029/2021gl092555

Whitt, D., Cherian, D., Holmes, R., Bachman, S., Lien, R.-C., Large, W., & Moum, J. (2022). Simulation and scaling of the turbulent vertical heat transport and deep-cycle turbulence across the equatorial Pacific cold tongue. *Journal of Physical Oceanography*, *52*(5), 981–1014. https://doi.org/10.1175/jpo-d-21-0153.1

Yuan, J., Liang, J.-H., Chassignet, E. P., Zavala-Romero, O., Wan, X., & Cronin, M. F. (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*(9), e2024MS004405. https://doi.org/10.1029/2024ms004405

Yuval, J., & O'Gorman, P. A. (2023). Neural-network parameterization of subgrid momentum transport in the atmosphere. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2023MS003606. https://doi.org/10.1029/2023ms003606

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376. https://doi.org/10.1029/2020gl088376

Zhang, C., Perezhogin, P., Adcroft, A., & Zanna, L. (2025). Addressing out-of-sample issues in multi-layer convolutional neural-network parameterization of mesoscale eddies applied near coastlines. *Journal of Advances in Modeling Earth Systems*, *17*, e2024MS004819. https://doi.org/10.1029/2024MS004819

Zhou, S., Dong, J., Xu, F., Jing, Z., & Dong, C. (2024). A neural network-based submesoscale vertical heat flux parameterization and its implementation in regional ocean modeling system (ROMS). *arXiv preprint arXiv:2403.05028*.