

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

3 Gregory LeClaire Wagner¹, Adeline Hillier¹, Navid C. Constantinou^{2,3},
4 Simone Silvestri¹, Andre Souza¹, Keaton J. Burns¹, Chris Hill¹,
5 Jean-Michel Campin¹, John Marshall¹, and Raffaele Ferrari¹

6 ¹Massachusetts Institute of Technology, Cambridge, MA, USA

7 ²University of Melbourne, Parkville, VIC, Australia

8 ³ARC 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 — because these “microscale” eddies are smaller than a grid cell and
 38 therefore must be represented more approximately. This is a problem because these small yet
 39 mighty eddies mix heat and carbon deep into the ocean, and thus help keep the atmosphere
 40 from getting too hot, and too rich in CO₂. In this paper, we propose a new model component
 41 called “CATKE” (pronounced *kăt-kee*) that does a decent job at approximately incorporating
 42 the effect of such relatively small ocean eddies in global ocean models. CATKE stands for
 43 “Convective Adjustment and Turbulent Kinetic Energy”. Basically, CATKE keeps track of
 44 the *energy* of small-scale turbulence — a measure of how vigorous it is, and thus how much
 45 it mixes the ocean — to predict ocean mixing rates.

46 **1 Introduction**

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

60 This paper documents the development, calibration, and preliminary validation of a
 61 new parameterization for vertical mixing by ocean microscale turbulence. Our goal is to use
 62 the new parameterization in a GPU-based climate model that is automatically calibrated
 63 to observations, reports quantified uncertainties, and has an ocean component with a high,
 64 $O(10\text{ km})$ or finer resolution that fully resolves ocean mesoscale turbulence. The dynamical
 65 core of the GPU-based ocean component is described by Silvestri et al. (2024). In service of

this ultimate goal, the work documented in this paper prioritizes not just accurate predictions, but also efficiency on GPUs in high-resolution configurations. We also invest in automated calibration that constrains all of the parameterization’s free parameters to 21 large eddy simulations (LESs) simultaneously, accounting for the peculiarities of our specific numerical implementation of the parameterization in a single column model. The 21 LES we use to calibrate and the additional 14 LES we use to validate the parameterization are described in section 2.

Our new parameterization, which we call “CATKE”, uses a downgradient formulation that estimates eddy diffusivities in terms of a prognostic turbulent kinetic energy (TKE) variable and a diagnostic mixing length with a novel dynamic convective adjustment (CA) component. CATKE is a “one-equation” model (because it includes an additional equation for TKE) that bears resemblance to a family of battle-tested parameterizations long used in European climate models (Gaspar et al., 1990; Blanke & Delecluse, 1993; Kuhlbrodt et al., 2018; Madec et al., 2017; Gutjahr et al., 2021; Jungclaus et al., 2022). One-equation downgradient parameterizations are appropriate for high-resolution ocean modeling and amenable to GPU performance optimization due to their spatially-local formulation. In contrast, the main benefit of “ K -profile” schemes used in many global ocean models — accommodating hours-long time steps (Reichl & Hallberg, 2018) — is not realized in high-resolution simulations that require short time-steps anyways to resolve advection by mesoscale turbulence. Moreover, K -profile schemes achieve this time-step flexibility by solving nonlinear algebraic equations to determine boundary layer depth diagnostically (Large et al., 1994; Reichl & Hallberg, 2018; Reichl & Li, 2019), which may require significant optimization to achieve good performance on GPU-like systems (as experienced by Zhang et al., 2020). As for two-equation or “ $k-\epsilon$ ”-type models (Mellor & Yamada, 1982; Kantha & Clayson, 1994; Canuto et al., 2001; Umlauf & Burchard, 2003; Harcourt, 2015), CATKE is less expensive merely by having one fewer prognostic variable. The primary downside of any downgradient parameterization is unavoidable biases when instantaneously non-local, non-downgradient fluxes dominate, such as during free convection.

We therefore devote special attention to free convection during CATKE’s formulation, which is described in section 3, to minimize this downgradient bias and assess its importance. Section 3.1.5 describes CATKE’s diagnostic convective length scale and primary novelty, which uses dimensional analysis (Deardorff, 1970) to predict the convective boundary layer depth in terms of the *local* TKE in order to estimate a dynamically evolving convective diffusivity. This improves on the constant “convective adjustment” diffusivity typically used with one-equation parameterizations in ocean climate models (typically $0.1 \text{ m}^2 \text{ s}^{-1}$; Madec et al., 2017; Gutjahr et al., 2021; Jungclaus et al., 2022), which cannot describe how the convective mixing rate *varies* with both boundary layer depth and the intensity of the destabilizing surface buoyancy flux over the wide range of conditions observed in Earth’s ocean. As a result, CATKE might be able to represent scenarios where mixing competes with other dynamics such as submesoscale restratification. We also implement different mixing lengths for momentum, tracer, TKE, and the TKE dissipation rate in shear-driven turbulence that all vary as a function of the local gradient Richardson number. This contrasts with typical approaches that estimate the TKE diffusivity as a constant multiple of the eddy viscosity (Blanke & Delecluse, 1993; Madec et al., 2017; Umlauf & Burchard, 2003), or which allow only the tracer mixing length to vary with Richardson number (Blanke & Delecluse, 1993; Madec et al., 2017).

CATKE’s formulation could not be realized without an effective method for constraining CATKE’s free parameters against observational or LES data. Section 4 describes how we calibrate CATKE’s free parameters by minimizing the error between 21 variously-forced LES and the predictions of the LES data made by forward CATKE-parameterized single column simulations. Because this calibration method is posed in terms of forward simulations, rather than an *a priori* analysis of parameters or isolated subcomponents of the parameterization, it is sometimes called “*a posteriori*” calibration (Duraisamy, 2021; Frezat et al., 2022).

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 validate CATKE by a variety of methods 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 observations or other measurements. 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. 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 turbulent kinetic energy 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 of these 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.

We conclude in section 6 with comments about future efforts to calibrate CATKE against more comprehensive data sets and future 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 calibrate CATKE’s free parameters against satellite and in-situ ocean observations.

2 Large eddy simulations of turbulent mixing beneath surface waves

We begin by concretely defining the parameterization problem that drives the cyclical process of formulating, calibrating, and validating CATKE. In this paper, the 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 & Fox-Kemper, 2016),

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

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

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

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

We conduct 35 LES of (1)–(3) forced by constant, horizontally-uniform fluxes of momentum and buoyancy in a $512 \text{ m} \times 512 \text{ m} \times 256 \text{ m}$ horizontally-periodic domain with $O(1 \text{ m})$ resolution using Oceananigans ([Ramadhan et al., 2020](#)). All 35 LES are initialized with the same piecewise-constant density stratification given in equation A1, which has a weakly-stratified near-surface layer, a more strongly stratified middle layer, and a weakly-stratified lower layer. The surface momentum flux or “wind stress” τ_x is defined via \mathbf{F}_u in (1) as

$$\mathbf{F}_u = -\partial_z [\tau_x \delta(z)] \hat{x}, \quad (4)$$

where $\delta(z)$ is a delta function concentrate at $z = 0$, such that negative stress $\tau_x < 0$ forces a current in the $+x$ -direction. Two types of buoyancy fluxes are used: a destabilizing surface flux $J_b > 0$ representing cooling or heat loss, which is defined via F_b in equation (3) via

$$F_b = -\partial_z [J_b \delta(z)]. \quad (5)$$

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

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

In (6), $I(z)$ is the buoyancy flux profile associated with penetrating solar insolation, $J_b < 0$ is the surface solar insolation, ϵ_1 is the fraction of penetrating radiation absorbed over the vertical scale λ_1 , and $(1 - \epsilon_1)$ is the remaining fraction absorbed over λ_2 . All simulations 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 et al., 2022](#)).

The forcing strength for each case is rationalized by categorizing the LES into 6-, 12-, 24-, 48-, and 72-hour “suites” according to their duration. Because all the LES are initialized identically and run until the boundary layer is roughly half the depth of the domain, duration indicates forcing strength: the 6-hour-suite are the most strongly forced and the 72-hour suite simulations are the most weakly forced. The intermediately-forced 12-, 24-, and 48-hour suites are used for calibration. The 35 LES are divided into 5 “suites” with 7 cases each, according to their duration and the intensity of the surface fluxes: the 6-hour suite exhibits extreme forcing, while the 72-hour suite exhibits relatively weak forcing. Each suite consists of 7 physical scenarios that represent different forcing regimes:

- “free convection”, which has pure destabilizing buoyancy forcing and no winds,
- “weak wind strong cooling”,
- “medium wind medium cooling”,
- “strong wind weak cooling”,
- “strong wind”, with no buoyancy forcing,
- “strong wind no rotation” with no buoyancy forcing and $f = 0$.
- “strong wind and sunny” with penetrative heating, wind forcing, and $f = 0$.

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

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

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

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

2.1 The single column context

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

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

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

Next, we derive a set of equations that governs the horizontally-averaged zonal momentum $u(z, t)$, meridional momentum $v(z, t)$, and any tracer $c(z, t)$ by taking a horizontal 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 , with superscript L is omitted to simplify the notation. Lateral fluxes vanish from (9)–(11) due to horizontal homogeneity. No terms 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

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	$+4.0 \times 10^{-7}$	4.0×10^{-4}	+833	15
12 hour	mid wind mid cooling	$+3.2 \times 10^{-7}$	6.0×10^{-4}	+667	17
12 hour	strong wind weak cooling	$+2.0 \times 10^{-7}$	8.0×10^{-4}	+417	20
12 hour	strong wind	0	9.0×10^{-4}	0	21
12 hour	strong wind no rotation	0	6.0×10^{-4}	0	17
12 hour	strong wind and sunny	-5.0×10^{-7}	9.0×10^{-4}	-1042	21
24 hour	free convection	$+2.4 \times 10^{-7}$	0	+500	0
24 hour	weak wind strong cooling	$+2.0 \times 10^{-7}$	3.0×10^{-4}	+417	13
24 hour	mid wind mid cooling	$+1.6 \times 10^{-7}$	4.5×10^{-4}	+333	16
24 hour	strong wind weak cooling	$+1.0 \times 10^{-7}$	5.9×10^{-4}	+208	17
24 hour	strong wind	0	6.8×10^{-4}	0	18
24 hour	strong wind no rotation	0	3.0×10^{-4}	0	13
24 hour	strong wind and sunny	-3.0×10^{-7}	4.5×10^{-4}	-625	16
48 hour	free convection	$+1.2 \times 10^{-7}$	0	+250	0
48 hour	weak wind strong cooling	$+1.0 \times 10^{-7}$	2.0×10^{-4}	+208	11
48 hour	mid wind mid cooling	$+8.0 \times 10^{-8}$	3.4×10^{-4}	+167	14
48 hour	strong wind weak cooling	$+5.0 \times 10^{-8}$	3.8×10^{-4}	+104	15
48 hour	strong wind	0	4.5×10^{-4}	0	16
48 hour	strong wind no rotation	0	1.6×10^{-4}	0	10
48 hour	strong wind and sunny	-1.0×10^{-7}	2.0×10^{-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 Coriolis parameter $f = 10^{-4} \text{ s}^{-1}$ except “strong wind no rotation” and “strong wind and sunny”, which use $f = 0$. The “suite” indicates simulation duration. J_b is the surface buoyancy flux, τ_x is the kinematic momentum flux (momentum flux divided by ocean reference density), $Q \approx \rho_o c_p J_b / (ag)$ is the heat flux associated with J_b , and u_{10} is an estimate of the 10-meter wind speed associated with τ_x according to equation A4 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 τ_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.

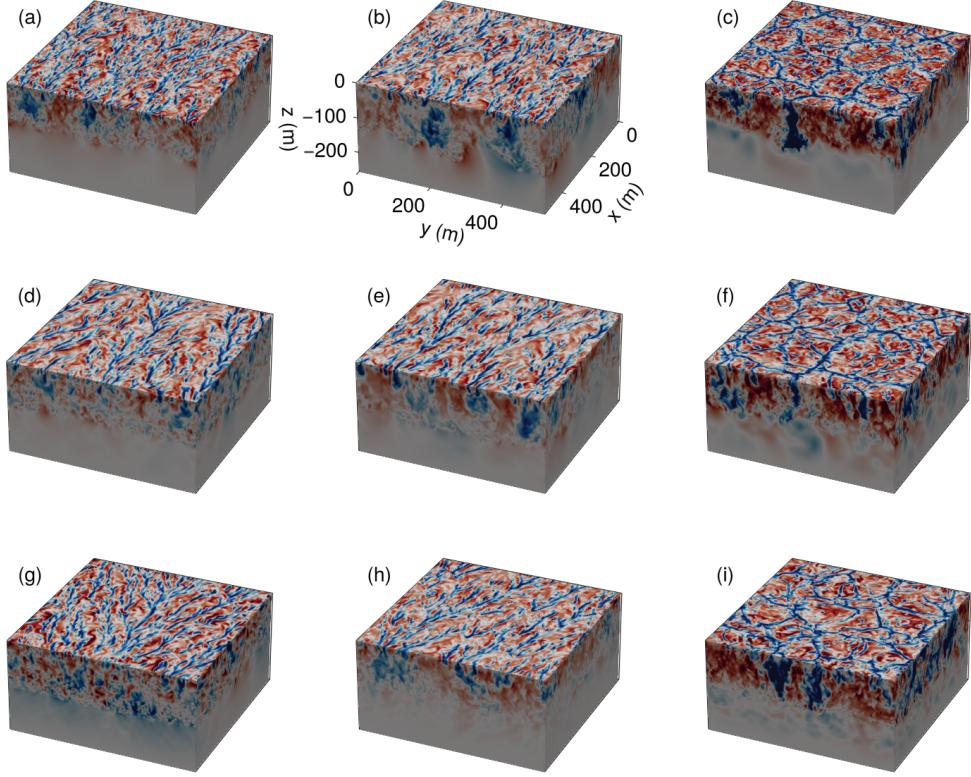


Figure 1. Visualization of vertical velocity in 9 of 35 large eddy simulations (LES) of the ocean surface boundary layer used in this paper, forced variously by winds, surface waves, and heat fluxes. All LES, which are summarized in tables 1 and 2 and described in more detail in Appendix A, are initialized 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 $\frac{1}{2} \max |w|$, which for each panel 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 m s^{-1} .

260 conditions, and potentially, additional auxiliary variables. For example, the parameterization
 261 described in the next section uses a downgradient formulation $\overline{w'c'} \sim \partial_z c$ to predict vertical
 262 tracer and momentum fluxes.

263 2.2 Connection to the regional and global ocean modeling context

264 Our LES, and the models that predict the horizontal average of the LES, may be
 265 described as “single column models”. This nomenclature reflects the notion that the models
 266 simulate the vertical redistribution of momentum and tracers by turbulent motions in a
 267 single column of a three-dimensional ocean model. Indeed, we envision that the single
 268 column context is generalized to a large-scale ocean simulation merely by adding advection
 269 by motions somewhat larger than the scale of the LES domain. This approach relies on two

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	$+8.0 \times 10^{-7}$	5.0×10^{-4}	+1666	16
6 hour	mid wind mid cooling	$+6.4 \times 10^{-7}$	8.0×10^{-4}	+1333	20
6 hour	strong wind weak cooling	$+4.0 \times 10^{-7}$	1.2×10^{-3}	+833	23
6 hour	strong wind	0	1.4×10^{-3}	0	24
6 hour	strong wind no rotation	0	1.1×10^{-3}	0	22
6 hour	strong wind and sunny	-6.0×10^{-7}	1.5×10^{-3}	-1250	25
72 hour	free convection	$+8.7 \times 10^{-8}$	0	+181	0
72 hour	weak wind strong cooling	$+7.5 \times 10^{-8}$	1.8×10^{-4}	+156	11
72 hour	mid wind mid cooling	$+6.0 \times 10^{-8}$	2.9×10^{-4}	+125	13
72 hour	strong wind weak cooling	$+3.8 \times 10^{-8}$	3.4×10^{-4}	+79	14
72 hour	strong wind	0	4.1×10^{-4}	0	15
72 hour	strong wind no rotation	0	1.1×10^{-4}	0	9
72 hour	strong wind and sunny	-5.0×10^{-8}	1.3×10^{-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.

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 the horizontal ones. 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 the 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 outlined in this paper, this 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 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.

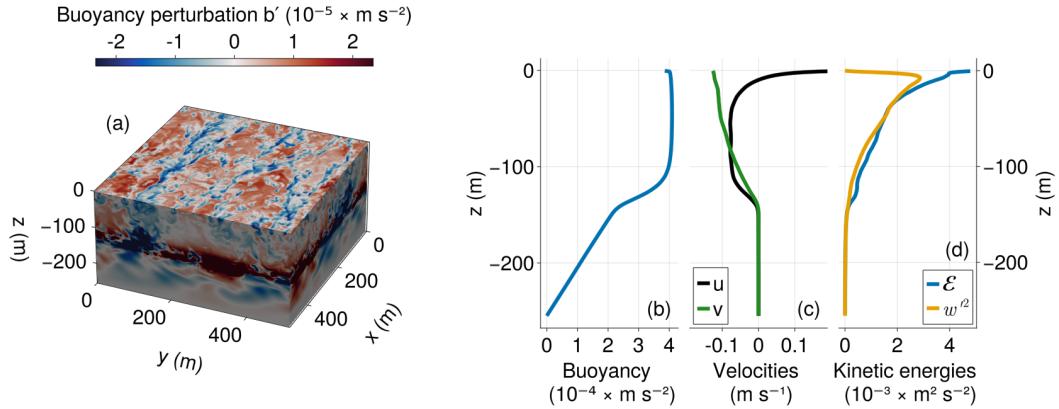


Figure 2. Illustration of horizontally-averaged data from the 12-hour strong wind, weak cooling LES. Panel (a) shows the buoyancy perturbation b' . Note the colorbar is strongly saturated to illustrate boundary layer structure; the buoyancy perturbation is particularly large at the base of the boundary 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, $\overline{w'^2}$.

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),

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

where e is the turbulent kinetic energy, \sqrt{e} is the turbulent velocity scale, and ℓ_ψ 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 ℓ_ψ , the form $\kappa_\psi = \ell_\psi \sqrt{e}$ follows from dimensional analysis. CATKE invokes three mixing lengths and three eddy diffusivities for horizontal velocities (ℓ_u and κ_u), tracers (ℓ_c and κ_c), and turbulent kinetic energy (ℓ_e and κ_e).

With (12), the single column equations become

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

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

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

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,

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

where $\bar{F}_b = -\partial_z I$ corresponds to heating within the water column due to penetrating solar radiation, I . 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).

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

$$321 \quad \mathcal{E} \stackrel{\text{def}}{=} \frac{1}{2} \overline{|u'|^2} = \frac{1}{2} \left(\overline{u'^2} + \overline{v'^2} + \overline{w'^2} \right). \quad (17)$$

322 We postulate a close relationship between e in (12) and the subgrid kinetic energy, \mathcal{E} .
 323 However, this is a relationship rather than an identity, because \mathcal{E} has contributions from
 324 motions that are unrelated to the eddy diffusivity in (12). For example, internal waves
 325 generated by convective plumes make a significant contribution to \mathcal{E} below the base of
 326 boundary layer, despite that there is no mixing there. We note further that if the kinetic
 327 energy and mixing length are actually known, the inexact relationship between \mathcal{E} and e
 328 manifests through a “correlation coefficient” (G. I. Taylor, 1922) that appears in formulations
 329 like (12). We therefore define e as a *latent variable* which is linked to the averaged velocity
 330 and tracer fields via (12), rather than as corresponding directly to the observable, but less
 331 relevant quantity (17). This interpretation has important implications for calibration: rather
 332 than using the discrepancy between LES-derived \mathcal{E} and e to estimate free parameters, we
 333 only use the error in momentum and buoyancy profiles — which are strongly affected by e
 334 through (12) — to constrain the free parameters that govern the evolution of e . In other
 335 words, e can only be observed indirectly via the evolution of momentum and buoyancy.
 336 Interpreting e as a latent variable rather than as the actual subgrid kinetic energy \mathcal{E} is also
 337 proposed by Kolmogorov (see Spalding, 1991) and Saffman (1970).

338 Though we define e as a latent variable that is linked to u, v, c solely via (12), we
 339 nevertheless postulate a similarity between e and \mathcal{E} on physical grounds — where there is
 340 turbulence, there will be mixing — and following a litany of prior work (Saffman, 1970;
 341 Gaspar et al., 1990; Spalding, 1991; Umlauf & Burchard, 2003), use the evolution equation
 342 for \mathcal{E} to derive a model for the evolution of e . An equation describing the evolution of \mathcal{E} can
 343 be derived from (1), including the molecular stress divergence $\nu \nabla^2 (\mathbf{U}^L - \mathbf{U}^S)$ (we include
 344 the Stokes drift term here for completeness, though it does not contribute to the equation
 345 for \mathcal{E}). The result is

$$346 \quad \partial_t \mathcal{E} = - \underbrace{\partial_z (\overline{w' \mathcal{E}'} + \overline{w' p'})}_{\text{transport}} - \underbrace{\overline{u' w'} \cdot \partial_z \mathbf{u}}_{\text{shear production}} + \underbrace{\overline{w' b'}}_{\text{buoyancy flux}} - \underbrace{\nu \overline{|\nabla \mathbf{u}'|^2}}_{\text{dissipation}}, \quad (18)$$

347 where ν is the kinematic viscosity, p is kinematic pressure (dynamic pressure divided by a
 348 reference density) and $\mathcal{E}' = \frac{1}{2} |\mathbf{u}'|^2 - \mathcal{E}$. Note that because \mathbf{u} is the horizontally-averaged
 349 Lagrangian-mean velocity, the shear production term in (18) represents the total transfer of
 350 kinetic energy from the average \mathbf{u} to the fluctuations \mathbf{u}' — including the so-called “Stokes
 351 production” term (McWilliams et al., 1997). Inspired by (18), we formulate an equation for
 352 e consisting of terms that mirror each term in equation (18):

$$353 \quad \partial_t e = \underbrace{\partial_z (\kappa_e \partial_z e)}_{\text{transport}} + \underbrace{\kappa_u |\partial_z \mathbf{u}|^2}_{\text{shear production}} - \underbrace{\kappa_c N^2}_{\text{buoyancy flux}} - \underbrace{\frac{e^{3/2}}{\ell_D}}_{\text{dissipation}}, \quad (19)$$

354 where $|\partial_z \mathbf{u}|^2 = (\partial_z u)^2 + (\partial_z v)^2$ is the square vertical shear of the horizontally-averaged
 355 velocity field \mathbf{u} (note that $w = 0$ because of horizontal homogeneity), κ_e is the vertical
 356 diffusivity of e , ℓ_D is the “dissipation length scale”, and we have labeled the corresponding
 357 terms in (18) and (19). The shear production and buoyancy flux terms are formulated
 358 by applying the eddy diffusivity hypothesis (12) to their corresponding expressions in
 359 equation (18). Like in the budget for \mathcal{E} , the shear production term in (19) represents the
 360 total shear production including both “Eulerian” and “Stokes” production. We assume that
 361 the transport of e , which helps to deepen boundary layers by modeling turbulence spreading
 362 away from turbulence-generating regions, can be modeled with an eddy diffusivity $\kappa_e = \ell_e \sqrt{e}$.
 363 Finally, to model the dissipation of e we introduce the dissipation length scale ℓ_D , which

364 has a similar form to the mixing lengths ℓ_u , ℓ_c , and ℓ_e . The expression $e^{3/2}/\ell_D$ follows on
 365 dimensional grounds.

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

$$371 J_e \stackrel{\text{def}}{=} -\kappa_e \partial_z e \Big|_{z=0} = -\mathbb{C}_J^{\text{shear}} u_*^3 - \mathbb{C}_J^{\text{conv}} w_\Delta^3, \quad \text{where } w_\Delta^3 \stackrel{\text{def}}{=} \Delta z \max(J_b, 0), \quad (20)$$

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

$$378 u_* \stackrel{\text{def}}{=} (\tau_x^2 + \tau_y^2)^{1/4}, \quad (21)$$

379 defined in terms of the zonal and meridional kinematic momentum fluxes τ_x and τ_y (wind
 380 stresses divided by reference water density). The boundary condition (20) differs from
 381 boundary conditions used in the TKE-based models described by Blanke and Delecluse
 382 (1993) and Madec et al. (2017), which prescribe TKE (rather than prescribing TKE flux),
 383 and depend only on the friction velocity u_* .

384 The surface flux formulation in (20) introduces the notation

$$385 \mathbb{C}_{\text{component}}^{\text{label}} \quad (22)$$

386 for two free parameters $\mathbb{C}_J^{\text{shear}}$ and $\mathbb{C}_J^{\text{conv}}$, where “label” indicates the parameter’s role and
 387 “component” refers to the variable or component to which the parameter associates.

388 3.1 Turbulence length scale model

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

$$393 \ell_\psi = \max(\ell_\psi^{\text{conv}}, \ell_\psi^{\text{shear}}), \quad (23)$$

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

397 3.1.1 Shear turbulence length scale

398 To represent shear dominated turbulence either in strong stratification or near the ocean
 399 surface, we use the length scale

$$400 \ell_\psi^{\text{shear}} = \mathbb{S}_\psi(Ri) \min\left(\frac{\sqrt{e}}{N_+}, \mathbb{C}^s d\right), \quad \text{where } N_+^2 \stackrel{\text{def}}{=} \max(0, \partial_z b) \quad (24)$$

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

406 We use (24) for ℓ_c^{shear} , ℓ_u^{shear} , and ℓ_e^{shear} . For the dissipation length scale ℓ_D^{shear} , we use

$$407 \quad \ell_D = \frac{1}{\mathbb{S}_D(Ri)} \min \left(\frac{\sqrt{e}}{N_+}, \mathbb{C}^s d \right), \quad (25)$$

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

411 The stability functions \mathbb{S}_ψ and $1/\mathbb{S}_D$ in (24)–(25) modulate each length scale with the
412 stably-stratified Richardson number

$$413 \quad Ri \stackrel{\text{def}}{=} \frac{\partial_z b}{|\partial_z \mathbf{u}|^2}, \quad (26)$$

414 which, among other meanings, indicates the role of shear production in turbulent mixing.
415 The stability functions give CATKE a turbulent Prandtl number,

$$416 \quad Pr(Ri) \stackrel{\text{def}}{=} \frac{\kappa_u}{\kappa_c} = \frac{\mathbb{S}_u(Ri)}{\mathbb{S}_c(Ri)}, \quad (27)$$

417 that depends on Ri .

418 We propose a four-part functions $\mathbb{S}_\psi(Ri)$,

$$419 \quad \mathbb{S}_\psi(Ri) = \begin{cases} \mathbb{C}_\psi^- & \text{when } Ri < 0, \\ \mathbb{C}_\psi^0 & \text{when } 0 \leq Ri \leq \mathbb{C}_{Ri}^0, \\ \mathbb{C}_\psi^0 + \left(\mathbb{C}_\psi^\infty - \mathbb{C}_\psi^0 \right) \frac{Ri - \mathbb{C}_{Ri}^0}{\mathbb{C}_{Ri}^\delta} & \text{when } \mathbb{C}_{Ri}^0 < Ri < \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta, \\ \mathbb{C}_\psi^\infty & \text{when } Ri \geq \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta. \end{cases} \quad (28)$$

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

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

426 The stability function (28) plays a similar role as the more elaborate stability functions used
427 in two-equation models (Burchard & Bolding, 2001), which are derived from an second-
428 moment closure. The stability functions in equation (28) are plotted in the left panel of
429 figure 3 (see section 4 for how the parameters are obtained via calibration to LES).

430 The four shear length scales introduce 15 free parameters: \mathbb{C}^s , \mathbb{C}_{Ri}^δ , and \mathbb{C}_{Ri}^0 used in all
431 four length scales, along with 12 additional parameters associated with the coefficients \mathbb{C}_ψ^- ,
432 \mathbb{C}_ψ^0 and \mathbb{C}_ψ^∞ for each length scale respectively.

433 3.1.2 Turbulent Prandtl and Schmidt numbers in stably stratified shear 434 turbulence

435 Note that CATKE's Pr in (27) is a rational function of Ri , slightly different from the
436 piecewise linear formulation proposed by Blanke and Delecluse (1993) and Madec et al.
437 (2017). In particular,

$$438 \quad Pr = \begin{cases} \mathbb{C}_u^- / \mathbb{C}_c^- & Ri < 0 \\ \mathbb{C}_u^0 / \mathbb{C}_c^0 & 0 \leq Ri \leq \mathbb{C}_{Ri}^0 \\ \frac{\mathbb{C}_u^0 + \mu_u (Ri - \mathbb{C}_{Ri}^0)}{\mathbb{C}_c^0 + \mu_c (Ri - \mathbb{C}_{Ri}^0)} & \mathbb{C}_{Ri}^0 < Ri < \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta \\ \mathbb{C}_u^\infty / \mathbb{C}_c^\infty & Ri \geq \mathbb{C}_{Ri}^0 + \mathbb{C}_{Ri}^\delta \end{cases}, \quad (29)$$

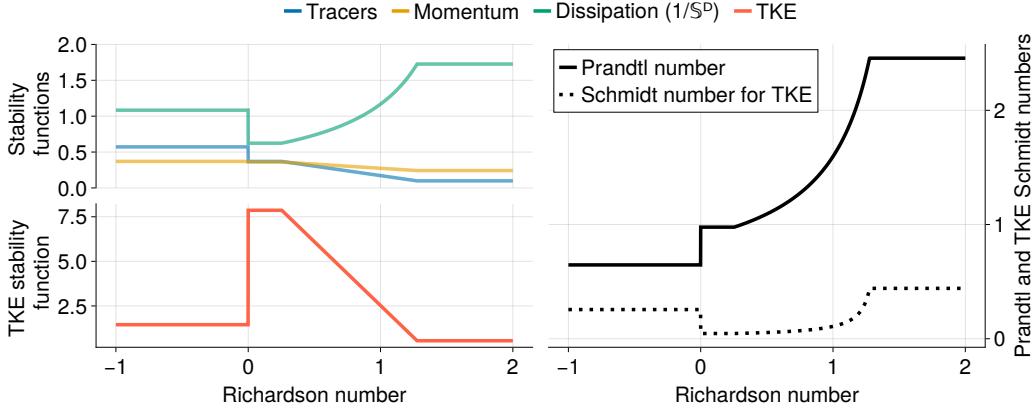


Figure 3. Stability functions (left panel), and Prandtl numbers and Schmidt numbers (right panel). The stability functions for tracers, momentum, and TKE are given by \mathbb{S}_ψ in (28). The stability function for dissipation length scale is $1/\mathbb{S}_D$. The Prandtl number is $\mathbb{S}_u/\mathbb{S}_c$ and the Schmidt number for TKE is $\mathbb{S}_u/\mathbb{S}_e$.

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 $Sc \stackrel{\text{def}}{=} \kappa_u/\kappa_e$. The Prandtl number and Schmidt number for calibrated parameters are visualized in the right panel figure 3.

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

To interpret CATKE's near-surface mixing length $\ell_\psi \sim d$, we consider neutrally-stratified ($\partial_z b = 0$), quasi-equilibrium ($\partial_t u \approx \partial_t e \approx 0$), non-rotating ($f = 0$) near-surface turbulence driven by wind stress $\tau = \tau_x \hat{x}$. We hypothesize that CATKE possesses a similarity solution in this scenario,

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

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

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

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

Note that in neutral conditions,

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

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

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

Notice that e is independent of d in this expression. This means that neglecting turbulent transport in (31) in the context of the similarity hypothesis (30) is at least self-consistent, though this assumption may fail when applied over significant portions of the boundary layer. Next, integrating the quasi-equilibrium x -momentum equation $0 \approx \partial_z (\kappa_u \partial_z u)$ from $z = 0$ to $z = -d$ yields

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

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

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

474 3.1.4 Steady-state gradient Richardson number for stably stratified shear 475 turbulence

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

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

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

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

where r is a rate and Ri^\dagger is the steady-state Richardson number,

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

When the Richardson number $Ri = Ri^\dagger$ equals the steady-state value Ri^\dagger , the shear production of TKE is perfectly balanced by TKE destruction via buoyancy flux and dissipation. But if $Ri < Ri^\dagger$, then $r > 0$ — and TKE will grow. Conversely, if $Ri > Ri^\dagger$ then $r < 0$ and TKE will decay. Finally we note that the functions \mathbb{S}_ψ , defined in (28), depend on Ri . 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 $Ri^\dagger = \mathbb{C}_u^\infty / (\mathbb{C}_c^\infty + \mathbb{C}_D^\infty)$.

492 3.1.5 Convective turbulence length scale

To formulate a length scale for free convection, we divide the freely convecting boundary layer into two regions: a “convecting layer” with unstable $N^2 < 0$, and a “penetration layer” with thickness δ . In the penetration layer, $N^2(z) > 0$ but $N^2(z + \delta) < 0$, where we note that the vertical coordinate z increases upwards and is defined such that $z < 0$. (We use “penetration layer” rather than “entrainment layer” used by Deardorff (1970) because it is less likely to be confused with other types of “entrainment”.) Our formulation for the convective length scale models both rapid mixing in the convective layer as well as entrainment into the boundary layer from below by plumes plunging through the convecting layer into the stably-stratified penetration layer below.

Our dynamic length scale for mixing in the convective layer is based on a dimensional analysis first proposed by Deardorff (1970) that links the turbulent velocity \sqrt{e} (m s^{-1}),

504 surface buoyancy flux J_b (m^2/s^3), and convective layer depth, h (m),

$$505 \quad \sqrt{e} \sim (h J_b)^{1/3}. \quad (39)$$

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

$$508 \quad t_{\text{mix}} \sim \left(\frac{h^2}{J_b} \right)^{1/3}. \quad (40)$$

509 But if we represent convection as a diffusive process with diffusivity κ_c , then we also have
510 that

$$511 \quad t_{\text{mix}} \sim \frac{h^2}{\kappa_c}. \quad (41)$$

512 Equating (40) and (41) yields a scaling relation for the convective diffusivity κ_c .

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

521 To capture t_{mix} consistently between (40) and (41) over the convective region where
522 $N^2 < 0$, we introduce a dynamic convective mixing length scale ℓ_ψ^h that scales with h ,

$$523 \quad \ell_\psi^h \stackrel{\text{def}}{=} \mathbb{C}_\psi^h \frac{e^{3/2}}{\tilde{J}_b + J_b^{\min}} \sim h, \quad (42)$$

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

$$528 \quad \partial_t \tilde{J}_b = \underbrace{\left(\frac{J_b}{\ell_D^2(z=0)} \right)^{1/3}}_{\sim t_{\text{mix}}^{-1}} (J_b - \tilde{J}_b), \quad (43)$$

529 where ℓ_D is the dissipation length scale and $(\ell_D^2/J_b)^{1/3} \sim t_{\text{mix}}$ scales with the instantaneous
530 convective mixing time. Equation (43) relaxes \tilde{J}_b to J_b over the time-scale t_{mix} as defined
531 by (40), and therefore effectively acts to average J_b in time. We use the dissipation length
532 scale ℓ_D in (43) rather than the tracer mixing length ℓ_c because we hypothesize that convective
533 turbulence evolution time-scale is most closely related to the time-scale for turbulent kinetic
534 energy dissipation rather than the time-scale for tracer mixing. In quasi-equilibrium, $\tilde{J}_b \approx J_b$.
535 Because $\ell_\psi^h \sim h$, CATKE's convective tracer diffusivity scales with $\kappa_c \sim h\sqrt{e}$.

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

$$543 \quad \overline{w' b'}|_{z=z_p} \sim -J_b \quad \text{such that} \quad h^2 \sim \frac{J_b t}{N^2}, \quad (44)$$

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

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

$$546 \quad \ell_\psi^p \stackrel{\text{def}}{=} \mathbb{C}_c^p \frac{\tilde{J}_b}{N^2\sqrt{e} + J_b^{\min}}, \quad (45)$$

547 which is applied at the height $z_p < 0$ defined via

$$548 \quad N^2(z_p) > 0 \quad \text{and} \quad N^2(z_p + \delta) < 0, \quad (46)$$

549 where δ is the thickness of the penetration layer. At $z = z_p$, (45) produces $\overline{w'b'} = -\ell_c^p \sqrt{e} N^2 \approx$
550 $-\mathbb{C}_c^p J_b$ in accordance with the empirical law in (44). Our numerical implementation of the
551 convective mixing length uses $\delta = \Delta z$ where Δz is the grid spacing at z_p . This assumes that
552 the entrainment layer is thinner than the grid spacing: when $\delta > \Delta z$, CATKE solutions may
553 exhibit a “thin entrainment layer bias” even if the boundary layer deepening rate is correct.

554 Finally, because e is much larger in shear turbulence than in convective turbulence with
555 similar mixing rates, the scaling (42) will greatly overestimate the mixing length when e is
556 produced by both convection and shear. To limit the impact of the convective mixing length
557 in the presence of shear, we use an estimate of the flux Richardson number,

$$558 \quad \widetilde{Ri}_f \stackrel{\text{def}}{=} \frac{d\sqrt{e}|\partial_z \mathbf{u}|^2}{\tilde{J}_b + J_b^{\min}}, \quad (47)$$

559 where $d = -z$ is depth, which measures the relative contribution of shear production
560 (the numerator) versus buoyancy flux (the denominator) to the TKE budget in unstable
561 stratification. We then use this estimate to reduce the convective mixing length by

$$562 \quad \epsilon_{sp} \stackrel{\text{def}}{=} \max(0, 1 - \mathbb{C}^{sp} \widetilde{Ri}_f), \quad (48)$$

563 where \mathbb{C}^{sp} is a free parameter. The reduction factor (48) is used in lieu of more detailed
564 understanding of how shear acts to limit turbulence correlation scales during convection.
565 Note that the numerator in (47) estimates shear production using the mixing length d , which
566 is appropriate for shear-driven turbulent mixing. This formulation means that the free
567 convection length scale is more limited at depth, where convective plumes are less connected
568 to destabilizing surface buoyancy fluxes.

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

$$570 \quad \ell_\psi^{\text{conv}}(z) = \epsilon_{sp} \begin{cases} \ell_\psi^h & \text{if } N^2 < 0 \text{ and } J_b > 0, \\ \ell_\psi^p & \text{if } N^2 > 0, N^2(z + \Delta z) < 0, \text{ and } J_b > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (49)$$

571 Figure 4 illustrates the behavior of the convective length scale predicted by CATKE in (49)
572 for three free convection cases with surface buoyancy fluxes $J_b = 9.6 \times 10^{-7}$, 2.4×10^{-7} ,
573 and $8.8 \times 10^{-8} \text{ m}^2 \text{s}^{-3}$ integrated for 6, 24, and 72 hours respectively, using the initial
574 buoyancy profile in equation (A1), which is also used for all our LES. Figure 4(a) shows
575 CATKE-simulated buoyancy profiles after integrating for 6, 24, and 72 hours. Figure 4(b)
576 shows that stronger forcing cases have greater levels of turbulent kinetic energy. Figure 4(c)
577 shows the tracer mixing length, which above $z = -100$ meters is dominated by the convective
578 mixing length. Though each case has different TKE and different surface buoyancy flux,
579 they nevertheless predict similar tracer mixing lengths which are $O(100)$ meters and thus
580 similar to the boundary layer depth, corroborating the dimensional analysis in equation (39).
581 Figure 4(d) shows the eddy diffusivity for the three cases — unlike a typical constant-
582 diffusivity convective adjustment model, CATKE’s “convective adjustment diffusivity” varies
583 depending on the strength of the surface buoyancy flux. Because the predicted mixing length
584 is similar for all three cases, the tracer diffusivity varies with the surface buoyancy flux due
585 to variation in the turbulent kinetic energy.

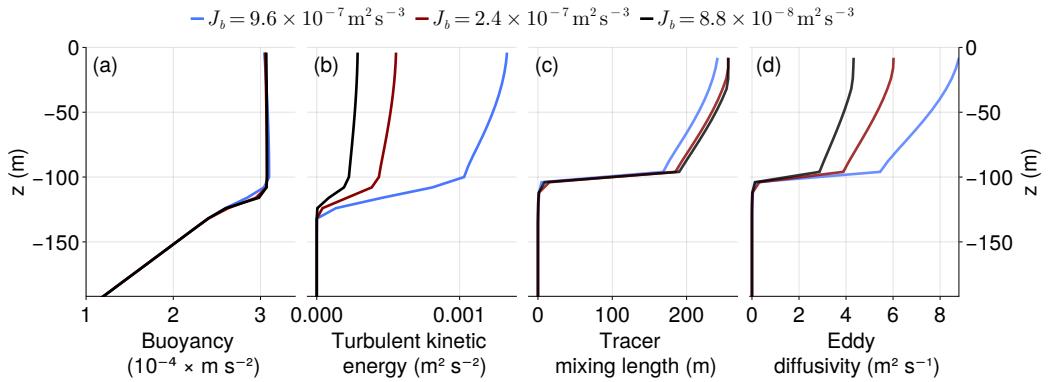


Figure 4. CATKE mixing length and eddy diffusivity during free convection for three cases with boundary layer depth $h \approx 100$ m. (a) CATKE-predicted buoyancy profiles for the three cases, (b) profiles turbulent kinetic energy, e , (c) tracer mixing lengths ℓ_c , (d) tracer eddy diffusivities κ_c . 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 $^{-2}$.

4 A posteriori calibration against large eddy simulations

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

4.1 The importance of *a posteriori* calibration

Explicitly minimizing simulation bias distinguishes *a posteriori* calibration from other methods that minimize other biases that are only indirectly related to simulation bias — 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 critically on additional and often problematically strong 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 parameterized single column equation for buoyancy b ,

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

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 ξ that represents spatial and temporal discretization errors. We additionally define the ideal or “perfect” solution as b^\dagger . When equation (50) is integrated forward to predict the evolution of b , fluctuations away from the perfect solution b^\dagger inevitably develop due both to structural errors in \mathcal{J} and because of the discretization error ξ , leading to an error $\varepsilon = b - b^\dagger$ 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 $\mathcal{J}(b^\dagger)$ in terms of the *perfect* b^\dagger , 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 $\mathcal{J}(b) - \mathcal{J}(b^\dagger)$ 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 leverage this feature to realize a key innovation of this work: we explicitly minimize spatial discretization error by including single-column simulations with 2-, 4-, and 8-meter resolution in our loss function.

634 4.2 Ensemble Kalman Inversion for *a posteriori* calibration

The downside of *a posteriori* calibration is that nonlinear inverse problems are difficult to solve. In this work we use an ensemble-based, gradient-free method called Ensemble Kalman Inversion (EKI; Iglesias et al., 2013). A major advantage of EKI is that it does not require a gradient or adjoint of the CATKE-parameterized single column model. Instead, EKI only requires the ability to evaluate the loss functions for an ensemble of free parameters. The EKI algorithm can be construed either as the integration of a dynamical system or as an iterative scheme for repeatedly refining an initial distribution of free parameter values.

EKI minimizes the “EKI objective function” Φ , defined as

$$643 \quad \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}) \stackrel{\text{def}}{=} \|\mathcal{M}^{-1/2} [\mathcal{G}(\mathbb{C}) - \mathcal{Y}] \|^2, \quad (51)$$

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

We note that the near-surface dynamics in the LES seems uncertain. For example, the LES profiles exhibit strong unstable near-surface buoyancy gradients for strongly-forced convective cases. Though these features are robust to changes in LES resolution (see Appendix A), we are unsure whether the simple implicit LES turbulence closure is missing crucial turbulent mixing processes important near a wavy, bubbly, broken ocean surface. We therefore omit the top 8 meters of the LES domain from \mathcal{Y} to avoid overconstraining parameters based on the most uncertain elements of the LES data.

EKI finds a set of optimal parameters $\mathbb{C} = \mathbb{C}_*$ that minimize $\Phi(\mathcal{G}, \mathcal{Y}, \mathbb{C})$ in (51) by evolving an ensemble of parameter sets using the algorithm described in Appendix C. In this work we use relatively large ensembles with 1000 members. This means that every EKI iteration requires performing up to $21 \times 3 \times 1000 = 63,000$ single column simulations,

662 corresponding to 21 LES cases and 3 vertical resolutions for every ensemble member. To
 663 make the calibration as efficient as possible, we implement CATKE in Oceananigans and
 664 leverage a feature that permits us to integrate an ensemble of single column models in parallel
 665 in the configuration of a single three-dimensional simulation on a GPU. As a result, each
 666 EKI iteration requires evolving 9 effectively three-dimensional simulations (3 resolutions
 667 for each of the 12-, 24- and 48-hour suites). On an Nvidia Titan V GPU and with 1,000
 668 ensemble members, a single EKI iteration takes 40–50 seconds, and the entire calibration
 669 takes 4–6 hours. In the course of this work we have performed complete calibrations of
 670 CATKE’s parameters hundreds of times — to experiment with new formulations, new
 671 numerical schemes, and to tweak the calibration setup. This workflow represents a new
 672 “calibration-based” paradigm in parameterization development, where physical formulation or
 673 numerical implementation changes are tested against the baseline by comparing predictions
 674 for independently calibrated parameterizations. The 23 calibrated free parameters that
 675 correspond to the version of CATKE described in this paper and the previously described
 676 LES are listed in table 3.

677 5 Validation

678 We next assess CATKE’s ability to make accurate predictions in a single column context
 679 with the free parameters listed in table 3. First, we derive quantities with well-understood
 680 physical interpretations from CATKE’s free parameters, and evaluate whether their calibrated
 681 values are close to expected or directly measured values reported in the literature. Second,
 682 we compare CATKE-parameterized simulations both to the 21 constant-forcing LES used
 683 for calibration and to an additional 12 constant-forcing LES that are both more strongly
 684 and more weakly forced than the calibration LES. Third, we conduct a 34-day CATKE-
 685 parameterized simulation of equatorial deep-cycle turbulence using the dataset provided
 686 by Whitt et al. (2022), and then compare the results to the LES used therein. This third
 687 validation context is useful because it involves both time-dependent surface forcing, solar
 688 insolation, and lateral flux divergences derived from a high resolution tropical GCM. Finally,
 689 we evaluate CATKE’s sensitivity to vertical resolution and time-step size. These all provide
 690 a measure of confidence in CATKE’s ability to not only represent the LES data used for
 691 calibration but also to extrapolate to differently-forced conditions, time-dependent surface
 692 forcing, and GCM-like contexts that include lateral flux divergences from for example, the
 693 advection of momentum, temperature, and salinity. All of this said, we maintain a caveat
 694 that CATKE should still be assessed, and likely recalibrated, in a regional or global context
 695 that is more similar to the context in which CATKE is intended to be used.

696 5.1 Derived quantities

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

701 5.1.1 Steady-state Richardson number

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

$$704 Ri^\dagger \stackrel{\text{def}}{=} \frac{\mathbb{C}_u^0}{\mathbb{C}_c^0 + \mathbb{C}_D^0} \approx 0.18, \quad (52)$$

705 which lies in the “near-neutral” stability function regime, since $\mathbb{C}_{Ri}^0 = 0.25 > Ri^\dagger$. $Ri^\dagger = 0.18$
 706 is somewhat less than the 0.23 used by Blanke and Delecluse (1993), or the celebrated value
 707 $Ri = 1/4$ that determines the stability of a laminar stratified shear layer. In section 5.3, we

Symbol	Description	Optimal value	Bounds
$\mathbb{C}_J^{\text{shear}}$	Wind stress TKE surface flux	3.18	(0, 2)
$\mathbb{C}_J^{\text{conv}}$	Convective TKE surface flux	0.38	(0, 2)
\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^∞	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^∞	Velocity mixing for high Ri	0.24	(0, 2)
\mathbb{C}_e^h	TKE free convection scale	3.64	(0, 8)
\mathbb{C}_e^-	TKE transport for negative Ri	1.44	(0, 8)
\mathbb{C}_e^0	TKE transport for near-neutral Ri	7.86	(0, 8)
\mathbb{C}_e^∞	TKE transport for high Ri	0.55	(0, 8)
\mathbb{C}_D^h	Dissipation free convection scale	3.25	(0, 8)
\mathbb{C}_D^-	Dissipation scale for negative Ri	0.92	(0, 8)
\mathbb{C}_D^0	Dissipation scale for near-neutral Ri	1.60	(0, 8)
\mathbb{C}_D^∞	Dissipation scale for high Ri	0.58	(0, 8)
\mathbb{C}_{Ri}^0	Stability function transitional Ri	0.25	(0, 2)
\mathbb{C}_{Ri}^δ	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.

708 find that Ri^\dagger is a crucial parameter controlling mixing in forced stably-stratified turbulence,
 709 and that LES tend to exhibit Ri in the range 0.2–0.23.

710 5.1.2 Near-surface similarity constant

711 Section 3.1.3 shows how a near-surface similarity constant — analogous to the von
 712 Kármán constant for turbulence near rigid non-wavy walls — may be computed from the
 713 near-wall and momentum stability function parameters. In terms of the parameters in table 3
 714 from (35) we find that

$$715 \quad \varkappa = \mathbb{C}^s \left[(\mathbb{C}_u^0)^3 / \mathbb{C}_D^0 \right]^{1/4} \approx 0.47, \quad (53)$$

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

720 A similar wave-induced enhancement to the similarity constant is proposed by Samelson
 721 (2022). However, Samelson (2022) models the enhancement as a function of wind at ten
 722 meters height, u_{10} . In our case, the LES are forced with varying u_{10} , but constant Langmuir
 723 number $La \approx 0.3$ (see table 1 for a summary of the LES cases). Thus we must either
 724 hypothesize that surface waves can be modeled with a La -dependent enhancement of \varkappa , or
 725 that CATKE is missing physics. Either way, we are unable to proceed further in determining
 726 wave-induced enhancements to \varkappa without LES that vary both u_{10} and La , so we save such
 727 considerations for future work.

728 5.1.3 The turbulent Prandtl number

729 The turbulent Prandtl number is defined as

$$730 \quad Pr \stackrel{\text{def}}{=} \frac{\kappa_u}{\kappa_c}, \quad (54)$$

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

- 732 • $Pr_c \approx 0.77$ for weakly-sheared convection,
- 733 • $Pr_- \approx 0.65$ for unstably-stratified shear turbulence,
- 734 • $Pr_0 \approx 0.98$ for near-neutral shear turbulence,
- 735 • $Pr_\infty \approx 2.46$ for strongly-stratified shear turbulence.

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

743 5.1.4 The turbulent Schmidt number

744 Calibration determines that $Sc = 0.26$ for unstably-stratified shear turbulence with
 745 $Ri < 0$, 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
 746 result, TKE is transported much more rapidly than momentum or tracers in shear-dominated
 747 turbulence, and similarly to momentum or tracers in convective or weakly-sheared stratified
 748 turbulence. Rapid TKE diffusion relative to momentum or tracer diffusion introduces an
 749 “implicitly non-local” element to CATKE’s mixing predictions, because TKE transport can
 750 generate mixing in a region that is displaced from the region of TKE generation.

Symbol	Value	Description
Ri^\dagger	0.18	Steady-state gradient Richardson number
\varkappa	0.47	Near-neutral near-surface similarity constant
Pr_0	0.98	Near-neutral turbulent Prandtl number ($Ri \rightarrow 0$)
Pr_∞	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 ($Ri \rightarrow -\infty$)
Γ_0	0.23	Near-neutral mixing coefficient ($Ri \rightarrow 0$)
Γ_∞	0.17	Strongly-stratified mixing coefficient ($Ri \rightarrow \infty$)
Sc_0	0.046	Near-neutral turbulent TKE Schmidt number ($Ri \rightarrow 0$)
Sc_∞	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.

5.1.5 Stratified turbulence mixing coefficient

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

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

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

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

In this section, we validate CATKE’s ability to make predictions both for within and outside the range of surface forcings used for calibration. To add context to this validation exercise and connect with other studies, we include a comparison with predictions from the K -profile parameterization (KPP; Large et al., 1994), and the “Langmuir turbulence” second-moment closure (SMC-LT) described by Harcourt (2015), whose results depend additionally on the Stokes drift profile we used for LES. All simulations, including those with KPP and SMC-LT, use staggered vertical grids with 128 points, in a 256-meter deep domain and thus with 2-meter vertical resolution. We use a 5-minute time step for CATKE, a 2-minute time step for KPP, and a 1-second time-step for SMC-LT. Such a short time-step was used for SMC-LT because we observed that the results were sensitive to time steps 20 seconds and longer for the strong forcing cases.

We should treat these comparisons with some caution, because KPP or SMC-LT were calibrated to somewhat different datasets than what we use for CATKE. That said, we find

777 that for every constant-forcing case, CATKE predicts the boundary layer depth simulated by
 778 LES — both inside and outside the training dataset — more accurately than either KPP or
 779 SMC-LT. This is an important result because boundary layer is a key metric determining the
 780 short-term sensitivity of climate predictions (Gregory, 2000; Held et al., 2010). Moreover and
 781 by design of the calibration problem (because we omit the upper 4 meters of the LES profiles
 782 from the error estimate), CATKE predicts more well-mixed near-surface profiles during
 783 convection, and thus warmer sea surface temperatures, than either KPP or SMC-LT. With
 784 this broad summary of CATKE’s main successes stated, we focus the subsequent discussion
 785 for each case on CATKE’s biases and areas to focus on for future improvements.

786 5.2.1 Constant forcing validation: free convection

787 We begin with the free convection cases plotted in figure 5. The free convection cases
 788 represent some of the best predictions of KPP and SMC-LT. Boundary layer depth is well-
 789 predicted by all parameterizations to within 10 meters, with perhaps the greatest bias coming
 790 from SMC-LT in the weakly-forced 72-hour case — despite that KPP has known structural
 791 biases for representing free convection (Souza et al., 2020). Oddly, for the more strongly
 792 forced cases, a large portion of the KPP profiles are stably stratified within the boundary
 793 layer, and capped by a very strong unstable stratification near the surface. Of the three,
 794 CATKE’s convective mixing length most capably keeps the boundary layer nearly-neutrally
 795 stratified during strong free convection.

796 For near-surface buoyancy (and equivalently sea surface temperature, or SST) the three
 797 parameterizations make somewhat different predictions. For example, CATKE predicts
 798 a nearly-mixed boundary layer due to its convective mixing length, which means that it
 799 predicts a warmer SST. On the other hand KPP, SMC-LT, and the LES all predict layers
 800 (of varying thickness) of unstable stratification next to the surface, and thereby also predict
 801 substantially colder SST than CATKE. Caution is probably warranted when interpreting
 802 the LES results, however: our LES may exhibit spuriously reduced mixing near the upper
 803 boundary where the simulated scale of turbulent eddies shrinks significantly below the grid
 804 scale. Addressing this uncertainty in the LES data requires the use of observations of the
 805 near-surface temperature profiles to inform modifications to the LES, which is left for future
 806 work.

807 The buoyancy profiles in figure 5 reveal bias in CATKE’s predictions of the detailed
 808 structure of the lower half of the convecting boundary layer. One contribution to this bias
 809 is well-known: in free convection, buoyancy fluxes in the lower half of the boundary layer
 810 are upgradient. In order to accurately capture the boundary layer depth, CATKE must
 811 accurately predict the buoyancy flux — and therefore cannot avoid erroneously predicting
 812 a slightly unstably stratified buoyancy profile where in the LES the profile is either nearly
 813 mixed or actually slightly stably stratified. No amount of calibration or additional free
 814 parameters can fix this bias given CATKE’s downgradient formulation — the only recourse
 815 is to introduce a non-downgradient, and therefore non-local, contribution to CATKE’s
 816 fluxes. For example, CATKE could be augmented with a mass flux scheme in the manner of
 817 Siebesma et al. (2007); Giordani et al. (2020). Other alternatives include evolving fluxes
 818 directly as in Garanaik et al. (2024), or adding additional tracer variance equations and
 819 computing non-gradient fluxes in terms of those (Legay et al., 2024). But even this may not
 820 be sufficient — for example, even though KPP has a non-local model for fluxes, it still has
 821 significant biases in convective boundary layer buoyancy structure.

822 To investigate CATKE’s free convection bias further, figure 6 compares CATKE’s
 823 predictions of the forced passive tracer profile with LES. This comparison reveals that while
 824 CATKE generally models the tracer profile well (except for the extreme, extrapolating, 6-hour
 825 case in panel a), CATKE tends to overmix especially in the lower part of the boundary
 826 layer, where the LES profiles exhibit a slight peak and a bit more shape. Thus in addition
 827 to lacking a non-local contribution to fluxes, CATKE also overpredicts mixing to some

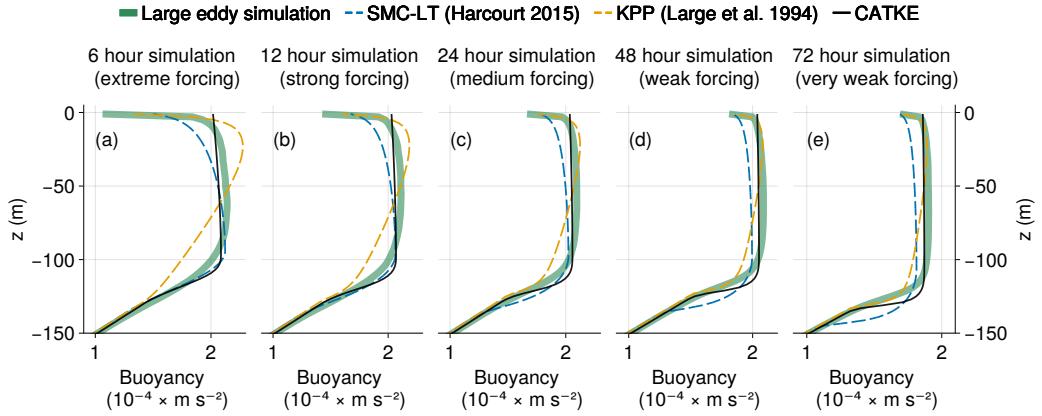


Figure 5. A four-way comparison for the “free convection” constant forcing cases described in 1 and Appendix A between LES, CATKE, the K -profile parameterization (KPP Large et al., 1994), and the Langmuir turbulence second moment closure described by Harcourt (2015) (SMC-LT). KPP and SMC-LT are implemented in the General Ocean Turbulence Model (GOTM, Umlauf & Burchard, 2005). Panels (a)–(e) show free convection for 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 W m^{-2} . The initial condition is density stratified with a depth-varying buoyancy gradient that varies between 10^{-6} s^{-2} and $2 \times 10^{-5}\text{ s}^{-2}$.

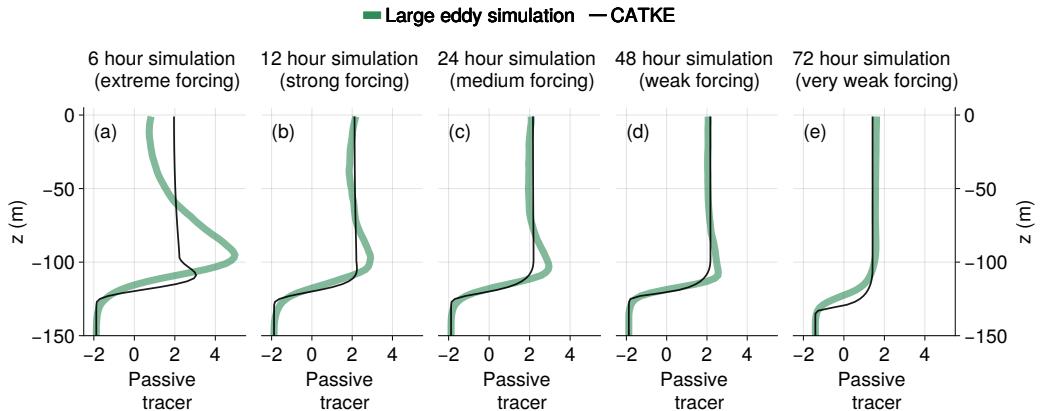


Figure 6. Comparison between the forced passive tracer profile simulated by LES and CATKE for free convection. The passive tracer forcing, which is described in appendix A2, is a Gaussian centered on $z = -96\text{ m}$ and 8 m wide. The strength of the forcing depends on the suite: the 6-, 12-, 24-, 48-, and 72-hour suites use 15 minute, 30 minute, 1 hour, 2 hour, and 4 hour forcing time scales, respectively.

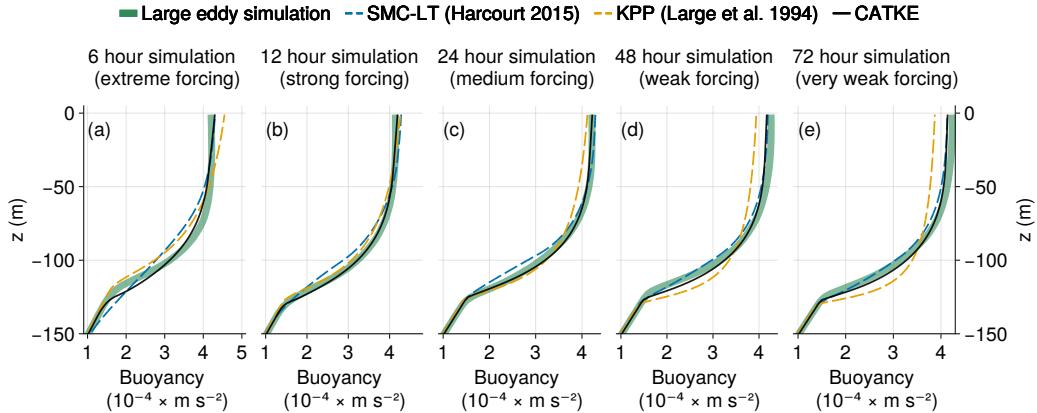


Figure 7. A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “strong wind, no rotation” constant forcing cases described in table 1 and Appendix A. The strong wind, no rotation cases are forced by surface stresses that correspond roughly to $9\text{--}22 \text{ m s}^{-1}$ atmospheric wind at a height of 10 m. See figure 5.

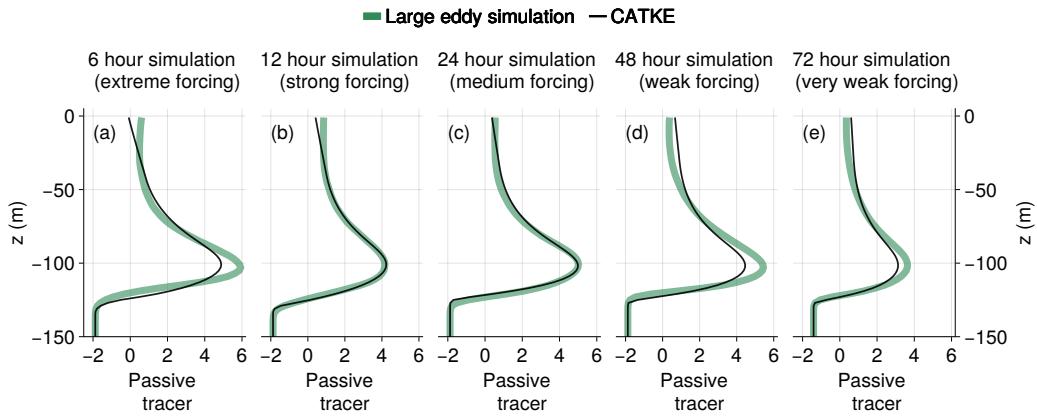


Figure 8. Comparison between the forced passive tracer profile simulated by LES and CATKE for strong wind, no rotation. See figure 6.

degree, especially near the base of the boundary layer. Solving this bias could simultaneously motivate adding non-local contributions to convective fluxes as well as modifying the depth structure of the convective mixing length.

5.2.2 Constant forcing validation: shear-driven turbulence

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

A comparison between LES, SMC-LT, KPP, and CATKE for the strong wind, no rotation case is shown in figure 7. All parameterizations make similar and good predictions for boundary layer depth and surface temperature, except for SMC-LT in the 6-hour case, where it overmixes slightly. A comparison between CATKE and LES simulations of the

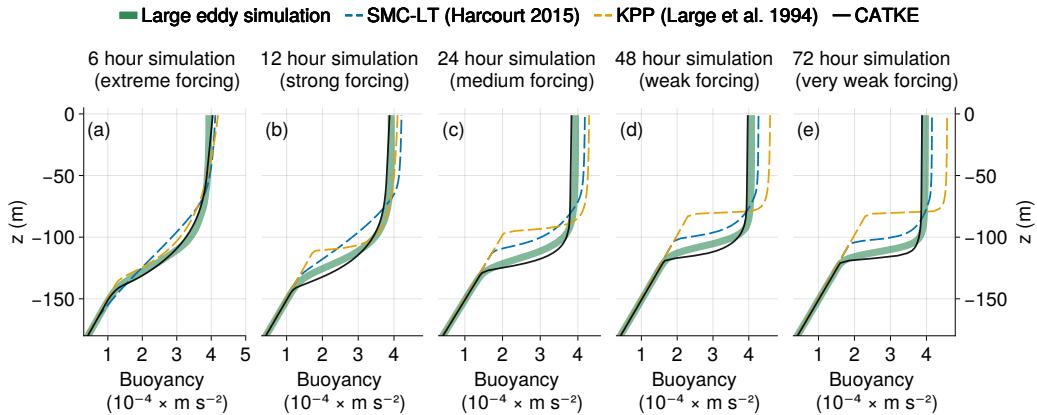


Figure 9. A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the strong wind constant forcing cases described in table 1 and Appendix A. The strong wind cases are rotating with Coriolis parameter $f = 10^{-4} \text{ s}^{-1}$ and forced by surface stresses that correspond roughly to $15\text{--}24 \text{ m s}^{-1}$ atmospheric wind at 10 meters height. See figure 5.

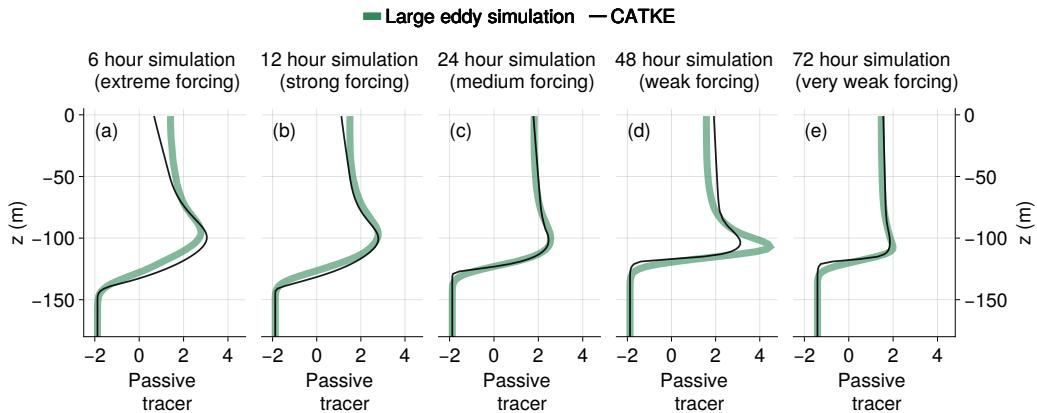


Figure 10. Comparison between the forced passive tracer profile simulated by LES and CATKE for strong wind. See figure 6

forced passive tracer for the strong wind, no rotation case is shown in figure 8, 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 9 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 good but exhibit a tendency to overmix slightly, resulting in boundary layers that are approximately 5% too deep. Figure 10 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.

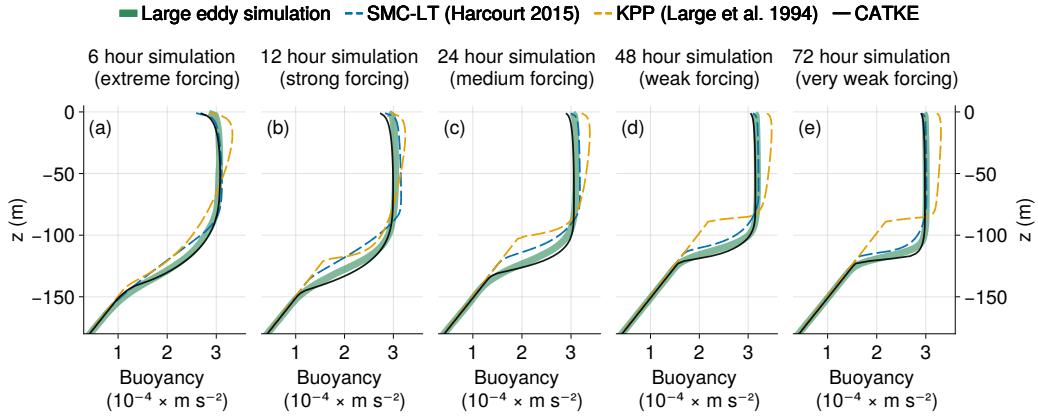


Figure 11. A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “strong wind, weak cooling” constant forcing cases described in table 1 and Appendix A. The strong wind weak cooling cases are rotating with Coriolis parameter $f = 10^{-4} s^{-1}$, forced by surface stresses that correspond roughly to $14\text{--}23 m s^{-1}$ atmospheric wind at 10 meters height, and destabilizing buoyancy fluxes that correspond roughly to heat fluxes between $79\text{--}833 W m^{-2}$. See figure 5.

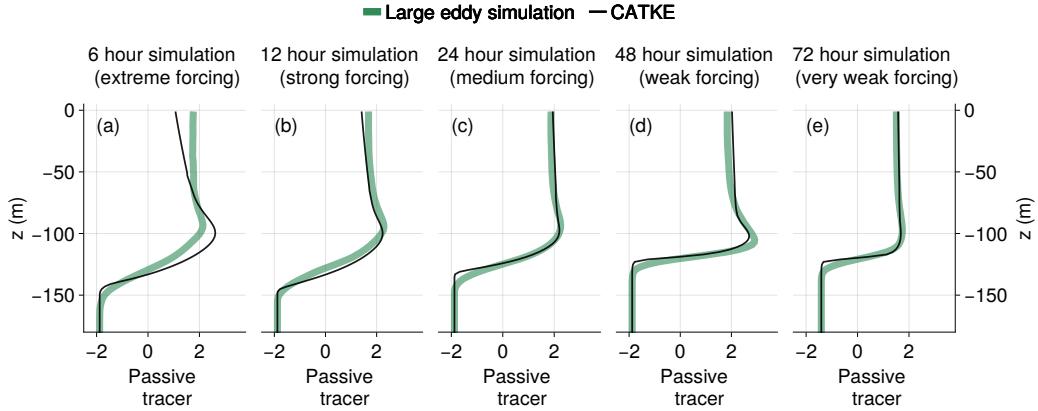


Figure 12. Comparison between the forced passive tracer profile simulated by LES and CATKE for strong wind, weak cooling. See figure 6.

852

5.2.3 Constant forcing validation: mixed shear and convective turbulence

853

CATKE simulations are also accurate 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 11, 13, and 15. KPP exhibits significant shallow bias for all cases. SMC-LT exhibits less shallow bias than KPP, but still more than CATKE. Because KPP and SMC-LT also predict a spuriously strong unstable buoyancy gradient near the surface (compared to the present LES), the SST biases are more variable. CATKE, on the other hand, makes good predictions for all cases except in the weak wind, strong cooling cases where it overmixes.

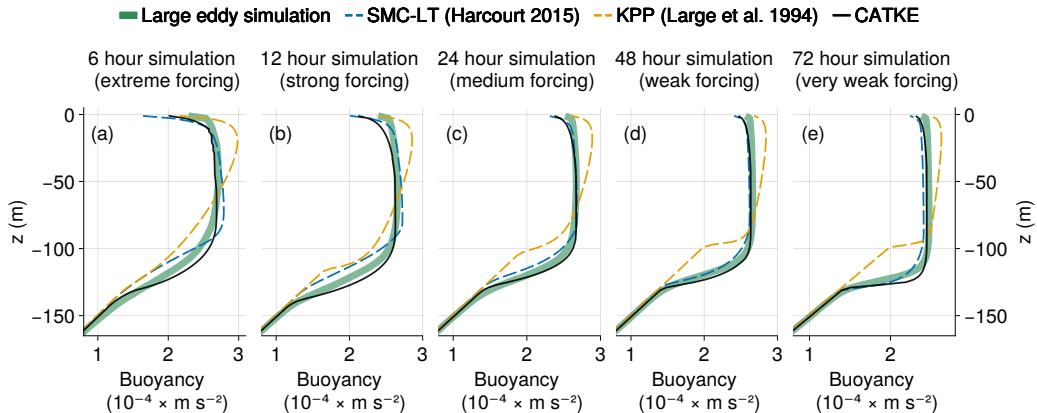


Figure 13. A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “mid wind, mid cooling” constant forcing cases described in table 1 and Appendix A. The mid wind mid cooling cases are rotating with Coriolis parameter $f = 10^{-4} \text{ s}^{-1}$, forced by surface stresses that correspond roughly to $13\text{--}20 \text{ m s}^{-1}$ atmospheric wind at 10 meters height, and destabilizing buoyancy fluxes that correspond roughly to heat fluxes between $125\text{--}1333 \text{ W m}^{-2}$. See figure 5.

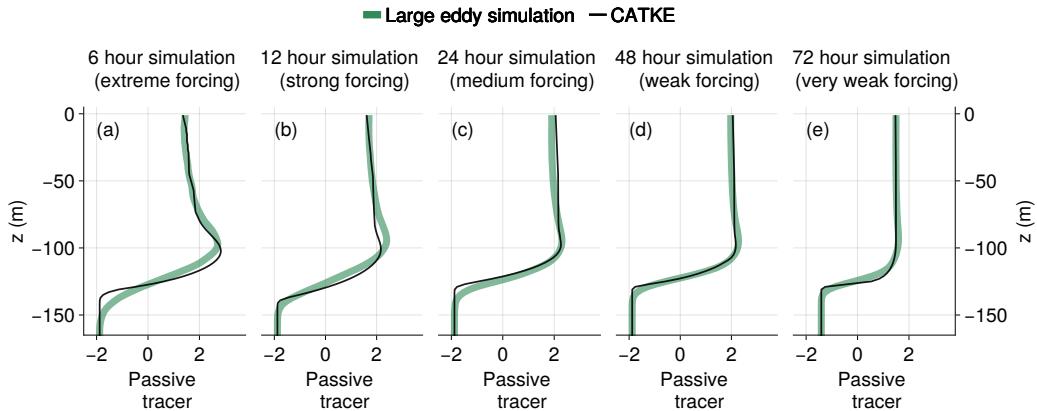


Figure 14. Comparison between the forced passive tracer profile simulated by LES and CATKE for mid wind, mid cooling. See figure 6.

Figures 12, 14, and 16 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 11 and the tracer profiles in figure 12. This shows that the most difficult cases are free convection and “weak wind, strong cooling”—the cases where convective dynamics dominate.

The “weak winds, strong cooling” case is the most challenging for CATKE. For this case, the 72-hour LES is forced by 156 W m^{-2} equivalent heat flux and 11 m s^{-1} 10-meter atmospheric winds, while the 6-hour LES is forced by 1666 W m^{-2} and 16 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

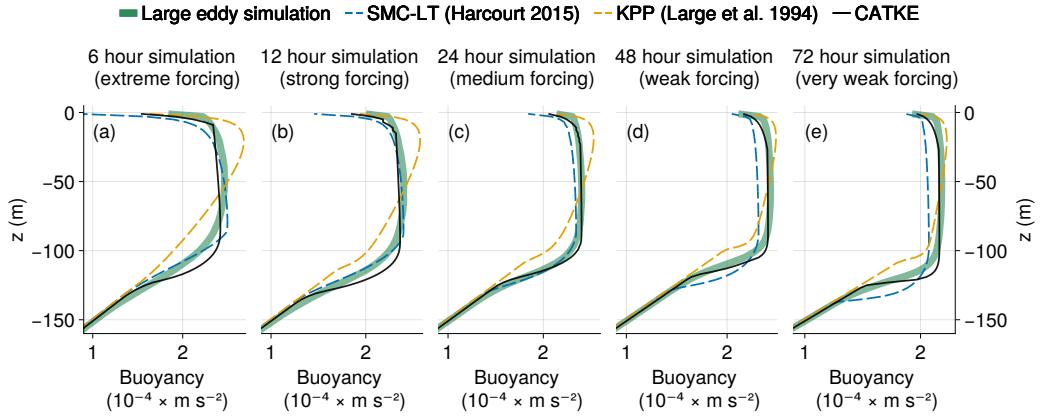


Figure 15. A four-way comparison between LES and three turbulence closures (CATKE, KPP, and SMC-LT) for the “weak wind, strong cooling” constant forcing cases described in table 1 and Appendix A. The weak wind strong cooling cases are rotating with Coriolis parameter $f = 10^{-4} \text{ s}^{-1}$, forced by surface stresses that correspond roughly to $11\text{--}16 \text{ m s}^{-1}$ atmospheric wind at 10 meters height, and destabilizing buoyancy fluxes that correspond roughly to heat fluxes between $156\text{--}1666 \text{ W m}^{-2}$. See figure 5.

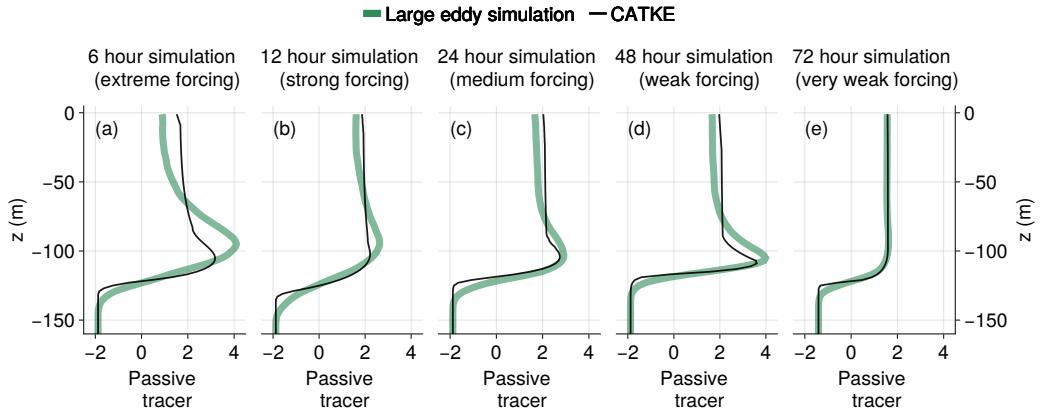


Figure 16. Comparison between the forced passive tracer profile simulated by LES and CATKE for weak wind, strong cooling. See figure 6. See figure 6.

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.

880 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 depth, even in convective dominated cases where an analysis of tracer profiles suggests that CATKE tends to overmix. Fixing CATKE's convective 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 an ambiguity in comparing the three: 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? We cannot answer this question conclusively. While KPP has known structural biases (see, for example, Souza et al., 2020), the formulation of SMC-LT is seemingly more general than CATKE. We therefore suspect that *a posteriori* calibration of SMC-LT will allow it to make predictions that are as or more accurate than CATKE. And until this calibration is performed, any judgments about the biases of SMC-LT must be taken with a grain of salt.

900 5.3 Deep cycle turbulence in the tropics

We next 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).

Figure 17 illustrates the complex dynamics of this situation by showing vertical kinetic energy from the LES, TKE from CATKE, and Ri from days 8 to 13 of the time-series. In this topical turbulence scenario, a combination of wind stress and stabilizing solar insolation in daytime produces a shallow, stably-stratified jet in the upper ~ 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 the boundary layer eventually leads to shear instability which spans the entire, roughly 100 m depth of the region below the mixed layer. More often than not, the turbulence “pulsates”—initial bursts of turbulence mix momentum and buoyancy and thus decay rather quickly, only to 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 17a. Figure 17b shows that CATKE exhibits a qualitatively similar bursting behavior, though the timing of the bursts are sometimes misrepresented. Moreover, inspection of the Richardson number plotted in figures 17c and d

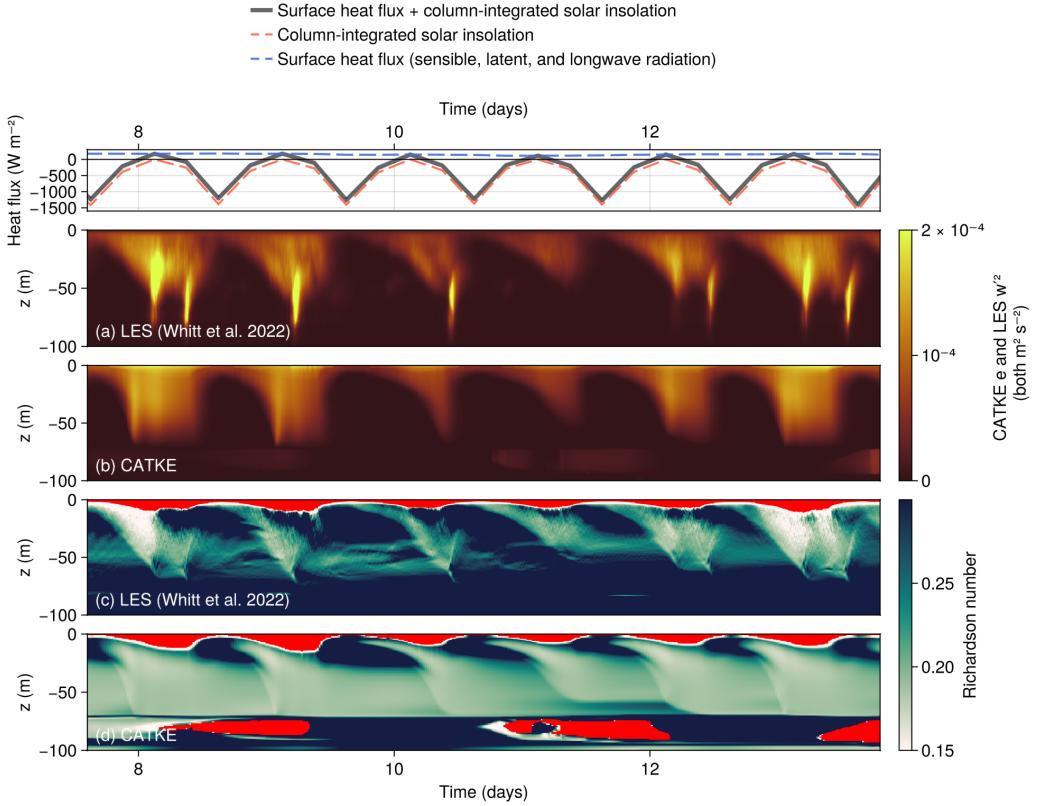


Figure 17. Overview of the tropical turbulence validation case. Panels show: (a) the forcing and heat fluxes, (b) vertical kinetic energy $\overline{w^2}$ from the LES described by Whitt et al. (2022), (c) CATKE’s TKE variable, (d) the Richardson number computed from the horizontally-averaged LES momentum 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.

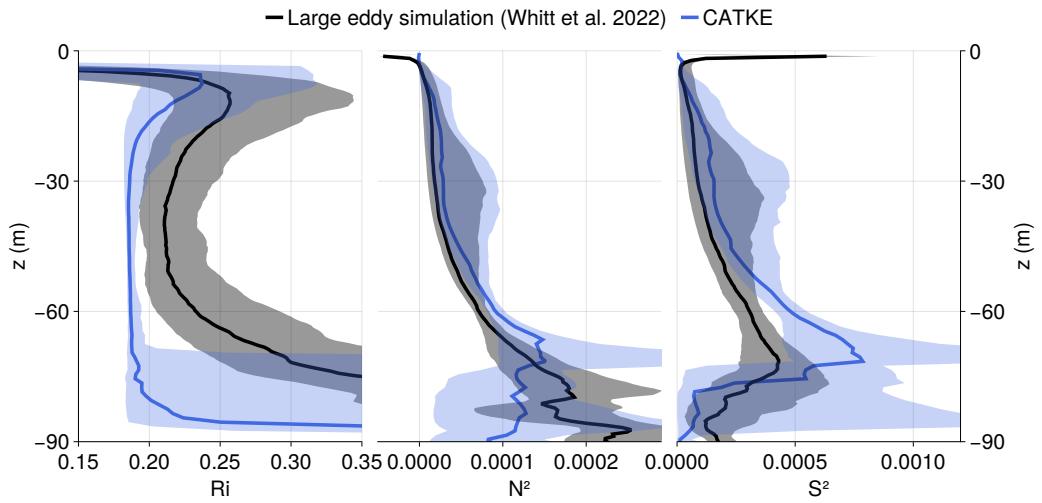


Figure 18. Median Ri , shear (S^2), and buoyancy frequency N^2 at each depth computed from 34 days of realistic equatorial turbulence simulated by LES and CATKE. The LES Ri is computed in terms of the horizontally-averaged shear and buoyancy. Shading shows the range between the 25% and 75% quantiles. CATKE’s prediction of Ri is narrowly peaked around its steady-state Richardson number, $Ri^\dagger = 0.18$. This reveals a bias in CATKE: the median Ri in the LES is more variable and in particular, does not reach values as low as 0.18. Turning to the buoyancy gradient and shear, it seems that CATKE overpredicts both. This reflects complexity: apparently CATKE undermixes both momentum and buoyancy, but exhibits *more bias* for momentum mixing, which permits the development of lower Ri than observed in the LES. Given that CATKE has already been calibrated to cases that presumably exhibit similar stratified shear mixing physics as this tropical turbulence case, fixing the Ri , N^2 , and S^2 biases may require changing the formulation of CATKE’s stability functions.

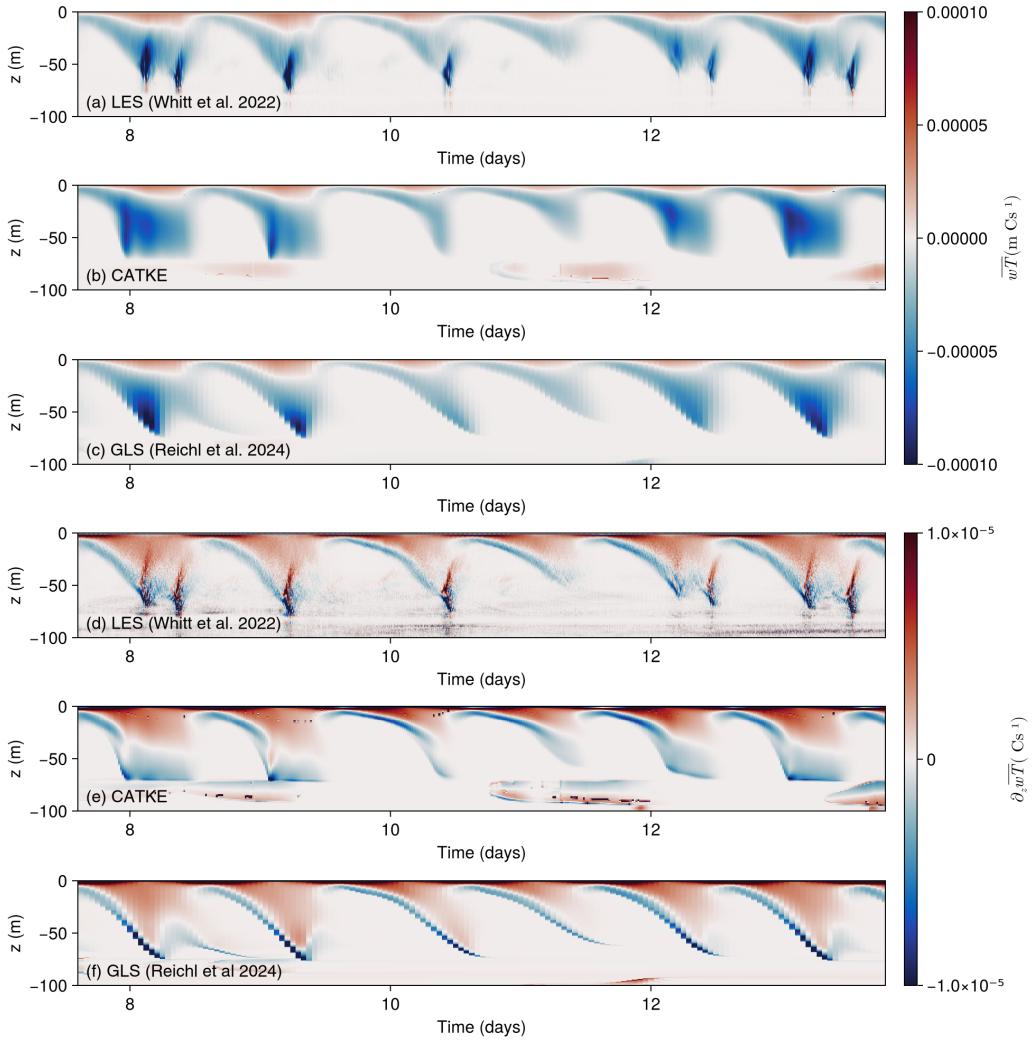


Figure 19. A comparison of the vertical temperature flux and vertical temperature flux divergence in tropical turbulence between LES (Whitt et al., 2022), CATKE, and the Generic Length Scale (GLS) turbulence closure as reported by Reichl et al. (2024).

reveals that CATKE sometimes underpredicts, and sometimes overpredicts the Richardson number. Figure 18 investigates this 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 are striking: while the Ri in the LES 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, we argue that CATKE's predictions should be interpreted as relatively accurate. To make this point, figure 19 compares the vertical temperature flux and flux divergence between the LES, CATKE, and a third single column GOTM run that uses the “Generic Length Scale” (GLS) closure reported by Reichl et al. (2024). GLS is a second-moment closure similar to SMC-LT (Umlauf &

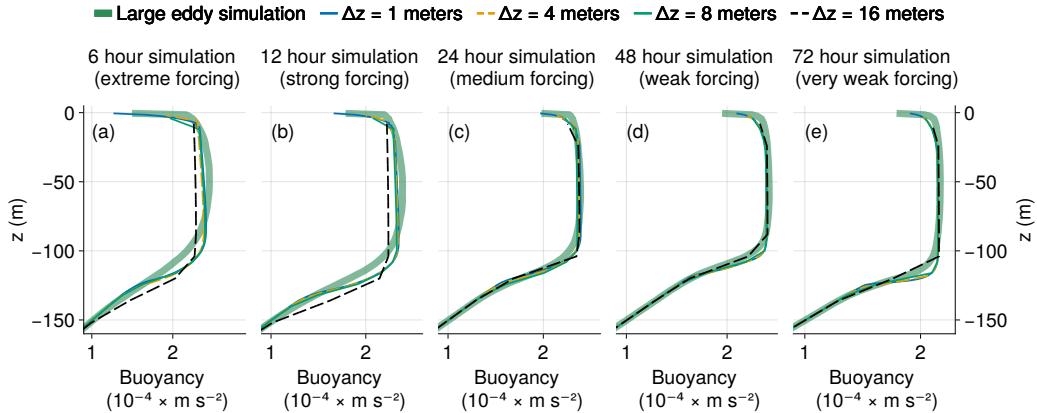


Figure 20. Illustration of sensitivity of CATKE predictions to vertical resolution for the weak wind, strong cooling case. Four vertical resolutions are shown: 1, 4, 8, and 16 meters. CATKE’s calibration explicitly minimized errors between LES and CATKE simulations at 2, 4, and 8 meter resolution, such 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.

Burchard, 2003), which is used to facilitate a comparison with Reichl et al. (2024). For some reason, the bursting behavior observed in both the LES and CATKE solutions is absent from GLS — suggesting that CATKE may hold an edge over GLS (at least, the GLS implemented in GOTM with default free parameter choices) in modeling intermittent forced stratified shear turbulence. The vertical structure of the flux divergences in CATKE are also more similar to the LES than the GLS solution.

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 20 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”. Based off the results in figure 20, we preliminarily conclude that CATKE is insensitive to vertical resolutions 8 meters and finer. At 16 meter resolution, CATKE’s predictions are still good compared to the biases observed for KPP and SMC-LT, but nevertheless start to deviate from the higher-resolution solutions and, in particular, tend to overmix. It may be that with such a coarse resolution, it simply is not possible to resolve the structure of strongly-stratified entrainment layers at the base of the boundary layer.

The sensitivity of CATKE’s predictions to time-step size — at a vertical resolution of 1 meter — are shown in figure 20. Note that CATKE requires a smaller time step for finer 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 21, 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

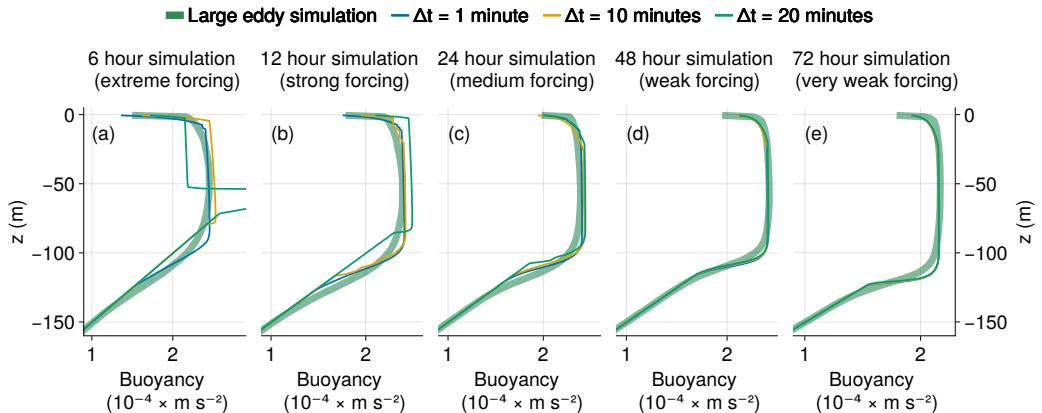


Figure 21. Sensitivity of CATKE predictions to time step for 1 meter vertical resolution for the weak wind, strong cooling case. At 1 meter resolution and in the strong forcing conditions of the 12- and 6-hour suites, CATKE solutions show time-step dependence for time steps longer than 1 minute. To enable 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 22.

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 22, 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.

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.

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

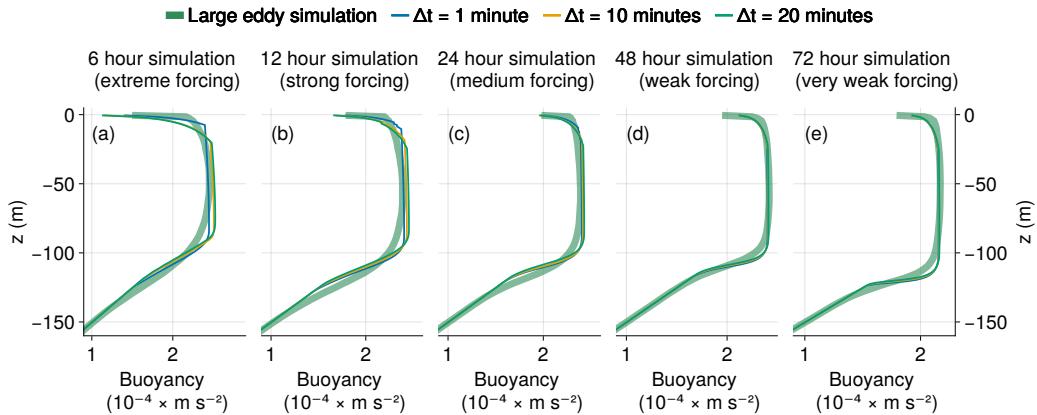


Figure 22. A comparison between LES and CATKE-parameterized single column simulations at 1 meter vertical resolution and three different momentum and tracer time-steps, when turbulent kinetic energy is substepped with a 30 second time step using the scheme described in Appendix B. For the 6-hour suite, the time-step dependence is greatly reduced compared to the non-substepped case shown in figure 21, 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.

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 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).

Our calibration to a relatively limited range of LES cases reported in this paper 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 high-resolution data, 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 limited LES context used 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 robust prior estimates of CATKE’s parameters in preparation of the final goal of calibrating CATKE in a global context, by minimizing the mismatch between predictions of the ocean climate state and relevant observations with global or near-global coverage. 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.

1021 **Appendix A A synthetic dataset generated by large eddy simulations**

1022 We use a synthetic dataset to calibrate and assess CATKE consisting of 35 idealized
 1023 large eddy simulations (LES) of the ocean surface boundary layer with imposed constant
 1024 surface fluxes of temperature and momentum and a simple surface wave field.

1025 **A1 Initial conditions**

1026 The LES are initialized from rest with zero velocity and the piecewise-linear buoyancy
 1027 stratification

$$1028 b(z, t=0) = \begin{cases} N_1^2 z & \text{for } z > -h_1, \\ N_2^2 z + (N_2^2 - N_1^2) h_1 & \text{for } -h_2 > z > -h_1, \\ N_3^2 z + (N_3^2 - N_2^2) h_1 + (N_2^2 - N_1^2) h_1 & \text{for } z < -h_2, \end{cases} \quad (\text{A1})$$

1029 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 $h_2 = 72 \text{ m}$.

1030 **A2 Passive tracer forcing**

1031 We additionally simulate the evolution of a passive tracer c which is forced by

$$1032 F_c(z) = \omega_+ e^{-(z-z_c)^2/2\lambda_c^2} - \omega_-, \quad (\text{A2})$$

1033 where z_c is the depth of the forcing, λ_c is the width of the forcing, ω_+ is an inverse forcing
 1034 time-scale that varies between each suite, and ω_- is chosen so that F_c has zero mean, that is

$$1035 \omega_- \stackrel{\text{def}}{=} \frac{\omega_+}{L_z} \int_{-L_z}^0 e^{-(z-z_c)^2/2\lambda_c^2} dz \\ 1036 \approx \omega_+ \frac{\lambda_c \sqrt{2\pi}}{L_z}, \quad (\text{A3})$$

1037 where L_z is the depth of the domain. The approximation of the integral holds when the
 1038 forcing is far from boundaries, or when $-L_z \ll z_c - \lambda_c$ and $0 \gg z_c + \lambda_c$. We use $z_c = -96 \text{ m}$
 1039 and $\lambda_c = 8 \text{ m}$ for all cases. For the forcing time scale ω_+^{-1} , we use 15 minutes, 30 minutes,
 1040 1 hour, 2 hours, and 4 hours for the 6, 12, 24, 48, and 72 hour suites, respectively.

1041 **A3 Constant-flux boundary conditions**

1042 The 35 simulations differ in their boundary conditions and Stokes drift. The 35
 1043 simulations, which have different boundary conditions and S are organized into 5 “suites”,
 1044 each of which has 7 cases: free convection, weak wind strong cooling, medium wind medium
 1045 cooling, strong wind weak cooling, strong wind, strong wind no rotation, and strong wind
 1046 and sunny. The suites differ by both forcing strength and duration, simulating 6, 12, 24,
 1047 48, and 72 hours of boundary layer turbulence respectively. The forcing strength is chosen
 1048 for each suite and case so that the boundary layer deepens to roughly half the depth of the
 1049 domain; for example, the “6-hour suite” has the strongest forcing, and the “72-hour suite”
 1050 has the weakest forcing. “Strong wind no rotation” and “strong wind and sunny” use $f = 0$,
 1051 while the rest use the Coriolis parameter $f = 10^{-4} \text{ s}^{-1}$. The surface fluxes for the 35 LES
 1052 are summarized in tables 1 and 2. To draw a connection between the LES suites and real
 1053 air-sea flux conditions, tables 1 and 2 provide an estimate of heat fluxes Q for each case,
 1054 as well as an estimate of the atmospheric wind at 10 meters height using similarity theory
 1055 (reduced to the case of neutral buoyancy fluxes for simplicity, see Large and Yeager (2009)),

$$1056 u_{10} = \sqrt{\frac{|\tau_a|}{c_{10}}}, \quad \text{where} \quad c_{10} = \left(\frac{\kappa}{\log(10/\ell_r)} \right)^2, \quad \text{and} \quad \ell_r = 0.011 \frac{|\tau_a|}{g}, \quad (\text{A4})$$

1057 where ℓ_r is the Charnock roughness length given gravitational acceleration $g = 9.81 \text{ m s}^{-2}$,
 1058 $\kappa = 0.4$ 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 $\rho_a = 1.2 \text{ kg m}^{-3}$.

1061 A4 Stokes drift model

1062 For all wind-forced cases, we additionally impose a surface wave field with a surface
 1063 Stokes drift amounting to a constant “Langmuir number” $La = \sqrt{u_*}/U^S(z=0) \approx 0.3$. Our
 1064 Stokes drift prescription models a surface wave field with the friction-number-dependent
 1065 peak wavenumber

$$1066 \quad k_p = C_k \frac{g}{u_*^2}, \quad (\text{A5})$$

1067 where $u_* = \sqrt{|\tau_x|}$ is the water-side friction velocity, g is gravitational acceleration, and we
 1068 use $C_k = 10^{-6}$.

1069 We follow [Lenain and Pizzo \(2020\)](#) to estimate the depth-profiles of Stokes drift and
 1070 Stokes drift shear. The Stokes drift beneath a spectrum of deep-water waves is

$$1071 \quad U^S(z) = 2 \int_{k_p}^{k_i} e^{2kz} k \sqrt{gk} \chi(k) dk, \quad (\text{A6})$$

1072 where $\chi(k)$ is a one-dimensional wave spectrum that neglects “directional spreading”. The
 1073 spectrum $\chi(k)$ is divided into an “equilibrium range” just above the spectral peak k_p , and a
 1074 “saturation range” at even higher wavenumbers:

$$1075 \quad \chi(k) = \begin{cases} \frac{C_\beta}{2\sqrt{g}} a_* k^{-5/2} & \text{for } k_p < k < k_n \text{ (equilibrium),} \\ C_B k^{-3} & \text{for } k_n < k < k_i \text{ (saturation),} \end{cases} \quad (\text{A7})$$

1076 where k_n is a transition wavenumber between equilibrium and saturation ranges, k_i is an
 1077 upper wavenumber cutoff above which waves are assumed to be isotropic and there do not
 1078 contribute to Stokes drift. $a_* = u_* \sqrt{\rho_o/\rho_a}$ is the air-side friction velocity defined in terms
 1079 of the water-side friction velocity u_* , a reference air density $\rho_a = 1.2 \text{ kg m}^{-3}$ and ocean
 1080 density $\rho_o = 1024 \text{ kg m}^{-3}$. Wavenumbers *below* the spectral peak k_p are assumed too weak
 1081 to contribute appreciably to Stokes drift.

1082 Both the transition wavenumber k_n and the isotropic wavenumber k_i decrease with
 1083 increasing u_* :

$$1084 \quad k_n \stackrel{\text{def}}{=} C_r g a_*^{-2}, \quad (\text{A8})$$

$$1085 \quad k_i \stackrel{\text{def}}{=} C_i g a_*^{-2}, \quad (\text{A9})$$

1084 where $C_r = 9.7 \times 10^{-3}$ and $C_i = 0.072$.

1085 The Stokes drift is

$$1086 \quad U^S(z) = C_\beta a_* \int_{k_p}^{k_n} \frac{e^{2kz}}{k} dk + 2C_B \sqrt{g} \int_{k_n}^{k_i} k^{-3/2} e^{2kz} dk. \quad (\text{A10})$$

1087 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
 1088 function, we find

$$1089 \quad U^S(z) = C_\beta a_* [\text{Ei}(2k_n z) - \text{Ei}(2k_p z)] + 2C_B \sqrt{g} [v(k_n) - v(k_i)], \quad (\text{A11})$$

1090 and

$$1091 \quad \partial_z U^S = 2C_\beta a_* \int_{k_p}^{k_n} e^{2kz} dk + 4C_B \sqrt{g} \int_n^I \frac{e^{2kz}}{\sqrt{k}} dk, \quad (\text{A12})$$

$$1092 \quad = C_\beta a_* \frac{e^{2k_p z} - e^{2k_n z}}{|z|} + 2C_B \sqrt{\frac{2\pi g}{|z|}} \left[\text{erf} \left(\sqrt{2k_n |z|} \right) - \text{erf} \left(\sqrt{2k_i |z|} \right) \right], \quad (\text{A13})$$

1093 for the Stokes shear.

1094 **A5 LES uncertainty: effects of resolution and Stokes drift**

1095 All LES use 2 meter horizontal resolution and a stretched vertical resolution that varies
 1096 from 0.8 meters in the upper half of domain to 2.3 meters at the bottom. We refer to this
 1097 as “1 meter” vertical resolution. To account for the effects of resolution on the 35 LES used
 1098 as synthetic observations in this paper, we run 70 additional LES on coarser grids with
 1099 double (“2 meter”) and quadruple (“4 meter”) resolution, and use these to estimate the
 1100 observational uncertainty used in calibration (see 4 for more details). The effect of resolution
 1101 depends on forcing strength: for the 6 and 12 hour suite, the results are nearly identical for 1-
 1102 and 2-meter vertical resolution. Figure A5 shows the results for 4 cases in the 12 hour suite.
 1103 Note that in the free convection case, the first two grid points exhibit a strong unstable
 1104 stratification in the 12 hour suite. We attribute this to an artificial reduction of mixing near
 1105 the top boundary of the LES. It might be possible to address this artificially-strong unstable
 1106 mean stratification by introducing, for example, a surface-concentrated eddy diffusivity.
 1107 However, because the LES are used only for training CATKE and thus matter mostly in their
 1108 predicted boundary layer depth, we choose instead to ignore the top 4 m when computing
 1109 the LES-CATKE discrepancy during calibration.

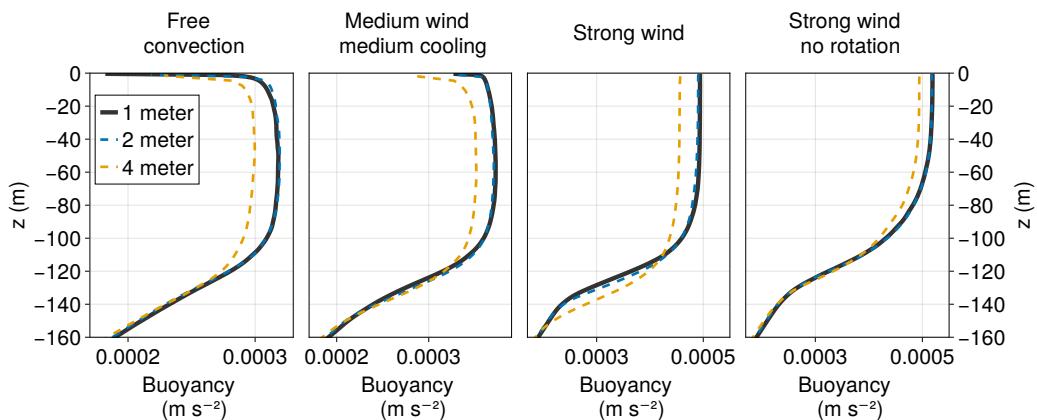


Figure A1. Resolution dependence of 12-hour LES.

1110 Figure A5 shows the resolution dependence of the 24-hour suite. These LES show
 1111 slightly more resolution dependence than the 12-hour suite, especially for cases forced by a
 1112 combination of wind and cooling. This indicates that our LES data for more weakly forced
 1113 cases are *less certain* than the strongly forced cases. Interestingly, we find that CATKE
 1114 exhibits the least bias for the weakly forced cases than for the strongly forced cases. This
 1115 means that the bias exhibited in the strongly-forced cases is real bias, while the weakly
 1116 forced cases may be interpreted as exhibiting essentially no bias.

1117 The LES also use an “implicit closure” technique whereby advection is discretized with
 1118 a 9th order weighted essentially non-oscillatory scheme (or WENO for short) and no explicit
 1119 subgrid-scale closure is added.

1120 **A6 Effect of Stokes drift on LES results**

1121 Next we turn to the effect that including the Stokes drift profile described in section A4
 1122 has on our LES results. The inclusion of Stokes drift in our LES is an attempt to make
 1123 them slightly more realistic. In other words, we hypothesize that calibrating CATKE to LES
 1124 without surface waves would generally lead to a shallow bias in mixed layer depth prediction

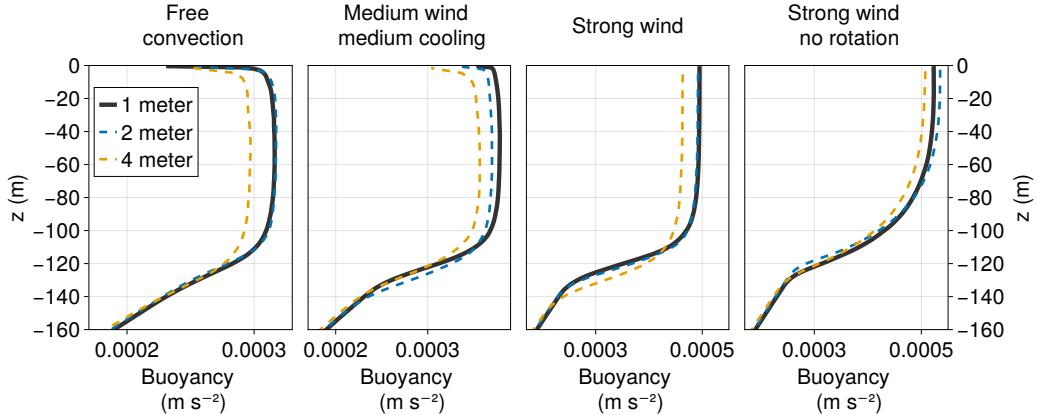


Figure A2. Resolution dependence of 24-hour LES.

with CATKE — since surface waves are always present above real wind-forced ocean surface boundary layers.

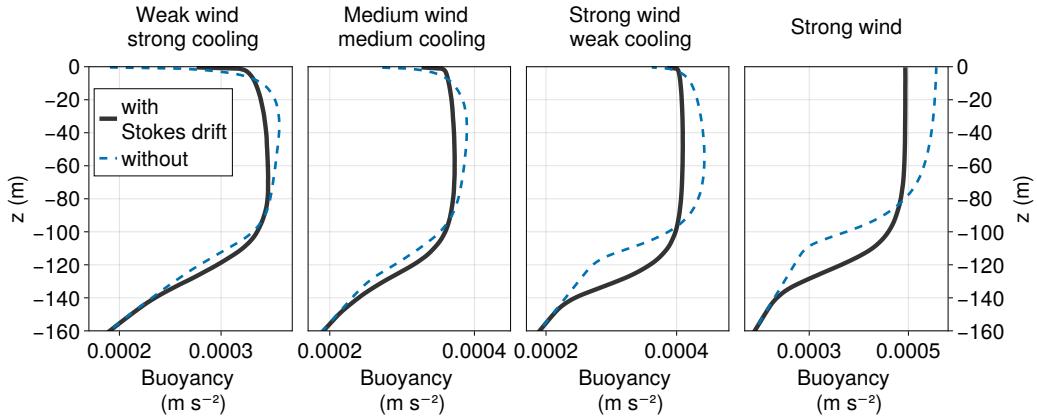


Figure A3. Stokes drift dependence of 12-hour LES.

This notion is corroborated by figure A6, which shows the horizontally-averaged buoyancy profiles for 4 LES in the 12 hour suite, with and without Stokes drift. As expected, the inclusion of Stokes drift produces more mixing and makes the boundary layer deeper. The effect of Stokes drift is minor in the case of weak and medium winds (leftmost and second from left panels). In the strong wind (and rotating) case, the inclusion of Stokes drift makes the boundary layer 20 meters deeper, or around 20% of the total. In the strong wind, no rotation case, the case without Stokes drift completely fails to transition to the turbulence. (A small amount of cooling would probably be required to produce turbulence in the strong wind, no rotation case without Stokes drift.)

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

The time discretization is a little non-trivial since we step forward velocity and tracers first, then step forward TKE and also use substepping/split-explicit scheme for TKE. In the

single column case, we integrate equations (13)–(15) with the backward Euler scheme

$$\frac{u^{n+1} - u^n}{\Delta t} = \partial_z (\kappa_u^n \partial_z u^{n+1}), \quad (B1)$$

$$\frac{v^{n+1} - v^n}{\Delta t} = \partial_z (\kappa_u^n \partial_z v^{n+1}), \quad (B2)$$

$$\frac{c^{n+1} - c^n}{\Delta t} = \partial_z (\kappa_c^n \partial_z c^{n+1}), \quad (B3)$$

where Δt is the time step and the superscripts n or $n + 1$ indicate the time-level at which the quantity is evaluated. For the TKE equation (19), we introduce a substepping scheme that uses M short time-steps $\Delta t/M$ to integrate e between n to $n + 1$,

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

where the superscripts m and $m + 1$ denote the substep level. The buoyancy flux $\overline{w'b'}^m$ in (B4) is discretized in time using the conditionally-implicit “Patankar trick” (Burchard, 2002), such that

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

which improves the stability of (B4) and keeps e from becoming too negative. Note that shear production is not updated during substepping. The time discretization of the shear production term in (B4), which incorporates shear measured at the time step n and $n + 1$, also follows Burchard (2002) and requires an algorithm that stores the velocity field at time step n , stepping forward momentum and tracers, and then substepping forward e .

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

Appendix C A posteriori calibration

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

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

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

1. *Observations* \mathcal{Y} with M discrete elements \mathcal{Y}_m . In this paper, each \mathcal{Y}_m represents a state variable like velocity U or temperature Θ at a particular depth and time, computed from LES data by horizontal averaging and vertical coarse-graining, and then normalized and shifted to have zero mean and unit variance.
2. A *parameter set* \mathbb{C} with P free parameter values \mathbb{C}_p .

- 1180 3. A *forward map* $\mathcal{G}(\mathbb{C})$ whose elements $\mathcal{G}_m(\mathbb{C})$ predict the observation \mathcal{Y}_m . $\mathcal{G}(\mathbb{C})$ represents a *model* that maps a parameter set \mathbb{C} to the space of observations \mathcal{Y} . In our case, constructing $\mathcal{G}(\mathbb{C})$ requires forward evaluations of 63 single column models parameterized by \mathbb{C} , each predicting the evolution of horizontally-averaged variables in 21 LES at 2-, 4-, and 8-meter resolution.
- 1181 4. Random Gaussian *uncertainty* $\eta \sim \mathcal{N}(0, \mathcal{M})$ with covariance \mathcal{M} associated with both $\mathcal{G}_m(\mathbb{C})$ and \mathcal{Y}_m . η conflates uncertainty in \mathcal{Y} with “structural” uncertainty associated with imperfect forward maps \mathcal{G} .

1188 The elements of \mathcal{Y} are the discrete values of the horizontally-averaged temperature
 1189 and velocity fields output from 21 LES coarse-grained to three grids with uniform 2-, 4-,
 1190 and 8-meter spacing. Each physical field is shifted, normalized, and weighted before being
 1191 assembled into \mathcal{Y} . Each forward map $G(\mathbb{C})$ involves $3 \times 21 = 63$ simulations to find U , V ,
 1192 and Θ profiles for each LES case at the two model vertical resolutions.

1193 C1 Ensemble Kalman dynamics

1194 Ensemble Kalman Inversion uses a dynamical system that governs the evolution of an
 1195 ensemble of N parameter sets, or “particles”, $\mathbf{C} \stackrel{\text{def}}{=} [\mathbb{C}^1, \mathbb{C}^2, \dots, \mathbb{C}^N]$. Here the superscript
 1196 ω denotes the “particle index”, which varies across the ensemble: \mathbb{C}_p^ω is the p^{th} parameter
 1197 value of the ω^{th} particle.

1198 Each parameter set \mathbb{C}^ω obeys the ordinary differential equation

$$1199 \quad \frac{d}{dT} \mathbb{C}^\omega = -\mathcal{K}(\mathbf{C}, \mathbf{G}) \Gamma^{-1} (\mathcal{G}^\omega - \mathcal{Y}), \quad (\text{C2})$$

1200 where $\mathcal{G}^\omega \stackrel{\text{def}}{=} \mathcal{G}(\mathbb{C}^\omega)$ is the forward map computed with the parameter set \mathbb{C}^ω , and T is
 1201 the “pseudotime”. The matrix $\mathcal{K}(\mathbf{C}, \mathbf{G})$ in (C2) is the covariance matrix estimated from
 1202 ensemble statistics at pseudotime T , thus coupling the evolution of the ensemble. For two
 1203 “ensemble matrices” \mathbf{A} and \mathbf{B} , where \mathbf{A} for example is constructed from an ensemble of
 1204 vectors $[A_i^1, A_i^2, \dots, A_i^N]$, the elements $\mathcal{K}_{ij}(\mathbf{A}, \mathbf{B})$ are defined

$$1205 \quad \mathcal{K}_{ij}(\mathbf{A}, \mathbf{B}) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\omega=1}^N (A_i^\omega - \langle A \rangle_i)(B_j^\omega - \langle B \rangle_j), \quad \text{with} \quad \langle C \rangle_i \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\omega=1}^N C_i^\omega. \quad (\text{C3})$$

1206 For nearly-linear maps $\mathcal{G}_m(\mathbb{C}) \approx H_{mp}\mathbb{C}_p$, (C2) reduces to

$$1207 \quad \frac{d}{dT} \mathbb{C}^\omega \approx -\mathcal{K}(\mathbf{C}, \mathbf{C}) \nabla_{\mathbb{C}} \Phi^\omega, \quad (\text{C4})$$

1208 where $\mathcal{K}_{pq}(\mathbf{C}, \mathbf{C})$ is the $P \times P$ parameter-parameter covariance matrix (Kovachki & Stuart,
 1209 2019). The “EKI objective” Φ^ω associated with parameter set ω appears in (C4), where

$$1210 \quad \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}) \stackrel{\text{def}}{=} \|\mathcal{M}^{-1/2} [\mathcal{G}(\mathbb{C}) - \mathcal{Y}] \|^2, \quad (\text{C5})$$

1211 and $\Phi^\omega \stackrel{\text{def}}{=} \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}^\omega)$. Φ in (C5) is a functional of \mathcal{G} that measures the uncertain discrepancy
 1212 between $\mathcal{G}(\mathbb{C}) - \mathcal{Y}$. The system (C4) minimizes Φ using gradient descent preconditioned
 1213 with $\mathcal{K}(\mathbf{C}, \mathbf{C})$, where the gradients $\nabla_{\mathbb{C}} \Phi$ are estimated from the parameter ensemble.

1214 We integrate the EKI dynamical system (C2) in using a forward Euler discretization,

$$1215 \quad \mathbb{C}^\omega|_{n+1} = \mathbb{C}^\omega|_n - \Delta T \left[\mathcal{K}(\mathbf{C}, \mathbf{G}) \mathcal{M}^{-1} (\mathcal{G}^\omega - \mathcal{Y}) \right]_n, \quad (\text{C6})$$

1216 where n is the pseudotime iteration, ΔT is a pseudotime step size, and $\omega \in [1, N_e]$ is the
 1217 “ensemble index” out of an ensemble with N_e members. The adaptive step size ΔT is chosen
 1218 at each iteration according to Kovachki and Stuart (2019). The initial parameter sets \mathbb{C}^ω at
 1219 $T = 0$ are generated by randomly sampling the priors listed in table 3.

1220 EKI is practical for two reasons: *(i)* it does not require explicit gradients of \mathcal{G} with
 1221 respect to parameters \mathbb{C} , and *(ii)* the forward map evaluations \mathcal{G}^ω — the most expensive
 1222 part of integrating (C2) — are independent and thus easily parallelized. Reason *(i)* means
 1223 EKI is applicable to any simulation framework with changeable parameters \mathbb{C} . Reason *(ii)*
 1224 means that considerable yet distributed resources can be leveraged efficiently: given sufficient
 1225 distributed resources, the cost of a single EKI iteration depends only on the cost of a single
 1226 forward map evaluation, independent of ensemble size. This parallelizability benefits small
 1227 problems such as calibration in a single column context and is decisive for large problems
 1228 like global ocean calibration.

1229 C2 Uncertainty covariance

1230 We associate the uncertainty \mathcal{M} with the numerical fidelity of the large eddy simulations
 1231 by defining

$$1232 \mathcal{M} = \text{cov}([\mathcal{Y}^{1m} \mathcal{Y}^{2m} \mathcal{Y}^{4m}]) , \quad (\text{C7})$$

1233 where $\mathcal{Y}^{1m}, \mathcal{Y}^{2m}, \mathcal{Y}^{4m}$ denote observations obtained from LES with 1-, 2-, and 4-meter vertical
 1234 resolution.

1235 C3 Constrained and unconstrained parameters

1236 The dynamics (C6) require normally-distributed parameters \mathbb{C}_p , which precludes the
 1237 imposition of strict bounds such as non-negativity. We therefore introduce the forward and
 1238 inverse transforms,

$$1239 \mathbb{C}_p = \log \frac{b - \tilde{\mathbb{C}}_p}{\tilde{\mathbb{C}}_p - a} \quad \text{and} \quad \tilde{\mathbb{C}}_p = a + \frac{b - a}{1 + \exp(\mathbb{C}_p)} , \quad (\text{C8})$$

1240 between “constrained” physical parameters $\tilde{\mathbb{C}}$ that are bounded between (a, b) , and uncon-
 1241 strained parameters \mathbb{C} . The transformation (C8) implies that if \mathbb{C}_p is normally-distributed
 1242 then $\tilde{\mathbb{C}}$ is bounded by (a, b) with a scaled, shifted logit-normal distribution.

1243 We denote the scaled, shifted logit-normal distribution bounded by (a, b) as $\mathcal{B}(a, b)$ and
 1244 use it to model the distribution of all of CATKE’s free parameters. The distributions $\mathcal{B}(a, b)$
 1245 formulated so their corresponding normal distributions have zero mean and unit variance.
 1246 When integrating (C6), the normally-distributed parameter sets \mathbb{C}^ω are transformed into
 1247 their physical space counterparts $\tilde{\mathbb{C}}^\omega$ via (C8) when evaluating $\mathcal{G}^\omega = \mathcal{G}(\mathbb{C}^\omega)$ and thus solving
 1248 the single column equations (13)–(15) and (19).

1249 C4 Failure criterion handling

1250 Poor parameter choices \mathbb{C}^ω often lead to failed simulations of the single column system
 1251 (13)–(15) and (19). In that case the forward map \mathcal{G}^ω is not informative and must be
 1252 ignored when performing the Euler step (C6).

1253 We first define the median and the “median absolute deviation” of the EKI objective
 1254 samples, $\Phi^\omega \stackrel{\text{def}}{=} \Phi(\mathcal{G}, \mathcal{Y}; \mathbb{C}^\omega)$,

$$1255 \tilde{\Phi} \stackrel{\text{def}}{=} \text{median}(\Phi^\omega) \quad \text{and} \quad s \stackrel{\text{def}}{=} \text{median}(|\Phi^\omega - \tilde{\Phi}|) , \quad (\text{C9})$$

1256 We mark a particle ω as “failed” if

$$1257 \Phi^\omega > \tilde{\Phi} + 3s . \quad (\text{C10})$$

1258 This excludes both non-finite and just “particularly anomalous” Φ^ω .

1259 **Open Research Section**

1260 This work relied on the open-source software LESbrary.jl (Wagner et al., 2023) and
 1261 Oceananigans.jl (Ramadhan et al., 2020) to run the LES, Oceananigans.jl to run calibration
 1262 simulations, and ParameterEstimocean.jl (Wagner et al., 2022) and EnsembleKalmanPro-
 1263 cesses.jl (Dunbar et al., 2022) for the Ensemble Kalman Inversion. Visualizations were made
 1264 using Makie.jl (Danisch & Krumbiegel, 2021). Scripts for performing the calibration are avail-
 1265 able at the GitHub repository github.com/glwagner/SingleColumnModelCalibration.jl.

1266 **Acknowledgments**

1267 This project is supported by Schmidt Sciences, LLC and by the National Science Foundation
 1268 grant AGS-1835576. N.C.C. is additionally supported by the Australian Research Council
 1269 under DECRA Fellowship DE210100749 and the Center of Excellence for the Weather of
 1270 the 21st Century CE230100012. Without implying their endorsement, we would also like to
 1271 acknowledge stimulating and useful conversations with Bruno Deremble, Alex Legay, Qing
 1272 Li, Brodie Pearson, Brandon Reichl, Roger Samelson, and Bill Young.

1273 **References**

- 1274 Belcher, S. E., Grant, A. L., Hanley, K. E., Fox-Kemper, B., Van Roekel, L., Sullivan, P. P.,
 1275 ... others (2012). A global perspective on Langmuir turbulence in the ocean surface
 1276 boundary layer. *Geophysical Research Letters*, 39(18). doi: 10.1029/2012GL052932
- 1277 Blanke, B., & Delecluse, P. (1993). Variability of the tropical Atlantic Ocean simulated
 1278 by a general circulation model with two different mixed-layer physics. *Journal of*
 1279 *Physical Oceanography*, 23(7), 1363–1388. doi: 10.1175/1520-0485(1993)023(1363:
 1280 VOTTAO)2.0.CO;2
- 1281 Boccaletti, G., Ferrari, R., & Fox-Kemper, B. (2007). Mixed layer instabilities and restratifi-
 1282 cation. *Journal of Physical Oceanography*, 37(9), 2228–2250.
- 1283 Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network
 1284 parametrization trained by coarse-graining. *Journal of Advances in Modeling Earth*
 1285 *Systems*, 11(8), 2728–2744.
- 1286 Burchard, H. (2002). Energy-conserving discretisation of turbulent shear and buoyancy
 1287 production. *Ocean Modelling*, 4(3-4), 347–361. doi: 10.1016/S1463-5003(02)00009-4
- 1288 Burchard, H., & Bolding, K. (2001). Comparative analysis of four second-moment turbulence
 1289 closure models for the oceanic mixed layer. *Journal of Physical Oceanography*, 31(8),
 1290 1943–1968. doi: 10.1175/1520-0485(2001)031<1943:CAOFSM>2.0.CO;2
- 1291 Canuto, V. M., Howard, A., Cheng, Y., & Dubovikov, M. (2001). Ocean turbulence.
 1292 Part I: One-point closure model—momentum and heat vertical diffusivities. *Journal*
 1293 *of Physical Oceanography*, 31(6), 1413–1426.
- 1294 Caulfield, C.-c. P. (2020). Open questions in turbulent stratified mixing: do we even
 1295 know what we do not know? *Physical Review Fluids*, 5(11), 110518. doi: 10.1103/
 1296 PhysRevFluids.5.110518
- 1297 Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., & Stuart, A. M. (2021). Calibrate,
 1298 emulate, sample. *Journal of Computational Physics*, 424, 109716. doi: 10.1016/
 1299 j.jcp.2020.109716
- 1300 Craik, A. D. D., & Leibovich, S. (1976). A rational model for Langmuir circulations. *Journal*
 1301 *of Fluid Mechanics*, 73(3), 401–426. doi: 10.1017/S0022112076001420
- 1302 Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization
 1303 for Julia. *Journal of Open Source Software*, 6(65), 3349. doi: 10.21105/joss.03349
- 1304 Deardorff, J. W. (1970). Convective velocity and temperature scales for the unstable planetary
 1305 boundary layer and for Rayleigh convection. *Journal of the Atmospheric Sciences*,
 1306 27(8), 1211–1213. doi: 10.1175/1520-0469(1970)027<1211:CVATSF>2.0.CO;2
- 1307 Dunbar, O. R., Lopez-Gomez, I., Garbuno-Íñigo, A., Huang, D. Z., Bach, E., & Wu, J.
 1308 (2022). EnsembleKalmanProcesses.jl: Derivative-free ensemble-based model calibration.
 1309 *Journal of Open Source Software*, 7(80), 4869. doi: 10.21105/joss.04869

- 1310 Duraisamy, K. (2021). Perspectives on machine learning-augmented Reynolds-averaged and
 1311 large eddy simulation models of turbulence. *Physical Review Fluids*, *6*(5), 050504.
- 1312 DuVivier, A. K., Large, W. G., & Small, R. J. (2018). Argo observations of the deep mixing
 1313 band in the Southern Ocean: A salinity modeling challenge. *Journal of Geophysical
 1314 Research: Oceans*, *123*(10), 7599–7617.
- 1315 Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., & Lguensat, R. (2022). A posteriori
 1316 learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in
 1317 Modeling Earth Systems*, *14*(11).
- 1318 Garanaik, A., Pereira, F. S., Smith, K., Robey, R., Li, Q., Pearson, B., & Van Roekel, L.
 1319 (2024). A new hybrid mass-flux/high-order turbulence closure for ocean vertical mixing.
 1320 *Journal of Advances in Modeling Earth Systems*, *16*(1), e2023MS003846.
- 1321 Gardiner, C. W. (2021). *Elements of stochastic methods*. AIP Publishing Melville, NY,
 1322 USA.
- 1323 Gaspar, P., Grégoris, Y., & Lefevre, J.-M. (1990). A simple eddy kinetic energy model for
 1324 simulations of the oceanic vertical mixing: Tests at station Papa and long-term upper
 1325 ocean study site. *Journal of Geophysical Research: Oceans*, *95*(C9), 16179–16193.
- 1326 Giordani, H., Bourdallé-Badie, R., & Madec, G. (2020). An eddy-diffusivity mass-flux
 1327 parameterization for modeling oceanic convection. *Journal of Advances in Modeling
 1328 Earth Systems*, *12*(9), e2020MS002078.
- 1329 Gregg, M. C., D'Asaro, E. A., Riley, J. J., & Kunze, E. (2018). Mixing efficiency in the
 1330 ocean. *Annual Review of Marine Science*, *10*, 443–473.
- 1331 Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent
 1332 climate change. *Climate Dynamics*, *16*, 501–515.
- 1333 Groeskamp, S., Griffies, S. M., Iudicone, D., Marsh, R., Nurser, A. G., & Zika, J. D. (2019).
 1334 The water mass transformation framework for ocean physics and biogeochemistry.
 1335 *Annual review of marine science*, *11*, 271–305.
- 1336 Gutjahr, O., Brügmann, N., Haak, H., Jungclaus, J. H., Putrasahan, D. A., Lohmann,
 1337 K., & von Storch, J.-S. (2021). Comparison of ocean vertical mixing schemes in
 1338 the Max Planck Institute Earth System Model (MPI-ESM1.2). *Geoscientific Model
 1339 Development*, *14*(5), 2317–2349.
- 1340 Häfner, D., Nuterman, R., & Jochum, M. (2021). Fast, cheap, and turbulent—global ocean
 1341 modeling with GPU acceleration in python. *Journal of Advances in Modeling Earth
 1342 Systems*, *13*(12), e2021MS002717.
- 1343 Harcourt, R. R. (2015). An improved second-moment closure model of Langmuir turbulence.
 1344 *Journal of Physical Oceanography*, *45*(1), 84–103.
- 1345 Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010).
 1346 Probing the fast and slow components of global warming by returning abruptly to
 1347 preindustrial forcing. *Journal of Climate*, *23*(9), 2418–2427.
- 1348 Huang, N. E. (1979). On surface drift currents in the ocean. *Journal of Fluid Mechanics*,
 1349 *91*(1), 191–208.
- 1350 Iglesias, M. A., Law, K. J., & Stuart, A. M. (2013). Ensemble Kalman methods for inverse
 1351 problems. *Inverse Problems*, *29*(4), 045001.
- 1352 Jungclaus, J. H., Lorenz, S. J., Schmidt, H., Brovkin, V., Brügmann, N., Chegini, F., ...
 1353 others (2022). The ICON earth system model version 1.0. *Journal of Advances in
 1354 Modeling Earth Systems*, *14*(4), e2021MS002813. doi: 10.1029/2021MS002813
- 1355 Kantha, L. H., & Clayson, C. A. (1994). An improved mixed layer model for geophysical
 1356 applications. *Journal of Geophysical Research: Oceans*, *99*(C12), 25235–25266.
- 1357 Kovachki, N. B., & Stuart, A. M. (2019). Ensemble Kalman inversion: a derivative-free
 1358 technique for machine learning tasks. *Inverse Problems*, *35*(9), 095005.
- 1359 Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., ... others
 1360 (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation
 1361 for global climate. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2865–2888.
- 1362 Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review
 1363 and a model with a nonlocal boundary layer parameterization. *Reviews of geophysics*,
 1364 *32*(4), 363–403.

- 1365 Large, W. G., & Yeager, S. G. (2009). The global climatology of an interannually varying
 1366 air-sea flux data set. *Climate dynamics*, 33, 341–364.
- 1367 Legay, A., Deremble, B., & Burchard, H. (2024). Derivation and implementation of a
 1368 non-gradient term to improve the oceanic convection representation within the $k-\epsilon$
 1369 parameterization.
- 1370 Lenain, L., & Pizzo, N. (2020). The contribution of high-frequency wind-generated surface
 1371 waves to the Stokes drift. *Journal of Physical Oceanography*, 50(12), 3455–3465.
- 1372 Li, D. (2019). Turbulent Prandtl number in the atmospheric boundary layer—where are we
 1373 now? *Atmospheric Research*, 216, 86–105.
- 1374 Li, G., & Xie, S.-P. (2014). Tropical biases in CMIP5 multimodel ensemble: The excessive
 1375 equatorial Pacific cold tongue and double ITCZ problems. *Journal of Climate*, 27(4),
 1376 1765–1780.
- 1377 Li, Q., Reichl, B. G., Fox-Kemper, B., Adcroft, A. J., Belcher, S. E., Danabasoglu, G., ...
 1378 others (2019). Comparing ocean surface boundary vertical mixing schemes including
 1379 Langmuir turbulence. *Journal of Advances in Modeling Earth Systems*, 11(11), 3545–
 1380 3592. doi: 10.1029/2019MS001810
- 1381 Luyten, J., Pedlosky, J., & Stommel, H. (1983). The ventilated thermocline. *Journal of
 1382 Physical Oceanography*, 13(2), 292–309.
- 1383 Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D.,
 1384 ... others (2017). NEMO ocean engine.
- 1385 McDougall, T. J., & Barker, P. M. (2011). Getting started with TEOS-10 and the Gibbs
 1386 Seawater (GSW) oceanographic toolbox. *Scor/Iapso WG*, 127, 1–28.
- 1387 McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal
 1388 Society A: Mathematical, Physical and Engineering Sciences*, 472(2189), 20160117.
- 1389 McWilliams, J. C., Huckle, E., & Shchepetkin, A. F. (2009). Buoyancy effects in a stratified
 1390 Ekman layer. *Journal of Physical Oceanography*, 39(10), 2581–2599.
- 1391 McWilliams, J. C., Sullivan, P. P., & Moeng, C.-H. (1997). Langmuir turbulence in the
 1392 ocean. *Journal of Fluid Mechanics*, 334, 1–30.
- 1393 Melet, A. V., Hallberg, R., & Marshall, D. P. (2022). Chapter 2 - The role of ocean mixing
 1394 in the climate system. In M. Meredith & A. Naveira Garabato (Eds.), *Ocean mixing*
 1395 (p. 5-34). Elsevier. doi: 10.1016/B978-0-12-821512-8.00009-8
- 1396 Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for
 1397 geophysical fluid problems. *Reviews of Geophysics*, 20(4), 851–875.
- 1398 Omand, M. M., D'Asaro, E. A., Lee, C. M., Perry, M. J., Briggs, N., Cetinić, I., & Mahadevan,
 1399 A. (2015). Eddy-driven subduction exports particulate organic carbon from the spring
 1400 bloom. *Science*, 348(6231), 222–225.
- 1401 Prandtl, L., Wieselsberger, C., & Betz, A. (1925). *Results of the aerodynamic research
 1402 institute in Göttingen* (No. 1). R. Oldenburg.
- 1403 Price, J. F., Weller, R. A., & Pinkel, R. (1986). Diurnal cycling: Observations and models
 1404 of the upper ocean response to diurnal heating, cooling, and wind mixing. *Journal of
 1405 Geophysical Research: Oceans*, 91(C7), 8411–8427.
- 1406 Ramadhan, A., Wagner, G. L., Hill, C., Campin, J.-M., Churavy, V., Besard, T., ... Ferrari,
 1407 R. (2020). Oceananigans.jl: Fast and friendly geophysical fluid dynamics on GPUs.
 1408 *The Journal of Open Source Software*, 4(44), 1965. doi: 10.21105/joss.01965
- 1409 Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in
 1410 neural network parameterizations: General algorithms and lorenz 96 case study (v1.
 1411 0). *Geoscientific Model Development*, 13(5), 2185–2196.
- 1412 Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid
 1413 processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39),
 1414 9684–9689.
- 1415 Reffray, G., Bourdalle-Badie, R., & Calone, C. (2015). Modelling turbulent vertical mixing
 1416 sensitivity using a 1-D version of NEMO. *Geoscientific Model Development*, 8(1),
 1417 69–86.
- 1418 Reichl, B. G., & Hallberg, R. (2018). A simplified energetics based planetary boundary layer
 1419 (ePBL) approach for ocean climate simulations. *Ocean Modelling*, 132, 112–129.

- 1420 Reichl, B. G., & Li, Q. (2019). A parameterization with a constrained potential energy
 1421 conversion rate of vertical mixing due to Langmuir turbulence. *Journal of Physical*
 1422 *Oceanography*, 49(11), 2935–2959.
- 1423 Reichl, B. G., Wittenberg, A. T., Griffies, S. M., & Adcroft, A. (2024). Improved equatorial
 1424 upper ocean vertical mixing in the NOAA/GFDL OM4 model. *Authorea Preprints*.
- 1425 Saffman, P. G. (1970). A model for inhomogeneous turbulent flow. *Proceedings of the Royal*
 1426 *Society of London. A. Mathematical and Physical Sciences*, 317(1530), 417–433.
- 1427 Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J., Bracegirdle, T. J., & Wang,
 1428 Z. (2013). Assessment of Southern Ocean mixed-layer depths in CMIP5 models:
 1429 Historical bias and forcing response. *Journal of Geophysical Research: Oceans*, 118(4),
 1430 1845–1862.
- 1431 Samelson, R. (2022). Wind drift in a homogeneous equilibrium sea. *Journal of Physical*
 1432 *Oceanography*, 52(9), 1945–1967. doi: 10.1175/JPO-D-22-0017.1
- 1433 Siebesma, A. P., Soares, P. M., & Teixeira, J. (2007). A combined eddy-diffusivity mass-flux
 1434 approach for the convective boundary layer. *Journal of the Atmospheric Sciences*,
 1435 64(4), 1230–1248. doi: 10.1175/JAS3888.1
- 1436 Silvestri, S., Wagner, G. L., Constantinou, N. C., Hill, C. N., Campin, J.-M., Souza,
 1437 A. N., ... Ferrari, R. (2024). A GPU-based ocean dynamical core for routine
 1438 mesoscale-resolving climate simulations. *Earth and Space Science Open Archive*. doi:
 1439 10.22541/essoar.171708158.82342448/v1
- 1440 Smyth, W., & Moum, J. (2013). Marginal instability and deep cycle turbulence in the
 1441 eastern equatorial Pacific Ocean. *Geophysical Research Letters*, 40(23), 6181–6185.
 1442 doi: 10.1002/2013GL058403
- 1443 Smyth, W., Pham, H., Moum, J., & Sarkar, S. (2017). Pulsating turbulence in a marginally
 1444 unstable stratified shear flow. *Journal of Fluid Mechanics*, 822, 327–341.
- 1445 Souza, A. N., He, J., Bischoff, T., Waruszewski, M., Novak, L., Barra, V., ... Schneider, T.
 1446 (2023). The flux-differencing discontinuous Galerkin method applied to an idealized
 1447 fully compressible nonhydrostatic dry atmosphere. *Journal of Advances in Modeling*
 1448 *Earth Systems*, 15(4), e2022MS003527. doi: 10.1029/2022MS003527
- 1449 Souza, A. N., Wagner, G. L., Ramadhan, A., Allen, B., Churavy, V., Schloss, J., ... Ferrari,
 1450 R. (2020). Uncertainty quantification of ocean parameterizations: Application to the
 1451 K-profile-parameterization for penetrative convection. *Journal of Advances in Modeling*
 1452 *Earth Systems*, 12(12), e2020MS002108. doi: 10.1029/2020MS002108
- 1453 Spalding, D. B. (1991). Kolmogorov's two-equation model of turbulence. *Proceedings of the*
 1454 *Royal Society of London. Series A: Mathematical and Physical Sciences*, 434(1890),
 1455 211–216. doi: 10.1098/rspa.1991.0089
- 1456 Suzuki, N., & Fox-Kemper, B. (2016). Understanding Stokes forces in the wave-averaged
 1457 equations. *Journal of Geophysical Research: Oceans*, 121(5), 3579–3596. doi: 10.1002/
 1458 2015JC011563
- 1459 Taylor, G. I. (1922). Diffusion by continuous movements. *Proceedings of the London*
 1460 *Mathematical Society*, 2(1), 196–212. doi: 10.1112/plms/s2-20.1.196
- 1461 Taylor, J. R., & Thompson, A. F. (2023). Submesoscale dynamics in the upper ocean. *Annual*
 1462 *Review of Fluid Mechanics*, 55, 103–127. doi: 10.1146/annurev-fluid-031422-095147
- 1463 Thomas, L. N., Tandon, A., & Mahadevan, A. (2008). Submesoscale processes and dynamics.
 1464 In *Ocean modeling in an eddying regime* (p. 17–38). American Geophysical Union
 1465 (AGU). doi: 10.1029/177GM04
- 1466 Umlauf, L., & Burchard, H. (2003). A generic length-scale equation for geophysical
 1467 turbulence models. *Journal of Marine Research*, 61(2), 235–265. Retrieved from
 1468 https://elischolar.library.yale.edu/journal_of_marine_research/9
- 1469 Umlauf, L., & Burchard, H. (2005). Second-order turbulence closure models for geophysical
 1470 boundary layers. A review of recent work. *Continental Shelf Research*, 25(7-8), 795–827.
 1471 doi: 10.1016/j.csr.2004.08.004
- 1472 Van Roekel, L., Adcroft, A. J., Danabasoglu, G., Griffies, S. M., Kauffman, B., Large, W. G.,
 1473 ... Schmidt, M. (2018). The KPP boundary layer scheme for the ocean: Revisiting its
 1474 formulation and benchmarking one-dimensional simulations relative to LES. *Journal of*

- 1475 *Advances in Modeling Earth Systems*, 10(11), 2647–2685. doi: 10.1029/2018MS001336
- 1476 Wagner, G. L., Hillier, A., & Constantinou, N. C. (2022). *ParameterEstimocean.jl v0.14.2: Julia package for parameter estimation of turbulence closures for ocean models using Ensemble Kalman Inversion*. doi: 10.5281/zenodo.5762810
- 1477 Wagner, G. L., Ramadhan, A., Chor, T., Constantinou, N. C., Hillier, A., Lee, X. K., & Li, Q. (2023). *LESbrary.jl: v0.1.1: Oceananigans large eddy simulation (LES) data for calibrating parameterizations* [Dataset]. doi: 10.5281/zenodo.7997002
- 1478 Whitt, D. B., Cherian, D. A., Holmes, R. M., Bachman, S. D., Lien, R.-C., Large, W. G., & Moum, J. N. (2022). Simulation and scaling of the turbulent vertical heat transport and deep-cycle turbulence across the equatorial Pacific cold tongue. *Journal of Physical Oceanography*, 52(5), 981–1014. doi: 10.1175/JPO-D-21-0153.1
- 1480 Williams, R. G. (1991). The role of the mixed layer in setting the potential vorticity of the main thermocline. *Journal of Physical Oceanography*, 21(12), 1803–1814. doi: 10.1175/1520-0485(1991)021<1803:TRO>2.0.CO;2
- 1482 Wunsch, C., & Ferrari, R. (2004). Vertical mixing, energy, and the general circulation of the oceans. *Annual Review of Fluid Mechanics*, 36(1), 281-314. doi: 10.1146/annurev.fluid.36.050802.122121
- 1484 Zhang, S., Fu, H., Wu, L., Li, Y., Wang, H., Zeng, Y., ... Guo, Y. (2020). Optimizing high-resolution community earth system model on a heterogeneous many-core supercomputing platform. *Geoscientific Model Development*, 13(10), 4809–4829. doi: 10.5194/gmd-13-4809-2020
- 1486
1488
1490
1492
1494
1495