

# 1 Formulation and calibration of CATKE, a one-equation 2 parameterization for microscale ocean mixing

3 Gregory LeClaire Wagner<sup>1</sup>, Adeline Hillier<sup>1</sup>, Navid C. Constantinou<sup>2,3</sup>,  
4 Simone Silvestri<sup>1</sup>, Andre Souza<sup>1</sup>, Keaton J. Burns<sup>1</sup>, Chris Hill<sup>1</sup>,  
5 Jean-Michel Campin<sup>1</sup>, John Marshall<sup>1</sup>, and Raffaele Ferrari<sup>1</sup>

6 <sup>1</sup>Massachusetts Institute of Technology, Cambridge, MA, USA

7 <sup>2</sup>University of Melbourne, Parkville, VIC, Australia

8 <sup>3</sup>ARC Center of Excellence for the Weather of the 21st Century, Parkville, VIC, Australia

## 9 **Key Points:**

- 10 We describe a new parameterization called CATKE with a convective adjustment  
(CA) component and prognostic turbulent kinetic energy (TKE).
- 11 We use Ensemble Kalman Inversion to calibrate CATKE's free parameters against 21  
idealized large eddy simulations (LES).
- 12 • We validate CATKE by interpreting its free parameters and comparing to additional  
idealized and realistic LES.

16    **Abstract**

17    We describe CATKE, a parameterization for fluxes associated with small-scale or “microscale”  
 18    ocean turbulent mixing on scales between 1 and 100 meters. CATKE uses a downgradient  
 19    formulation that depends on a prognostic turbulent kinetic energy (TKE) variable and  
 20    a diagnostic mixing length scale that includes a dynamic convective adjustment (CA)  
 21    component. With its dynamic convective mixing length, CATKE predicts not just the depth  
 22    spanned by convective plumes but also the characteristic convective mixing timescale, an  
 23    important aspect of turbulent convection not captured by simpler static convective adjustment  
 24    schemes. As a result, CATKE can describe the competition between convection and other  
 25    processes such as shear-driven mixing and baroclinic restratification. To calibrate CATKE,  
 26    we use Ensemble Kalman Inversion to minimize the error between 21 large eddy simulations  
 27    (LES) and predictions of the LES data by CATKE-parameterized single column simulations  
 28    at three different vertical resolutions. We find that CATKE makes accurate predictions  
 29    of both idealized and realistic LES compared to microscale turbulence parameterizations  
 30    commonly used in climate models.

31    **Plain Language Summary**

32    Turbulence is everywhere in the Earth’s ocean, from ephemeral swirls no bigger than  
 33    a fingertip to gigantic eddies larger than Iceland. Ocean models used in climate studies  
 34    simulate currents by dividing the ocean into grid cells between 10 and 100 kilometers wide.  
 35    As a result, ocean models do a decent job simulating eddies that are significantly larger than  
 36    a single grid cell. But models do far worse at incorporating the effects of eddies that are  
 37    person- to building-sized, which are smaller than a grid cell and therefore must be represented  
 38    more approximately. This is a problem because these small yet mighty eddies mix heat  
 39    and carbon deep into the ocean, and thus help keep the atmosphere from getting too hot,  
 40    and too rich in CO<sub>2</sub>. In this paper, we propose a new model component called “CATKE”  
 41    (pronounced *kăt-kee*) that approximately incorporates the effect of small eddies in global  
 42    ocean models. CATKE stands for “Convective Adjustment and Turbulent Kinetic Energy”,  
 43    and keeps track of the *energy* of small-scale turbulence — a measure of how vigorous it is,  
 44    and thus how much it mixes the ocean — to predict ocean mixing rates.

45    **1 Introduction**

46    Vertical mixing by “microscale” ocean turbulence, with scales between 1 and 100 meters,  
 47    is an important process affecting, for example, ocean uptake of atmospheric heat and  
 48    carbon (Price et al., 1986; Large et al., 1994; Omand et al., 2015), the structure of the ocean  
 49    interior (Luyten et al., 1983; Williams, 1991), and ocean circulation on decadal to millennial  
 50    time-scales (Wunsch & Ferrari, 2004; Melet et al., 2022). In large-scale ocean models — from  
 51    regional models covering tens of kilometers to global ocean models — microscale turbulent  
 52    vertical fluxes are approximately modeled by parameterizations. Imperfect predictions by  
 53    turbulence parameterizations contribute to biases in tropical sea surface temperature (G. Li  
 54    & Xie, 2014), Southern Ocean boundary layer depth (Sallée et al., 2013; DuVivier et al.,  
 55    2018), and water mass transformation rates (Groeskamp et al., 2019). These errors degrade  
 56    the accuracy of climate projections that depend on accurate air-sea fluxes (sensitive to sea  
 57    surface temperature, Large et al., 1994) and the effective heat capacity of the upper ocean  
 58    (which scales with the boundary layer depth, Gregory, 2000; Held et al., 2010).

59    This paper documents the development, calibration, and preliminary validation of a  
 60    new parameterization for vertical mixing by ocean microscale turbulence. Our goal is to use  
 61    the new parameterization in a GPU-based climate model that is automatically calibrated to  
 62    observations, reports quantified uncertainties, and has an ocean component with  $O(10\text{ km})$   
 63    or finer resolution that resolves ocean mesoscale turbulence. The dynamical core of the  
 64    GPU-based ocean component is described by Silvestri, Wagner, Constantinou, et al. (2024).  
 65    In service of this ultimate goal, the work documented in this paper prioritizes not just

66 accurate predictions, but also efficiency on GPUs in high-resolution configurations. We also  
 67 invest in automated calibration that constrains all of the parameterization’s free parameters  
 68 to 21 large eddy simulations (LESs) simultaneously, accounting for the peculiarities of our  
 69 specific numerical implementation of the parameterization in a single column model. The  
 70 21 LES we use to calibrate and the additional 14 LES we use to validate the parameterization  
 71 are described in section 2. Uncertainty quantification, an important step for a future  
 72 re-calibration that leverages global-scale observations, is left for future work.

73 Our new parameterization, which we call “CATKE”, uses a downgradient formulation  
 74 that estimates eddy diffusivities in terms of a prognostic turbulent kinetic energy (TKE)  
 75 variable and a diagnostic mixing length with a novel dynamic convective adjustment (CA)  
 76 component. CATKE is a “one-equation” model (because it includes an additional equation  
 77 for TKE) that bears resemblance to a family of battle-tested parameterizations long used  
 78 in European climate models (Gaspar et al., 1990; Blanke & Delecluse, 1993; Kuhlbrodt et  
 79 al., 2018; Madec et al., 2017; Gutjahr et al., 2021; Jungclaus et al., 2022). One-equation  
 80 downgradient parameterizations are appropriate for high-resolution ocean modeling and  
 81 amenable to GPU performance optimization due to their spatially-local formulation. In  
 82 contrast, the main feature of “ $K$ -profile” schemes used in many global ocean models —  
 83 accommodating hours-long time steps (Reichl & Hallberg, 2018) by implicitly time-averaging  
 84 mixing physics — does not benefit and may even degrade high-resolution simulations that  
 85 resolve relatively fast mesoscale and submesoscale processes. Moreover,  $K$ -profile schemes  
 86 achieve time-step flexibility by solving nonlinear algebraic equations for boundary layer  
 87 depth (Large et al., 1994; Reichl & Hallberg, 2018; Reichl & Li, 2019), which may require  
 88 significant optimization to achieve good performance on GPU-like systems (see by Zhang  
 89 et al., 2020). As for two-equation “ $k-\epsilon$ ”-type models (Mellor & Yamada, 1982; Kantha &  
 90 Clayson, 1994; Canuto et al., 2001; Umlauf & Burchard, 2003; Harcourt, 2015), or equations  
 91 with even more than two prognostic variables (Garanaik et al., 2024; Legay et al., 2024),  
 92 CATKE is less expensive merely by having one fewer prognostic variable. CATKE therefore  
 93 serves as a high-performance “baseline” whose accuracy must be surpassed to justify the use  
 94 of more expensive parameterizations.

95 The downsides of downgradient parameterizations include unavoidable biases when non-  
 96 local, non-downgradient fluxes dominate, such as during free convection (Large et al., 1994;  
 97 Legay et al., 2024). We therefore devote special attention to free convection during CATKE’s  
 98 formulation, which is described in section 3, to minimize this downgradient bias and assess its  
 99 importance. Section 3.1.5 describes CATKE’s diagnostic convective length scale and primary  
 100 novelty, which uses dimensional analysis (Deardorff, 1970) to estimate a dynamically evolving  
 101 convective diffusivity in terms of the local TKE. This improves upon constant “convective  
 102 adjustment” diffusivities typically used with one-equation parameterizations in ocean climate  
 103 models (typically  $0.1 \text{ m}^2 \text{ s}^{-1}$ ; Madec et al., 2017; Gutjahr et al., 2021; Jungclaus et al., 2022),  
 104 which cannot describe how the convective mixing rate varies with both boundary layer depth  
 105 and the intensity of the destabilizing surface buoyancy flux. As a result, CATKE might be  
 106 able to represent scenarios where mixing competes with other dynamics such as submesoscale  
 107 restratification. We also implement different mixing lengths for momentum, tracer, TKE,  
 108 and the TKE dissipation rate in shear-driven turbulence that all vary as a function of the  
 109 local gradient Richardson number. This contrasts with typical approaches that estimate  
 110 the TKE diffusivity as a constant multiple of the eddy viscosity (Blanke & Delecluse, 1993;  
 111 Madec et al., 2017; Umlauf & Burchard, 2003), or which allow only the tracer mixing length  
 112 to vary with Richardson number (Blanke & Delecluse, 1993; Madec et al., 2017).

113 CATKE’s formulation could not be realized without an effective method for constraining  
 114 CATKE’s free parameters against observational or LES data. Section 4 describes how we  
 115 use automatic, *a posteriori* calibration (Duraisamy, 2021; Frezat et al., 2022) to estimate  
 116 CATKE’s free parameters by minimizing the error between 21 variously-forced LES and  
 117 the predictions of the LES data made by forward CATKE-parameterized single column  
 118 simulations. Because *a posteriori* calibration computes errors based on simulated time-series,

it can incorporate numerical errors that accumulate during time stepping and can leverage even indirect observational data if it can be computed from model output. For example, we leverage *a posteriori* calibration to specifically minimize CATKE’s dependence on vertical resolution. We solve the calibration problem using Ensemble Kalman Inversion (EKI; see Iglesias et al., 2013), which does not require gradients of the error with respect to free parameters. We argue that automatic, EKI-based, *a posteriori* calibration is crucial not only for CATKE’s development, but for any parameterization development effort that seeks the simplest possible model that can adequately simulate available data. Without automatic calibration, we cannot generally tell whether bias has to do with structural error — which can only be addressed by formulation changes, possibly increasing model complexity — or because of poorly chosen parameters, which does not justify increasing model complexity.

We validate CATKE in various ways in section 5. We first diagnose quantities with known physical interpretations such as CATKE’s steady-state Richardson number and “similarity layer constant” (analogous to the von Kármán constant) in terms of CATKE’s calibrated free parameters, and assess their consistency with values reported in the literature. Second, we compare CATKE’s predictions versus idealized LES, both including those used in calibration and additional LES that are more strongly and more weakly forced than the calibration cases. In this way we test whether CATKE can reproduce the training data as well as CATKE’s capacity for extrapolation. Third, we compare CATKE predictions to LES of a long 34-day deep cycle turbulence case, which is forced by realistic winds, heat fluxes, salinity fluxes, solar insolation, and lateral flux divergences derived from a regional ocean model (Whitt et al., 2022). This case illustrates CATKE’s ability to extrapolate to cases with time-dependent forcing. Fourth, we evaluate the sensitivity of CATKE’s predictions to vertical resolution and time-step size. After finding that CATKE can be sensitive to time steps longer than 1 minute if the forcing is very strong and the vertical resolution is 1 meter or finer, we describe a split-explicit substepping scheme for TKE that nearly eliminates time step sensitivity while preserving the ability to step forward momentum and tracers with a relatively long time step.

We also compare CATKE to the  $K$ -profile parameterization (KPP; Large et al., 1994) and the second-moment closure of Langmuir turbulence (Langmuir Turbulence Second Moment Closure, or “SMC-LT”; Harcourt, 2015), which are implemented in the General Ocean Turbulence Model (GOTM; see Umlauf & Burchard, 2005; Q. Li et al., 2019). CATKE outperforms both in almost all cases — though the results must be taken with a grain of salt, because both KPP and SMC-LT have been calibrated to different data. Despite this caveat, the comparison contributes context to CATKE’s small but finite biases versus constant forcing LES.

In section 6, we conclude with a discussion about future efforts to calibrate CATKE against more comprehensive data sets, and model development efforts to capture physics not considered in this work, such as the effect of surface wave fields that vary independently from winds and the modulation of turbulence by lateral density fronts. The most important piece of future work is the construction of a global calibration context to further refine CATKE’s free parameters using satellite and in-situ ocean observations.

## 161 2 Large eddy simulations of turbulent mixing beneath surface waves

We begin by defining the parameterization problem that drives the cyclical process of formulating, calibrating, and validating CATKE. In this paper, the parameterization problem is posed by comparing high-fidelity and three-dimensional large eddy simulations (LES) of turbulent mixing with one-dimensional parameterized models for the horizontally-averaged dynamics of the LES. Our LES integrate the rotating, wave-averaged Boussinesq equations simplified for a steady surface wave field (Craik & Leibovich, 1976; Huang, 1979; Suzuki &

168 Fox-Kemper, 2016),

$$\partial_t \mathbf{U}^L + (\mathbf{U}^L \cdot \nabla) \mathbf{U}^L + (f \hat{\mathbf{z}} - \nabla \times \mathbf{U}^S) \times \mathbf{U}^L + \nabla P = B \hat{\mathbf{z}} + \partial_t \mathbf{U}^S + \mathbf{F}_u, \quad (1)$$

$$\nabla \cdot \mathbf{U}^L = 0, \quad (2)$$

$$\partial_t C + (\mathbf{U}^L \cdot \nabla) C = -\nabla \cdot \mathbf{J}_c + F_c, \quad (3)$$

172 where  $\mathbf{U}^L = (U^L, V^L, W^L)$  is the Lagrangian-mean velocity,  $\mathbf{U}^S$  is the Stokes drift associated  
 173 with surface waves (which are always steady and oriented in the  $\hat{\mathbf{x}}$ -direction in this paper),  
 174  $P$  is Eulerian-mean kinematic pressure,  $B$  is Eulerian-mean buoyancy,  $f$  is the Coriolis  
 175 parameter,  $\mathbf{F}_u$  is a momentum forcing term representing surface wind stress,  $C$  is any  
 176 tracer such as temperature or salinity, and  $F_c$  is forcing term for  $C$  representing boundary  
 177 conditions, solar insolation, and other imposed body forcing. The Lagrangian-mean velocity  
 178  $\mathbf{U}^L$  is defined as the sum of the Eulerian-mean velocity and Stokes drift, and setting  $\mathbf{U}^S = 0$   
 179 reduces equation (1) to the ordinary Navier–Stokes equations. Note that we have neglected  
 180 molecular diffusion from (1) and (3), as well as diffusion by a hypothetical LES closure, to  
 181 simplify the ensuing discussion. In this work we use buoyancy  $B$  itself as a tracer, which is  
 182 tantamount to using a linear equation of state with a single constituent.

183 We conduct 35 LES of (1)–(3) forced by constant, horizontally-uniform fluxes of mo-  
 184 mentum and buoyancy in a  $512 \text{ m} \times 512 \text{ m} \times 256 \text{ m}$  horizontally-periodic domain with  $O(1 \text{ m})$   
 185 resolution using Oceananigans (Ramadhan et al., 2020). Grid-scale dissipation of kinetic  
 186 energy and tracer variance is implicitly provided by a Weighted, Essentially Non-Oscillatory  
 187 (WENO, Shu, 2020) advection scheme. The advantages of this approach are described by  
 188 Pressel et al. (2017). All 35 LES are initialized with the same piecewise-constant density  
 189 stratification given in equation A1, which has a weakly-stratified near-surface layer, a more  
 190 strongly stratified middle layer, and a weakly-stratified lower layer. The surface momentum  
 191 flux or “wind stress”  $\tau_x$  is defined via  $\mathbf{F}_u$  in (1) as

$$192 \mathbf{F}_u = -\partial_z [\tau_x \mathcal{H}(z)] \hat{\mathbf{x}}, \quad \text{where } \mathcal{H}(z) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 \end{cases} \quad (4)$$

193 is a Heaviside function. Negative stress  $\tau_x < 0$  forces a current in the  $+x$ -direction. Two  
 194 types of buoyancy fluxes are used: a destabilizing surface flux  $J_b > 0$  representing cooling or  
 195 heat loss, which is defined via  $F_b$  in equation (3) via

$$196 F_b = -\partial_z [J_b \mathcal{H}(z)]. \quad (5)$$

197 We also include 5 LES forced by both wind stress and stabilizing buoyancy forcing that  
 198 represents heating by solar insolation. In these “sunny” cases, the flux divergence of buoyancy  
 199  $F_b$  is given by

$$200 F_b = -\partial_z I, \quad \text{where } I(z) = J_b \left[ \epsilon_1 e^{z/\lambda_1} + (1 - \epsilon_1) e^{z/\lambda_2} \right]. \quad (6)$$

201 In (6),  $I(z)$  is the buoyancy flux profile associated with penetrating solar insolation,  $J_b < 0$   
 202 is the surface solar insolation,  $\epsilon_1$  is the fraction of penetrating radiation absorbed over the  
 203 vertical scale  $\lambda_1$ , and  $(1 - \epsilon_1)$  is the remaining fraction absorbed over  $\lambda_2$ . All simulations  
 204 use  $\epsilon_1 = 0.6$ ,  $\lambda_1 = 1 \text{ m}$ , and  $\lambda_2 = 16 \text{ m}$  (see for example the solar insolation used by Whitt  
 205 et al., 2022).

206 The LES are organized by duration into 6-, 12-, 24-, 48-, and 72-hour “suites”. Because  
 207 all the LES are initialized identically and run until the boundary layer is roughly half the  
 208 depth of the domain, duration indicates forcing strength: the 6-hour-suite are the most  
 209 strongly forced and the 72-hour suite simulations are the most weakly forced. So that we can  
 210 validate CATKE’s ability to extrapolate outside the training dataset, only intermediately-  
 211 forced 12-, 24-, and 48-hour suites are used for calibration. The 35 LES are divided into 5  
 212 “suites” with 7 cases each, according to their duration and the intensity of the surface fluxes:

213 the 6-hour suite exhibits extreme forcing, while the 72-hour suite exhibits relatively weak  
214 forcing. Each suite consists of 7 physical scenarios that represent different forcing regimes:

- 215 • “free convection”, which has pure destabilizing buoyancy forcing and no winds,
- 216 • “weak wind strong cooling”,
- 217 • “medium wind medium cooling”,
- 218 • “strong wind weak cooling”,
- 219 • “strong wind”, with no buoyancy forcing,
- 220 • “strong wind no rotation” with no buoyancy forcing and  $f = 0$ .
- 221 • “strong wind and sunny” with penetrative heating, wind forcing, and  $f = 0$ .

222 The “strong wind no rotation” and “strong wind and sunny” are non-rotating with  $f = 0$ ,  
223 and the rest are rotating with Coriolis parameter  $f = 10^{-4} \text{ s}^{-1}$ . The range of buoyancy  
224 fluxes roughly corresponds to cooling between 156–2000  $\text{W m}^{-2}$  or heating by penetrating  
225 solar insolation between 104–1250  $\text{W m}^{-2}$ , and the momentum fluxes correspond to 10-meter  
226 atmospheric winds of approximately 9–25  $\text{m s}^{-1}$  and oriented in the  $\hat{\mathbf{x}}$ -direction. The fluxes  
227 associated with each case are summarized in tables 1 and 2.

228 In any LES with wind forcing, we also include the effect of wind-driven surface waves  
229 through an estimate of  $\partial_z \mathbf{U}^S = \partial_z U^S \hat{\mathbf{x}}$  in (1) for equilibrium waves (Lenain & Pizzo, 2020).  
230 The equilibrium wave model depends on the peak wavenumber of the surface wave field,  
231 which is chosen so that the Langmuir number  $La$  is

$$232 La \stackrel{\text{def}}{=} \sqrt{\frac{u_*}{U^S(z=0)}} \approx 0.3, \quad (7)$$

233 close to the peak of its global distribution (Belcher et al., 2012). In (7),  $u_*$  is the friction  
234 velocity computed from the surface wind stress (here  $u_* = \sqrt{|\boldsymbol{\tau}_x|}$ , where  $\boldsymbol{\tau} = \tau_x \hat{\mathbf{x}}$  is the wind  
235 stress). All LES are initialized from rest with  $\mathbf{U}^L = 0$ . The LES also include a forced passive  
236 tracer, providing additional information about the time scales of mixing in the interior of  
237 the boundary layer. The initial density stratification, numerical methods, Stokes drift model,  
238 effects of including Stokes drift, and the sensitivity of the LES to resolution are described  
239 in Appendix A. Out of the 35 LES cases, 21 are used for calibration, while another 14 are  
240 reserved for validation. Figure 1 visualizes vertical velocity in 9 of the 35 cases.

## 241 2.1 The single column context

242 We would like to develop a model that can predict the horizontally-averaged momentum  
243 and buoyancy simulated by the LES. We therefore decompose all three-dimensional variables  $\Psi$   
244 in (1)–(3) into a horizontally-averaged component  $\psi \stackrel{\text{def}}{=} \bar{\Psi}$  and a fluctuation  $\psi'$  such that,

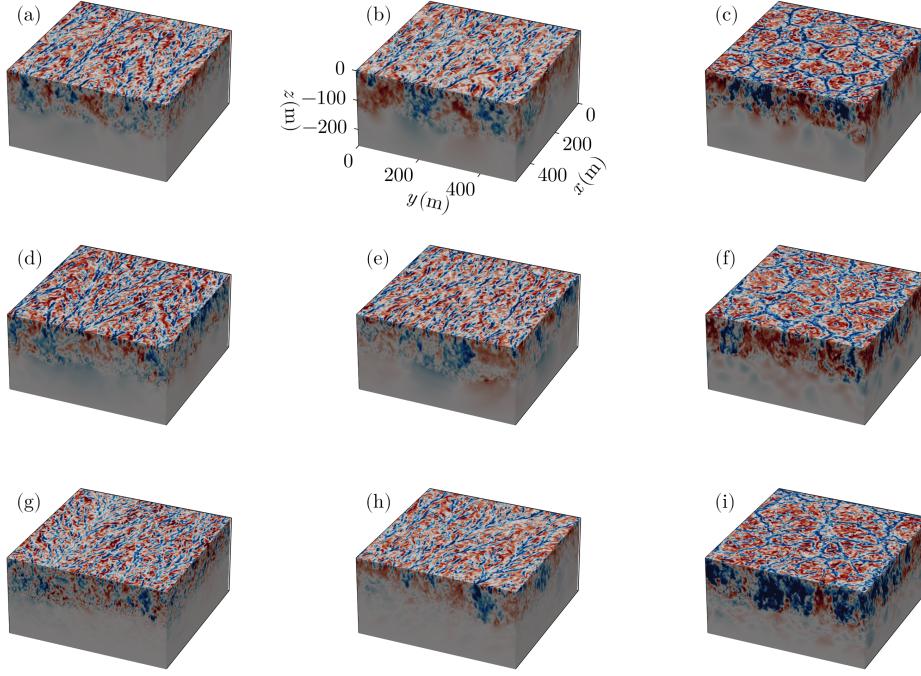
$$245 \Psi(x, y, z, t) = \underbrace{\bar{\Psi}(z, t)}_{\stackrel{\text{def}}{=} \psi(z, t)} + \psi'(x, y, z, t), \quad (8)$$

246 where the overline  $(\bar{})$  denotes a horizontal average, and  $\Psi \in (U^L, V^L, W^L, C)$  includes the  
247 velocity components  $U^L, V^L, W^L$ , and tracer concentrations  $C$ . Note that the horizontal  
248 average of (2) and the horizontal homogeneity of our LES implies that  $w^L = 0$  and  $W^L = w'$   
249 and thus the vertical momentum equation reduces to a statement of wave-modified hydrostatic  
250 balance. Figure 2 shows horizontally-averaged buoyancy, velocity, and kinetic energy profiles  
251 alongside a three-dimensional visualization of the buoyancy perturbation  $b'$  for the 12-hour  
252 strong wind, weak cooling case.

253 Next, we derive a set of equations that governs the horizontally-averaged zonal mo-  
254 mentum  $u(z, t)$ , meridional momentum  $v(z, t)$ , and any tracer  $c(z, t)$  by taking a horizontal

Suite	Case	$J_b$ ( $\text{m}^2 \text{s}^{-3}$ )	$ \tau_x $ ( $\text{m}^2 \text{s}^{-2}$ )	$Q$ ( $\frac{\text{W}}{\text{m}^2}$ )	$u_{10}$ ( $\frac{\text{m}}{\text{s}}$ )
12 hour	free convection	$+4.8 \times 10^{-7}$	0	+1000	0
12 hour	weak wind strong cooling	$+4.0 \times 10^{-7}$	$4.0 \times 10^{-4}$	+833	15
12 hour	mid wind mid cooling	$+3.2 \times 10^{-7}$	$6.0 \times 10^{-4}$	+667	17
12 hour	strong wind weak cooling	$+2.0 \times 10^{-7}$	$8.0 \times 10^{-4}$	+417	20
12 hour	strong wind	0	$9.0 \times 10^{-4}$	0	21
12 hour	strong wind no rotation	0	$6.0 \times 10^{-4}$	0	17
12 hour	strong wind and sunny	$-5.0 \times 10^{-7}$	$9.0 \times 10^{-4}$	-1042	21
24 hour	free convection	$+2.4 \times 10^{-7}$	0	+500	0
24 hour	weak wind strong cooling	$+2.0 \times 10^{-7}$	$3.0 \times 10^{-4}$	+417	13
24 hour	mid wind mid cooling	$+1.6 \times 10^{-7}$	$4.5 \times 10^{-4}$	+333	16
24 hour	strong wind weak cooling	$+1.0 \times 10^{-7}$	$5.9 \times 10^{-4}$	+208	17
24 hour	strong wind	0	$6.8 \times 10^{-4}$	0	18
24 hour	strong wind no rotation	0	$3.0 \times 10^{-4}$	0	13
24 hour	strong wind and sunny	$-3.0 \times 10^{-7}$	$4.5 \times 10^{-4}$	-625	16
48 hour	free convection	$+1.2 \times 10^{-7}$	0	+250	0
48 hour	weak wind strong cooling	$+1.0 \times 10^{-7}$	$2.0 \times 10^{-4}$	+208	11
48 hour	mid wind mid cooling	$+8.0 \times 10^{-8}$	$3.4 \times 10^{-4}$	+167	14
48 hour	strong wind weak cooling	$+5.0 \times 10^{-8}$	$3.8 \times 10^{-4}$	+104	15
48 hour	strong wind	0	$4.5 \times 10^{-4}$	0	16
48 hour	strong wind no rotation	0	$1.6 \times 10^{-4}$	0	10
48 hour	strong wind and sunny	$-1.0 \times 10^{-7}$	$2.0 \times 10^{-4}$	-208	11

**Table 1.** Summary of surface boundary conditions for LES used to calibrate CATKE. All LES are initialized with the buoyancy profile described in equation (A1) and use the traditional  $f$ -plane approximation with Coriolis parameter  $f = 10^{-4} \text{ s}^{-1}$ , except “strong wind no rotation” and “strong wind and sunny”, which omit Coriolis forces entirely. The “suite” indicates simulation duration.  $J_b$  is the surface buoyancy flux,  $\tau_x$  is the kinematic momentum flux (momentum flux divided by ocean reference density),  $Q \approx \rho_o c_p J_b / (\alpha g)$  is the heat flux associated with  $J_b$ , and  $u_{10}$  is an estimate of the 10-meter wind speed associated with  $\tau_x$  according to equation A5 using reference density  $\rho_o = 1024 \text{ kg m}^{-3}$ , seawater heat capacity  $c_p = 3991 \text{ J } ^\circ\text{C}^{-1}$ , thermal expansion coefficient  $\alpha = 2 \times 10^{-4} \text{ }^\circ\text{C}^{-1}$ , gravitational acceleration  $g = 9.81 \text{ m s}^{-2}$  are used for  $Q$  and  $u_{10}$ . When the surface buoyancy flux is negative ( $J_b < 0$ ),  $J_b$  represents  $J_b = I(z = 0)$ , where  $I(z)$  is the buoyancy flux associated with penetrating solar insolation in equation 6. The forcing in equation (3) is then defined as  $F_b = -\partial_z I$ . All fluxes use the convention that a positive flux carries quantities upwards, out of the ocean, which means a negative  $\tau_x$  drives currents in the  $+\hat{x}$  direction and a positive buoyancy flux cools the ocean by extracting buoyancy. Additional LES used to validate CATKE are summarized in table 2.



**Figure 1.** Visualization of vertical velocity  $w$  in 9 of 35 large eddy simulations (LES) of the ocean surface boundary layer used in this paper, forced variously by winds, surface waves, and heat fluxes. All LES, which are summarized in tables 1 and 2 and described in more detail in [Appendix A](#), are initialized with the same density stratification. (a)–(c) show strongly-forced LES after just 6 hours of simulation, (d)–(f) show LES driven by medium-strength forcing after 24 hours, and (g)–(i) show weakly forced LES after 72 hours. (a), (d), and (g) show a purely wind and wave driven case, (b), (e), (h) are forced by a mixture of winds, waves, and cooling, and (c), (f), and (i) are “free convection” cases forced only by cooling with no winds and waves. All simulations are rotating with Coriolis parameter  $f = 10^{-4} \text{ s}^{-1}$ . The colorscale for each panel saturates at  $\frac{1}{2} \max |w|$ . For each panel,  $\max |w|$  is (a) 0.26, (b) 0.29, (c) 0.086, (d) 0.20, (e) 0.23, (f) 0.070, (g) 0.056, (h) 0.14, and (i)  $0.041 \text{ m s}^{-1}$ .

average of (1) and (3) to obtain,

$$\partial_t u - fv = -\partial_z \overline{w'u'} + \bar{F}_u, \quad (9)$$

$$\partial_t v + fu = -\partial_z \overline{w'v'} + \bar{F}_v, \quad (10)$$

$$\partial_t c = -\partial_z \overline{w'c'} + \bar{F}_c, \quad (11)$$

where  $u, v$  represent the horizontal average of the horizontal Lagrangian-mean velocities  $U^L$ ,  $V^L$ , and the superscript L is omitted to simplify notation. Lateral fluxes vanish from (9)–(11) due to horizontal homogeneity. No Stokes-drift-dependent terms enter into (9)–(11) because  $\mathbf{U}^S(z)$  is horizontally uniform. Figure 2 illustrates the horizontally-averaged buoyancy, velocity, and turbulent kinetic energy for the 12-hour strong wind, weak cooling case.

The parameterization problem may now be stated: we seek a parameterization that predicts the vertical fluxes  $\overline{w'u'}$ ,  $\overline{w'v'}$ , and  $\overline{w'c'}$  in terms of the resolved state  $u, v, c$ , boundary conditions, and potentially, additional auxiliary variables. For example, the parameterization

Suite	Case	$J_b$ ( $\text{m}^2 \text{s}^{-3}$ )	$ \tau_x $ ( $\text{m}^2 \text{s}^{-2}$ )	$Q$ ( $\frac{\text{W}}{\text{m}^2}$ )	$u_{10}$ ( $\frac{\text{m}}{\text{s}}$ )
6 hour	free convection	$+9.6 \times 10^{-7}$	0	+2000	0
6 hour	weak wind strong cooling	$+8.0 \times 10^{-7}$	$5.0 \times 10^{-4}$	+1666	16
6 hour	mid wind mid cooling	$+6.4 \times 10^{-7}$	$8.0 \times 10^{-4}$	+1333	20
6 hour	strong wind weak cooling	$+4.0 \times 10^{-7}$	$1.2 \times 10^{-3}$	+833	23
6 hour	strong wind	0	$1.4 \times 10^{-3}$	0	24
6 hour	strong wind no rotation	0	$1.1 \times 10^{-3}$	0	22
6 hour	strong wind and sunny	$-6.0 \times 10^{-7}$	$1.5 \times 10^{-3}$	-1250	25
72 hour	free convection	$+8.7 \times 10^{-8}$	0	+181	0
72 hour	weak wind strong cooling	$+7.5 \times 10^{-8}$	$1.8 \times 10^{-4}$	+156	11
72 hour	mid wind mid cooling	$+6.0 \times 10^{-8}$	$2.9 \times 10^{-4}$	+125	13
72 hour	strong wind weak cooling	$+3.8 \times 10^{-8}$	$3.4 \times 10^{-4}$	+79	14
72 hour	strong wind	0	$4.1 \times 10^{-4}$	0	15
72 hour	strong wind no rotation	0	$1.1 \times 10^{-4}$	0	9
72 hour	strong wind and sunny	$-5.0 \times 10^{-8}$	$1.3 \times 10^{-4}$	-104	9

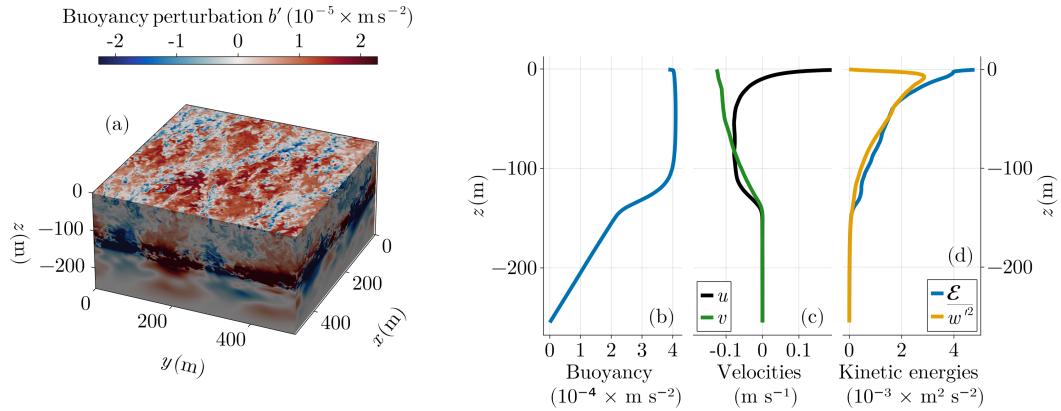
**Table 2.** Summary of surface boundary conditions for LES used to validate CATKE. See table 1 for a description and a summary of the LES used to calibrate CATKE.

described in the next section uses a downgradient formulation  $\overline{w'c'} \sim \partial_z c$  to predict vertical tracer and momentum fluxes.

## 2.2 Connection to the regional and global ocean modeling context

Our LES, and the models that predict the horizontal average of the LES, may be described as “single column models”. This nomenclature reflects the notion that the models simulate the vertical redistribution of momentum and tracers by turbulent motions in a single column of a three-dimensional ocean model. Indeed, we envision that the single column context is generalized to a large-scale ocean simulation merely by adding advection by motions somewhat larger than the scale of the LES domain. This approach relies on two key assumptions. First, the microscale turbulence must be horizontally homogeneous so as to ignore lateral flux divergences. Second, there must be a scale separation between microscale turbulence and larger-scale motions so that interactions between the two can be ignored.

For typical oceanic situations, the first assumption is likely satisfied because vertical gradients are much larger than horizontal ones on the scales of a “single column model” and thus the vertical flux divergences dominate over horizontal divergences. In other words the ocean is more homogeneous in the horizontal than in the vertical on scales of  $O(100 \text{ m})$ . The second assumption is more problematic especially near the ocean surface and bottom boundaries. While microscale turbulence does not significantly interact with mesoscale geostrophic eddies with scales of  $O(10\text{--}100 \text{ km})$ , there is growing evidence of interactions between submesoscale frontal dynamics with scales of  $O(100 \text{ m} \text{--} 10 \text{ km})$  and microscale turbulence (see reviews by Thomas et al., 2008; McWilliams, 2016; J. R. Taylor & Thompson, 2023). Frontal instabilities are also effective at restratifying the ocean boundary layers during time of weak microscale turbulence (see for example Boccaletti et al., 2007). These interactions are presently ignored in the formulation of microscale turbulence parameterizations, but they are an obvious direction for future development of CATKE. Following the approach



**Figure 2.** Illustration of horizontally-averaged data from the 12-hour strong wind, weak cooling LES. Panel (a) shows the buoyancy perturbation  $b'$ . Note the colorbar is strongly saturated to illustrate boundary layer structure; the buoyancy perturbation is particularly large at the base of the boundary layer, where the horizontally-averaged buoyancy gradient is also strong. (b) shows the horizontally-averaged buoyancy  $b$ , (c) shows the horizontally-averaged velocities  $u, v$ , and (d) shows the horizontally-averaged fluctuation kinetic energy,  $\mathcal{E} \stackrel{\text{def}}{=} (\overline{u'^2} + \overline{v'^2} + \overline{w'^2}) / 2$  and horizontally-averaged vertical velocity variance,  $\overline{w'^2}$ .

outlined in this paper, such an effort will require generating a library of simulations which resolve microscale turbulence in the presence of ocean fronts, extending CATKE to include those physics, and then calibrating the extended CATKE against the new library of those simulations.

Similarly, microscale turbulent mixing in the ocean interior requires considering multi-scale dynamics. For example, internal waves generated by surface winds and tide-bathymetry interactions produce a direct cascade of internal wave energy to progressively smaller scales until wave breaking finally transfers energy to microscale turbulence. Incorporating the physics of turbulent mixing driven by internal wave breaking is another area for future development.

### 3 CATKE formulation

CATKE models the horizontally-averaged vertical fluxes  $\overline{w'\psi'}$  appearing on the right side of (9)–(11) with a downgradient, mixing length formulation (Prandtl et al., 1925),

$$\overline{w'\psi'} \approx -\underbrace{\ell_\psi \sqrt{e}}_{\stackrel{\text{def}}{=} K_\psi} \partial_z \psi, \quad (12)$$

where  $e$  is the turbulent kinetic energy,  $\sqrt{e}$  is the turbulent velocity scale, and  $\ell_\psi$  is the mixing length for the horizontally-averaged variable  $\psi(z, t)$ . After choosing to parameterize turbulent transport with eddy diffusion that depends on the turbulent velocity  $\sqrt{e}$  and mixing length  $\ell_\psi$ , the form  $K_\psi = \ell_\psi \sqrt{e}$  follows from dimensional analysis. CATKE invokes three mixing lengths and three eddy diffusivities for horizontal velocities ( $\ell_u$  and  $K_u$ ), tracers ( $\ell_c$  and  $K_c$ ), and turbulent kinetic energy ( $\ell_e$  and  $K_e$ ).

With (12), the single column equations become

$$\partial_t u - fv = \partial_z (K_u \partial_z u) + \bar{F}_u, \quad (13)$$

$$\partial_t v + fu = \partial_z (K_u \partial_z v) + \bar{F}_v, \quad (14)$$

$$\partial_t c = \partial_z (K_c \partial_z c) + \bar{F}_c. \quad (15)$$

In this paper we use a linear equation of state that relates density to a single thermodynamic constituent, such that the buoyancy  $b$  is just another tracer,

$$\partial_t b = \partial_z (K_c \partial_z b) + \bar{F}_b. \quad (16)$$

The buoyancy gradient  $N^2 \stackrel{\text{def}}{=} \partial_z b$  appears in many of the scaling arguments central to CATKE's formulation, where  $N$  is often referred to as the "buoyancy frequency". Note that in more realistic simulations of seawater,  $b$  and  $N^2$  are functions of geopotential height, mean temperature, and mean salinity through the empirically-determined seawater equation of state (McDougall & Barker, 2011).

Next we turn to the estimation of the turbulent kinetic energy  $e$ , and thus the turbulent velocity scale  $\sqrt{e}$  in (12). For this we first introduce the kinetic energy of the subgrid velocity field,  $\mathcal{E}$ , defined in terms of the velocity fluctuations  $(u', v', w')$ ,

$$\mathcal{E} \stackrel{\text{def}}{=} \frac{1}{2} \overline{|u'|^2} = \frac{1}{2} \left( \overline{u'^2} + \overline{v'^2} + \overline{w'^2} \right). \quad (17)$$

We postulate a close relationship between  $e$  in (12) and the subgrid kinetic energy,  $\mathcal{E}$ . However, this is a relationship rather than an identity, because  $\mathcal{E}$  has contributions from motions that are unrelated to the eddy diffusivity in (12). For example, internal waves generated by convective plumes make a significant contribution to  $\mathcal{E}$  below the base of boundary layer, despite that there is no mixing there. Moreover, even if the kinetic energy and mixing length are known, a correlation coefficient is still required to compute the eddy diffusivity in (12) (G. I. Taylor, 1922). We therefore interpret  $e$  as a *latent variable* whose sole purpose is to enable accurate computation of the eddy diffusivity in (12), rather conflating  $e$  with the observable but less relevant quantity  $\mathcal{E}$ . This interpretation has implications for calibration: we do not use discrepancy between LES-derived  $\mathcal{E}$  and  $e$  to constrain CATKE's free parameters. Instead, we only use the discrepancies between LES and model-predicted variables  $u$ ,  $v$ , and  $c$ . CATKE's  $e$  is therefore free to deviate from  $\mathcal{E}$  if this produces more accurate eddy diffusivities and thus more accurate predictions of  $u$ ,  $v$ ,  $c$ . Interpreting  $e$  as a latent variable rather than as the subgrid kinetic energy  $\mathcal{E}$  is also proposed by Kolmogorov (see Spalding, 1991) and Saffman (1970).

Though we define  $e$  as a latent variable, we still expect similarity between  $e$  and  $\mathcal{E}$  on physical grounds — where there is turbulence, there will be mixing — and following prior work (Saffman, 1970; Gaspar et al., 1990; Spalding, 1991; Umlauf & Burchard, 2003), use the evolution equation for  $\mathcal{E}$  to formulate a model for the evolution of  $e$ . An equation describing the evolution of  $\mathcal{E}$  can be derived from (1), including the molecular stress divergence  $\nu \nabla^2 (\mathbf{U}^L - \mathbf{U}^S)$  (we include the Stokes drift term here for completeness, though it does not contribute to the equation for  $\mathcal{E}$ ). The result is

$$\partial_t \mathcal{E} = \underbrace{-\partial_z (\overline{w' \mathcal{E}'} + \overline{w' p'} - \nu \partial_z \mathcal{E})}_{\text{transport}} - \underbrace{\overline{\mathbf{u}' \mathbf{w}' \cdot \partial_z \mathbf{u}}}_{\text{shear production}} + \underbrace{\overline{w' b'}}_{\text{buoyancy flux}} - \underbrace{\nu \overline{|\nabla \mathbf{u}'|^2}}_{\text{dissipation}}, \quad (18)$$

where  $\nu$  is the kinematic viscosity,  $p$  is kinematic pressure (dynamic pressure divided by a reference density) and  $\mathcal{E}' = \frac{1}{2} |\mathbf{u}'|^2 - \mathcal{E}$ . Because  $\mathbf{u}$  is the horizontally-averaged Lagrangian-mean velocity, the shear production term in (18) represents the total transfer of kinetic energy from the average  $\mathbf{u}$  to the fluctuations  $\mathbf{u}'$ , including the so-called "Stokes production" term (McWilliams et al., 1997). Again following prior work (Saffman, 1970; Gaspar et al.,

356 1990; Spalding, 1991; Umlauf & Burchard, 2003) we write the equation for  $e$  using terms  
 357 that mirror each term in equation (18):

$$358 \quad \partial_t e = \underbrace{\partial_z (K_e \partial_z e)}_{\text{transport}} + \underbrace{K_u |\partial_z \mathbf{u}|^2}_{\text{shear production}} - \underbrace{K_c N^2}_{\text{buoyancy flux}} - \underbrace{\frac{e^{3/2}}{\ell_D}}_{\text{dissipation}}, \quad (19)$$

359 where  $|\partial_z \mathbf{u}|^2 = (\partial_z u)^2 + (\partial_z v)^2$  is the square vertical shear of the horizontally-averaged  
 360 velocity field  $\mathbf{u}$  ( $w = 0$  because of horizontal homogeneity),  $K_e$  is the vertical diffusivity of  $e$ ,  
 361  $\ell_D$  is the “dissipation length scale”, and we have labeled the corresponding terms in (18)  
 362 and (19). The shear production and buoyancy flux terms are formulated by applying the  
 363 eddy diffusivity hypothesis (12) to their corresponding expressions in equation (18). Like in  
 364 the budget for  $\mathcal{E}$ , the shear production term in (19) represents the total shear production  
 365 including both “Eulerian” and “Stokes” production.

366 Even with perfect predictions of  $u, v, c$  — and therefore perfect shear production and  
 367 buoyancy flux —  $\mathcal{E}$  and  $e$  can still differ because of the approximate transport and dissipation  
 368 terms in (19). In particular, we assume in (19) that the transport of  $e$ , which helps to  
 369 deepen boundary layers by modeling turbulence spreading away from turbulence-generating  
 370 regions, can be modeled with an eddy diffusivity  $K_e = \ell_e \sqrt{e}$ . To model the dissipation of  $e$   
 371 we introduce the dissipation length scale  $\ell_D$ , which has a similar form to the mixing lengths  
 372  $\ell_u$ ,  $\ell_c$ , and  $\ell_e$ . The expression  $e^{3/2}/\ell_D$  in (19) follows on dimensional grounds.

373 Equation (19) requires boundary conditions. We impose a no-flux condition on  $e$  at  
 374 the bottom. (Extending CATKE to describe the bottom boundary layer in the future may  
 375 require imposing a different bottom boundary condition.) At  $z = 0$ , we parameterize subgrid  
 376 production of  $e$  by wind stress and destabilizing buoyancy fluxes across the uppermost cell  
 377 interface with

$$378 \quad J_e \stackrel{\text{def}}{=} -K_e \partial_z e|_{z=0} = -\mathbb{C}_J^{\text{shear}} u_*^3 - \mathbb{C}_J^{\text{conv}} w_\Delta^3, \quad \text{where } w_\Delta^3 \stackrel{\text{def}}{=} \Delta z \max(J_b, 0), \quad (20)$$

379 and  $\mathbb{C}_J^{\text{shear}}$  and  $\mathbb{C}_J^{\text{conv}}$  are constant, non-dimensional free parameters,  $J_b$  is the surface  
 380 buoyancy flux defined such that  $J_b > 0$  removes buoyancy and thus causes convection,  $\Delta z$  is  
 381 the distance between the top of the ocean domain and the first interior cell interface, and  
 382  $w_\Delta^2$  is the convective TKE scale that follows from a balance between buoyant production  
 383 and dissipation estimated using the grid spacing  $\Delta z$  as a length scale.  $u_*$  in (20) is the  
 384 ocean-side friction velocity,

$$385 \quad u_* \stackrel{\text{def}}{=} (\tau_x^2 + \tau_y^2)^{1/4}, \quad (21)$$

386 defined in terms of the zonal and meridional kinematic momentum fluxes  $\tau_x$  and  $\tau_y$  (wind  
 387 stresses divided by reference water density). Note that other TKE-based models (Blanke &  
 388 Delecluse, 1993; Madec et al., 2017) prescribe surface TKE (rather than TKE flux), and do  
 389 not depend on the surface buoyancy flux  $J_b$ .

390 Equation (20) introduces the notation

$$391 \quad \mathbb{C}_{\text{component}}^{\text{label}} \quad (22)$$

392 for two free parameters  $\mathbb{C}_J^{\text{shear}}$  and  $\mathbb{C}_J^{\text{conv}}$ , where “label” indicates the parameter’s role and  
 393 “component” associates the parameter with a variable or model component.

### 394 3.1 Turbulence length scale model

395 We decompose the four length scales  $\ell_\psi \in (\ell_u, \ell_c, \ell_e, \ell_D)$  into a shear-dominated length  
 396 scale  $\ell_\psi^{\text{shear}}$  limited by density-stratification and surface distance, and a convection-dominated  
 397 length scale  $\ell_\psi^{\text{conv}}$  limited by the depth of the convective boundary layer. At any time and  
 398 location, the maximum of these two length scales is chosen as the mixing length via

$$399 \quad \ell_\psi = \max(\ell_\psi^{\text{conv}}, \ell_\psi^{\text{shear}}), \quad (23)$$

encapsulating a sharp separation between turbulence regimes. We next describe a length scale formulation that can be calibrated to predict turbulent fluxes associated with the kinds of flows plotted in figure 1.

### 3.1.1 Shear turbulence length scale

To represent shear-dominated turbulence, we use the length scale

$$\ell_\psi^{\text{shear}} = \mathbb{S}_\psi(Ri) \min \left( \frac{\sqrt{e}}{N_+}, \mathbb{C}^s d \right), \quad \text{where } N_+^2 \stackrel{\text{def}}{=} \max(0, \partial_z b) \quad (24)$$

with  $d$  the distance to the ocean surface,  $\mathbb{C}^s$  a free parameter (“ $s$ ” for “surface”), and  $\mathbb{S}_\psi$  a “stability function” defined below.  $\sqrt{e}/N$  is the vertical distance traversed by a patch of turbulence expending all its kinetic energy  $e$  to mix the uniform stratification  $N$ . Blanke and Delecluse (1993) point out that  $\sqrt{e}/N$  is a local approximation to the more complete but computationally-expensive length scale proposed by Gaspar et al. (1990).

We use (24) for  $\ell_c^{\text{shear}}$ ,  $\ell_u^{\text{shear}}$ , and  $\ell_e^{\text{shear}}$ . For the dissipation length scale  $\ell_D^{\text{shear}}$ , we use

$$\ell_D^{\text{shear}} = \frac{1}{\mathbb{S}_D(Ri)} \min \left( \frac{\sqrt{e}}{N_+}, \mathbb{C}^s d \right), \quad (25)$$

so that the stability function for the dissipation length scale is  $1/\mathbb{S}_D$ . The alternative formulation in (25) yields a tight connection between  $\mathbb{S}_D$ ’s free parameters and  $e$  dissipation, and facilitates the physical interpretation of CATKE’s parameters.

The stability functions  $\mathbb{S}_\psi(Ri)$  and  $1/\mathbb{S}_D(Ri)$  in (24)–(25) depend on the gradient Richardson number,

$$Ri \stackrel{\text{def}}{=} \frac{\partial_z b}{|\partial_z \mathbf{u}|^2}, \quad (26)$$

which means that each diffusivity  $K_\psi$  also depends explicitly on  $Ri$ . More specifically, we hypothesize that  $K_u$ ,  $K_c$ , and  $K_e$  are all explicit functions of  $|\partial_z \mathbf{u}|^2$  in addition to  $N^2$ ,  $e$ , and the wall-distance  $d$ . CATKE is therefore more expressive than the closure described by Blanke and Delecluse (1993), wherein  $K_u$  and  $K_e$  do not depend explicitly on  $|\partial_z \mathbf{u}|^2$ . Second-moment closures also define  $K_u$  and  $K_c$  that depend on  $|\partial_z \mathbf{u}|^2$ , in addition to  $N^2$ ,  $e$ , and the dissipation rate  $\epsilon$  (see, for example Burchard & Bolding, 2001).  $Ri$ -dependent stability functions also allow CATKE to capture, in some form, the well-known dependence between  $Ri$  and the turbulent Prandtl number (D. Li, 2019; C. Caulfield, 2021)

$$Pr(Ri) \stackrel{\text{def}}{=} \frac{K_u}{K_c} = \frac{\mathbb{S}_u(Ri)}{\mathbb{S}_c(Ri)}. \quad (27)$$

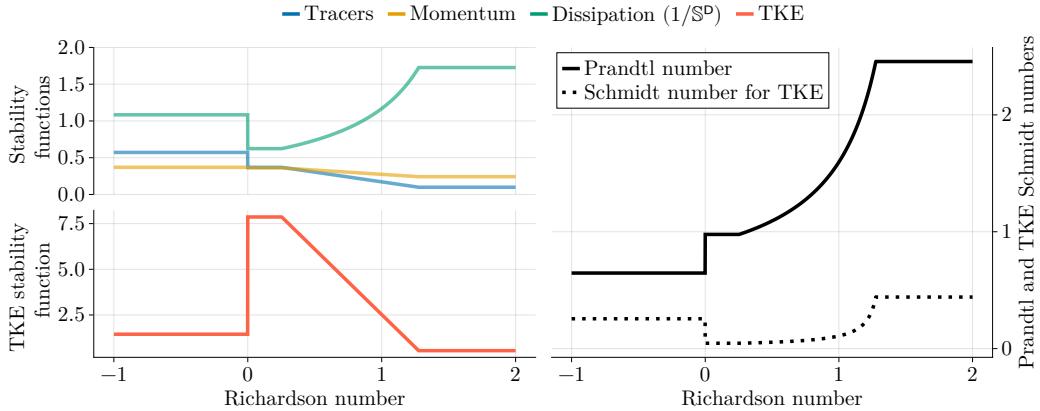
We balance expressiveness and parsimony with four-part  $\mathbb{S}_\psi(Ri)$ ,

$$\mathbb{S}_\psi(Ri) = \begin{cases} \mathbb{C}_\psi^- & \text{when } Ri < 0, \\ \mathbb{C}_\psi^0 & \text{when } 0 \leq Ri \leq \mathbb{C}_{Ri}^0, \\ \mathbb{C}_\psi^0 + \left( \mathbb{C}_\psi^\infty - \mathbb{C}_\psi^0 \right) \frac{Ri - \mathbb{C}_{Ri}^0}{\mathbb{C}_{Ri}^\delta} & \text{when } \mathbb{C}_{Ri}^0 < Ri < \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta, \\ \mathbb{C}_\psi^\infty & \text{when } Ri \geq \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta. \end{cases} \quad (28)$$

In (28), the parameter  $\mathbb{C}_{Ri}^0$  is the “transition  $Ri$ ”. The four regions of the stability function are:

- Constant  $\mathbb{S}_\psi = \mathbb{C}_\psi^-$  for unstably-stratified shear turbulence with  $Ri < 0$ .
- Constant  $\mathbb{S}_\psi = \mathbb{C}_\psi^0$  for near-neutral turbulence with  $0 \leq Ri \leq \mathbb{C}_{Ri}^0$ .
- Linearly-varying from  $\mathbb{C}_\psi^0$  to  $\mathbb{C}_\psi^\infty$  as  $Ri$  increases from  $\mathbb{C}_{Ri}^0$  to  $\mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta$ .
- Constant  $\mathbb{S}_\psi = \mathbb{C}_\psi^\infty$  when high  $Ri > \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta$ .

The stability function (28) plays a similar role as the more elaborate stability functions used in two-equation models (Burchard & Bolding, 2001), which are derived from a second-moment closure. The stability functions in equation (28) are plotted in the left panel of figure 3 (see section 4 for how the parameters are obtained via calibration to LES). Note that the form of the stability functions in (28) imply that  $Pr$  is constant in the limit  $Ri \rightarrow 0$  and  $Ri \rightarrow \infty$ , which Venayagamoorthy and Stretch (2010) argue is inconsistent with direct numerical simulation data. An extensive exploration of different formulations for  $\mathbb{S}_\psi$  is beyond the scope of the present work but remains an important direction for future research.



**Figure 3.** Stability functions (left panel), and Prandtl numbers and Schmidt numbers (right panel), computed with parameters calibrated against large eddy simulations as described in section 4. The stability functions for tracers, momentum, and TKE are given by  $\mathbb{S}_\psi$  in (28). The stability function for dissipation length scale is  $1/\mathbb{S}_D$ . The Prandtl number is  $\mathbb{S}_u/\mathbb{S}_c$  and the Schmidt number for TKE is  $\mathbb{S}_u/\mathbb{S}_e$ .

The four shear length scales introduce 15 free parameters:  $\mathbb{C}^s$ ,  $\mathbb{C}_{Ri}^\delta$ , and  $\mathbb{C}_{Ri}^0$  used in all four length scales, along with 12 additional parameters associated with the coefficients  $\mathbb{C}_\psi^-$ ,  $\mathbb{C}_\psi^0$  and  $\mathbb{C}_\psi^\infty$  for each length scale respectively.

### 3.1.2 Turbulent Prandtl and Schmidt numbers in stably stratified shear turbulence

CATKE's  $Pr$  in (27) is a rational function of  $Ri$ , slightly different from the piecewise linear formulation proposed by Blanke and Delecluse (1993) and Madec et al. (2017). In particular,

$$Pr = \begin{cases} \mathbb{C}_u^-/\mathbb{C}_c^- & Ri < 0 \\ \mathbb{C}_u^0/\mathbb{C}_c^0 & 0 \leq Ri \leq \mathbb{C}_{Ri}^0 \\ \frac{\mathbb{C}_u^0 + \mu_\psi(Ri - \mathbb{C}_{Ri}^0)}{\mathbb{C}_c^0 + \mu_c(Ri - \mathbb{C}_{Ri}^0)} & \mathbb{C}_{Ri}^0 < Ri < \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta \\ \mathbb{C}_u^\infty/\mathbb{C}_c^\infty & Ri \geq \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta \end{cases}, \quad (29)$$

where  $\mu_\psi \stackrel{\text{def}}{=} (\mathbb{C}_\psi^\infty - \mathbb{C}_\psi^0)/\mathbb{C}_{Ri}^\delta$ . Similarly, the Schmidt number for TKE transport in stably-stratified shear turbulence is  $Sc \stackrel{\text{def}}{=} K_u/K_e$ . The Prandtl number and Schmidt number for calibrated parameters are visualized in the right panel figure 3.

456      **3.1.3 Neutral, self-similar, wave-modulated, non-rotating, near-surface mixing**  
 457

458      To interpret CATKE's mixing length near the surface in neutrally-stratified ( $\partial_z b = 0$ )  
 459      conditions, when  $\ell_\psi \sim d$ , we consider quasi-equilibrium ( $\partial_t u \approx \partial_t e \approx 0$ ), non-rotating ( $f = 0$ )  
 460      near-surface turbulence driven by wind stress  $\boldsymbol{\tau} = \tau_x \hat{\mathbf{x}}$ . We suppose that the CATKE-  
 461      parameterized single column equations (13)–(15) and (19) possess a similarity solution in  
 462      this scenario (Von Kármán, 1931),

463      
$$\partial_z u \approx \frac{u_*}{\kappa d}, \quad (30)$$

464      where  $u_*$  is the friction velocity defined in equation (21) (here simply  $\sqrt{|\tau_x|}$ ),  $d = -z$  is the  
 465      distance to the surface, and  $\kappa$  is a constant parameter. If the ocean surface were rigid,  $\kappa$   
 466      could be interpreted as the von Kármán constant. But because the LES we use in this paper  
 467      include surface wave effects,  $\kappa$  has a slightly different interpretation — as a “wave-modified”  
 468      similarity layer constant, perhaps, as proposed by Samelson (2022).

469      To express  $\kappa$  in terms of CATKE's free parameters, we begin by assuming a balance  
 470      between shear production and dissipation and neglecting diffusive turbulent transport to  
 471      simplify (19) to

472      
$$K_u (\partial_z u)^2 \approx \frac{e^{3/2}}{\ell_D}. \quad (31)$$

473      Note that in neutral conditions,

474      
$$K_u = \mathbb{C}_u^0 \mathbb{C}^s d \sqrt{e}, \quad \text{and} \quad \ell_D = \frac{\mathbb{C}^s}{\mathbb{C}_D^0} d. \quad (32)$$

475      Inserting (30) and (32) into (31) and rearranging, we find an expression that relates the  
 476      constant  $\kappa$ ,  $u_*$ , and  $e$ ,

477      
$$\frac{u_*^2}{e} \approx \kappa^2 \frac{\mathbb{C}_D^0}{\mathbb{C}_u^0 (\mathbb{C}^s)^2}. \quad (33)$$

478      Notice that  $e$  is independent of  $d$  in this expression. This means that neglecting turbulent  
 479      transport in (31) in the context of the similarity hypothesis (30) is self-consistent. Next,  
 480      integrating the quasi-equilibrium  $x$ -momentum equation  $0 \approx \partial_z (K_u \partial_z u)$  from  $z = 0$  to  
 481       $z = -d$  yields

482      
$$\partial_z u \approx \frac{u_*}{d} \underbrace{\frac{u_*}{\mathbb{C}_u^0 \mathbb{C}^s \sqrt{e}}}_{=1/\kappa}, \quad (34)$$

483      where we have used the neutral momentum diffusivity in (32) and the friction velocity  
 484      definition  $-K_u \partial_z u|_{z=0} = u_*$ . Equation (34) identifies  $\kappa$  by comparison to (30). We next  
 485      use (33) to eliminate  $u_*/\sqrt{e}$  and obtain an expression for CATKE's wave-modified similarity  
 486      layer constant  $\kappa$ ,

487      
$$\kappa \stackrel{\text{def}}{=} \mathbb{C}^s \left[ (\mathbb{C}_u^0)^3 / \mathbb{C}_D^0 \right]^{1/4}. \quad (35)$$

488      **3.1.4 Steady-state gradient Richardson number for stably stratified shear**  
 489      **turbulence**

490      CATKE's dependence on the stable length scale  $\ell \sim \sqrt{e}/N$  is associated with a steady-  
 491      state gradient Richardson number in stably-stratified shear turbulence (Blanke & Delecluse,  
 492      1993). To see this, we first note that in stable stratification and far from boundaries, the  
 493      mixing and dissipation length scales become

494      
$$\ell_\psi = \mathbb{S}_\psi \frac{\sqrt{e}}{N} \quad \text{for} \quad \psi \in (u, c, e) \quad \text{and} \quad \ell_D = \frac{1}{\mathbb{S}_D} \frac{\sqrt{e}}{N}. \quad (36)$$

495 Inserting (36) into (19) and neglecting turbulent transport (equivalently, assuming spatially-  
 496 uniform  $e$ ) yields

$$497 \quad \partial_t e = N(\mathbb{S}_c + \mathbb{S}_D) \underbrace{\left( \frac{Ri^\dagger}{Ri} - 1 \right)}_{\stackrel{\text{def}}{=} r} e, \quad (37)$$

498 where  $r$  is a rate, and

$$499 \quad Ri^\dagger \stackrel{\text{def}}{=} \frac{\mathbb{S}_u}{\mathbb{S}_c + \mathbb{S}_D}. \quad (38)$$

500 When  $Ri = Ri^\dagger$ , the shear production of TKE is perfectly balanced by TKE destruction  
 501 via buoyancy flux and dissipation, such that  $r = 0$  and  $\partial_t e = 0$ . We therefore call  $Ri^\dagger$   
 502 the “steady-state Richardson number”. If  $Ri < Ri^\dagger$ , then TKE and mixing will increase,  
 503 while if  $Ri > Ri^\dagger$  then TKE will decay and mixing will be suppressed. As a result — and  
 504 as illustrated in section 5.3 and figure 12 —  $Ri$  is driven towards  $Ri^\dagger$  in forced stratified  
 505 shear turbulence. Finally we note that the functions  $\mathbb{S}_\psi$ , defined in (28), depend on  $Ri$ .  
 506 For example if  $Ri < \mathbb{C}_{Ri}^0$ , then  $Ri^\dagger = \mathbb{C}_u^0 / (\mathbb{C}_c^0 + \mathbb{C}_D^0)$ . But if  $Ri^\dagger > \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta$ , then  
 507  $Ri^\dagger = \mathbb{C}_u^\infty / (\mathbb{C}_c^\infty + \mathbb{C}_D^\infty)$ .

### 508 3.1.5 Convective turbulence length scale

509 To formulate a length scale for free convection, we divide the freely convecting boundary  
 510 layer into two regions: a “convecting layer” with unstable  $N^2 < 0$ , and a “penetration  
 511 layer” with thickness  $\delta$ . In the penetration layer,  $N^2(z) > 0$  but  $N^2(z + \delta) < 0$ , where we  
 512 note that the vertical coordinate  $z$  increases upwards and is defined such that  $z < 0$ . We  
 513 use “penetration layer” rather than “entrainment layer” used by Deardorff (1970) to avoid  
 514 confusion with lateral entrainment.) Our formulation for the convective length scale models  
 515 both rapid mixing in the convective layer as well as entrainment into the boundary layer from  
 516 below by plumes plunging through the convecting layer into the stably-stratified penetration  
 517 layer below.

518 Our dynamic length scale for mixing in the convective layer is based on a dimensional  
 519 analysis first proposed by Deardorff (1970) that links the turbulent velocity  $\sqrt{e}$  ( $\text{m s}^{-1}$ ),  
 520 surface buoyancy flux  $J_b$  ( $\text{m}^2/\text{s}^3$ ), and convective layer depth,  $h$  (m),

$$521 \quad \sqrt{e} \sim (h J_b)^{1/3}. \quad (39)$$

522 Recasting (39) in terms of a time-scale  $t_{\text{mix}} \sim h/\sqrt{e}$  for convective mixing over the depth  $h$   
 523 yields

$$524 \quad t_{\text{mix}} \sim \left( \frac{h^2}{J_b} \right)^{1/3}. \quad (40)$$

525 But if we represent convection as a diffusive process with diffusivity  $K_c$ , then we also have  
 526 that

$$527 \quad t_{\text{mix}} \sim \frac{h^2}{K_c}. \quad (41)$$

528 Equating (40) and (41) yields a scaling relation for the convective diffusivity  $K_c$ .

529 Now consider convection driven by constant destabilizing buoyancy fluxes  $J_b$  and  
 530 increasing  $h(t)$ : according to (40), the mixing time then evolves according to  $t_{\text{mix}} \sim h^{2/3}$ . On  
 531 the other hand, if we instead we impose a *constant*  $K_c$  — a commonly used parameterization  
 532 when  $N^2 < 0$  (Madec et al., 2017; Kuhlbrodt et al., 2018; Gutjahr et al., 2021; Jungclaus  
 533 et al., 2022) — then (41) implies that, spuriously,  $t_{\text{mix}} \sim h^2$ . Thus, constant convective  
 534 adjustment diffusivities inaccurately exhibit  $t_{\text{mix}} \sim h^2$  and may produce bias when convection  
 535 competes with other processes such as lateral restratification, or biogeochemical production  
 536 and destruction.

To capture  $t_{\text{mix}}$  consistently between (40) and (41) over the convective region where  $N^2 < 0$ , we introduce a dynamic convective mixing length scale  $\ell_\psi^h$  that scales with  $h$ ,

$$\ell_\psi^h \stackrel{\text{def}}{=} \mathbb{C}_\psi^h \frac{e^{3/2}}{\tilde{J}_b + J_b^{\min}} \sim h, \quad (42)$$

where  $J_b^{\min}$  is chosen small enough to have no impact on CATKE-parameterized solutions, and  $\tilde{J}_b$  is an estimate of the slowly-evolving part of the buoyancy flux  $J_b$  averaged over time-scales  $t \sim t_{\text{mix}}$ . We compute  $\tilde{J}_b$  by integrating

$$\partial_t \tilde{J}_b = \underbrace{\left( \frac{J_b}{\ell_D^2(z=0)} \right)^{1/3} (J_b - \tilde{J}_b)}_{\sim t_{\text{mix}}^{-1}}, \quad (43)$$

where  $\ell_D$  is the dissipation length scale and  $(\ell_D^2/J_b)^{1/3} \sim t_{\text{mix}}$  scales with the instantaneous convective mixing time. Equation (43) relaxes  $\tilde{J}_b$  to  $J_b$  over  $t_{\text{mix}}$ . We use the dissipation length scale  $\ell_D$  in (43) rather than a mixing length because we hypothesize that the convective turbulence evolution time-scale is most closely related to the time-scale for turbulent kinetic energy dissipation rather than a mixing time-scale. In quasi-equilibrium,  $\tilde{J}_b \approx J_b$ . Because  $\ell_\psi^h \sim h$ , CATKE's convective tracer diffusivity scales with  $K_c \sim h\sqrt{e}$ .

The second objective of our convective mixing length formulation is to correctly predict the evolution of  $h$ . For this we introduce a model for “penetrative mixing” *below* the convective mixed layer associated with convective plumes that plunge through the mixed layer and penetrate into the strongly stratified region below. The “empirical law of convection” (Large et al., 1994; Siebesma et al., 2007; Van Roekel et al., 2018; Souza et al., 2020, 2023) is the observation, robust across a wide range of convective conditions, that penetrative fluxes at the penetration level  $z_p$  scale with

$$\overline{w'b'}|_{z=z_p} \sim -J_b \quad \text{such that} \quad h^2 \sim \frac{J_b t}{N^2}, \quad (44)$$

for initially-constant buoyancy gradient  $N^2$  and constant buoyancy flux  $J_b$ .

To ensure that CATKE reproduces (44), we introduce a “penetrative mixing length”,

$$\ell_\psi^p \stackrel{\text{def}}{=} \mathbb{C}_c^p \frac{\tilde{J}_b}{N^2 \sqrt{e} + J_b^{\min}}, \quad (45)$$

which is applied within the aforementioned penetration layer at the depth  $z_p$ , defined via

$$N^2(z_p) > 0 \quad \text{and} \quad N^2(z_p + \delta) < 0, \quad (46)$$

where  $\delta$  is the thickness of the penetration layer. At  $z = z_p$ , (45) produces  $\overline{w'b'} = -\ell_\psi^p \sqrt{e} N^2 \approx -\mathbb{C}_c^p J_b$  in accordance with the empirical law in (44). Our numerical implementation of the convective mixing length uses  $\delta = \Delta z$  where  $\Delta z$  is the grid spacing at  $z_p$ . This assumes that the entrainment layer is thinner than the grid spacing: when  $\delta > \Delta z$ , CATKE solutions may exhibit a “thin entrainment layer bias” even if the boundary layer deepening rate is correct.

The scaling  $h \sim e^{3/2}/J_b$  is an overestimate when  $e$  is produced by both shear and convective buoyancy flux. Since the total mixing length  $\ell_\psi$  takes the maximum between the convective and shear mixing lengths, blending the length scales in a mixed turbulence regime requires a way to reduce the convective mixing length in the presence of significant shear production. For this purpose we introduce an estimate of the flux Richardson number in near-neutral conditions,

$$\widetilde{Ri}_f \stackrel{\text{def}}{=} \frac{d\sqrt{e}|\partial_z \mathbf{u}|^2}{\tilde{J}_b + J_b^{\min}}, \quad (47)$$

575 where  $d = -z$  is depth.  $\widetilde{Ri}_f$  in (47) measures the relative contribution of shear production  
 576 (the numerator) versus buoyancy flux (the denominator) to the TKE budget in unstable  
 577 stratification. We then use this estimate to reduce the convective mixing length by

$$578 \quad \epsilon_{sp} \stackrel{\text{def}}{=} \max(0, 1 - \mathbb{C}^{sp} \widetilde{Ri}_f), \quad (48)$$

579 where  $\mathbb{C}^{sp}$  is a free parameter. The reduction factor (48) may also be interpreted as modeling  
 580 how shear disrupts coherent plumes and thereby reduces convective turbulence correlation  
 581 scales. Note that the numerator in (47) estimates shear production using the mixing length  
 582  $d$ , which is appropriate for shear-driven turbulent mixing. This formulation means that  
 583 the free convection length scale is more limited at depth, where convective plumes are less  
 584 connected to destabilizing surface buoyancy fluxes.

585 Putting (42), (45), and (48) together yields the piecewise parameterization

$$586 \quad \ell_\psi^{\text{conv}}(z) = \epsilon_{sp} \begin{cases} \ell_\psi^h & \text{if } N^2 < 0 \text{ and } J_b > 0, \\ \ell_\psi^p & \text{if } N^2 > 0, N^2(z + \Delta z) < 0, \text{ and } J_b > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (49)$$

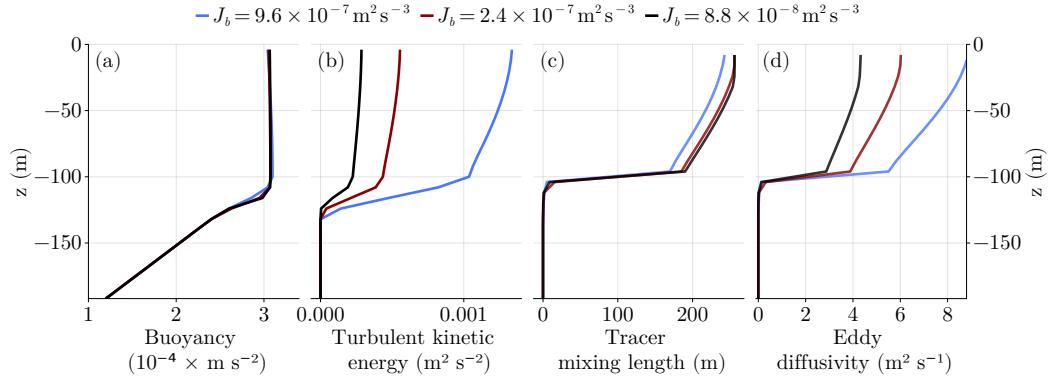
587 Figure 4 illustrates the behavior of the convective length scale predicted by CATKE in (49)  
 588 for three free convection cases with surface buoyancy fluxes  $J_b = 9.6 \times 10^{-7}$ ,  $2.4 \times 10^{-7}$ , and  
 589  $8.8 \times 10^{-8} \text{ m}^2 \text{s}^{-3}$  integrated for 6, 24, and 72 hours respectively, using the initial buoyancy  
 590 profile in equation (A1), which is also used for all our LES. The parameters used to make  
 591 figure 4 are automatically calibrated to large eddy simulations, as described in section 4.  
 592 Figure 4(a) shows CATKE-simulated buoyancy profiles after integrating for 6, 24, and  
 593 72 hours. Figure 4(b) shows that stronger forcing cases have greater levels of turbulent  
 594 kinetic energy. Figure 4(c) shows the tracer mixing length, which above  $z = -100$  meters is  
 595 dominated by the convective mixing length. Though each case has different TKE and different  
 596 surface buoyancy flux, they nevertheless predict similar tracer mixing lengths, corroborating  
 597 the dimensional analysis in equation (39). (We also note that the mixing lengths are twice  
 598 the boundary layer depth. We discuss this and other possible biases in free convection  
 599 further in section 4.) Figure 4(d) shows the eddy diffusivity for the three cases — unlike a  
 600 typical constant-diffusivity convective adjustment model, CATKE’s “convective adjustment  
 601 diffusivity” varies depending on the strength of the surface buoyancy flux. Because the  
 602 predicted mixing length is similar for all three cases, the tracer diffusivity varies with the  
 603 surface buoyancy flux due to variation in the turbulent kinetic energy.

## 604 4 *A posteriori* calibration against large eddy simulations

605 We calibrate CATKE’s 23 free parameters in an *a posteriori* (Duraisamy, 2021; Frezat et  
 606 al., 2022) single-column context using horizontally-averaged data from 21 LES described in  
 607 section 2 and Appendix A. *A posteriori* calibration estimates free parameters by minimizing  
 608 the error between LES data —  $b(z, t)$ ,  $u(z, t)$ ,  $v(z, t)$ , and the forced passive tracer  $c(z, t)$   
 609 extracted from solutions of (1)–(3) — and single column simulations of  $b$ ,  $u$ ,  $v$ , and  $c$  in (13)–  
 610 (15) that use CATKE as a parameterization. The minimization is computed over the whole  
 611 time series and thus in *a posteriori* calibration free parameters are determined by directly  
 612 minimizing simulation bias. In this way, *a posteriori* calibration incorporates numerical and  
 613 other errors that accumulate during a simulation. Moreover, *a posteriori* calibration can  
 614 leverage any observational data computable from the predicted solution, even only indirectly  
 615 informative data. For example, in this work we calibrate elements of the TKE equation  
 616 using only horizontally-averaged momentum and buoyancy profiles derived from LES.

### 617 4.1 The importance of *a posteriori* calibration

618 Explicitly minimizing simulation bias distinguishes *a posteriori* calibration from other  
 619 methods that minimize other biases that are only indirectly related to simulation bias



**Figure 4.** CATKE mixing length and eddy diffusivity during free convection for three cases with boundary layer depth  $h \approx 100$  m. (a) CATKE-predicted buoyancy profiles for the three cases, (b) profiles turbulent kinetic energy,  $e$ , (c) tracer mixing lengths  $\ell_c$ , (d) tracer eddy diffusivities  $K_c$ . The buoyancy fluxes  $J_b$  correspond to heat fluxes  $Q \approx 2000, 500$ , and  $183$  W m $^{-2}$  using  $Q \approx \rho_o c_p J_b / \alpha g$  and  $\rho_o = 1024$  kg m $^{-3}$ ,  $c_p = 3991$  J °C $^{-1}$ ,  $\alpha = 2 \times 10^{-4}$  °C $^{-1}$ , and  $g = 9.81$  m s $^{-2}$ .

— for example by attempting to compute free parameters directly from data, usually by considering subcomponents of the parameterization in isolation (examples may be found in Umlauf & Burchard, 2003; Reichl & Li, 2019). These latter methods are called “*a priori*” (Duraisamy, 2021), because they hinge on additional (often problematic) hypotheses — such as an assumption of structurally perfect, unbiased parameterization (permitting a direct computation of free parameters from limited data), or an assumption that free parameters are uncorrelated with one another (permitting free parameters to be determined in isolated contexts, rather than leveraging all data simultaneously).

To illustrate the pitfalls of *a priori* calibration, we consider integrating a CATKE-parameterized single column equation for buoyancy  $b$ ,

$$\partial_t b = -\partial_z \mathcal{J}(b; \mathbb{C}) + \underbrace{\xi}_{\text{CATKE}} + \underbrace{\text{noisy error}}_{\text{.}} \quad (50)$$

In (50), we include two terms: (*i*) the divergence of a parameterized flux  $\mathcal{J}$  that depends on both the simulated buoyancy  $b$  (omitting here for simplicity other aspects of the state such as  $u$  or  $v$ ) and a set of free parameters  $\mathbb{C}$ , and (*ii*) an explicit “error” term  $\xi$  that represents spatial and temporal discretization errors. We additionally define the ideal or “perfect” solution as  $\hat{b}$ . When equation (50) is integrated forward to predict the evolution of  $b$ , fluctuations away from the perfect solution  $\hat{b}$  inevitably develop due both to structural errors in  $\mathcal{J}$  and because of the discretization error  $\xi$ , leading to an error  $= b - \hat{b}$  that grows as  $\sqrt{t}$  (see, for example Gardiner, 2021).

This error accumulation is potentially fatal for *a-priori*-calibrated parameterizations: because the parameters  $\mathbb{C}$  are determined by evaluating  $\mathcal{J}(\hat{b})$  in terms of the *perfect*  $\hat{b}$ , while the predictions  $\mathcal{J}(b)$  made in terms of the noisy  $b$  are unconstrained by the calibration procedure. At best, the unconstrained predictions  $\mathcal{J}(b)$  are inaccurate. At worst, however, the errors  $\mathcal{J}(b) - \mathcal{J}(\hat{b})$  self-amplify without bound, thwarting prediction altogether (Rasp et al., 2018; Brenowitz & Bretherton, 2019; Rasp, 2020).

*A posteriori* calibration avoids all of these pitfalls by definition, since  $\mathcal{J}(b, \mathbb{C}_*)$  computed in terms of the simulated  $b$  and optimal parameters  $\mathbb{C}_*$  is explicitly constrained by minimizing the discrepancy between  $\mathcal{J}(b, \mathbb{C})$  and data. Put differently: *a posteriori* calibration “teaches”  $\mathcal{J}$  how to make accurate, stable predictions in terms of potentially noisy inputs  $b$ . We

649 leverage this feature to realize a key innovation of this work: we explicitly minimize spatial  
 650 discretization error by including single-column simulations with 2-, 4-, and 8-meter resolution  
 651 in our loss function.

## 652 4.2 Ensemble Kalman Inversion for *a posteriori* calibration

653 To solve the nonlinear inverse problem posed by *a posteriori* calibration, we use an  
 654 ensemble-based method called Ensemble Kalman Inversion (EKI; Iglesias et al., 2013). An  
 655 advantage of EKI is that it is gradient-free, requiring only the ability to run an ensemble of  
 656 simulations with different parameters. The EKI algorithm can be construed either as the  
 657 integration of a dynamical system or as an iterative scheme for repeatedly refining an initial  
 658 distribution of free parameter values.

659 EKI minimizes the objective function

$$660 \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}) \stackrel{\text{def}}{=} \|\mathcal{M}^{-1/2} [\mathcal{G}(\mathbb{C}) - \mathcal{Y}] \|_2^2, \quad (51)$$

661 where  $\mathcal{Y}$  denotes a vector of observational data,  $\mathcal{G}(\mathbb{C})$  denotes a parameterized prediction of  
 662 the observations made with a set of free parameters  $\mathbb{C}$ , and  $\mathcal{M}$  is a matrix that represents the  
 663 uncertainty of  $\mathcal{Y}$ .  $\Phi$  measures the discrepancy between  $\mathcal{G}(\mathbb{C})$  and  $\mathcal{Y}$  given uncertainty  $\mathcal{M}$ . The  
 664 data  $\mathcal{Y}$  is extracted from 21 of the LES described in table 1 that have intermediate surface  
 665 forcing, each coarse-grained three times to 2-, 4-, and 8-meter vertical resolution, respectively.  
 666  $\mathcal{G}$  is constructed by assembling  $21 \times 3 = 63$  single column simulations, representing a  
 667 prediction of each of the 21 LES cases at the three vertical resolutions.

668 We note that the near-surface dynamics in the LES seem uncertain. For example, the  
 669 LES profiles exhibit strong unstable near-surface buoyancy gradients for strongly-forced  
 670 convective cases, indicating that turbulent mixing is suppressed near the top of the LES  
 671 domain. These features are robust to changes in LES resolution (see Appendix A) and  
 672 may represent real physics, since the scale of turbulent motions is restricted by proximity  
 673 to the ocean surface. However, it is also plausible that the LES are missing important  
 674 mixing processes near a wavy, bubbly, broken ocean surface, such as wave breaking, or  
 675 unresolved surface-wave-turbulence interactions. We therefore omit the top 4 meters of the  
 676 LES domain from the data vector  $\mathcal{Y}$ , and thereby avoid overconstraining parameters with  
 677 the most uncertain elements of the LES data.

678 EKI finds a set of optimal parameters  $\mathbb{C} = \mathbb{C}_*$  that minimize  $\Phi(\mathcal{G}, \mathcal{Y}, \mathbb{C})$  in (51) by  
 679 evolving an ensemble of parameter sets using the algorithm described in Appendix C. In  
 680 this work we use relatively large ensembles with 1000 members. This means that every EKI  
 681 iteration requires  $21 \times 3 \times 1000 = 63,000$  single column simulations, for 21 LES cases and 3  
 682 vertical resolutions. To make the calibration as efficient as possible, we implement CATKE  
 683 in Oceananigans and leverage a feature that permits us to integrate an ensemble of single  
 684 column models in parallel in the configuration of a single three-dimensional simulation on  
 685 a GPU. As a result, each EKI iteration requires evolving 9 effectively three-dimensional  
 686 simulations (3 resolutions for each of the 12-, 24- and 48-hour suites). On an Nvidia Titan  
 687 V GPU and with 1,000 ensemble members, a single EKI iteration takes 40-50 seconds,  
 688 and the entire calibration takes 4-6 hours. In the course of this work we have performed  
 689 complete calibrations of CATKE's parameters hundreds of times — to experiment with new  
 690 formulations, new numerical schemes, and to tweak the calibration setup. This workflow  
 691 represents a new “calibration-based” paradigm in parameterization development, where  
 692 physical formulation or numerical implementation changes are tested against the baseline by  
 693 comparing predictions for independently calibrated parameterizations. The 23 calibrated  
 694 free parameters that correspond to the version of CATKE described in this paper and the  
 695 previously described LES are listed in table 3.

Symbol	Description	Optimal value	Bounds
$\mathbb{C}_J^{\text{shear}}$	Wind stress TKE surface flux	3.18	(0, 8)
$\mathbb{C}_J^{\text{conv}}$	Convective TKE surface flux	0.38	(0, 8)
$\mathbb{C}^s$	Near-surface mixing scale	1.13	(0, 2)
$\mathbb{C}_c^h$	Tracer free convection scale	4.79	(0, 8)
$\mathbb{C}_c^-$	Tracer mixing for negative $Ri$	0.57	(0, 2)
$\mathbb{C}_c^0$	Tracer mixing for near-neutral $Ri$	0.37	(0, 2)
$\mathbb{C}_c^\infty$	Tracer mixing for high $Ri$	0.098	(0, 2)
$\mathbb{C}_c^p$	Tracer free entrainment scale	0.11	(0, 2)
$\mathbb{C}_u^h$	Momentum free convection scale	3.71	(0, 8)
$\mathbb{C}_u^-$	Velocity mixing for negative $Ri$	0.37	(0, 2)
$\mathbb{C}_u^0$	Velocity mixing for near-neutral $Ri$	0.36	(0, 2)
$\mathbb{C}_u^\infty$	Velocity mixing for high $Ri$	0.24	(0, 2)
$\mathbb{C}_e^h$	TKE free convection scale	3.64	(0, 10)
$\mathbb{C}_e^-$	TKE transport for negative $Ri$	1.44	(0, 10)
$\mathbb{C}_e^0$	TKE transport for near-neutral $Ri$	7.86	(0, 10)
$\mathbb{C}_e^\infty$	TKE transport for high $Ri$	0.55	(0, 10)
$\mathbb{C}_D^h$	Dissipation free convection scale	3.25	(0, 10)
$\mathbb{C}_D^-$	Dissipation scale for negative $Ri$	0.92	(0, 10)
$\mathbb{C}_D^0$	Dissipation scale for near-neutral $Ri$	1.60	(0, 10)
$\mathbb{C}_D^\infty$	Dissipation scale for high $Ri$	0.58	(0, 10)
$\mathbb{C}_{Ri}^0$	Stability function transitional $Ri$	0.25	(0, 2)
$\mathbb{C}_{Ri}^\delta$	Stability function $Ri$ width	1.02	(0, 2)
$\mathbb{C}^{sp}$	Sheared plume scale	0.50	(0, 2)

**Table 3.** A summary of CATKE’s free parameters. Note that “near-neutral  $Ri$ ” means  $Ri \leq \mathbb{C}_{Ri}^0$ , while “high  $Ri$ ” means  $Ri \geq \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta$ . The bounds limit the values a parameter can take during calibration, using the method described in C3. The prior distributions for each parameter span the range between the bounds.

696    **5 Validation**

697    We next assess CATKE’s ability to make accurate predictions in a single column  
 698    context with the free parameters listed in table 3. First, we derive quantities with well-  
 699    understood physical interpretations from CATKE’s free parameters, and evaluate whether  
 700    their calibrated values are close to values reported in the literature. Second, we compare  
 701    CATKE-parameterized simulations both to the 21 constant-forcing LES used for calibration  
 702    and to an additional 12 constant-forcing LES that are both more strongly and more weakly  
 703    forced than the calibration LES. Third, we conduct a 34-day CATKE-parameterized simula-  
 704    tion of equatorial deep-cycle turbulence using the dataset provided by Whitt et al. (2022),  
 705    and then compare the results to the LES used therein. This third validation context is useful  
 706    because it involves both time-dependent surface forcing, solar insolation, and lateral flux  
 707    divergences derived from a high resolution tropical GCM. Finally, we evaluate CATKE’s  
 708    sensitivity to vertical resolution and time-step size. These all provide a measure of confidence  
 709    in CATKE’s ability to not only represent the LES data used for calibration but also to  
 710    extrapolate to differently-forced conditions, time-dependent surface forcing, and GCM-like  
 711    contexts that include interactions with other parameterizations and lateral flux divergences  
 712    from for example, the advection of momentum, temperature, and salinity. All of this said, we  
 713    maintain a caveat that CATKE should still be assessed, and likely recalibrated, in a regional  
 714    or global context involving lateral fluxes and interactions with other model components.

715    **5.1 Derived quantities**

716    Table 4 shows several quantities that can be derived or computed in terms of CATKE’s  
 717    calibrated free parameters. There is unknown uncertainty in these estimates, so the precise  
 718    values must be taken with a grain of salt. Uncertainty quantification, using the methodology  
 719    proposed by Cleary et al. (2021) for example, is left for future work.

720    **5.1.1 Steady-state Richardson number**

721    Section 3.1.4 shows how a steady-state  $Ri$  may be derived from CATKE’s TKE equation.  
 722    From the parameters in table 3, we find that

$$723 \quad Ri^\dagger \stackrel{\text{def}}{=} \frac{\mathbb{C}_u^0}{\mathbb{C}_c^0 + \mathbb{C}_D^0} \approx 0.18, \quad (52)$$

724    which lies in the “near-neutral” stability function regime, since  $\mathbb{C}_{Ri}^0 = 0.25 > Ri^\dagger$ .  $Ri^\dagger = 0.18$   
 725    is somewhat less than the 0.23 used by Blanke and Delecluse (1993), or the value  $Ri = 1/4$   
 726    that determines the stability of a laminar stratified shear layer. In section 5.3, we find that  
 727     $Ri^\dagger$  is a crucial parameter controlling mixing in forced stably-stratified turbulence, and that  
 728    LES tend to exhibit  $Ri$  in the range 0.2–0.23.

729    **5.1.2 Near-surface similarity constant**

730    Section 3.1.3 shows how a near-surface similarity constant — analogous to the von  
 731    Kármán constant for turbulence near rigid non-wavy walls — may be computed from the  
 732    near-wall and momentum stability function parameters. From table 3 and equation (35) we  
 733    find that

$$734 \quad \kappa = \mathbb{C}^s \left[ (\mathbb{C}_u^0)^3 / \mathbb{C}_D^0 \right]^{1/4} \approx 0.47, \quad (53)$$

735    which is slightly higher than the rigid-wall von Kármán constant value of 0.4. A slightly  
 736    higher similarity constant is consistent with the notion that surface waves act to increase  
 737    the coherence of turbulent motions, which increases mixing lengths and suppresses turbulent  
 738    kinetic energy dissipation.

739    A similar wave-induced enhancement to the similarity constant is proposed by Samelson  
 740    (2022). However, Samelson (2022) models the enhancement as a function of wind at ten

741 meters height,  $u_{10}$ . In our case, the LES are forced with varying  $u_{10}$ , but constant Langmuir  
 742 number  $La \approx 0.3$  (see table 1 for a summary of the LES cases). Thus we must either  
 743 hypothesize that surface waves can be modeled with a  $La$ -dependent enhancement of  $\kappa$ , or  
 744 that CATKE is missing physics. We are unable to proceed further in determining wave-  
 745 induced enhancements to  $\kappa$  without LES that vary  $La$ , so we save such considerations for  
 746 future work.

### 747 5.1.3 The turbulent Prandtl number

748 The turbulent Prandtl number is defined as

$$749 \quad Pr \stackrel{\text{def}}{=} \frac{K_u}{K_c}, \quad (54)$$

750 which is derived for CATKE in section 3.1.1. For various regimes of turbulence we obtain

- 751 •  $Pr_c \approx 0.77$  for weakly-sheared convection,
- 752 •  $Pr_- \approx 0.65$  for unstably-stratified shear turbulence,
- 753 •  $Pr_0 \approx 0.98$  for near-neutral shear turbulence,
- 754 •  $Pr_\infty \approx 2.46$  for strongly-stratified shear turbulence.

755 A turbulent  $Pr$  that increases from less than unity to above unity as  $Ri$  crosses zero is  
 756 consistent with laboratory and DNS studies (for example, D. Li, 2019), as well as typical  
 757 two-equation models (for example, Burchard & Bolding, 2001). On the other hand, one-  
 758 equation models (Blanke & Delecluse, 1993; Madec et al., 2017) often prescribe  $Pr$  to a  
 759 value of 10 or higher as  $Ri$  tends to infinity. It is unlikely that our boundary layer LES are  
 760 informative for such high  $Ri$  mixing, so more LES are needed to assess and perhaps refine  
 761 CATKE’s stability function to capture very high  $Ri$  regimes.

### 762 5.1.4 The turbulent Schmidt number

763 Calibration determines that  $Sc = 0.26$  for unstably-stratified shear turbulence with  
 764  $Ri < 0$ , and then varies between  $0.046 < Sc < 0.44$  as  $Ri$  increases from 0 to  $\mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta$ . As a  
 765 result, TKE is transported much more rapidly than momentum or tracers in shear-dominated  
 766 turbulence, and similarly to momentum or tracers in convective or weakly-sheared stratified  
 767 turbulence. Rapid TKE diffusion relative to momentum or tracer diffusion introduces an  
 768 “implicitly non-local” element to CATKE’s mixing predictions, because TKE transport can  
 769 generate mixing in a region that is displaced from the region of TKE generation.

### 770 5.1.5 Stratified turbulence mixing coefficient

771 The “mixing coefficient” — the ratio between buoyancy flux and dissipation in stably-  
 772 stratified turbulence (Gregg et al., 2018; C.-c. P. Caulfield, 2020) — measures the relative  
 773 level of TKE converted to potential energy in the process of mixing buoyancy vs TKE  
 774 dissipation. Using (19) and assuming stably-stratified turbulence far from boundaries such  
 775 that  $\ell_c = \mathbb{S}_c \sqrt{e}/N$ ,  $\ell_D = \sqrt{e}/(\mathbb{S}_D N)$ , and  $K_c = \mathbb{S}_c e/N$ , we find that

$$776 \quad \Gamma \stackrel{\text{def}}{=} -\frac{\text{buoyancy flux}}{\text{dissipation}} = \frac{\mathbb{S}_c}{\mathbb{S}_D}. \quad (55)$$

777 The free parameters in table 3 imply that the mixing coefficient  $\Gamma$  varies between  $\Gamma_0 \approx 0.26$   
 778 for near-neutral turbulence and  $\Gamma_\infty \approx 0.17$  for strongly-stratified (shear-free) turbulence.  
 779 The latter is applicable to internal wave breaking, where an extensive literature suggests  
 780 that  $\Gamma_\infty \approx 0.2$  (Gregg et al., 2018).

Symbol	Value	Description
$Ri^\dagger$	0.18	Steady-state gradient Richardson number
$\kappa$	0.47	Near-neutral near-surface similarity constant
$Pr_0$	0.98	Near-neutral turbulent Prandtl number ( $Ri \rightarrow 0$ )
$Pr_\infty$	2.46	Strongly-stratified turbulent Prandtl number ( $Ri \rightarrow \infty$ )
$Pr_-$	0.65	Unstably-stratified shear turbulence Prandtl number ( $Ri < 0$ )
$Pr_c$	0.77	Free convection turbulent Prandtl number ( $Ri \rightarrow -\infty$ )
$\Gamma_0$	0.23	Near-neutral mixing coefficient ( $Ri \rightarrow 0$ )
$\Gamma_\infty$	0.17	Strongly-stratified mixing coefficient ( $Ri \rightarrow \infty$ )
$Sc_0$	0.046	Near-neutral turbulent TKE Schmidt number ( $Ri \rightarrow 0$ )
$Sc_\infty$	0.44	Strongly-stratified turbulent TKE Schmidt number ( $Ri \rightarrow \infty$ )
$Sc_-$	0.26	Unstably-stratified shear turbulence TKE Schmidt number ( $Ri < 0$ )
$Sc_c$	1.02	Free convection turbulent TKE Schmidt number ( $Ri \rightarrow -\infty$ )

**Table 4.** A summary of parameters and non-dimensional numbers derived from CATKE’s calibrated free parameters.

## 781            5.2 Validation against constant-forcing LES and comparison with other 782            parameterizations

783            In this section, we validate CATKE’s ability to make predictions both within and outside  
784            the range of surface forcings used for calibration. To add context to this validation exercise  
785            and connect with other studies, we include a comparison with predictions from the  $K$ -profile  
786            parameterization (KPP; Large et al., 1994), and the “Langmuir turbulence” second-moment  
787            closure (SMC-LT) described by Harcourt (2015), whose results depend additionally on the  
788            Stokes drift profile we used for LES. All simulations, including those with KPP and SMC-LT,  
789            use staggered vertical grids with 128 cells, in a 256-meter deep domain with 2-meter vertical  
790            resolution. We use a 2-minute time step for CATKE and KPP, and a 1-second time-step for  
791            SMC-LT. Such a short time-step was used for SMC-LT because we observed that the results  
792            were sensitive to time steps 20 seconds and longer for the strong forcing cases.

793            We should treat these comparisons with some caution: KPP or SMC-LT were calibrated  
794            to different datasets than what we use for CATKE. Moreover, uncertainty in the accuracy of  
795            LES profiles near the surface — where CATKE, KPP, and SMC-LT often exhibit significant  
796            discrepancies — prevent firm conclusions about near-surface biases. That said, we find by  
797            manual inspection that for every constant-forcing case, CATKE predicts boundary layer  
798            depth simulated by LES — both inside and outside the training dataset — more accurately  
799            than either KPP or SMC-LT. This is an important result because boundary layer depth is  
800            a key metric determining the short-term sensitivity of climate predictions (Gregory, 2000;  
801            Held et al., 2010). With this broad summary of CATKE’s main successes stated, we focus  
802            the subsequent discussion for each case on CATKE’s biases and areas to focus on for future  
803            improvements.

### 804            5.2.1 Constant forcing validation: free convection

805            We begin with the free convection cases plotted in figure 5. The free convection cases  
806            represent some of the best predictions of KPP and SMC-LT. Boundary layer depth is  
807            well-predicted by all parameterizations to within 10 meters, with perhaps the greatest bias

808 coming from SMC-LT in the weakly-forced 72-hour case — despite that KPP has known  
 809 structural biases for representing free convection (Souza et al., 2020). A large portion of  
 810 the KPP profiles are stably-stratified within the boundary layer in our most strongly-forced  
 811 convective cases. This bias, which is a known issue with KPP (see section 8.6.3 in Griffies,  
 812 2015), is particularly prominent in the cases we consider due to the strength of our forcing  
 813 and the weakness of our underlying stratification. Of the three, CATKE exhibits the most  
 814 well-mixed boundary layers under very strong forcing due to its convective mixing length.

815 For near-surface buoyancy (and equivalently sea surface temperature, or SST) the three  
 816 parameterizations make different predictions. For example, CATKE predicts a warmer  
 817 SST because of its near-neutral boundary layer profile. On the other hand KPP, SMC-LT,  
 818 and the LES all exhibit layers of unstable stratification next to the surface, and thereby  
 819 also predict substantially colder SST than CATKE. Such upper boundary layer structure  
 820 sensitively depends on a description of how mixing is suppressed (or not) close to the  
 821 ocean surface. Unfortunately, we are unsure how far to trust the LES results, which  
 822 may be missing important processes associated with wave breaking or unresolved wave-  
 823 turbulence interactions. Addressing near-surface uncertainties in the LES data, and thereby  
 824 coming to stronger conclusions about the relative fidelity of CATKE, KPP, and SMC-LT,  
 825 requires observations of near-surface boundary layer structure to either validate or motivate  
 826 improvements to the LES. We leave this for future work.

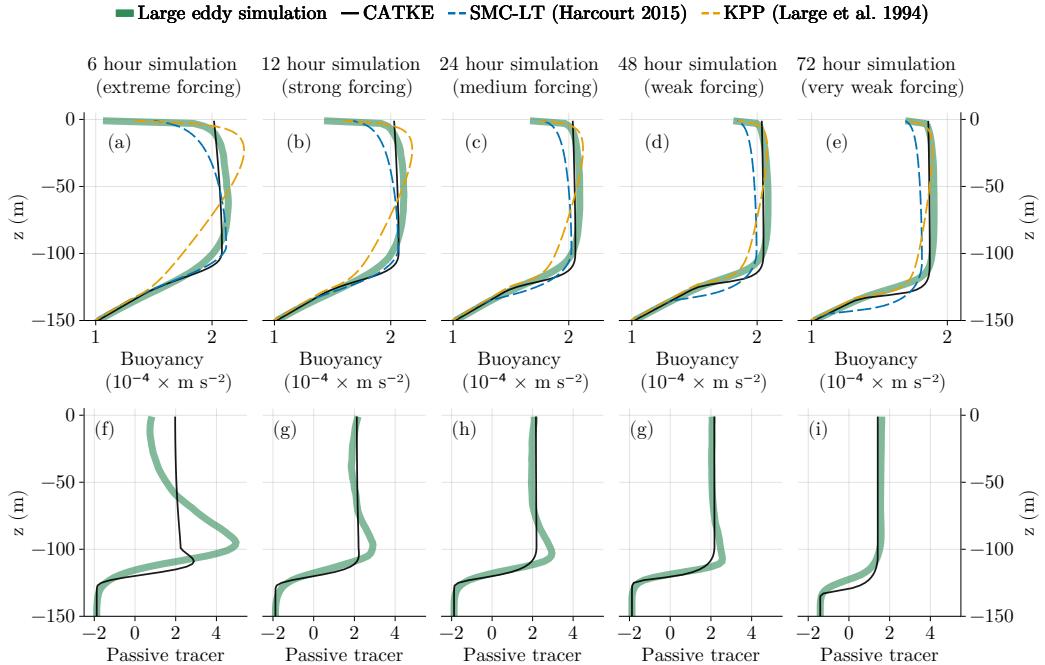
827 The buoyancy profiles in figure 5 reveal bias in CATKE’s predictions of the detailed  
 828 structure of the lower half of the convecting boundary layer. One contribution to this bias  
 829 is well-known: in free convection, buoyancy fluxes in the lower half of the boundary layer  
 830 are upgradient. In order to accurately capture the boundary layer depth, CATKE must  
 831 accurately predict the buoyancy flux — and therefore cannot avoid erroneously predicting  
 832 a slightly unstably stratified buoyancy profile where in the LES the profile is either nearly  
 833 mixed or actually slightly stably stratified. No amount of calibration or additional free  
 834 parameters can fix this bias given CATKE’s downgradient formulation. The only solution is  
 835 to introduce a non-downgradient, non-local contribution to CATKE’s fluxes. For example,  
 836 CATKE could be augmented with a mass flux scheme in the manner of Siebesma et al. (2007);  
 837 Giordani et al. (2020). Other alternatives include evolving fluxes directly as in Garanaik et  
 838 al. (2024), or adding prognostic tracer variances (Legay et al., 2024).

839 To investigate CATKE’s free convection bias further, figure 5 compares CATKE’s  
 840 predictions of the forced passive tracer profile with LES. This comparison reveals that while  
 841 CATKE generally models the tracer profile well (except for the extreme, extrapolating, 6-hour  
 842 case in panel a), CATKE tends to overmix especially in the lower part of the boundary layer,  
 843 where the LES tracer profiles exhibit a slight peak and stronger gradients. Thus in addition  
 844 to lacking a non-local contribution to fluxes, CATKE also overpredicts mixing to some degree,  
 845 especially near the base of the boundary layer. The overprediction of mixing may be related  
 846 to an overprediction of the tracer mixing length exhibited by figure 4. Addressing this bias  
 847 could motivate adding non-local contributions to convective fluxes as well as modifying the  
 848 depth structure of the convective mixing length.

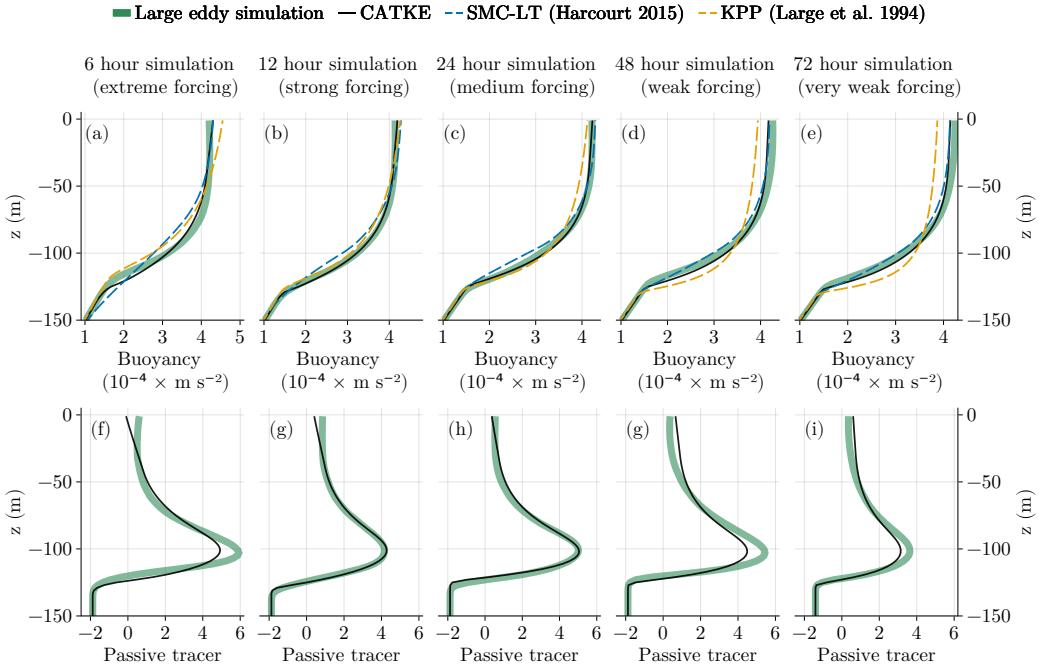
### 849 5.2.2 Constant forcing validation: shear-driven turbulence

850 We next turn to pure shear- or wind-driven turbulence. We have two such cases, one  
 851 without rotation and thus representing near-equatorial mixing, and a second with a Coriolis  
 852 parameter of  $f = 10^{-4} \text{ s}^{-1}$  corresponding to a latitude of about 43°N. The wind forcing that  
 853 would produce the momentum flux applied to the strong wind, no rotation cases spans from  
 854 9–22  $\text{m s}^{-1}$ . The wind forcing in the strong wind (and rotating) cases spans 15–24  $\text{m s}^{-1}$ .

855 A comparison between LES, SMC-LT, KPP, and CATKE for the strong wind, no  
 856 rotation case is shown in figure 6. All parameterizations make similar and good predictions  
 857 for boundary layer depth and surface temperature, except for SMC-LT in the 6-hour case,  
 858 where it overmixes slightly. A comparison between CATKE and LES simulations of the



**Figure 5.** A four-way comparison for the “free convection” constant forcing cases described in 1 and Appendix A between LES, CATKE, the  $K$ -profile parameterization (KPP Large et al., 1994), and the Langmuir turbulence second moment closure described by Harcourt (2015) (SMC-LT). Both KPP and SMC-LT are implemented in the General Ocean Turbulence Model (GOTM, Burchard et al., 2006; Q. Li et al., 2019, 2021). Panel (a)–(e) show comparisons for free convection with forcing of decreasing strength corresponding to the 6-, 12-, 24-, 48-, and 72-hour suites, respectively. The free convection cases have no wind forcing and destabilizing buoyancy fluxes that correspond, roughly, to heat fluxes between 181 and 2000  $\text{W m}^{-2}$ . The initial condition is density stratified with a depth-varying buoyancy gradient that varies between  $10^{-6} \text{ s}^{-2}$  and  $2 \times 10^{-5} \text{ s}^{-2}$ . The passive tracer forcing, which is described in appendix A2, is a Gaussian centered on  $z = -96 \text{ m}$  and 8 m wide. The strength of the forcing depends on the suite: the 6-, 12-, 24-, 48-, and 72-hour suites use 15 minute, 30 minute, 1 hour, 2 hour, and 4 hour forcing time scales, respectively.



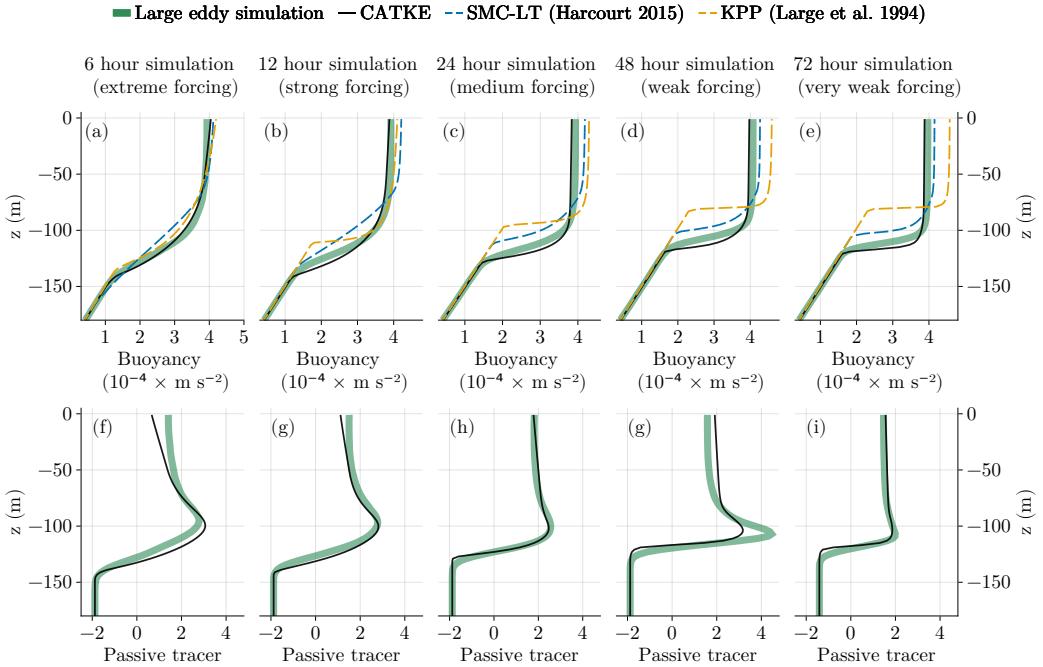
**Figure 6.** A comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “strong wind, no rotation” constant forcing cases described in table 1 and Appendix A. Surface stresses correspond to 9–22  $\text{m s}^{-1}$  10-meter winds. See figure 5.

forced passive tracer for the strong wind, no rotation case is shown in figure 6, revealing that CATKE fares far better for this case than for free convection, and more specifically exhibits a slight tendency to overmix near the base of the boundary layer and to undermix near the surface.

The strong wind case with rotation plotted in figure 7 proves more challenging for CATKE and extremely challenging for SMC-LT and KPP. For all forcing strength, SMC-LT and KPP exhibit serious shallow bias and warm SST bias. CATKE simulations, on the other hand, are better but still exhibit a tendency to overmix slightly, resulting in boundary layers that are approximately 5% too deep. Figure 7 compares CATKE and LES predictions of the forced passive tracer for the strong wind case, corroborating the “overmixing bias” especially for the 6- and 48-hour suites, while additionally revealing undermixing near the surface.

### 5.2.3 Constant forcing validation: mixed shear and convective turbulence

CATKE simulations are also more accurate than KPP or SMC-LT for cases involving both wind and destabilizing buoyancy forcing, which produces a mixed regime of turbulence with both shear and buoyant production of TKE. We have three mixed cases comprising a total of 15 LES with both wind and buoyancy forcing: strong wind, weak cooling, medium wind, weak cooling, and weak wind, strong cooling. Results for these 15 cases are shown in figures 8, 9, and 10. KPP exhibits significant shallow bias for all cases. SMC-LT exhibits less shallow bias than KPP, but still more than CATKE. CATKE’s worst performance is in the weak wind, strong cooling cases where it overmixes. Finally, we note that CATKE, KPP, and SMC-LT all make different predictions for the near-surface buoyancy gradient and sea surface temperature than the present LES. Given the uncertainty in the near-surface LES profiles, we leave interpretation of these discrepancies for future work.



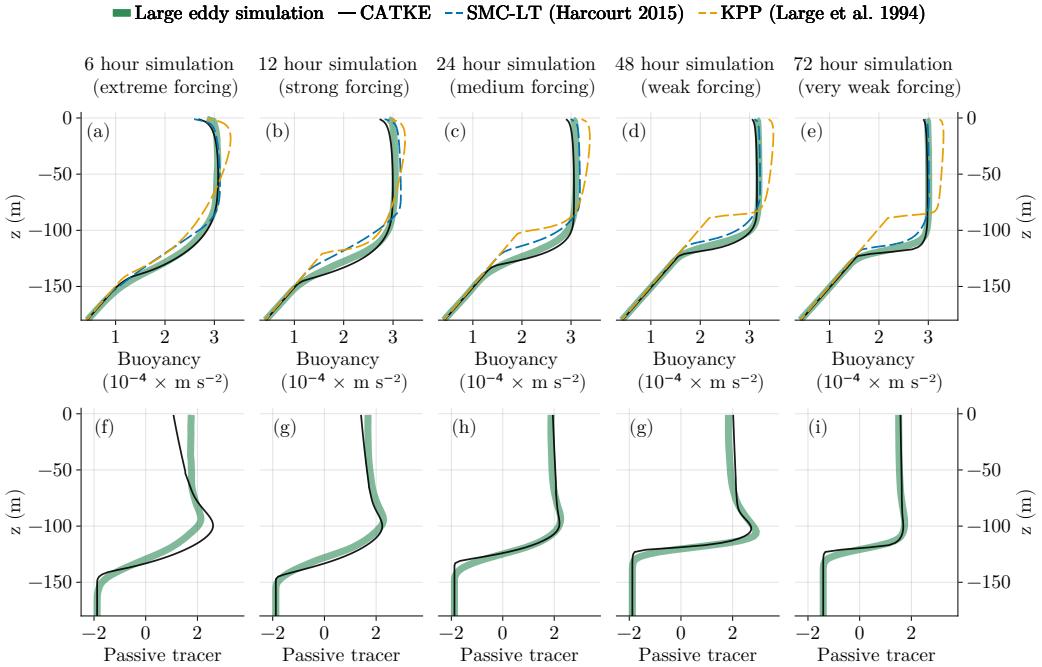
**Figure 7.** A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “strong wind” constant forcing cases described in table 1 and Appendix A. The Coriolis parameter is  $f = 10^{-4} \text{ s}^{-1}$  and surface stresses correspond to  $15\text{--}24 \text{ m s}^{-1}$  10-meter winds. See figure 5.

Figures 8, 9, and 10 also compare CATKE and LES predictions of the forced passive tracer for strong wind, weak cooling, mid wind mid cooling, and weak wind weak cooling cases. The most bias is exhibited in the weak wind strong cooling case, where it tends to overmix as exhibits in both the boundary layer depth in figure 8 and the tracer profiles in figure 8. This shows that the most difficult cases are free convection and “weak wind, strong cooling” — the cases where convective dynamics dominate.

In the “weak winds, strong cooling” case, the 72-hour LES is forced by  $156 \text{ W m}^{-2}$  equivalent heat flux and  $11 \text{ m s}^{-1}$  10-meter atmospheric winds, while the 6-hour LES is forced by  $1666 \text{ W m}^{-2}$  and  $16 \text{ m s}^{-1}$  10-meter winds. In the 6- and 12-hour cases, KPP exhibits a similar “stable stratification bias” as seen in free convection in figure 5. SMC-LT exhibits a shallow bias for the strongly forced cases and a deep biased for the weakly forced cases (and quite accurate predictions for the 24-hour case). CATKE also predicts a too-sharp entrainment layer that is much thinner than the broad entrainment layer observed in the LES in the 6- and 12-hour weak winds, strong cooling cases. These simulations are farthest from quasi-equilibrium in time and may exhibit strong non-locality. Despite CATKE’s errors for the 6-hour case, however, CATKE’s boundary layer depth predictions for the 24-, 48-, and 72-hour case are accurate.

#### 5.2.4 Constant forcing validation: summary

CATKE exhibits less bias than either KPP or SMC-LT across all cases, even when making predictions “outside” its training dataset. In particular, CATKE generates good predictions of boundary layer structure and depth, even in convective dominated cases where an analysis of tracer profiles suggests that CATKE tends to overmix. Fixing CATKE’s



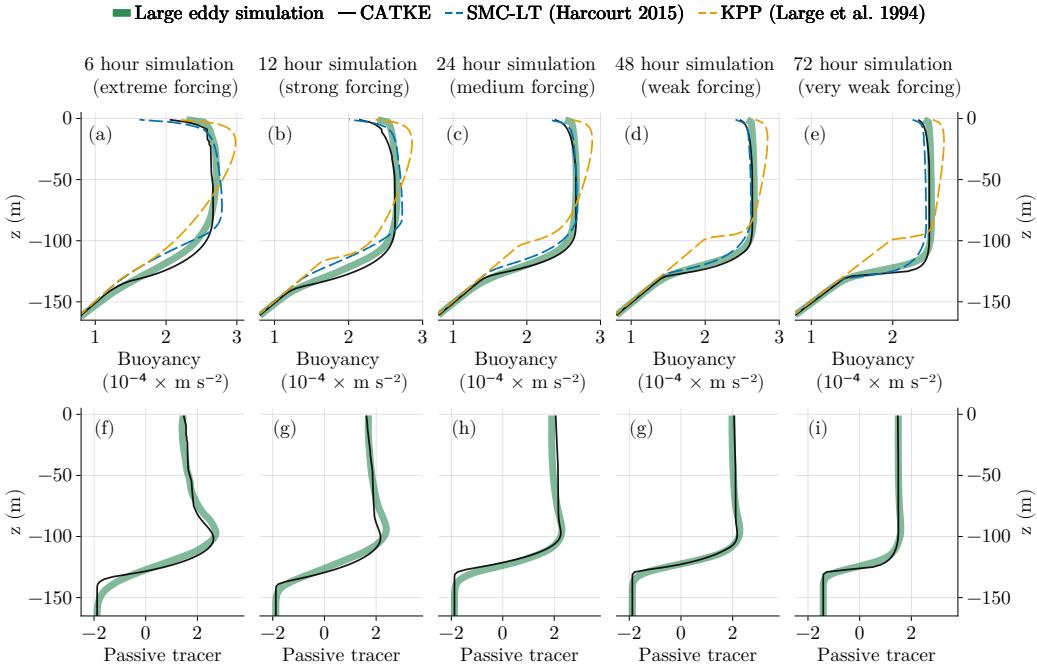
**Figure 8.** A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “strong wind, weak cooling” constant forcing cases described in table 1 and Appendix A. The Coriolis parameter is  $f = 10^{-4} \text{ s}^{-1}$ , surface stresses correspond to  $14\text{--}23 \text{ m s}^{-1}$  10-meter winds, and surface cooling ranges from  $79\text{--}833 \text{ W m}^{-2}$ . See figure 5.

convection biases will likely require additional work with both the convective mixing length, and CATKE’s stability function formulation for  $Ri < 0$ .

CATKE makes good predictions relative to KPP or SMC-LT in part because its formulation expresses reasonable physical hypotheses, but also because its parameters have been calibrated comprehensively to minimize bias across a wide range of physical scenarios and vertical resolutions. In particular, the simulations that CATKE has been trained on are more similar to the extrapolation test cases (the 6- and 72-hour cases) than the datasets that either KPP or SMC-LT have been trained on. This generates ambiguity: do KPP and SMC-LT exhibit greater bias because of structural issues with their formulation, or do they need to be recalibrated in a similar manner as CATKE? Answers prove elusive. While KPP has known structural biases (see, for example, Souza et al., 2020), the formulation of SMC-LT is seemingly more general than CATKE. Further understanding requires calibrating KPP and SMC-LT in the same way we calibrate CATKE.

### 5.3 Deep cycle turbulence in the tropics

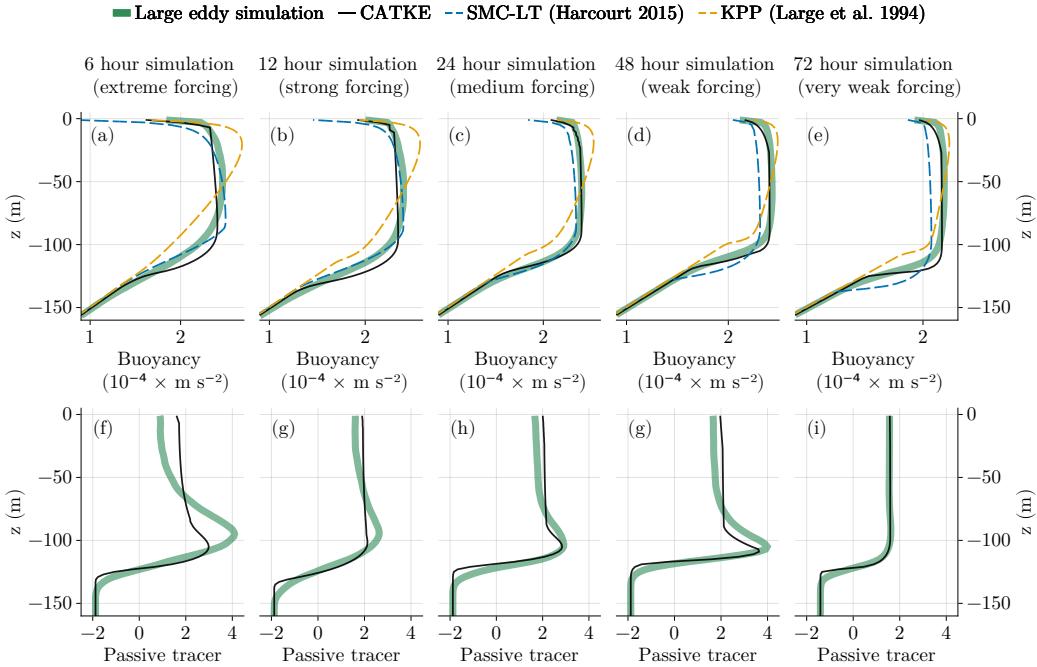
We turn to a validation case that requires significant extrapolation outside of the constant-forcing dataset: 34 days of deep cycle turbulence in the tropics forced by time-varying winds, surface heat fluxes, and surface freshwater fluxes, as well as lateral flux divergences derived from a regional ocean model. The scenario and LES that we use to validate the single column model simulations are described by Whitt et al. (2022). A comparison between the same LES and two other turbulence closures is also given by Reichl et al. (2024).



**Figure 9.** A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “mid wind, mid cooling” constant forcing cases described in table 1 and Appendix A. The Coriolis parameter is  $f = 10^{-4} \text{ s}^{-1}$ , surface stresses correspond to  $13\text{--}20 \text{ m s}^{-1}$  10-meter winds, and surface cooling ranges from  $125\text{--}1333 \text{ W m}^{-2}$ . See figure 5.

Figure 11 illustrates the complex dynamics of this tropical turbulence situation by showing vertical kinetic energy from the LES, TKE from CATKE, and  $Ri$  from days 8 to 13 of the time-series. A combination of wind stress and stabilizing solar insolation in daytime produces a shallow, stably-stratified jet in the upper  $\sim 10$  meters of the water column. As day turns to night, outgoing radiation starts to dominate the incoming solar insolation to reduce and eventually destabilize the upper part of the water column, producing turbulent mixing driven by a combination of convective buoyancy flux and shear. Momentum is thereby mixed downwards and injected into the stably stratified region below the base of the boundary layer. Remarkably, because the region below the boundary layer is close to marginally stable (Smyth & Moum, 2013), this nocturnal injection of momentum from above eventually leads to shear instability that spans the entire, roughly 100 m depth of the region below the mixed layer. More often than not, the turbulence “pulses” — initial bursts of turbulence mix momentum and buoyancy, decay, and precipitate a second and even a third burst of turbulence later on the evening (Smyth et al., 2017). The process, which is called “deep cycle turbulence”, repeats itself the next day.

The slow growth and intermittent bursting of turbulence at night is prominent in LES vertical kinetic energy shown in figure 11a. Figure 11b shows that CATKE exhibits a qualitatively similar bursting behavior, though the timing of the bursts are sometimes misrepresented. Moreover, inspection of figures 11c and d reveals that CATKE underpredicts the Richardson number,  $Ri$ . (Panel d also shows that CATKE exhibits regions of negative  $Ri$  below  $z = -70$  m which are absent from the LES. This deep unstable stratification, which can only be produced by the GCM-derived lateral flux divergences, also presents with other parameterizations, such as in the  $k\text{-}\epsilon$  solutions that underpin figure 13c and figure 14c. We are unsure why the lateral fluxes produce negative  $Ri$ , but do not investigate this issue



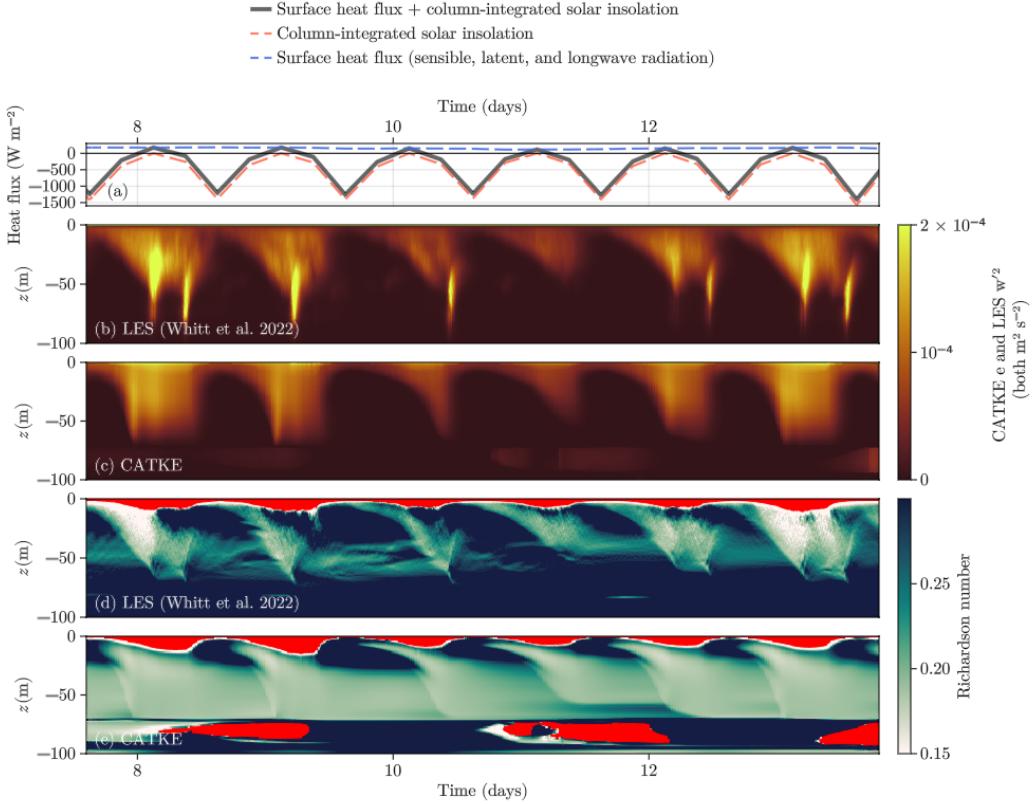
**Figure 10.** A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “weak wind, strong cooling” constant forcing cases described in table 1 and Appendix A. The Coriolis parameter is  $f = 10^{-4} \text{ s}^{-1}$ , surface stresses correspond to  $11\text{--}16 \text{ m s}^{-1}$  10-meter winds, and surface cooling ranges from  $156\text{--}1666 \text{ W m}^{-2}$ . See figure 5.

further here. Finally, we note that this issue is relatively less prominent outside days 8–13 within the total 34 day time-series.)

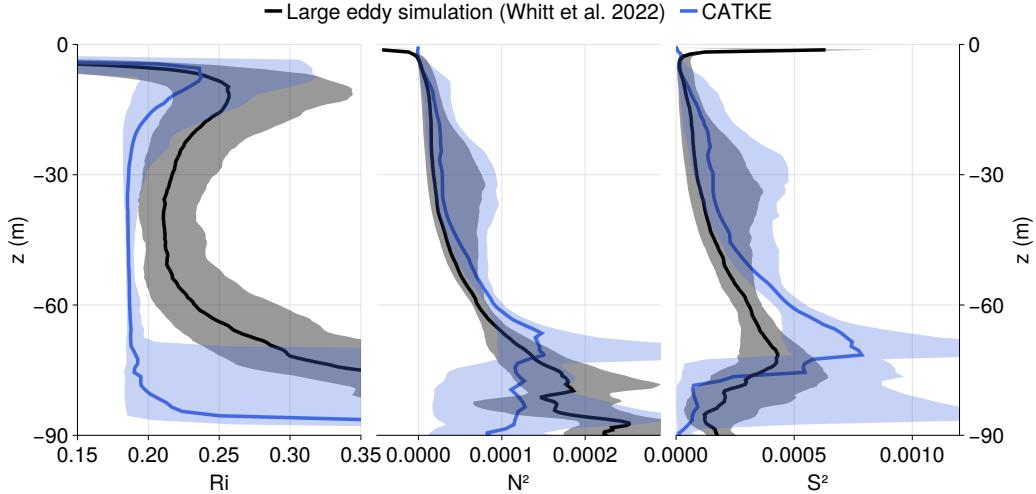
Figure 12 investigates the discrepancy between LES-derived and CATKE-based  $Ri$  further by plotting the median  $Ri$ ,  $N^2$ , and  $S^2$  and shading the range of values between the 25% and 75% quantiles. The  $Ri$  statistics in the left panel show that while the LES  $Ri$  is relatively variable with a broad peak around  $Ri \approx 0.21$ , CATKE’s  $Ri$  are narrowly concentrated around its steady state value 0.18. Turning to  $N^2$  (middle panel) and  $S^2$  (right panel), we see that the  $Ri$  bias is not straightforwardly associated with a bias in either  $N^2$  or  $S^2$  — both are slightly overpredicted (indicating undermixing), but nevertheless exhibit similar medians and ranges compared to the LES.

Despite the errors in burst timing and Richardson number, CATKE’s predictions have realistic qualities not shared by other closures. To show this, figures 13 and 14 compare the vertical temperature flux, and the time-derivative of the vertical temperature flux between the LES, CATKE, and the  $k$ - $\epsilon$  parameterization implemented in Oceananigans (Umlauf & Burchard, 2005; Ramadhan et al., 2020).  $k$ - $\epsilon$  is similar to SMC-LT except that, like the LES described by Whitt et al. (2022), it neglects surface wave effects. Note that the LES data has been smoothed with a moving average to reduce noise, which is especially distracting when computing the time derivative of the vertical flux.

In figure 13, which shows the period between days 8–13, both the LES and CATKE vertical fluxes exhibit vertically-coherent bursts, whereas  $k$ - $\epsilon$ ’s flux predictions are smoother and smeared out over the deep turbulence cycle. The vertical coherence of vertical flux maxima is even more pronounced in the time-derivative of the vertical fluxes plotted in panels d–f. Figure 14 shows the same data between days 28–34, during which the three solutions



**Figure 11.** Overview of the tropical turbulence validation case. Panels show: (a) surface heat fluxes and solar insolation, (b) vertical kinetic energy  $\overline{w'^2}$  from the LES described by Whitt et al. (2022), (c) CATKE’s TKE variable, (d) the Richardson number computed from the horizontally-averaged LES momentum and buoyancy profiles, and (e) the Richardson number predicted by CATKE. The shaded red areas in panels (d) and (e) indicate a negative Richardson number. Shown here are days 8–13 out of the entire 34-day time-series. The heat fluxes are negative during the day (heat going downwards, into the ocean) and positive at night (heat going up, out of the ocean). The LES vertical kinetic energy and CATKE turbulent kinetic energy exhibit intermittent bursting. In the deep region below the boundary layer where turbulent bursting occurs, LES-derived Richardson numbers get as low as 0.15. In the CATKE solution and in the same region, the Richardson number reaches a minimum of about 0.18.



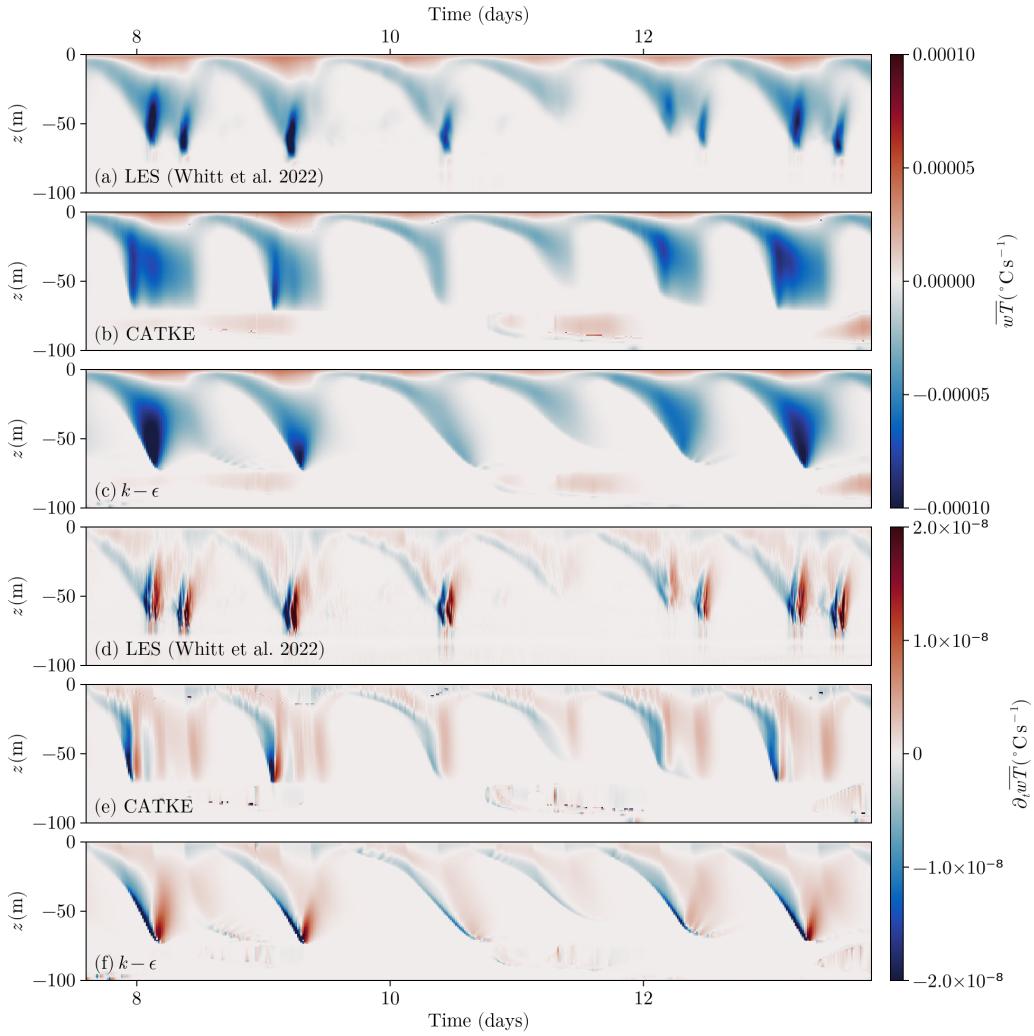
**Figure 12.** Median  $Ri = N^2/S^2$  (left panel), and buoyancy frequency  $N^2$  (middle), and shear  $S^2$  (right panel) at each depth computed from 34 days of realistic equatorial turbulence simulated by LES and CATKE. The LES  $Ri$  is computed in terms of the horizontally-averaged shear and buoyancy. Shading shows the range between the 25% and 75% quantiles. CATKE’s prediction of  $Ri$  is smaller and more narrowly distributed around its steady-state Richardson number  $Ri^\dagger = 0.18$  than the LES  $Ri$ . On the other hand, CATKE overpredicts both  $N^2$  and  $S^2$ , thus undermixing both momentum and buoyancy (with more momentum bias than buoyancy bias).

are more qualitatively distinguished. In particular, the time-derivative of the  $k-\epsilon$  fluxes shown in panel f of figure 14 exhibit sharp, progressively deepening interfaces and generally lack vertically-coherent features. Neither the LES (panel d) or CATKE solutions (panel e) possess these interfaces and instead exhibit vertically-coherent features. Despite their qualitative similarity to LES, however, the CATKE solutions misrepresent the magnitude and timing of the vertically-coherent bursts. Improving both CATKE and  $k-\epsilon$  will probably benefit from including time-dependent LES data with deep-cycle turbulence physics in a future calibration exercise.

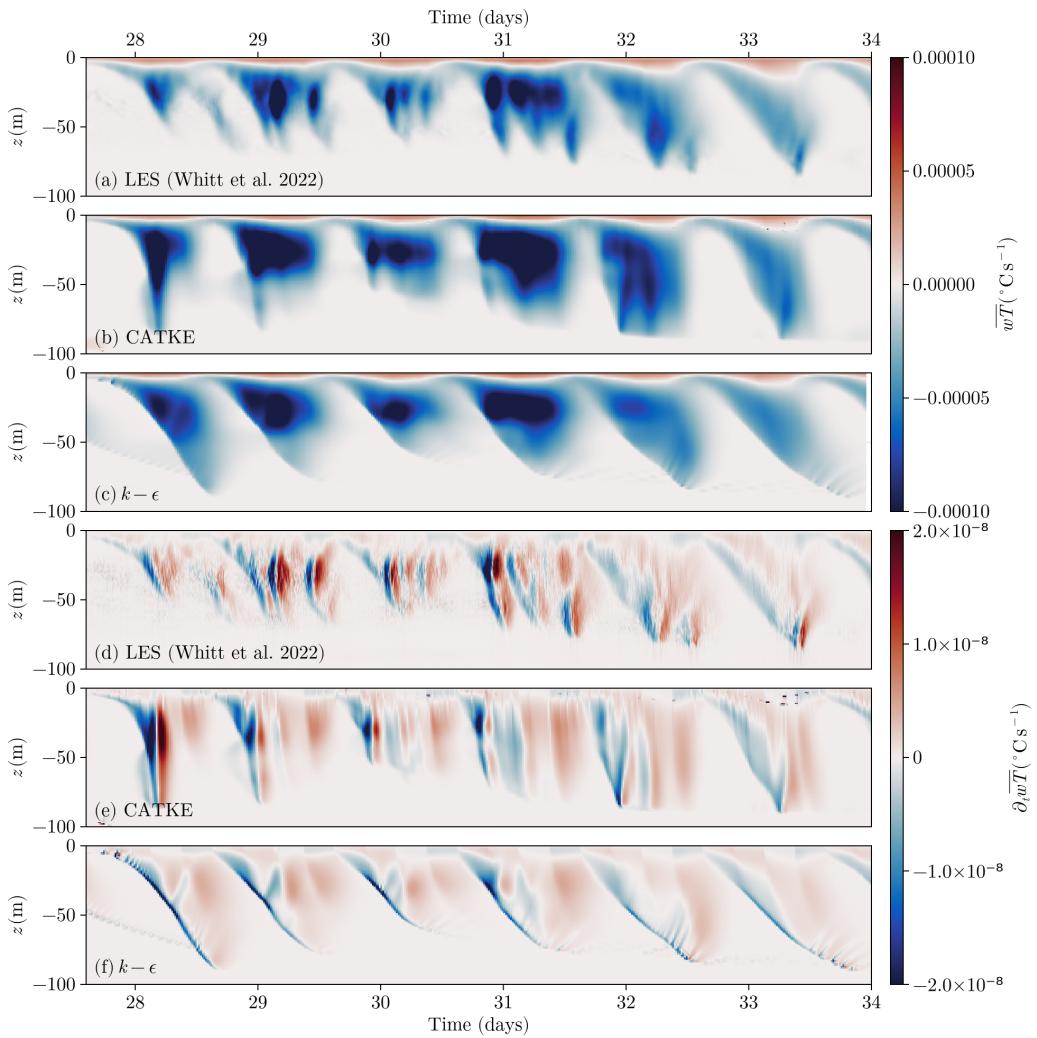
#### 5.4 Sensitivity to vertical resolution and time-step

Next we investigate the sensitivity of CATKE’s predictions to numerical parameters like vertical resolution and time-step size — a well-appreciated concern with ocean microscale parameterizations (Reffray et al., 2015; Van Roekel et al., 2018). The sensitivity of CATKE’s predictions to vertical resolutions ranging from 1 to 16 meters is shown in figure 15 for the weak wind, strong cooling case (the case for which CATKE exhibits the most bias). Recall that CATKE was calibrated using simulations with 2-, 4-, and 8-meter vertical resolution, such that 1 and 16 meters represent extrapolation in resolution. Based off the results in figure 15, we conclude that CATKE is insensitive to vertical resolutions 8 meters and finer. At 16 meter resolution, CATKE’s predictions are still better than KPP and SMC-LT, but nevertheless start to deviate from the higher-resolution CATKE solutions and, in particular, tend to overmix. It may be that with such a coarse resolution, the structure of strongly-stratified entrainment layers at the base of the boundary layer cannot be adequately resolved.

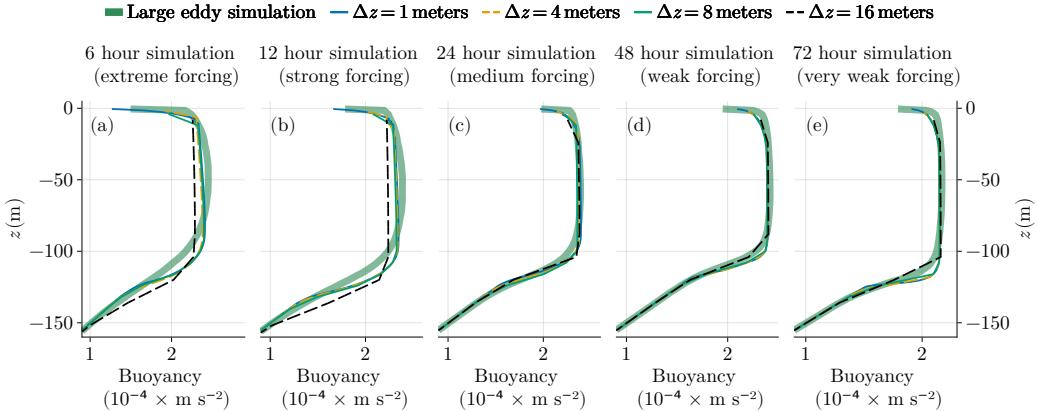
The sensitivity of CATKE’s predictions to time-step size — at a vertical resolution of 1 meter — are shown in figure 16. Note that CATKE requires a smaller time step for finer



**Figure 13.** A comparison of the vertical temperature flux and vertical temperature flux divergence in tropical turbulence between LES (Whitt et al., 2022), CATKE, and the  $k-\epsilon$  two-equation model (Umlauf & Burchard, 2005).



**Figure 14.** Vertical temperature flux and vertical temperature flux divergence as in figure 13, but showing days 28–34.



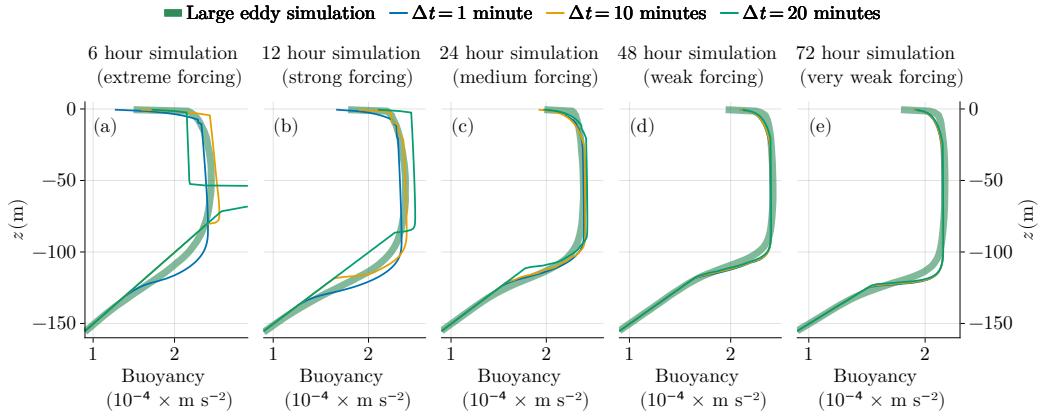
**Figure 15.** Illustration of sensitivity of CATKE predictions to vertical resolution for the weak wind, strong cooling case. Four vertical resolutions are shown: 1, 4, 8, and 16 meters. CATKE’s calibration explicitly minimized errors between LES and CATKE simulations at 2, 4, and 8 meter resolution, such that the 1 and 16 meter cases represent “extrapolation in resolution.” The predictions are converged for resolutions 8 meters and finer, but the 16 meter resolution results exhibit small discrepancies from the converged solutions.

vertical resolution. Put differently, smaller time-steps are required to resolve the evolution of TKE, momentum, and tracers, and associated vertical transmission of information, on finer grids. More strongly forced cases also require smaller time steps. Figure 16, and additional tests, show that with 1 meter vertical resolution, CATKE requires time-steps 2 minutes or shorter to resolve the dynamics associated with surface forcing as strong as that encountered in the 6-hour-suite. (A 5-minute time step is adequately converged for the 12-, 24-, 48-, and 72-hour suite, however.)

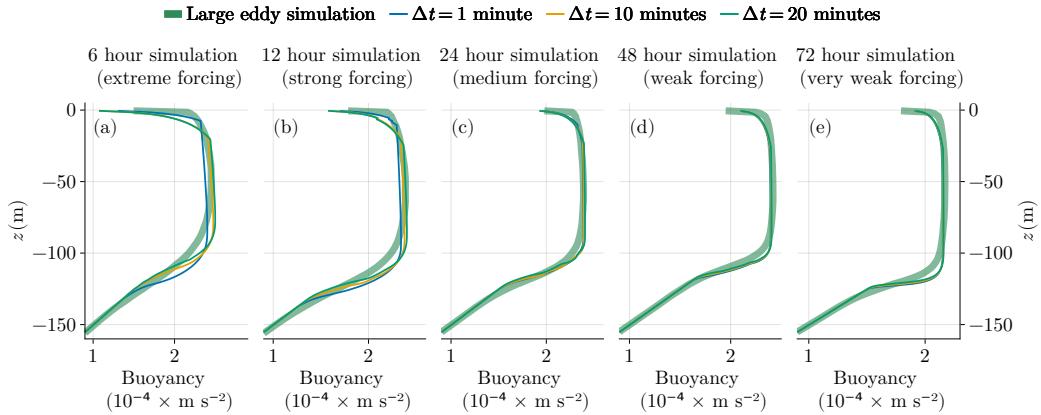
We address this sensitivity of CATKE’s predictions to time-step by implementing a novel split-explicit scheme that substeps the TKE using a short time-steps, while evolving momentum and tracers with a longer time-step. The details are given in Appendix B. The results are shown in figure 17, showing that CATKE generates converged predictions for momentum and tracer time-steps between 1 and 20 minutes when the TKE is substepped with a short 30 second time step. When using substepping, the TKE time-step can be configured according to the vertical resolution and strongest expected forcing over the duration of the simulation, while the momentum and tracer time-steps may be configured by other stability criteria, such as a CFL condition.

## 6 Discussion

This paper describes a novel one-equation parameterization for vertical fluxes by ocean microscale turbulence called “CATKE”. CATKE extends existing one-equation parameterizations (Blanke & Delecluse, 1993; Madec et al., 2017) with a dynamic model for convective adjustment capable of describing the wide range of convective mixing rates observed in the ocean surface boundary layer. CATKE’s 23 free parameters are calibrated against large eddy simulations accounting for discretization errors. We use *a posteriori* calibration, meaning that the CATKE parameters are calibrated to capture the full temporal evolution of the coarse-grained variables rather than, for example, matching the unresolved eddy fluxes. This approach improves both the accuracy and the stability of the calibrated parameterization.



**Figure 16.** Sensitivity of CATKE predictions to time step for 1 meter vertical resolution for the weak wind, strong cooling case. At 1 meter resolution and in the strong forcing conditions of the 12- and 6-hour suites, CATKE solutions show time-step dependence for time steps longer than 1 minute. To enable longer time steps for high vertical resolutions in the presence of strong forcing, the substepping scheme described in Appendix B is used and demonstrated in figure 17.



**Figure 17.** A comparison between LES and CATKE-parameterized single column simulations at 1 meter vertical resolution and three different momentum and tracer time-steps, when turbulent kinetic energy is substepped with a 30 second time step using the scheme described in Appendix B. For the 6-hour suite, the time-step dependence is greatly reduced compared to the non-substepped case shown in figure 16, but is not entirely converged. We suspect this is because even momentum and tracers require a time step shorter than 20 minutes for such strong forcing at high vertical resolution.

Our decision to develop a one-equation TKE-based parameterization rather than a *K*-profile parameterization (KPP, see Large et al., 1994; McWilliams et al., 2009; Van Roekel et al., 2018; Reichl & Hallberg, 2018; Reichl & Li, 2019) merits some discussion. KPPs have a major advantage over TKE-based parameterizations in coarse resolution ocean models (especially with different time-steps for momentum and tracer variables) because they admit time-steps as long as 2 hours (Reichl & Hallberg, 2018). In part, we are interested in one-equation parameterization because our focus is higher resolution, mesoscale-permitting and mesoscale-resolving simulations that require 1–10 minute time-steps to satisfy the advective numerical stability constraints of energetic solutions on relatively high-resolution grids. CATKE adds no additional time step constraints to such simulations, while offering some significant benefits: (*i*) dynamic prediction of diffusivity vertical structure versus prescription via “shape functions”; (*ii*) turbulent intensity growth and relaxation time scales or “memory”, and (*iii*) better computational performance on hardware with fine-grained parallelism such as Graphics Processing Units (GPUs) used for example by Oceananigans (Ramadhan et al., 2020; Silvestri, Wagner, Constantinou, et al., 2024) and Veros (Häfner et al., 2021), which are ill-suited for the nonlinear solvers for boundary layer depth common to KPP-type models (Zhang et al., 2020).

The automated calibration described in section 4 and Appendix C was repeated hundreds of times during the development of CATKE. We developed CATKE by starting with a simple formulation similar to the one described by Blanke and Delecluse (1993) — with no stability functions (and thus a constant Prandtl number) and no special convective mixing length. We then progressed, using calibration to justify increasing model complexity, to the presently described form with continuously *Ri*-dependent stability functions in equation 28 and the convective mixing length described in section 3.1.5. This development process represents a “knowledge discovery loop” (National Academies of Sciences, Engineering, and Medicine and others, 2022) with three steps: (*i*) formulation, (*ii*) calibration, and (*iii*) assessment. For complex, nonlinear models — and even in the relatively simple single column context of this paper — automatic calibration is essential to progress quickly from formulation to assessment, and then to discover and justify further improvements to formulation, thereby iteratively producing a high-quality, well-motivated, parsimonious parameterization.

Our calibration to a relatively limited range of LES cases reported in this paper (though extensive compared prior efforts in ocean turbulence parameterization development) is just the first step towards using CATKE for global ocean modeling and climate projection. In particular, our ultimate objective is more accurate climate predictions with quantified uncertainties. Addressing this ultimate goal requires first quantifying the uncertainty of CATKE’s free parameters relative to LES, using the calibration context presented in this work. Next, with prior parameter distributions in hand, CATKE’s free parameters must then be recalibrated concomitant with other climate model free parameters against global climate observations to account for physics missing from the LES in this work, and to account for interactions between CATKE and other components of the climate model.

A second future step is to further calibrate CATKE to a more comprehensive suite of LES forced with temporally-varying surface fluxes, surface wave fields with  $La \neq 0.3$ , and horizontal flux divergences (for example following Whitt et al., 2022). These calibrations against more comprehensive LES will provide better prior estimates of CATKE’s parameters in preparation of the final goal of calibrating CATKE in a global context. More comprehensive calibration to more LES and to observations in a global context will likely reveal deficiencies to be addressed by further development of CATKE’s formulation, such as accounting for the effect of surface waves on CATKE’s mixing and dissipation length scales.

1070 **Appendix A A synthetic dataset generated by large eddy simulations**

1071 We use a synthetic dataset to calibrate and assess CATKE consisting of 35 idealized  
 1072 large eddy simulations (LES) of the ocean surface boundary layer with imposed constant  
 1073 surface fluxes of temperature and momentum and a simple surface wave field.

1074 **A1 Initial conditions**

1075 The LES are initialized from rest with zero velocity and the piecewise-linear buoyancy  
 1076 stratification

$$1077 b(z, t=0) = \begin{cases} N_1^2 z & \text{for } z > -h_1, \\ N_2^2 z + (N_2^2 - N_1^2) h_1 & \text{for } -h_2 < z < -h_1, \\ N_3^2 z + (N_3^2 - N_2^2) h_2 + (N_2^2 - N_1^2) h_1 & \text{for } z < -h_2, \end{cases} \quad (\text{A1})$$

1078 with  $N_1^2 = N_3^2 = 2 \times 10^{-6} \text{ s}^{-2}$ ,  $N_2^2 = 10^{-5} \text{ s}^{-2}$ ,  $h_1 = 48 \text{ m}$ , and  $h_2 = 72 \text{ m}$ .

1079 **A2 Passive tracer forcing**

1080 We additionally simulate the evolution of a passive tracer  $c$  which is forced by

$$1081 F_c(z) = \omega_+ e^{-(z-z_c)^2/2\lambda_c^2} - \omega_-, \quad (\text{A2})$$

1082 where  $z_c$  is the depth of the forcing,  $\lambda_c$  is the width of the forcing,  $\omega_+$  is an inverse forcing  
 1083 time-scale that varies between each suite, and  $\omega_-$  is chosen so that  $F_c$  has zero mean, that is

$$\omega_- \stackrel{\text{def}}{=} \frac{\omega_+}{L_z} \int_{-L_z}^0 e^{-(z-z_c)^2/2\lambda_c^2} dz \quad (\text{A3})$$

$$\approx \omega_+ \frac{\lambda_c \sqrt{2\pi}}{L_z}, \quad (\text{A4})$$

1084 where  $L_z$  is the depth of the domain. The approximation in (A4) holds when the forcing is  
 1085 far from boundaries, or when  $-L_z \ll z_c - \lambda_c$  and  $0 \gg z_c + \lambda_c$ .

1086 To generate tracer gradients within the boundary layer, we use a relatively narrow  
 1087 forcing profile with  $\lambda_c = 8 \text{ m}$  centered at  $z_c = -96 \text{ m}$ , near the bottom of the boundary layer  
 1088 at the end of each simulation. We additionally use a forcing time scale  $\omega_+^{-1}$  that is similar to  
 1089 the typical mixing time-scale: 15 minutes, 30 minutes, 1 hour, 2 hours, and 4 hours for the  
 1090 6, 12, 24, 48, and 72 hour suites, respectively. These choices ensure a passive tracer profile  
 1091 that, unlike the well-mixed buoyancy profile, reveals the structure of turbulent tracer mixing  
 1092 within the boundary layer. The passive tracer data thus provides an important additional  
 1093 constraint on CATKE's prediction of the tracer mixing length,  $\ell_c$ .

1094 **A3 Constant-flux boundary conditions**

1095 The 35 simulations, which have different boundary conditions and Stokes drift are  
 1096 organized into 5 “suites”, each of which has 7 cases: free convection, weak wind strong  
 1097 cooling, medium wind medium cooling, strong wind weak cooling, strong wind, strong wind  
 1098 no rotation, and strong wind and sunny. The suites differ by both forcing strength and  
 1099 duration, simulating 6, 12, 24, 48, and 72 hours of boundary layer turbulence respectively.  
 1100 The forcing strength is chosen for each suite and case so that the boundary layer deepens  
 1101 to roughly half the depth of the domain; for example, the “6-hour suite” has the strongest  
 1102 forcing, and the “72-hour suite” has the weakest forcing. “Strong wind no rotation” and  
 1103 “strong wind and sunny” use  $f = 0$ , while the rest use the Coriolis parameter  $f = 10^{-4} \text{ s}^{-1}$ .  
 1104 The surface fluxes for the 35 LES are summarized in tables 1 and 2. To draw a connection  
 1105 between the LES suites and real air-sea flux conditions, tables 1 and 2 provide an estimate  
 1106 of heat fluxes  $Q$  for each case, as well as an estimate of the atmospheric wind at 10 meters

height using similarity theory (reduced to the case of neutral buoyancy fluxes for simplicity, see Large and Yeager (2009)),

$$u_{10} = \sqrt{\frac{|\tau_a|}{c_{10}}}, \quad \text{where } c_{10} = \left( \frac{\kappa}{\log(10/\ell_r)} \right)^2, \quad \text{and } \ell_r = 0.011 \frac{|\tau_a|}{g}, \quad (\text{A5})$$

where  $\ell_r$  is the Charnock roughness length given gravitational acceleration  $g = 9.81 \text{ m s}^{-2}$ ,  $\kappa = 0.4$  is the von Kármán constant, and  $\tau_a = \rho_o \tau_x / \rho_a$  is the atmospheric kinematic momentum flux given ocean reference density  $\rho_o = 1024 \text{ kg m}^{-3}$  and atmosphere density  $\rho_a = 1.2 \text{ kg m}^{-3}$ .

#### A4 Stokes drift model

For all wind-forced cases, we additionally impose a surface wave field with a surface Stokes drift amounting to a constant “Langmuir number”  $La = \sqrt{u_*/U^S(z=0)} \approx 0.3$ . Our Stokes drift prescription models a surface wave field with the friction-number-dependent peak wavenumber

$$k_p = C_k \frac{g}{u_*^2}, \quad (\text{A6})$$

where  $u_* = \sqrt{|\tau_x|}$  is the water-side friction velocity,  $g$  is gravitational acceleration, and we use  $C_k = 10^{-6}$ .

We follow Lenain and Pizzo (2020) to estimate the depth-profiles of Stokes drift and Stokes drift shear. The Stokes drift beneath a spectrum of deep-water waves is

$$U^S(z) = 2 \int_{k_p}^{k_i} e^{2kz} k \sqrt{gk} \chi(k) dk, \quad (\text{A7})$$

where  $\chi(k)$  is a one-dimensional wave spectrum that neglects “directional spreading”. The spectrum  $\chi(k)$  is divided into an “equilibrium range” just above the spectral peak  $k_p$ , and a “saturation range” at even higher wavenumbers:

$$\chi(k) = \begin{cases} \frac{C_\beta}{2\sqrt{g}} a_* k^{-5/2} & \text{for } k_p < k < k_n \quad (\text{equilibrium}), \\ C_B k^{-3} & \text{for } k_n < k < k_i \quad (\text{saturation}), \end{cases} \quad (\text{A8})$$

where  $k_n$  is a transition wavenumber between equilibrium and saturation ranges,  $k_i$  is an upper wavenumber cutoff above which waves are assumed to be isotropic and there do not contribute to Stokes drift.  $a_* = u_* \sqrt{\rho_o / \rho_a}$  is the air-side friction velocity defined in terms of the water-side friction velocity  $u_*$ , a reference air density  $\rho_a = 1.2 \text{ kg m}^{-3}$  and ocean density  $\rho_o = 1024 \text{ kg m}^{-3}$ . Wavenumbers *below* the spectral peak  $k_p$  are assumed too weak to contribute appreciably to Stokes drift.

Both the transition wavenumber  $k_n$  and the isotropic wavenumber  $k_i$  decrease with increasing  $u_*$ :

$$k_n \stackrel{\text{def}}{=} C_r g a_*^{-2}, \quad (\text{A9})$$

$$k_i \stackrel{\text{def}}{=} C_i g a_*^{-2}, \quad (\text{A10})$$

where  $C_r = 9.7 \times 10^{-3}$  and  $C_i = 0.072$ .

The Stokes drift is

$$U^S(z) = C_\beta a_* \int_{k_p}^{k_n} \frac{e^{2kz}}{k} dk + 2C_B \sqrt{g} \int_{k_n}^{k_i} k^{-3/2} e^{2kz} dk. \quad (\text{A11})$$

Noting that  $\int_{k_p}^{k_n} k^{-1} e^{2kz} dk = \text{Ei}(2k_n z) - \text{Ei}(2k_p z)$ , where Ei is the exponential integral function, we find

$$U^S(z) = C_\beta a_* [\text{Ei}(2k_n z) - \text{Ei}(2k_p z)] + 2C_B \sqrt{g} [v(k_n) - v(k_i)], \quad (\text{A12})$$

1143 and

$$1144 \quad \partial_z U^S = 2C_\beta a_* \int_{k_p}^{k_n} e^{2kz} dk + 4C_B \sqrt{g} \int_n^I \frac{e^{2kz}}{\sqrt{k}} dk, \quad (\text{A13})$$

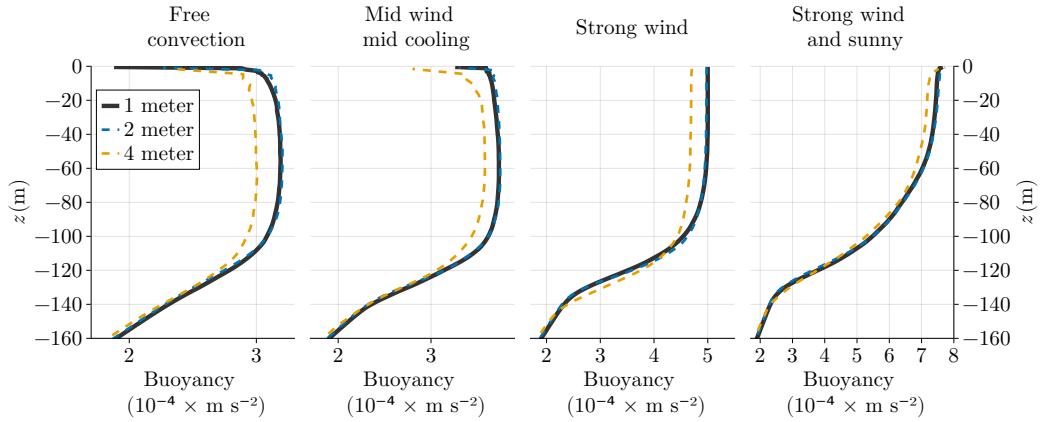
$$1145 \quad = C_\beta a_* \frac{e^{2k_p z} - e^{2k_n z}}{|z|} + 2C_B \sqrt{\frac{2\pi g}{|z|}} \left[ \operatorname{erf}(\sqrt{2k_n |z|}) - \operatorname{erf}(\sqrt{2k_i |z|}) \right], \quad (\text{A14})$$

1146 for the Stokes shear.

## 1147 A5 LES uncertainty: effects of resolution and Stokes drift

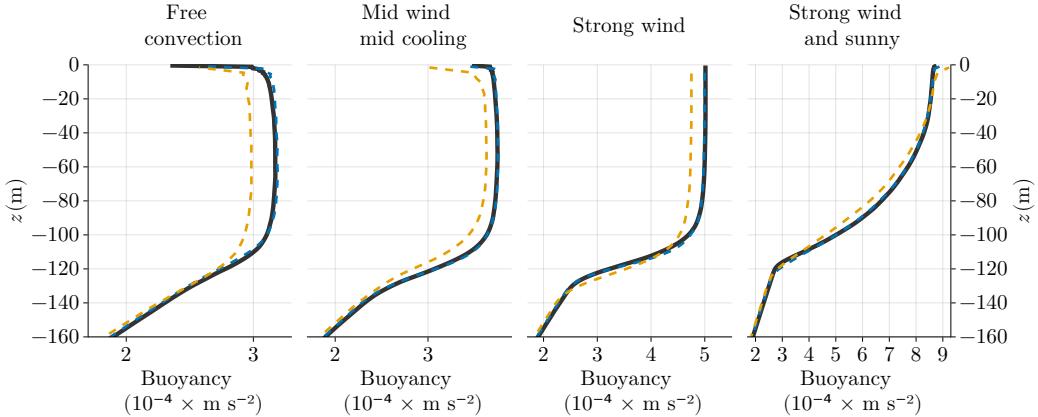
1148 All LES use 2 meter horizontal resolution and a stretched vertical resolution that varies  
 1149 from 0.8 meters in the upper half of domain to 2.3 meters at the bottom. We refer to this  
 1150 as “1 meter” vertical resolution. Our LES utilize an “implicit” model for subgrid fluxes  
 1151 whereby kinetic energy and tracer variance are solely dissipated by a minimally-diffusive  
 1152 9th-order Weighted, Essentially Non-Oscillatory (WENO) advection scheme (Shu, 2020).  
 1153 The advantages of using WENO-based implicit dissipation (and no explicit closure for subgrid  
 1154 turbulent fluxes) are discussed by Pressel et al. (2017) and Silvestri, Wagner, Campin, et al.  
 1155 (2024).

1156 To account for the effects of resolution on the 35 LES used as synthetic observations in  
 1157 this paper, we run 70 additional LES on coarser grids with double (“2 meter”) and quadruple  
 1158 (“4 meter”) resolution, and use these to estimate the observational uncertainty used in  
 1159 calibration (see 4 for more details). The effect of resolution depends on forcing strength: for  
 1160 the 6 and 12 hour suite, the results are nearly identical for 1- and 2-meter vertical resolution.  
 1161 Figure A1 shows the results for 4 cases in the 12 hour suite. Note that in the free convection  
 1162 case, the first two grid points exhibit a strong unstable stratification in the 12 hour suite. We  
 1163 attribute this to an artificial reduction of mixing near the top boundary of the LES. It might  
 1164 be possible to address this artificially-strong unstable mean stratification by introducing, for  
 1165 example, a surface-concentrated eddy diffusivity. However, because the LES are used only  
 1166 for training CATKE and thus matter mostly in their predicted boundary layer depth, we  
 1167 choose instead to ignore the top 4 m when computing the LES–CATKE discrepancy during  
 1168 calibration.



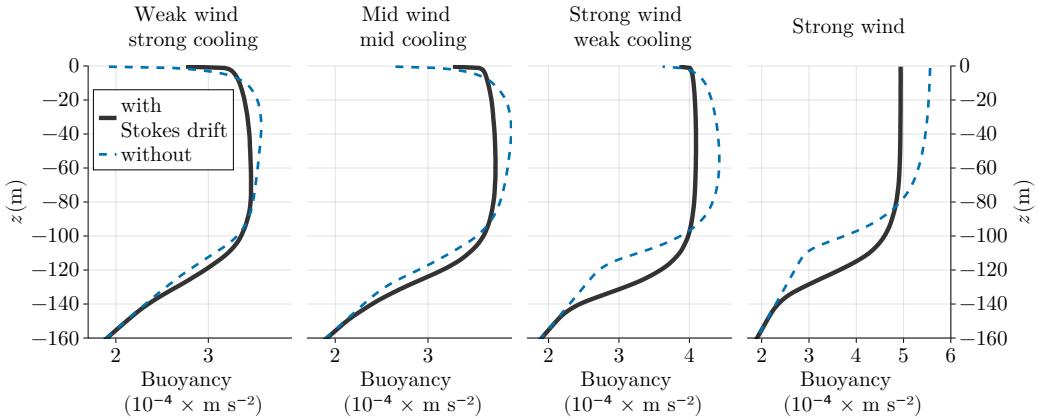
**Figure A1.** Resolution dependence of 12-hour LES.

1169 Figure A2 shows the resolution dependence of the 24-hour suite. These LES show  
 1170 slightly more resolution dependence than the 12-hour suite, especially for cases forced by a  
 1171 combination of wind and cooling. This indicates that our LES data for more weakly forced  
 1172 cases are *less certain* than the strongly forced cases.

**Figure A2.** Resolution dependence of 24-hour LES.

1173      **A6 Effect of Stokes drift on LES results**

1174      Next we turn to the effect that including the Stokes drift profile described in section A4  
 1175      has on our LES results. The inclusion of Stokes drift in our LES is an attempt to make  
 1176      them slightly more realistic. In other words, we hypothesize that calibrating CATKE to LES  
 1177      without surface waves would generally lead to a shallow bias in mixed layer depth prediction  
 1178      with CATKE — since surface waves are always present above real wind-forced ocean surface  
 1179      boundary layers.

**Figure A3.** Stokes drift dependence of 12-hour LES.

1180      This notion is corroborated by figure A3, which shows the horizontally-averaged buoyancy  
 1181      profiles for 4 LES in the 12 hour suite, with and without Stokes drift. As expected, the  
 1182      inclusion of Stokes drift produces more mixing and makes the boundary layer deeper. The  
 1183      effect of Stokes drift is minor in the case of weak and medium winds (leftmost and second  
 1184      from left panels). In the strong wind (and rotating) case, the inclusion of Stokes drift makes  
 1185      the boundary layer 20 meters deeper, or around 20% of the total. In the strong wind, no  
 1186      rotation case, the case without Stokes drift completely fails to transition to the turbulence.  
 1187      (A small amount of cooling would probably be required to produce turbulence in the strong  
 1188      wind, no rotation case without Stokes drift.)

## 1189 Appendix B Split-explicit turbulent kinetic energy time stepping and 1190 vertical discretization

1191 CATKE's time discretization is a little non-trivial since we step forward velocity and  
1192 tracers first, then step forward TKE and also use substepping/split-explicit scheme for TKE.  
1193 In the single column case, we integrate equations (13)–(15) with the backward Euler scheme

$$1194 \frac{u^{n+1} - u^n}{\Delta t} = \partial_z (K_u^n \partial_z u^{n+1}) + fv^n + \bar{F}_u^n, \quad (\text{B1})$$

$$1195 \frac{v^{n+1} - v^n}{\Delta t} = \partial_z (K_u^n \partial_z v^{n+1}) - fu^n + \bar{F}_v^n, \quad (\text{B2})$$

$$1196 \frac{c^{n+1} - c^n}{\Delta t} = \partial_z (K_c^n \partial_z c^{n+1}) + \bar{F}_c^n, \quad (\text{B3})$$

1197 where  $\Delta t = t^{n+1} - t^n$  and the superscripts  $n$  or  $n + 1$  indicate the time step at which the  
1198 quantity is evaluated. For the TKE equation (19), we introduce a substepping scheme that  
1199 uses  $M$  short time step sizes  $\Delta t/M$  to integrate  $e$  between  $n$  to  $n + 1$ ,

$$1200 \frac{e^{m+1} - e^m}{\Delta t/M} = \underbrace{\partial_z (K_e^m \partial_z e^{m+1})}_{\text{transport}} + \underbrace{K_u^m \frac{1}{2} (\partial_z \mathbf{u}^n + \partial_z \mathbf{u}^{n+1}) \cdot \partial_z \mathbf{u}^{n+1}}_{\text{shear production}} + \underbrace{\overline{w'b'}^m - \frac{\sqrt{e^m}}{\ell_D^m} e^{m+1}}_{\text{dissipation}}, \quad (\text{B4})$$

1201 where the superscripts  $m$  and  $m + 1$  denote the substep level. In practice, when using  
1202 substepping, we fix the time step size for the TKE equation,  $\Delta t_e$ , and compute the substep  
1203 number  $M = \text{ceil}(\Delta t/\Delta t_e)$  in terms of  $\Delta t_e$  and the momentum and tracer time step size,  
1204  $\Delta t$ .

1205 The buoyancy flux  $\overline{w'b'}^m$  in (B4) is discretized in time using the conditionally-implicit  
1206 “Patankar trick” (Burchard, 2002), such that

$$1207 \overline{w'b'}^m = \begin{cases} -K_c^n \partial_z b^{n+1} & \text{when } \partial_z b^{n+1} \leq 0 \\ -K_c^n \partial_z b^{n+1} \frac{e^{m+1}}{e^m} & \text{when } \partial_z b^{n+1} > 0 \end{cases} \quad (\text{B5})$$

1208 which improves the stability of (B4) and keeps  $e$  from becoming too negative due to numerical  
1209 errors associated with, for example, advection schemes with oscillatory errors. Note that  
1210 shear production is not updated during substepping. The time discretization of the shear  
1211 production term in (B4), which incorporates shear measured at the time step  $n$  and  $n + 1$ ,  
1212 also follows Burchard (2002) and requires an algorithm that stores the velocity field at time  
1213 step  $n$ , stepping forward momentum and tracers, and then substepping forward  $e$ .

1214 We spatially-discretize  $u$ ,  $v$ ,  $c$ , and  $e$  on a staggered vertical grid (not shown), with all  
1215 variables vertically located at cell centers — a deviation from Blanke and Delecluse (1993),  
1216 Burchard (2002), or Madec et al. (2017) who place  $u$ ,  $v$ ,  $c$  at vertical cell centers but TKE at  
1217 vertical cell interfaces where the diffusivity is computed (otherwise known as “ $w$ -locations”).  
1218 Because  $K_c$ ,  $K_c$ , and  $K_e$  are located at vertical cell interfaces, this discretization means  
1219 that  $e$  must be reconstructed from cell centers to cell interfaces to compute  $K_u$ ,  $K_c$ , and  
1220  $K_e$  according to (12). The vertical spatial discretization of the shear production term is  
1221 derived from the mean kinetic energy equation following Burchard (2002), but adapted to  
1222 our cell-centered placement of  $e$ . We use a tridiagonal solve to advance  $u$ ,  $v$ ,  $c$ ,  $e$  in (B1)–(B4)  
1223 over each time step of substep, treating both diffusion and linear terms in (B4) implicitly.

## 1224 Appendix C A posteriori calibration

1225 We use Ensemble Kalman Inversion (EKI; Iglesias et al., 2013) to calibrate CATKE.  
1226 EKI is a gradient-free and computationally inexpensive method for solving nonlinear inverse  
1227 problems. EKI supposes that a forward map  $\mathcal{G}(\mathbb{C})$  can predict uncertain observations  $\mathcal{Y}$   
1228 given a set of free parameters  $\mathbb{C}$ ,

$$1229 \mathcal{Y} = \mathcal{G}(\mathbb{C}) + \eta, \quad (\text{C1})$$

1230 where  $\eta \sim \mathcal{N}(0, \mathcal{M})$  is normally-distributed random uncertainty with covariance  $\mathcal{M}$ . Four  
1231 objects appear in the model-data relation (C1),

- 1232 1. *Observations*  $\mathcal{Y}$  with  $Q$  discrete elements  $\mathcal{Y}_q$ . In this paper, each  $\mathcal{Y}_q$  represents a state  
1233 variable like velocity  $U$  or buoyancy  $B$  at a particular depth and time, computed from  
1234 LES data by horizontal averaging and vertical coarse-graining, and then normalized  
1235 and shifted to have zero mean and unit variance.
- 1236 2. A *parameter set*  $\mathbb{C}$  with  $P$  free parameter values  $\mathbb{C}_p$ .
- 1237 3. A *forward map*  $\mathcal{G}(\mathbb{C})$  whose elements  $\mathcal{G}_q(\mathbb{C})$  predict the observation  $\mathcal{Y}_q$ .  $\mathcal{G}(\mathbb{C})$  rep-  
1238 presents a *model* that maps a parameter set  $\mathbb{C}$  to the space of observations  $\mathcal{Y}$ . In  
1239 our case, constructing  $\mathcal{G}(\mathbb{C})$  requires forward evaluations of 63 single column models  
1240 parameterized by  $\mathbb{C}$ , each predicting the evolution of horizontally-averaged variables  
1241 in 21 LES at 2-, 4-, and 8-meter resolution.
- 1242 4. Random Gaussian *uncertainty*  $\eta \sim \mathcal{N}(0, \mathcal{M})$  with covariance  $\mathcal{M}$  associated with both  
1243  $\mathcal{G}_q(\mathbb{C})$  and  $\mathcal{Y}_q$ .  $\eta$  conflates uncertainty in  $\mathcal{Y}$  with “structural” uncertainty associated  
1244 with imperfect forward maps  $\mathcal{G}$ .

1245 The elements of  $\mathcal{Y}$  are the discrete values of the horizontally-averaged temperature  
1246 and velocity fields output from 21 LES coarse-grained to three grids with uniform 2-, 4-,  
1247 and 8-meter spacing. Each physical field is shifted, normalized, and weighted before being  
1248 assembled into  $\mathcal{Y}$ . Each forward map  $G(\mathbb{C})$  involves  $3 \times 21 = 63$  simulations to find  $U$ ,  $V$ ,  
1249 and  $B$  profiles for each LES case at the three model vertical resolutions.

## 1250 C1 Ensemble Kalman dynamics

1251 Ensemble Kalman Inversion may be interpreted as a dynamical system that governs the  
1252 evolution of an ensemble of  $E$  parameter sets, or “particles”,  $\mathbf{C} \stackrel{\text{def}}{=} [\mathbb{C}^1, \mathbb{C}^2, \dots, \mathbb{C}^E]$ . Here  
1253 the superscript  $\alpha$  denotes the “particle index”, which varies across the ensemble:  $\mathbb{C}_p^\alpha$  is the  
1254  $p^{\text{th}}$  parameter value of the  $\alpha^{\text{th}}$  particle.

1255 Each parameter set  $\mathbb{C}^\alpha$  obeys the ordinary differential equation

$$1256 \frac{d}{dT} \mathbb{C}^\alpha = -\mathcal{K}(\mathbf{C}, \mathbf{G}) \mathcal{M}^{-1} (\mathcal{G}^\alpha - \mathcal{Y}), \quad (\text{C2})$$

1257 where  $\mathcal{G}^\alpha \stackrel{\text{def}}{=} \mathcal{G}(\mathbb{C}^\alpha)$  is the forward map computed with the parameter set  $\mathbb{C}^\alpha$ , and  $T$  is  
1258 the “pseudotime”. The matrix  $\mathcal{K}(\mathbf{C}, \mathbf{G})$  in (C2) is the covariance matrix estimated from  
1259 ensemble statistics at pseudotime  $T$ , thus coupling the evolution of the ensemble. For two  
1260 “ensemble matrices”  $\mathbf{A}$  and  $\mathbf{B}$ , where  $\mathbf{A}$  for example is constructed from an ensemble of  
1261 vectors  $[A_i^1, A_i^2, \dots, A_i^E]$ , the elements  $\mathcal{K}_{ij}(\mathbf{A}, \mathbf{B})$  are defined

$$1262 \mathcal{K}_{ij}(\mathbf{A}, \mathbf{B}) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\alpha=1}^N (A_i^\alpha - \langle A \rangle_i)(B_j^\alpha - \langle B \rangle_j), \quad \text{with} \quad \langle C \rangle_i \stackrel{\text{def}}{=} \frac{1}{E} \sum_{\alpha=1}^E C_i^\alpha. \quad (\text{C3})$$

1263 For nearly-linear maps  $\mathcal{G}_q(\mathbb{C}) \approx H_{pq} \mathbb{C}_p$ , (C2) reduces to

$$1264 \frac{d}{dT} \mathbb{C}^\alpha \approx -\mathcal{K}(\mathbf{C}, \mathbf{C}) \nabla_{\mathbb{C}} \Phi^\alpha, \quad (\text{C4})$$

1265 where  $\mathcal{K}_{ij}(\mathbf{C}, \mathbf{C})$  is the  $P \times P$  parameter-parameter covariance matrix (Kovachki & Stuart,  
1266 2019). The “EKI objective”  $\Phi^\alpha$  associated with parameter set  $\alpha$  appears in (C4), where

$$1267 \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}) \stackrel{\text{def}}{=} \|\mathcal{M}^{-1/2} [\mathcal{G}(\mathbb{C}) - \mathcal{Y}] \|^2, \quad (\text{C5})$$

1268 and  $\Phi^\alpha \stackrel{\text{def}}{=} \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}^\alpha)$ .  $\Phi$  in (C5) is a functional of  $\mathcal{G}$  that measures the uncertain discrepancy  
1269 between  $\mathcal{G}(\mathbb{C}) - \mathcal{Y}$ . The system (C4) minimizes  $\Phi$  using gradient descent preconditioned  
1270 with  $\mathcal{K}(\mathbf{C}, \mathbf{C})$ , where the gradients  $\nabla_{\mathbb{C}} \Phi$  are estimated from the parameter ensemble.

We integrate the EKI dynamical system (C2) in using a forward Euler discretization,

$$\mathbb{C}^\alpha|_{\nu+1} = \mathbb{C}^\alpha|_\nu - \Delta\mathcal{T} \left[ \mathcal{K}(\mathbf{C}, \mathbf{G}) \mathcal{M}^{-1} (\mathcal{G}^\alpha - \mathcal{Y}) \right]_\nu, \quad (\text{C6})$$

where  $\nu$  is the pseudotime iteration and  $\Delta\mathcal{T}$  is a pseudotime step size. The adaptive step size  $\Delta\mathcal{T}$  is chosen at each iteration according to Kovachki and Stuart (2019). The initial parameter sets  $\mathbb{C}^\alpha$  at  $\mathcal{T} = 0$  are generated by randomly sampling the priors listed in table 3.

EKI is practical for two reasons: (i) it does not require explicit gradients of  $\mathcal{G}$  with respect to parameters  $\mathbb{C}$ , and (ii) the forward map evaluations  $\mathcal{G}^\alpha$  — the most expensive part of integrating (C2) — are independent and thus easily parallelized. Reason (i) means EKI is applicable to any simulation framework with changeable parameters  $\mathbb{C}$ . Reason (ii) means that considerable yet distributed resources can be leveraged efficiently: given sufficient distributed resources, the cost of a single EKI iteration depends only on the cost of a single forward map evaluation, independent of ensemble size. This parallelizability benefits small problems such as calibration in a single column context.

## C2 Uncertainty covariance

We associate the uncertainty  $\mathcal{M}$  with the numerical fidelity of the large eddy simulations by defining

$$\mathcal{M} = \text{cov}([\mathcal{Y}^{1m} \mathcal{Y}^{2m} \mathcal{Y}^{4m}]), \quad (\text{C7})$$

where  $\mathcal{Y}^{1m}, \mathcal{Y}^{2m}, \mathcal{Y}^{4m}$  denote observations obtained from LES with 1-, 2-, and 4-meter vertical resolution.

## C3 Constrained and unconstrained parameters

The dynamics (C6) require normally-distributed parameters  $\mathbb{C}_p$ , which precludes the imposition of strict bounds such as non-negativity. We therefore introduce the forward and inverse transforms,

$$\mathbb{C}_p = \log \frac{b - \tilde{\mathbb{C}}_p}{\tilde{\mathbb{C}}_p - a} \quad \text{and} \quad \tilde{\mathbb{C}}_p = a + \frac{b - a}{1 + \exp(\mathbb{C}_p)}, \quad (\text{C8})$$

between “constrained” physical parameters  $\tilde{\mathbb{C}}$  that are bounded between  $(a, b)$ , and unconstrained parameters  $\mathbb{C}$ . The transformation (C8) implies that if  $\mathbb{C}_p$  is normally-distributed then  $\tilde{\mathbb{C}}$  is bounded by  $(a, b)$  with a scaled, shifted logit-normal distribution.

We denote the scaled, shifted logit-normal distribution bounded by  $(a, b)$  as  $\mathcal{B}(a, b)$  and use it to model the distribution of all of CATKE’s free parameters. The distributions  $\mathcal{B}(a, b)$  formulated so their corresponding normal distributions have zero mean and unit variance. When integrating (C6), the normally-distributed parameter sets  $\mathbb{C}^\alpha$  are transformed into their physical space counterparts  $\tilde{\mathbb{C}}^\alpha$  via (C8) when evaluating  $\mathcal{G}^\alpha = \mathcal{G}(\mathbb{C}^\alpha)$  and thus solving the single column equations (13)–(15) and (19).

## C4 Failure criterion handling

Poor parameter choices  $\mathbb{C}^\alpha$  often lead to failed simulations of the single column system (13)–(15) and (19). In that case the forward map  $\mathcal{G}^\alpha$  is not informative and must be ignored when performing the Euler step (C6).

We first define the median and the “median absolute deviation” of the EKI objective samples,  $\Phi^\alpha \stackrel{\text{def}}{=} \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}^\alpha)$ ,

$$\tilde{\Phi} \stackrel{\text{def}}{=} \text{median}(\Phi^\alpha) \quad \text{and} \quad s \stackrel{\text{def}}{=} \text{median}(|\Phi^\alpha - \tilde{\Phi}|), \quad (\text{C9})$$

We mark a particle  $\alpha$  as “failed” if

$$\Phi^\alpha > \tilde{\Phi} + 3s. \quad (\text{C10})$$

1313 This excludes both non-finite and just “particularly anomalous”  $\Phi^\alpha$ .

## 1314 Open Research Section

1315 This work relied on the open-source software LESbrary.jl (Wagner et al., 2023) and  
 1316 Oceananigans.jl (Ramadhan et al., 2020) to run the LES, Oceananigans.jl to run calibration  
 1317 simulations, and ParameterEstimocean.jl (Wagner et al., 2022) and EnsembleKalmanPro-  
 1318 cesses.jl (Dunbar et al., 2022) for the Ensemble Kalman Inversion. Visualizations were made  
 1319 using Makie.jl (Danisch & Krumbiegel, 2021). Scripts for performing the calibration are avail-  
 1320 able at the GitHub repository [github.com/glwagner/SingleColumnModelCalibration.jl](https://github.com/glwagner/SingleColumnModelCalibration.jl).

## 1321 Acknowledgments

1322 This project is supported by Schmidt Sciences, LLC and by the National Science Foundation  
 1323 grant AGS-1835576. N.C.C. is additionally supported by the Australian Research Council  
 1324 under DECRA Fellowship DE210100749 and the Center of Excellence for the Weather of  
 1325 the 21st Century CE230100012. Without implying their endorsement, we would also like  
 1326 to acknowledge useful conversations with Bruno Deremble, Alex Legay, Julien LeSommer,  
 1327 Qing Li, Brodie Pearson, Brandon Reichl, Roger Samelson, Luke Van Roekel, Dan Whitt,  
 1328 and Bill Young. Lastly, we like to acknowledge the three anonymous reviewers and Stephen  
 1329 Griffies for their constructive input that greatly improved the manuscript.

## 1330 References

- 1331 Belcher, S. E., Grant, A. L., Hanley, K. E., Fox-Kemper, B., Van Roekel, L., Sullivan, P. P.,  
 1332 ... others (2012). A global perspective on Langmuir turbulence in the ocean surface  
 1333 boundary layer. *Geophysical Research Letters*, 39(18). doi: 10.1029/2012GL052932
- 1334 Blanke, B., & Delecluse, P. (1993). Variability of the tropical Atlantic Ocean simulated  
 1335 by a general circulation model with two different mixed-layer physics. *Journal of  
 1336 Physical Oceanography*, 23(7), 1363–1388. doi: 10.1175/1520-0485(1993)023<1363:  
 1337 VOTTAO>2.0.CO;2
- 1338 Boccaletti, G., Ferrari, R., & Fox-Kemper, B. (2007). Mixed layer instabilities and restratifi-  
 1339 cation. *Journal of Physical Oceanography*, 37(9), 2228–2250.
- 1340 Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network  
 1341 parametrization trained by coarse-graining. *Journal of Advances in Modeling Earth  
 1342 Systems*, 11(8), 2728–2744.
- 1343 Burchard, H. (2002). Energy-conserving discretisation of turbulent shear and buoyancy  
 1344 production. *Ocean Modelling*, 4(3-4), 347–361. doi: 10.1016/S1463-5003(02)00009-4
- 1345 Burchard, H., & Bolding, K. (2001). Comparative analysis of four second-moment turbulence  
 1346 closure models for the oceanic mixed layer. *Journal of Physical Oceanography*, 31(8),  
 1347 1943–1968. doi: 10.1175/1520-0485(2001)031<1943:CAOFSM>2.0.CO;2
- 1348 Burchard, H., Bolding, K., Kühn, W., Meister, A., Neumann, T., & Umlauf, L. (2006).  
 1349 Description of a flexible and extendable physical–biogeochemical model system for the  
 1350 water column. *Journal of Marine systems*, 61(3-4), 180–211.
- 1351 Canuto, V. M., Howard, A., Cheng, Y., & Dubovikov, M. (2001). Ocean turbulence.  
 1352 Part I: One-point closure model—momentum and heat vertical diffusivities. *Journal  
 1353 of Physical Oceanography*, 31(6), 1413–1426.
- 1354 Caulfield, C. (2021). Layering, instabilities, and mixing in turbulent stratified flows. *Annual  
 1355 Review of Fluid Mechanics*, 53(1), 113–145.
- 1356 Caulfield, C.-c. P. (2020). Open questions in turbulent stratified mixing: do we even  
 1357 know what we do not know? *Physical Review Fluids*, 5(11), 110518. doi: 10.1103/  
 1358 PhysRevFluids.5.110518
- 1359 Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., & Stuart, A. M. (2021). Calibrate,  
 1360 emulate, sample. *Journal of Computational Physics*, 424, 109716. doi: 10.1016/  
 1361 j.jcp.2020.109716

- Craik, A. D. D., & Leibovich, S. (1976). A rational model for Langmuir circulations. *Journal of Fluid Mechanics*, 73(3), 401–426. doi: 10.1017/S0022112076001420
- Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia. *Journal of Open Source Software*, 6(65), 3349. doi: 10.21105/joss.03349
- 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. doi: 10.1175/1520-0469(1970)027<1211:CVATSF>2.0.CO;2
- Dunbar, O. R., Lopez-Gomez, I., Garbuno-Iñigo, A., Huang, D. Z., Bach, E., & Wu, J. (2022). EnsembleKalmanProcesses.jl: Derivative-free ensemble-based model calibration. *Journal of Open Source Software*, 7(80), 4869. doi: 10.21105/joss.04869
- Duraisamy, K. (2021). Perspectives on machine learning-augmented Reynolds-averaged and large eddy simulation models of turbulence. *Physical Review Fluids*, 6(5), 050504.
- DuVivier, A. K., Large, W. G., & Small, R. J. (2018). Argo observations of the deep mixing band in the Southern Ocean: A salinity modeling challenge. *Journal of Geophysical Research: Oceans*, 123(10), 7599–7617.
- Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., & Lguensat, R. (2022). A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems*, 14(11).
- 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.
- Gardiner, C. W. (2021). *Elements of stochastic methods*. AIP Publishing Melville, NY, USA.
- 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: Oceans*, 95(C9), 16179–16193.
- Giordani, H., Bourdallé-Badie, R., & Madec, G. (2020). An eddy-diffusivity mass-flux parameterization for modeling oceanic convection. *Journal of Advances in Modeling Earth Systems*, 12(9), e2020MS002078.
- Gregg, M. C., D'Asaro, E. A., Riley, J. J., & Kunze, E. (2018). Mixing efficiency in the ocean. *Annual Review of Marine Science*, 10, 443–473.
- Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dynamics*, 16, 501–515.
- Griffies, L. M. A. A. J. D. G. H. R. W. J. D. L. W. R. T., S. M. (2015). *Theory and numerics of the community ocean vertical mixing (cvmix) project* (Tech. Rep.).
- Groeskamp, S., Griffies, S. M., Iudicone, D., Marsh, R., Nurser, A. G., & Zika, J. D. (2019). The water mass transformation framework for ocean physics and biogeochemistry. *Annual review of marine science*, 11, 271–305.
- Gutjahr, O., Brüggemann, N., Haak, H., Jungclaus, J. H., Putrasahan, D. A., Lohmann, K., & von Storch, J.-S. (2021). Comparison of ocean vertical mixing schemes in the Max Planck Institute Earth System Model (MPI-ESM1.2). *Geoscientific Model Development*, 14(5), 2317–2349.
- Häfner, D., Nuterman, R., & Jochum, M. (2021). Fast, cheap, and turbulent—global ocean modeling with GPU acceleration in python. *Journal of Advances in Modeling Earth Systems*, 13(12), e2021MS002717.
- Harcourt, R. R. (2015). An improved second-moment closure model of Langmuir turbulence. *Journal of Physical Oceanography*, 45(1), 84–103.
- Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, 23(9), 2418–2427.
- Huang, N. E. (1979). On surface drift currents in the ocean. *Journal of Fluid Mechanics*, 91(1), 191–208.
- Iglesias, M. A., Law, K. J., & Stuart, A. M. (2013). Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4), 045001.
- Jungclaus, J. H., Lorenz, S. J., Schmidt, H., Brovkin, V., Brüggemann, N., Chegini, F., ...

- others (2022). The ICON earth system model version 1.0. *Journal of Advances in Modeling Earth Systems*, 14(4), e2021MS002813. doi: 10.1029/2021MS002813
- Kantha, L. H., & Clayson, C. A. (1994). An improved mixed layer model for geophysical applications. *Journal of Geophysical Research: Oceans*, 99(C12), 25235–25266.
- Kovachki, N. B., & Stuart, A. M. (2019). Ensemble Kalman inversion: a derivative-free technique for machine learning tasks. *Inverse Problems*, 35(9), 095005.
- Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., ... others (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, 10(11), 2865–2888.
- 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.
- Large, W. G., & Yeager, S. G. (2009). The global climatology of an interannually varying air-sea flux data set. *Climate dynamics*, 33, 341–364.
- Legay, A., Deremble, B., & Burchard, H. (2024). Derivation and implementation of a non-gradient term to improve the oceanic convection representation within the  $k-\epsilon$  parameterization.
- Lenain, L., & Pizzo, N. (2020). The contribution of high-frequency wind-generated surface waves to the Stokes drift. *Journal of Physical Oceanography*, 50(12), 3455–3465.
- Li, D. (2019). Turbulent prandtl number in the atmospheric boundary layer-where are we now? *Atmospheric Research*, 216, 86–105.
- Li, G., & Xie, S.-P. (2014). Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *Journal of Climate*, 27(4), 1765–1780.
- Li, Q., Bruggeman, J., Burchard, H., Klingbeil, K., Umlauf, L., & Bolding, K. (2021). Integrating cvmix into gotm (v6. 0): A consistent framework for testing, comparing, and applying ocean mixing schemes. *Geoscientific Model Development Discussions*, 2021, 1–30.
- Li, Q., Reichl, B. G., Fox-Kemper, B., Adcroft, A. J., Belcher, S. E., Danabasoglu, G., ... others (2019). Comparing ocean surface boundary vertical mixing schemes including Langmuir turbulence. *Journal of Advances in Modeling Earth Systems*, 11(11), 3545–3592. doi: 10.1029/2019MS001810
- Luyten, J., Pedlosky, J., & Stommel, H. (1983). The ventilated thermocline. *Journal of Physical Oceanography*, 13(2), 292–309.
- Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., ... others (2017). NEMO ocean engine.
- McDougall, T. J., & Barker, P. M. (2011). Getting started with TEOS-10 and the Gibbs Seawater (GSW) oceanographic toolbox. *Scor/Iapso WG*, 127, 1–28.
- McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2189), 20160117.
- McWilliams, J. C., Huckle, E., & Shchepetkin, A. F. (2009). Buoyancy effects in a stratified Ekman layer. *Journal of Physical Oceanography*, 39(10), 2581–2599.
- McWilliams, J. C., Sullivan, P. P., & Moeng, C.-H. (1997). Langmuir turbulence in the ocean. *Journal of Fluid Mechanics*, 334, 1–30.
- Melet, A. V., Hallberg, R., & Marshall, D. P. (2022). Chapter 2 - The role of ocean mixing in the climate system. In M. Meredith & A. Naveira Garabato (Eds.), *Ocean mixing* (p. 5-34). Elsevier. doi: 10.1016/B978-0-12-821512-8.00009-8
- Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, 20(4), 851–875.
- Omand, M. M., D'Asaro, E. A., Lee, C. M., Perry, M. J., Briggs, N., Cetinić, I., & Mahadevan, A. (2015). Eddy-driven subduction exports particulate organic carbon from the spring bloom. *Science*, 348(6231), 222–225.
- Prandtl, L., Wieselsberger, C., & Betz, A. (1925). *Results of the aerodynamic research institute in Göttingen* (No. 1). R. Oldenburg.
- Pressel, K. G., Mishra, S., Schneider, T., Kaul, C. M., & Tan, Z. (2017). Numerics and

- 1472 subgrid-scale modeling in large eddy simulations of stratocumulus clouds. *Journal of*  
 1473 *advances in modeling earth systems*, 9(2), 1342–1365.
- 1474 Price, J. F., Weller, R. A., & Pinkel, R. (1986). Diurnal cycling: Observations and models  
 1475 of the upper ocean response to diurnal heating, cooling, and wind mixing. *Journal of*  
 1476 *Geophysical Research: Oceans*, 91(C7), 8411–8427.
- 1477 Ramadhan, A., Wagner, G. L., Hill, C., Campin, J.-M., Churavy, V., Besard, T., ... Ferrari,  
 1478 R. (2020). Oceananigans.jl: Fast and friendly geophysical fluid dynamics on GPUs.  
 1479 *The Journal of Open Source Software*, 4(44), 1965. doi: 10.21105/joss.01965
- 1480 Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in  
 1481 neural network parameterizations: General algorithms and lorenz 96 case study (v1.  
 1482 0). *Geoscientific Model Development*, 13(5), 2185–2196.
- 1483 Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid  
 1484 processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39),  
 1485 9684–9689.
- 1486 Reffray, G., Bourdalle-Badie, R., & Calone, C. (2015). Modelling turbulent vertical mixing  
 1487 sensitivity using a 1-D version of NEMO. *Geoscientific Model Development*, 8(1),  
 1488 69–86.
- 1489 Reichl, B. G., & Hallberg, R. (2018). A simplified energetics based planetary boundary layer  
 1490 (ePBL) approach for ocean climate simulations. *Ocean Modelling*, 132, 112–129.
- 1491 Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential energy  
 1492 conversion rate of vertical mixing due to Langmuir turbulence. *Journal of Physical*  
 1493 *Oceanography*, 49(11), 2935–2959.
- 1494 Reichl, B. G., Wittenberg, A. T., Griffies, S. M., & Adcroft, A. (2024). Improved equatorial  
 1495 upper ocean vertical mixing in the NOAA/GFDL OM4 model. *Authorea Preprints*.
- 1496 Saffman, P. G. (1970). A model for inhomogeneous turbulent flow. *Proceedings of the Royal*  
 1497 *Society of London. A. Mathematical and Physical Sciences*, 317(1530), 417–433.
- 1498 Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J., Bracegirdle, T. J., & Wang,  
 1499 Z. (2013). Assessment of Southern Ocean mixed-layer depths in CMIP5 models:  
 1500 Historical bias and forcing response. *Journal of Geophysical Research: Oceans*, 118(4),  
 1501 1845–1862.
- 1502 Samelson, R. (2022). Wind drift in a homogeneous equilibrium sea. *Journal of Physical*  
 1503 *Oceanography*, 52(9), 1945–1967. doi: 10.1175/JPO-D-22-0017.1
- 1504 Shu, C.-W. (2020). Essentially non-oscillatory and weighted essentially non-oscillatory  
 1505 schemes. *Acta Numerica*, 29, 701–762.
- 1506 Siebesma, A. P., Soares, P. M., & Teixeira, J. (2007). A combined eddy-diffusivity mass-flux  
 1507 approach for the convective boundary layer. *Journal of the Atmospheric Sciences*,  
 1508 64(4), 1230–1248. doi: 10.1175/JAS3888.1
- 1509 Silvestri, S., Wagner, G. L., Campin, J.-M., Constantinou, N. C., Hill, C. N., Souza,  
 1510 A. N., & Ferrari, R. (2024). A new WENO-based momentum advection scheme for  
 1511 simulations of ocean mesoscale turbulence. *Earth and Space Science Open Archive*.  
 1512 doi: 10.22541/essoar.170110657.76489860/v2
- 1513 Silvestri, S., Wagner, G. L., Constantinou, N. C., Hill, C. N., Campin, J.-M., Souza,  
 1514 A. N., ... Ferrari, R. (2024). A GPU-based ocean dynamical core for routine  
 1515 mesoscale-resolving climate simulations. *Earth and Space Science Open Archive*. doi:  
 1516 10.22541/essoar.171708158.82342448/v1
- 1517 Smyth, W., & Moum, J. (2013). Marginal instability and deep cycle turbulence in the  
 1518 eastern equatorial Pacific Ocean. *Geophysical Research Letters*, 40(23), 6181–6185.  
 1519 doi: 10.1002/2013GL058403
- 1520 Smyth, W., Pham, H., Moum, J., & Sarkar, S. (2017). Pulsating turbulence in a marginally  
 1521 unstable stratified shear flow. *Journal of Fluid Mechanics*, 822, 327–341.
- 1522 Souza, A. N., He, J., Bischoff, T., Waruszewski, M., Novak, L., Barra, V., ... Schneider, T.  
 1523 (2023). The flux-differencing discontinuous Galerkin method applied to an idealized  
 1524 fully compressible nonhydrostatic dry atmosphere. *Journal of Advances in Modeling*  
 1525 *Earth Systems*, 15(4), e2022MS003527. doi: 10.1029/2022MS003527
- 1526 Souza, A. N., Wagner, G. L., Ramadhan, A., Allen, B., Churavy, V., Schloss, J., ... Ferrari,

- 1527 R. (2020). Uncertainty quantification of ocean parameterizations: Application to the  
 1528 K-profile-parameterization for penetrative convection. *Journal of Advances in Modeling  
 1529 Earth Systems*, 12(12), e2020MS002108. doi: 10.1029/2020MS002108
- 1530 Spalding, D. B. (1991). Kolmogorov's two-equation model of turbulence. *Proceedings of the  
 1531 Royal Society of London. Series A: Mathematical and Physical Sciences*, 434(1890),  
 1532 211–216. doi: 10.1098/rspa.1991.0089
- 1533 Suzuki, N., & Fox-Kemper, B. (2016). Understanding Stokes forces in the wave-averaged  
 1534 equations. *Journal of Geophysical Research: Oceans*, 121(5), 3579–3596. doi: 10.1002/  
 1535 2015JC011563
- 1536 Taylor, G. I. (1922). Diffusion by continuous movements. *Proceedings of the London  
 1537 Mathematical Society*, 2(1), 196–212. doi: 10.1112/plms/s2-20.1.196
- 1538 Taylor, J. R., & Thompson, A. F. (2023). Submesoscale dynamics in the upper ocean. *Annual  
 1539 Review of Fluid Mechanics*, 55, 103–127. doi: 10.1146/annurev-fluid-031422-095147
- 1540 National Academies of Sciences, Engineering, and Medicine and others. (2022). Automated  
 1541 research workflows for accelerated discovery: Closing the knowledge discovery loop.
- 1542 Thomas, L. N., Tandon, A., & Mahadevan, A. (2008). Submesoscale processes and dynamics.  
 1543 In *Ocean modeling in an eddying regime* (p. 17–38). American Geophysical Union  
 1544 (AGU). doi: 10.1029/177GM04
- 1545 Umlauf, L., & Burchard, H. (2003). A generic length-scale equation for geophysical  
 1546 turbulence models. *Journal of Marine Research*, 61(2), 235–265. Retrieved from  
 1547 [https://elischolar.library.yale.edu/journal\\_of\\_marine\\_research/9](https://elischolar.library.yale.edu/journal_of_marine_research/9)
- 1548 Umlauf, L., & Burchard, H. (2005). Second-order turbulence closure models for geophysical  
 1549 boundary layers. A review of recent work. *Continental Shelf Research*, 25(7–8), 795–827.  
 1550 doi: 10.1016/j.csr.2004.08.004
- 1551 Van Roekel, L., Adcroft, A. J., Danabasoglu, G., Griffies, S. M., Kauffman, B., Large, W. G.,  
 1552 ... Schmidt, M. (2018). The KPP boundary layer scheme for the ocean: Revisiting its  
 1553 formulation and benchmarking one-dimensional simulations relative to LES. *Journal of  
 1554 Advances in Modeling Earth Systems*, 10(11), 2647–2685. doi: 10.1029/2018MS001336
- 1555 Venayagamoorthy, S. K., & Stretch, D. D. (2010). On the turbulent prandtl number in  
 1556 homogeneous stably stratified turbulence. *Journal of fluid mechanics*, 644, 359–369.
- 1557 Von Kármán, T. (1931). *Mechanical similitude and turbulence* (No. 611). National Advisory  
 1558 Committee for Aeronautics.
- 1559 Wagner, G. L., Hillier, A., & Constantinou, N. C. (2022). *ParameterEstimocean.jl v0.14.2:  
 1560 Julia package for parameter estimation of turbulence closures for ocean models using  
 1561 Ensemble Kalman Inversion*. doi: 10.5281/zenodo.5762810
- 1562 Wagner, G. L., Ramadhan, A., Chor, T., Constantinou, N. C., Hillier, A., Lee, X. K., & Li,  
 1563 Q. (2023). *LESbrary.jl: v0.1.1: Oceananigans large eddy simulation (LES) data for  
 1564 calibrating parameterizations* [Dataset]. doi: 10.5281/zenodo.7997002
- 1565 Whitt, D. B., Cherian, D. A., Holmes, R. M., Bachman, S. D., Lien, R.-C., Large, W. G., &  
 1566 Moum, J. N. (2022). Simulation and scaling of the turbulent vertical heat transport  
 1567 and deep-cycle turbulence across the equatorial Pacific cold tongue. *Journal of Physical  
 1568 Oceanography*, 52(5), 981–1014. doi: 10.1175/JPO-D-21-0153.1
- 1569 Williams, R. G. (1991). The role of the mixed layer in setting the potential vorticity of  
 1570 the main thermocline. *Journal of Physical Oceanography*, 21(12), 1803–1814. doi:  
 1571 10.1175/1520-0485(1991)021<1803:TROTML>2.0.CO;2
- 1572 Wunsch, C., & Ferrari, R. (2004). Vertical mixing, energy, and the general circulation  
 1573 of the oceans. *Annual Review of Fluid Mechanics*, 36(1), 281–314. doi: 10.1146/  
 1574 annurev.fluid.36.050802.122121
- 1575 Zhang, S., Fu, H., Wu, L., Li, Y., Wang, H., Zeng, Y., ... Guo, Y. (2020). Optimizing  
 1576 high-resolution community earth system model on a heterogeneous many-core  
 1577 supercomputing platform. *Geoscientific Model Development*, 13(10), 4809–4829. doi:  
 1578 10.5194/gmd-13-4809-2020