Uncertainty quantification of ocean parameterizations: application to the K-Profile-Parameterization for penetrative convection

Andre Nogueira Souza[1,1,1], Gregory LeClaire Wagner[2,2,2], Ali Ramadhan[1,1,1], Valentin Churavy[1,1,1], Brandon Allen[1,1,1], James Schloss[1,1,1], Jean-Michel Campin[1,1,1], Chris Hill[3,3,3], Alan Edelman[1,1,1], John Marshall[1,1,1], Glenn R. Flierl[3,3,3], and Raffaele Ferrari[1,1,1]

[1]Massachusetts Institute of Technology
[2]Massachusetts Institution of Technology
[3]MIT

November 30, 2022

Abstract

Parameterizations of unresolved turbulent processes in the ocean compromise the fidelity of large-scale ocean models used in climate change projections. In this work, we use a Bayesian approach for evaluating and developing turbulence parameterizations by comparing parameterized models with observations or high-fidelity numerical simulations. The method obtains optimal parameter values, correlations, sensitivities, and, more generally, likely distributions of uncertain parameters. We demonstrate the approach by estimating the uncertainty of parameters in the popular 'K-Profile Parameterization', using an ensemble of large eddy simulations of turbulent penetrative convection in the ocean surface boundary layer. We uncover structural deficiencies and discuss their cause. We conclude by discussing the applicability of the approach to Earth system models.

Uncertainty quantification of ocean parameterizations: application to the K-Profile-Parameterization for penetrative convection

A. N. Souza[1], G. L. Wagner[1], A. Ramadhan[1], B. Allen[1], V. Churavy[1], J. Schloss[1], J. Campin[1], C. Hill[1], A. Edelman[1], J. Marshall[1], G. Flierl[1], R. Ferrari[1]

[1]Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, United States

Key Points:

- A Bayesian methodology is used to probe turbulence parameterizations and better understand their biases and uncertainties.
- Parameterization parameter distributions, learned using high-resolution simulations, should be used as prior distributions for climate modeling studies.

Corresponding author: A. N. Souza, andrenogueirasouza@gmail.com

Abstract

Parameterizations of unresolved turbulent processes often compromise the fidelity of large-scale ocean models. In this work, we argue for a Bayesian approach to the refinement and evaluation of turbulence parameterizations. Using an ensemble of large eddy simulations of turbulent penetrative convection in the surface boundary layer, we demonstrate the method by estimating the uncertainty of parameters in the convective limit of the popular 'K-Profile Parameterization'. We uncover structural deficiencies and propose an alternative scaling that overcomes them.

Plain Language Summary

Climate projections are often compromised by significant uncertainties which stem from the representation of physical processes that cannot be resolved – such as clouds in the atmosphere and turbulent swirls in the ocean – but which have to be parameterised. We propose a methodology for improving parameterizations in which they are tested against, and tuned to, high-resolution numerical simulations of subdomains that represent them more completely. A Bayesian methodology is used to calibrate the parameterizations against the highly resolved model, to assess their fidelity and identify shortcomings. Most importantly, the approach provides estimates of parameter uncertainty. While the method is illustrated for a particular parameterization of boundary layer mixing, it can be applied to any parameterization.

1 Introduction

Earth System Models (ESMs) require parameterizations for processes that are too small to resolve. Uncertainties arise both due to deficiencies in the scaling laws encoded in the parameterizations and the nonlinear interactions with resolved model components, sometimes leading to unanticipated and unphysical results. The first challenge can be addressed by improving the representation of the unresolved physics (e.g. Schneider, Lan, et al., 2017), while the second requires 'tuning' of the parameterizations when implemented in the full ESM (e.g. Hourdin et al., 2017). In this paper, we illustrate how to leverage recent advances in computation and uncertainty quantification to make progress toward the first challenge. Our focus will be on oceanic processes, but the approach can be applied to any ESM parameterization, provided that a high-resolution submodel can be constructed.

The traditional approach to the formulation of parameterizations of subgrid-scale processes is to derive scaling laws that relate the net effect of such processes to variables resolved by the ESMs. These scaling laws are then tested with either field observations (e.g. Price et al., 1986; Large et al., 1994), laboratory experiments (e.g. Deardorff et al., 1980; Cenedese et al., 2004) or results from a high resolution simulations (e.g. Wang et al., 1996; Harcourt, 2015; Reichl et al., 2016; Li & Fox-Kemper, 2017). Rarely are parameterizations tested over a wide range of possible scenarios due to the logistical difficulty and high cost of running many field experiments, the time necessary to change laboratory setups, and computational demand. The computational limitations have become much less severe over the last few years through a combination of new computer architectures such as Graphic Processing Units (GPUs; Besard et al., 2019), new languages that take advantage of these architectures (e.g Julia; Bezanson et al., 2017) and improved Large Eddy Simulation (LES) algorithms (Sullivan & Patton, 2011; Verstappen, 2018). Modern computational resources have opened up the possibility of running libraries of LES simulations to explore a vast range of possible scenarios. This paper discusses how such computational advances can be applied to assess parameterizations in ocean models.

LES simulations alone are not sufficient to formulate parameterizations. Statistical methods are needed to extract from the LES solutions the functional relationships

between small-scale processes and coarse variables available in ESMs. A common approach is to rely on well-established scaling laws and use the LES solutions to constrain the non-dimensional parameters that cannot be determined from first principles. In this approach, only a few LES simulations are necessary to find the optimal parameter values. However, it is rare that scaling laws and associated parameterizations perfectly capture the functional dependencies of large-scale variables – if they did, they would be referred to as solutions rather than parameterizations. In general, it is necessary to run a large ensemble of LES simulations to estimate optimal parameter values and test whether those values hold for different scenarios, thereby supporting the functional dependencies.

State-estimation, which has a long tradition in geophysics (Wunsch, 2006), has been used to constrain parameter values. A loss function is chosen to quantify the mismatch between the prediction of the parameterization and observations. Uncertain parameters are then adjusted to minimize the loss function. One can also estimate the standard deviation around the optimal values by computing the Hessian of the loss function (Thacker, 1989; Sraj et al., 2014).

An alternative approach, based on the seminal work of (Bayes, 1763) and its modern incarnation (Jaynes, 2003), is arguably better suited to constrain the transfer properties of turbulent processes. The Bayesian method allows one to estimate the entire joint probability distribution of all parameters. The method is a crucial extension over state-estimation, because the statistics of turbulent processes are generally far from Gaussian (Frisch, 1995) and thus are not fully characterized by the first and second moments alone. In the Bayesian approach, one defines a prior parameter distribution, based on physical considerations, and a 'likelihood function' which measures the mismatch between the parameterized prediction and the LES simulation. Based on this information, Bayes' formula shows how to compute the posterior distribution of the parameters consistent with the LES simulations and the parameterization. If the posterior distribution is narrow and peaked, then one can conclude that a unique set of parameters can be identified which can reproduce all LES results. In this limit, the Bayesian approach does not provide more information than state-estimation. However, the power of Bayes' formula is that it can reveal distinct parameter regimes, the existence of multiple maxima, relationships between parameters, and the likelihood of parameter values relative to optimal ones.

The Bayesian approach can also be used to test the functional dependence of the parameterization on large-scale variables. One estimates the posterior distribution on subsets of the LES simulations run for different scenarios. If the posterior probabilities for the different scenarios do not overlap, the functional form of the parameterization must be rejected. We will illustrate how this strategy can be used to improve the formulation of a parameterization.

Bayesian methods are particularly suited to constrain ESM parameterizations of subgrid-scale ocean processes. Most of these processes, such as boundary layer or geostrophic turbulence, are governed by well understood fluid dynamics and thermodynamics. Thus LES simulations provide credible solutions for the physics. The atmospheric problem is quite different where leading order subgrid-scale processes such as cloud microphysics are governed by poorly understood physics that may not be captured by LES simulations.

In this paper, we will apply Bayesian methods to constrain and improve a parameterization for the surface boundary layer turbulence that develops when air-sea fluxes cool the ocean. LES simulations that resolve all the relevant physics will be used as ground-truth to train the parameterization. Our paper is organized as follows: In section 2 we describe the physical setup and the LES model. In section 3 we introduce Bayesian parameter estimation for the parameters in the K-Profile Parameterization (KPP). We then perform the parameter estimation in the regime described by section 2 and show how

the Bayesian approach provides insight on how to improve the KPP parameterization. Finally, we end with a discussion in section 4.

2 Large eddy simulations and K-Profile Parameterization of penetrative convection

During winter, high latitude cooling induces near-surface mixing by convection which generates a 'mixed layer' of almost uniform temperature and salinity which can reach depths of hundreds of meters: - see (Marshall & Schott, 1999) for a review. At the base of the mixed layer, convective plumes can penetrate further into the stratified layer below – called the 'entrainment layer' – where plume-driven turbulent mixing between the mixed layer and stratification below cools the boundary layer. This process, in which the layer is cooled both at the surface and by turbulent mixing from the entrainment layer below, is called penetrative convection. Here we evaluate the ability of the K-Profile Parameterization (Large et al., 1994) to capture penetrative convection by comparing predictions based on it against large eddy simulations (LES) of idealized penetrative convection into a resting stratified fluid. It provides the context in which we outline the Bayesian approach to parameter estimation which we advocate.

2.1 Penetrative convection into a resting stratified fluid

We suppose a constant surface cooling $Q_h > 0$ to a resting, linearly stratified boundary layer with the initial state

$$\boldsymbol{u}|_{t=0} = 0 \text{ and } b|_{t=0} = 20\alpha g + N^2 z + \mathcal{N}(0, 10^{-10}\alpha g)\exp(4z/L_z), \tag{1}$$

where $z \in [-L_z, 0]$, $\boldsymbol{u} = (u, v, w)$ is the resolved velocity field simulated by LES, b is buoyancy, N^2 is the initial vertical buoyancy gradient, and $\mathcal{N}(0, \alpha g 10^{-10})$ is a Gaussian white noise process added to induce a transition to turbulence. The surface buoyancy flux Q_b is related to the imposed surface cooling Q_h, which has units $\mathrm{W\,m^{-2}}$, via

$$Q_b = \frac{\alpha g}{\rho_{\mathrm{ref}} c_p} Q_h, \tag{2}$$

where $\alpha = 2 \times 10^{-4}\,(^\circ\mathrm{C})^{-1}$ is the thermal expansion coefficient (assumed constant), $g = 9.81\,\mathrm{m\,s^{-2}}$ is gravitational acceleration, $\rho_{\mathrm{ref}} = 1035\,\mathrm{kg\,m^{-3}}$ is a reference density, and $c_p = 3993\,\mathrm{J/(kg\,^\circ C)}$ is the specific heat capacity. Our software and formulation of the large eddy simulation model is discussed in Appendix A.

Results from a large eddy simulation of turbulent penetrative convection in a domain $L_x = L_y = L_z = 100$ meters and $256 \times 256 \times 512$ grid cells, respectively, is presented in Figure 1. The resulting horizontally averaged temperature profiles are not affected by the domain size. The left panel shows the three-dimensional temperature field $\theta = \theta_0 + b/\alpha g$ associated with the buoyancy b, where $\theta_0 = 20^\circ\mathrm{C}$ is the surface temperature at $z = 0$. The right panel shows the horizontally averaged buoyancy profile

$$\bar{b}(z,t) \equiv \frac{1}{L_x L_y}\int_0^{Lx}\int_0^{Ly} b(x,y,z,t)\mathrm{d}x\mathrm{d}y. \tag{3}$$

The visualization reveals the two-part boundary layer produced by penetrative convection: close to the surface, cold and dense convective plumes organized by surface cooling sink and mix ambient fluid, producing a well-mixed layer that deepens in time. Below the mixed layer, the momentum carried by sinking convective plumes leads them to overshoot their level of neutral buoyancy (nominally, the depth of the mixed layer), 'penetrating' the stably stratified region below the surface mixed layer and generating the strongly stratified entrainment layer. The total depth of the boundary layer is h and includes the mixed layer and the entrainment layer of thickness Δh. Turbulent fluxes are assumed negligible below $z = -h$.

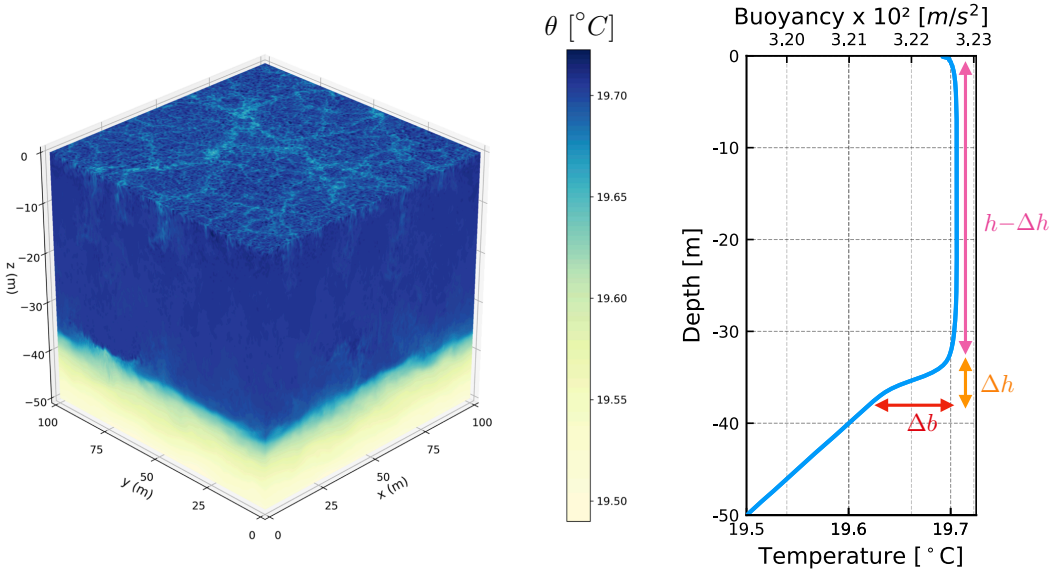

Figure 1. A 3D simulation of the LES model of the Boussinesq equations and its horizontal average at $t = 2$ days. The Δh region of the figure on the right corresponds to the entrainment layer, $h - \Delta h$ corresponds to the mixed layer, and h corresponds to the boundary layer depth.

In figure 2 we show the evolution of $h(t)$ defined as the first depth from the bottom where the stratification is equal to a weighted average of the maximum stratification and the initial stratification[1]. The dotted line confirms that the evolution after an initial transient is best fit by the formula,

$$h \simeq \sqrt{3.0 \frac{Q_b}{N^2} t}, \tag{4}$$

where N^2 is the initial stratification and the numerical factor is a best-fit parameter.

Equation 4 is easily understood through dimensional considerations (up to prefactors), but more information flows from an analysis of the horizontally-averaged buoyancy equation,

$$\partial_t \bar{b} = -\partial_z \left(\overline{wb} + \overline{q^{(z)}} \right), \tag{5}$$

where $\bar{b}$ is the horizontally averaged buoyancy, $\overline{wb}$ is the horizontally averaged vertical advective flux and $\overline{q^{(z)}}$ is the horizontally averaged vertical diffusive flux. Integrating the equation in time between $t' = 0$ and some later time $t' = t$, and in the vertical between the surface, where $q^{(z)} = -Q_b$, and the base of the entrainment layer where all turbulent fluxes vanish, one finds,

$$\int_{-h}^{0} \left[\bar{b}(z,t) - \bar{b}(z,0) \right] \, \mathrm{d}z = -Q_b t. \tag{6}$$

[1] The weights are 2/3 for the initial stratification N^2 and 1/3 for the maximum stratification N_m^2 so that h satisfies $\partial_z \bar{b}(-h) = 2N_b^2/3 + N_m^2/3$. This guarantees that h is a depth where the local stratification lies between the background stratification and the maximum stratification since it is defined as the *first* depth starting from the bottom that satisfies such a criteria.

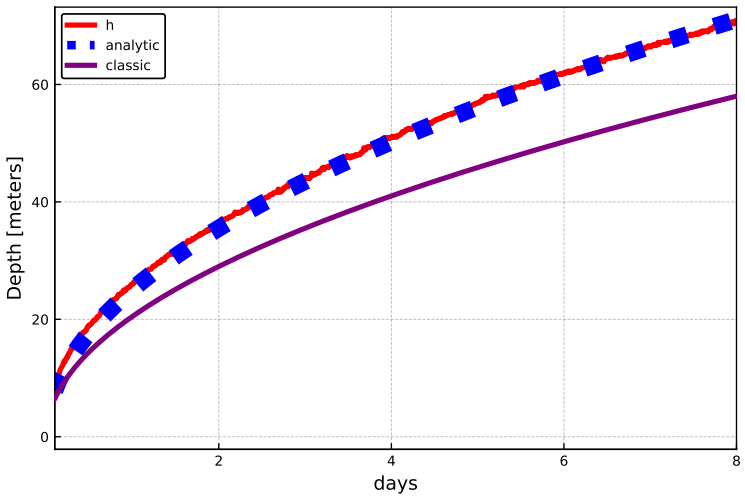

Figure 2. Boundary layer depth and its evolution in time after initial transients. The blue squares are the analytic scaling 4, the red line is an estimate of the boundary layer depth directly from the LES (described in the text), and the purple line is the classic scaling which ignores the entrainment layer 8.

Substituting $\bar{b}(z,0) = b_0 + N^2(z+h)$ and $\bar{b}(z,t) = b_0 + \Delta b$, an approximation of the profile shown in Fig. 1b except at very early times in the simulation, yields

$$\tfrac{1}{2}N^2h^2 - h\Delta b = Q_b t. \tag{7}$$

The first term on the left of equation 7 describes boundary layer deepening due to buoyancy loss at the surface, while the second term corresponds to the further cooling caused by turbulent mixing in the entrainment layer. Other authors have also arrived at a similar expression for the boundary layer depth upon taking into account turbulent entrainment. See, for example, Appendix F in (Van Roekel et al., 2018).

Ignoring turbulent mixing in the entrainment layer, i.e. setting $\Delta b = 0$, yields the deepening rate

$$h = \sqrt{2.0\frac{Q_b}{N^2}t}, \tag{8}$$

which differs by roughly 20% from the best fit expression 4 due to the effects of turbulent mixing in the entrainment layer. Equation 8 is the deepening rate associated with a convective adjustment parameterization and is known as the empirical law of free convection. We now review how these processes are represented in the KPP model.

2.2 The K-Profile Parameterization of penetrative convection

In penetrative convection in a horizontally-periodic domain, the K-Profile Parameterization models the horizontally-averaged temperature profile, $\bar{\theta}(z,t)$ with the coupled equations

$$\partial_t T = -\partial_z F\left(T, h; \boldsymbol{C}\right) \tag{9}$$

$$0 = \mathcal{D}(T, h; \boldsymbol{C}), \tag{10}$$

where $T(z,t)$ is the modeled temperature meant to approximate $\bar{\theta}(z,t)$, $h(t)$ is the boundary layer depth, $\boldsymbol{C} = \{C^S, C^N, C^D, C^H\}$ is a set of free parameters, $F(T, h; \boldsymbol{C})$ is the

parameterized temperature flux, and $\mathcal{D}(T, h; \boldsymbol{C})$ is a nonlinear constraint that determines the boundary layer depth at each time t. Our formulation, which isolates the four free parameters $\{C^S, C^N, C^D, C^H\}$, is superficially different but mathematically equivalent to the formulation in (Large et al., 1994) (see Appendix C for details). Finally, we emphasize that the K-Profile parameterization is deemed successful only if it accurately models the evolution of the entire observed temperature profile $\bar{\theta}(z,t)$, rather than, say, the boundary layer depth or the buoyancy jump across the base of the mixed layer.

The K-Profile Parameterization (KPP) represents F through the sum of a downgradient flux and a non-local flux term (Large et al., 1994),

$$F = -\underbrace{C^D \delta^{1/3} w_\star h \, \tfrac{z}{h} \left(1 + \tfrac{z}{h}\right)^2}_{\equiv K} \partial_z T + \underbrace{C^N Q^\theta \, \tfrac{z}{h} \left(1 + \tfrac{z}{h}\right)^2}_{\equiv \Phi}, \tag{11}$$

for $-h \le z \le 0$ and 0 otherwise, and $\delta = \min\{C^S, z/h\}$. Here $w_\star = (Q_b h)^{1/3}$ is the convective turbulent velocity scale, h is the boundary layer depth, $\tfrac{z}{h}\left(1 + \tfrac{z}{h}\right)^2$ is the 'K-profile' shape function (K is the namesake downgradient diffusivity of the K-Profile Parameterization) and Φ is a 'non-local' flux term that models convective boundary layer fluxes not described by downgradient diffusion.

The KPP model estimates the boundary layer depth h using the nonlinear constraint (10). The boundary layer geometry introduced in the right panel of figure 1 motivates the form of nonlinear constraint. The jump in buoyancy, Δb, is the difference between the buoyancy in the mixed layer and the base of the entrainment region. The buoyancy jump may thus be written in terms of the entrainment region thickness, Δh, and the entrainment region buoyancy gradient, N_e^2, as $\Delta b = N_e^2 \Delta h$. Using the plume theory outlined in Appendix B to motivate the scaling $\Delta h \propto w_\star / N_e$, we thus find

$$\tilde{C}^H = \frac{\Delta b}{w_\star N_e} \tag{12}$$

for some universal proportionality constant $\tilde{C}^H$. KPP posits that the boundary layer depth h is the first such depth from the surface at which equation 12 holds.

Large et al. (1994) estimate the mixed layer buoyancy with an average over the 'surface layer', $\frac{1}{C^S h} \int_{-C^S h}^{0} B(z) \mathrm{d}z$ where $B = \alpha g T$, and $0 < C^S < 1$ is a free parameter that defines the fractional depth of the surface layer relative to the total boundary layer depth, h. The buoyancy jump becomes, therefore

$$\Delta b = \frac{1}{C^S h} \int_{-C^S h}^{0} B(z) \mathrm{d}z - B(-h) \, . \tag{13}$$

Large et al. (1994) then express the stratification in the entrainment region, N_e, in terms of the stratification at the base of the boundary layer, such that

$$N_e \propto \sqrt{\max\left[0, \partial_z B(-h)\right]} \, . \tag{14}$$

The scaling in equation 14 introduces a new free parameter in addition to $\tilde{C}^H$; however because this free parameter is not independent from $\tilde{C}^H$, we combine the two into a new free parameter C^H, which we call the 'mixing depth parameter'. To prevent division by zero, the small dimensional constant $10^{-11} \mathrm{m}^2 \, \mathrm{s}^{-2}$ is added to the demoninator of equation 12 (Griffies et al., 2015). Combining equations 12, 13 and 14, we can write

$$0 = C^H - \frac{\frac{1}{C^S h} \int_{-C^S h}^{0} B(z) \mathrm{d}z - B(-h)}{(h Q_b)^{1/3} \sqrt{\max\left[0, \partial_z B(-h)\right]} + 10^{-11} \mathrm{m}^2 \mathrm{s}^{-2}} \, . \tag{15}$$

Equation 15 is the implicit nonlinear constraint in equation 10 that determines the boundary layer depth, h. In Appendix B we discuss the physical content of equation 15 for the case of penetrative convection.

The boundary layer depth criteria in equation 15 is often referred to as the bulk Richardson number criteria, because in mechanically forced turbulence the denominator is replaced by an estimate of the mean shear squared and C^H becomes a critical bulk Richardson number (Large et al., 1994). In penetrative convection there is no mean shear and C^H is not a Richardson number. See Appendix C for more details.

The representation of penetrative convection in KPP has four free parameters: the surface layer fraction C^{S}, the flux scalings C^N and C^D in equation 11, and the mixing depth parameter C^H in equation 15. Ranges for their default values are reported in (Large et al., 1994). We choose reference parameters within those ranges as

$$(C^S, C^N, C^D, C^H) = (0.1, 6.33, 1.36, 0.96). \tag{16}$$

These parameters are *not* the original set of independent parameters proposed by Large et al. (1994), but rather algebraic combinations thereof. Nevertheless, we emphasize that our formulation is mathematically identical to that proposed by Large et al. (1994). The mapping between the current set of parameters and the original are one-to-one, hence no information is lost in transforming from the current set of parameters to the original ones, see Appendix C for details. With regard to the numerical implementation, we do not use enhanced diffusivity as explained in the appendices of Large et al. (1994). Our objective is to calibrate the free parameters $\boldsymbol{C} = (C^S, C^N, C^D, C^H)$ by comparing KPP temperature profiles $T(z, t; \boldsymbol{C})$ with the LES output $\overline{\theta}(z, t)$.

3 Model calibration against LES solutions

We outline a Bayesian method for optimizing and estimating the uncertainty of the four free parameters through a comparison of the parameterization solution for $T(z, t; \boldsymbol{C})$ and the output $\overline{\theta}(z, t)$ of the LES simulations. First we introduce a loss function to quantify the parameterization-LES difference,

$$\mathcal{L}(\boldsymbol{C}) = \max_{t \in [t_1, t_2]} \left\{ \frac{1}{L_z} \int_{-L_z}^{0} \left[T(z, t; \boldsymbol{C}) - \overline{\theta}(z, t) \right]^2 dz \right\}. \tag{17}$$

We choose the square error in space to reduce the sensitivity to vertical fluctuations in the temperature profile. We take the maximum value of the squared error in time for $t \in [t_1, t_2]$ to guarantee that the temperature profile never deviates too far from the LES simulation at each instant in time. The parameterization is taken to be the KPP model given by equations 9 through 15, and the data are the horizontally averaged LES output. The initial time t_1 is chosen after the initial transition to turbulence of the LES simulations.

A natural way to extend the concept of loss functions to account for parameter uncertainty is to introduce a likelihood function for the parameters. Similar to how the form of the loss function is critical to the estimation of optimal parameters, the form of the likelihood function is critical for estimating the parameter uncertainties. The likelihood function quantifies what we mean by "good" or "bad" parameter choices. The Bayesian method uses this information to estimate parameter uncertainties. These estimates are only as good as the choice of likelihood function, much like optimal parameters are only as good as the choice of the loss function. See, for example, van Lier-Walqui, Vukicevic, & Posselt, 2012; Zedler, Kanschat, Korty, & Hoteit, 2012; Urrego-Blanco, Urban, Hunke, Turner, & Jeffery, 2016; Sraj, Zedler, Knio, Jackson, & Hoteit, 2016; Schneider, Teixeira, et al., 2017; Nadiga, Jiang, & Livescu, 2019; Morrison, van Lier-Walqui, Kumjian, & Prat, 2020 for definitions of likelihoods in various geophysical / fluid dynamical contexts. In Appendix D we discuss in detail the rationale for the choices made in this paper.

Following Schneider, Lan, et al. (2017) we introduce the likelihood function as the probability that parameter values explain the data $\mathbb{P}(\text{data}|\boldsymbol{C})$, as:

$$\mathbb{P}(\text{data}|\boldsymbol{C}) \propto \exp\left(-\frac{\mathcal{L}(\mathbf{C})}{\mathcal{L}_0}\right) \tag{18}$$

where $\mathcal{L}(\mathbf{C})$ is the loss function which depends both on data and parameters $\mathbf{C}$, and $\mathcal{L}_0 > 0$ is a *hyperparameter* associated with the likelihood function as opposed to a parameter in the parameterization. The posterior distribution, $\mathbb{P}(\boldsymbol{C}|\text{data})$, is then given by Bayes formula

$$\mathbb{P}(\boldsymbol{C}|\text{data}) \propto \mathbb{P}(\boldsymbol{C})\mathbb{P}(\text{data}|\boldsymbol{C}) \tag{19}$$

where $\mathbb{P}(\boldsymbol{C})$ is the prior distribution. In terms of probability densities, letting $\mathbb{P}(\boldsymbol{C}) \propto \rho^0(\boldsymbol{C})$ and $\mathbb{P}(\boldsymbol{C}|\text{data}) \propto \rho(\boldsymbol{C})$ denote our prior and posterior distributions for the parameters[2] $\boldsymbol{C}$, Bayes formula becomes

$$\rho(\mathbf{C}) \propto \rho^0(\boldsymbol{C}) \exp\left(-\frac{\mathcal{L}(\mathbf{C})}{\mathcal{L}_0}\right). \tag{20}$$

In our context Bayes' formula updates prior guesses about KPP parameter values and yields a posterior distrbution based on the LES data.

We choose the hyperparameter $\mathcal{L}_0$ as the minimum of the loss function $\mathcal{L}(\boldsymbol{C})$. The minimum is found using a modified simulated annealing procedure[3] (Kirkpatrick et al., 1983). Once the parameter values $\boldsymbol{C}^*$ that minimize the loss functions have been found, i.e. $\mathcal{L}_0 = \mathcal{L}(\boldsymbol{C}^*)$, the likelihood of any other parameter choice $\boldsymbol{C}^1$ is given by,

$$\rho(\boldsymbol{C}^1)/\rho(\boldsymbol{C}^*) = \exp\left(\frac{\mathcal{L}_0 - \mathcal{L}(\boldsymbol{C}^1)}{\mathcal{L}_0}\right). \tag{21}$$

For example, if the choice $\boldsymbol{C}^1$ increases the minimum of the loss function by a factor of two, i.e. $\mathcal{L}(\boldsymbol{C}^1) = 2\mathcal{L}_0$, then it is $1/e$ less likely. The probability distribution $\rho(\boldsymbol{C})$ is then sampled with a Random Walk Markov Chain Monte Carlo (RW-MCMC) algorithm (Metropolis et al., 1953), described further in Appendix E.

To illustrate our choices, as well as the RW-MCMC algorithm, we show a typical output from an RW-MCMC algorithm for a 2D probability distribution of the form in equation 18. We use the probability density function for the KPP parameterization presented in the next section, but keep two of the four parameters fixed (C^D and C^H) to reduce the problem from four to two parameters (C^N and C^S). The prior distributions for C^N and C^S are uniform over the ranges reported at the end of this section. The parameters C^D and C^H are set to the values that minimize the loss function. We show results for two arbitrary values of $\mathcal{L}_0$ for illustrative purposes. Starting from a poor initial guess, the RW-MCMC search proceeds towards regions of higher probability (lower loss function) by randomly choosing which direction to go. Once a region of high probability is found, in this case parameter values in the "blue" region, the parameters hover around the minimum of the loss function as suggested by the high values of the likelihood function. The orange hexagons represent the process of randomly walking towards the minimum of the loss function and correspond to the "burn-in" period. The burn-in period is often thrown away when calculating statistics since it corresponds to an initial transient before the RW-MCMC settles around the minimum of the likelihood function. We see that the choice of $\mathcal{L}_0$ does not change the overall structure of the probability distribution but does affect how far from optimal parameters the random walk is allowed to drift.

[2] The proportionality sign is introduced, because Bayes' formula applies to probabilities, while $\rho^0(\boldsymbol{C})$ is a probability density function.

[3] In simulated annealing one finds the minimum of the loss function decreasing $\mathcal{L}_0$ to zero as one explores the parameter space through a random walk. Here we keep updating $\mathcal{L}_0$ to the new local minimum every time the random walk stumbles on a set of parameters, for which $\mathcal{L}(\mathbf{C}) < \mathcal{L}_0$.

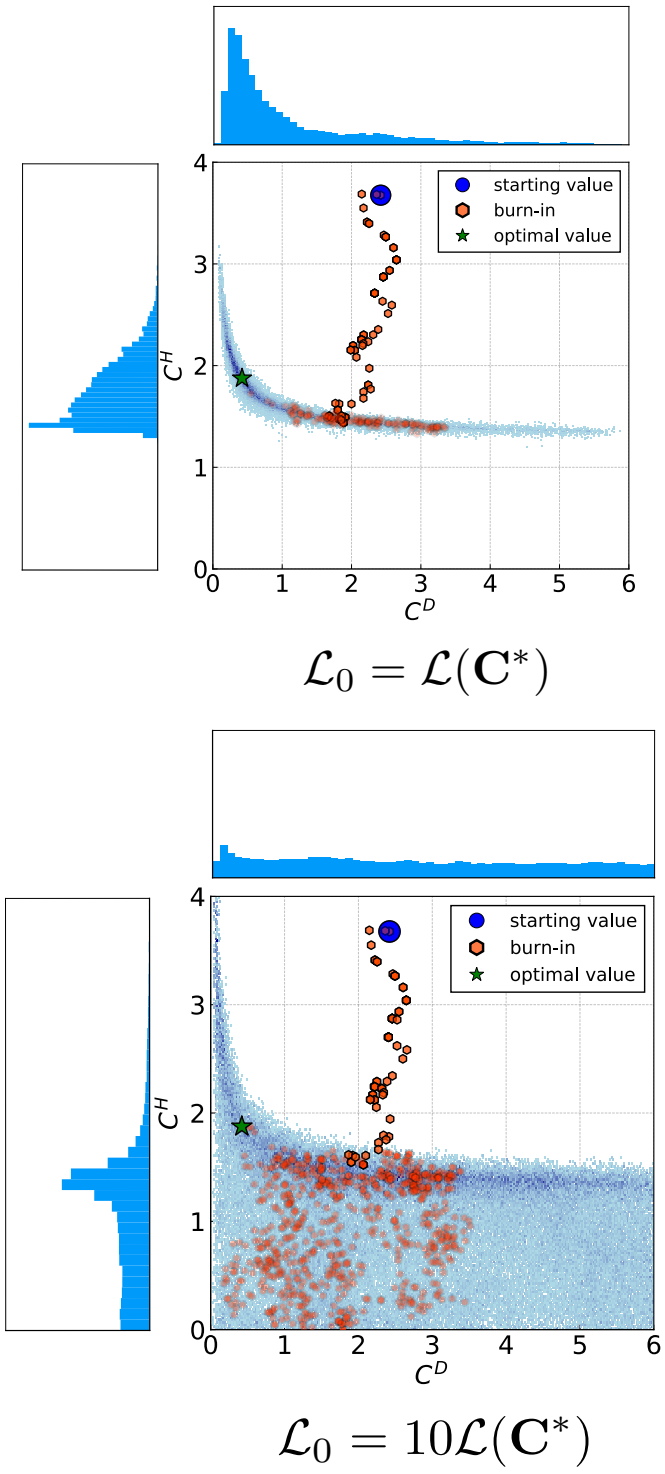

Figure 3. An example of a RW-MCMC search trajectory based on a sample probability distribution for KPP parameters using 10^5 RW-MCMC iterations. The trajectories starts from a region of very low probability (white areas) and moves toward progressively higher probabilities (the darker the blue shading, the higher the probability). The blue probability distributions on the left side and the top are the corresponding marginal distributions of C^H and C^D, respectively. The green star is the best known optimal of the probability distribution (i.e, the mode of the probability distribution). The value of $\mathcal{L}(\boldsymbol{C}^*)$ is the value of the loss function at the green star.

Parameterizations such as KPP exhibit a dependence on resolution in addition to nondimensional parameters. Here we perform all calculations for a vertical resolution $\Delta z = 6.25$ m and timestep $\Delta t = 10$ minutes representative of those used in state of the art ESMs. We do not use enhanced diffusivity as in (Large et al., 1994) for this resolution. The parameterization is relatively insensitive to halving Δz and Δt, for a fixed set of parameters, but the results are sensitive to doubling either one. Thus the optimal parameter values and their uncertainties are only appropriate for the resolution used for the calibration and would need to be updated especially if the parameterization was run at a coarser resolution. This dependence on resolution could be handled within the Bayesian method by introducing Δz and Δt as additional parameters in the probability distribution, but we do not pursue this approach.

The temporal window used to compute the loss function is from $t_1 = 0.25$ days (so as to eliminate initial transients in the LES) to the final simulation day chosen to be when $h \approx 70$ meters. We apply the Bayesian parameter estimation procedure to KPP using data from one LES simulation in section 3.1 and from multiple LES simulations using different initial stratifications in section 3.2. We use a uniform prior distributions for the KPP parameters over the following ranges:

$$0 \le C^S \le 1,\ 0 \le C^N \le 8,\ 0 \le C^D \le 6,\ \text{and } 0 \le C^H \le 5. \tag{22}$$

The surface layer fraction C^S, being a fraction, must stay between zero and one. The other parameter limits are chosen to span the whole range of physically plausible values around the reference values given in equation (16). The choice of uniform distributions is made to avoid favoring any particular value at the outset.

3.1 Calibration of KPP parameters from one LES simulation

In this section we apply the Bayesian calibration method to the LES simulation of penetrative convection described in section 2.1 and quantify uncertainties in KPP parameters in section 2.2. The horizontal averages from the LES simulations are compared with predictions from solutions of the KPP boundary layer scheme, equations 9 and 10. The boundary and initial conditions for KPP are taken to be the same as those for the LES simulation, i.e., 100 W/m^2 cooling at the top, $\partial_z T = 0.01$°C m^{-1} at the bottom, and an initial profile $T_p(z,0) = 20\text{°C} + 0.01\text{°C m}^{-1} z$.

To estimate the full probability distribution function, we use the RW-MCMC algorithm with 10^6 iterations to sample the probability distributions of the four KPP parameters (C^S, C^N, C^D, C^H). The large number of forward runs is possible because the forward model consists of a one-dimensional equation, namely, KPP in single column mode. The Markov chain leads to roughly 10^4 statistically independent samples as estimated using an autocorrelation length, see Sokal (1997). The RW-MCMC algorithm generates the entire four dimensional PDF, equation 18.

The parameter probability distribution can be used to choose an optimal set of KPP parameters. Of the many choices, we pick the most probable value of the four dimensional probability distribution, the mode, because it minimizes the loss function, see Appendix D for the detailed calculation. In figure 4a we show the horizontally averaged temperature profile from the LES simulation (continuous line) and the temperature profiles obtained running the KPP parameterization with reference and optimal parameters (squares and dots) at $t = 8$ days. The optimized temperature profiles are more similar to the LES simulation than the reference profiles especially in the entrainment region. Figure 4b confirms that the square root of the instantaneous loss function, the error, grows much faster with the reference parameters. The oscillations in the error are a consequence of the coarseness of the KPP model: only one grid point is being entrained at any given moment.

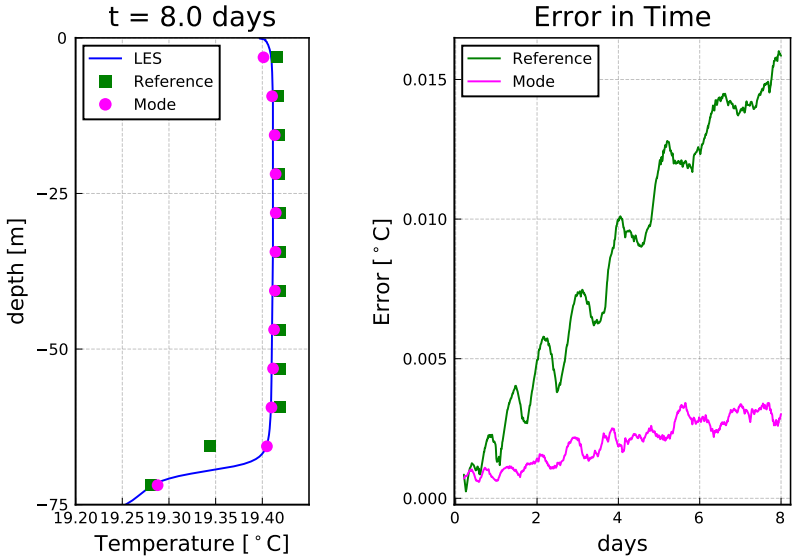

Figure 4. KPP and horizontally averaged LES temperature profiles for different point estimates of parameters at t=8 days as well as the error in time. In the left plot, the blue squares correspond to reference parameter choices, the red circles correspond to the optimized parameterization (using the mode of the probability distribution), and the blue line to the horizontally averaged LES solution, all at time t=8 days. On the right plot we show the instantaneous error at each moment in time. We see that the "optimal" parameter does indeed reduce the bias over the time period. The loss function is the largest square of the error over the time interval.

The improvement in boundary layer depth through optimization of the parameters is about 10%, or 10 m over 8 days. As discussed in section 2.1, the rate of deepening can be predicted analytically within 20% by simply integrating the buoyancy budget over time and depth and assuming that the boundary layer is well mixed everywhere, i.e. ignoring the development of enhanced stratification within an entrainment layer at the base of the mixed layer. KPP improves on this prediction by including a parameterization for the entrainment layer. The reference KPP parameters contribute a 10% improvement on the no entrainment layer prediction, and the optimized parameters contribute another 10%. While these may seem like modest improvements, they can prevent large biases for the boundary layer depth when integrated over a few months of cooling in winter rather than just 8 days. We will return to this point in the next section when we discuss structural deficiencies in the KPP formulation.

To visualize the probability distribution we focus on 2D marginal distributions, e.g.,

$$\rho_{2DM}(C^H, C^S) = \int\int \rho(\boldsymbol{C})\ \mathrm{d}C^D \mathrm{d}C^N, \tag{23}$$

along with the other five possible pairings, as well as the 1D marginal distributions such as

$$\rho_M(C^H) \equiv \iiint \rho(\boldsymbol{C})\ \mathrm{d}C^S \mathrm{d}C^D \mathrm{d}C^N, \tag{24}$$

and similarly for the other three parameters.

The marginal distribution can intuitively be thought of as the total of a parameter (or pair of parameters) while taking into account the total uncertainty of other pa-

rameters. Furthermore, the marginal distribution takes into account potential compensating effects that different parameters may have on one another. The marginal distribution does *not* capture the effect of individually varying a parameter while keeping all the other parameters fixed at a particular value[4]. That is an effect represented by a conditional distribution.

Constructing the marginal distributions only requires constructing histograms of the trajectories generated by the RW-MCMC algorithm. The 2D marginal distributions are visualized with heatmaps in figure 5 and the 1D marginal distributions of the corresponding parameters are shown along the outermost edges. For the 2D marginal distributions, the dark blue regions correspond to regions of high probability and the light blue regions are regions of low probability. The white space corresponds to regions that the RW-MCMC algorithm never visited. The 2D marginal distributions show that parameters must be changed in tandem with one another in order to correspond to a similar model output. Furthermore their structure is distinctly non-Gaussian.

The 1D marginal distribution of the mixing depth parameter C^H (the bottom left rectangular panel) is much more compact than that of the other three parameters suggesting that it is the most sensitive parameter. The mixing depth parameter's importance stems from its control over both the buoyancy jump across the entrainment layer and the rate-of-deepening of the boundary layer. (Again it may be useful to remember that C^H is often referred to as the bulk Richardson number in the KPP literature, even though it takes a different meaning in convective simulations, see Appendix C.) The parameters C^D and C^N set the magnitude of the local and nonlocal fluxes. Results are not sensitive to their specific values, as long as they are large enough to maintain a well-mixed layer. The value of the surface layer fraction C^S is peaked at lower values but is less sensitive to variations than C^D or C^H.

The uncertainties of the parameters can be used to infer the uncertainties of the temperature profile at each depth and time, predicted by KPP. To do this, we subsample the 10^6 parameter values down to 10^4 and evolve KPP forward in time for each set of parameter choices. We construct histograms for the temperature field at the final time for each location in space individually. We then stack these histograms to create a visual representation of the model uncertainty. This uncertainty quantifies the sensitivity of the parameterization with respect to parameter perturbations as defined by the parameter distributions.

The histogram of temperature profiles at time $t = 8$ days as calculated by both our prior distribution (uniform distribution) and the posterior distribution (as obtained from the RW-MCMC algorithm) is visualized in figure 6. We see that there is a reduction of the uncertainty in the temperature profile upon taking into account information gained from the LES simulation. The salient features of the posterior distribution temperature uncertainty are

1. 0-10 meter depth: There is some uncertainty associated with the vertical profile of temperature close to the surface.
2. 20-60 meter depth: The mean profile of temperature in the mixed layer is very well predicted by KPP.
3. 60-70 meter depth: The entrainment region contains the largest uncertainties.
4. 70-100 meter depth: There is virtually no uncertainty. The unstratified region below the boundary layer does not change from its initial value.

[4] That is, unless the other parameters have essentially delta function 1D marginal distributions.

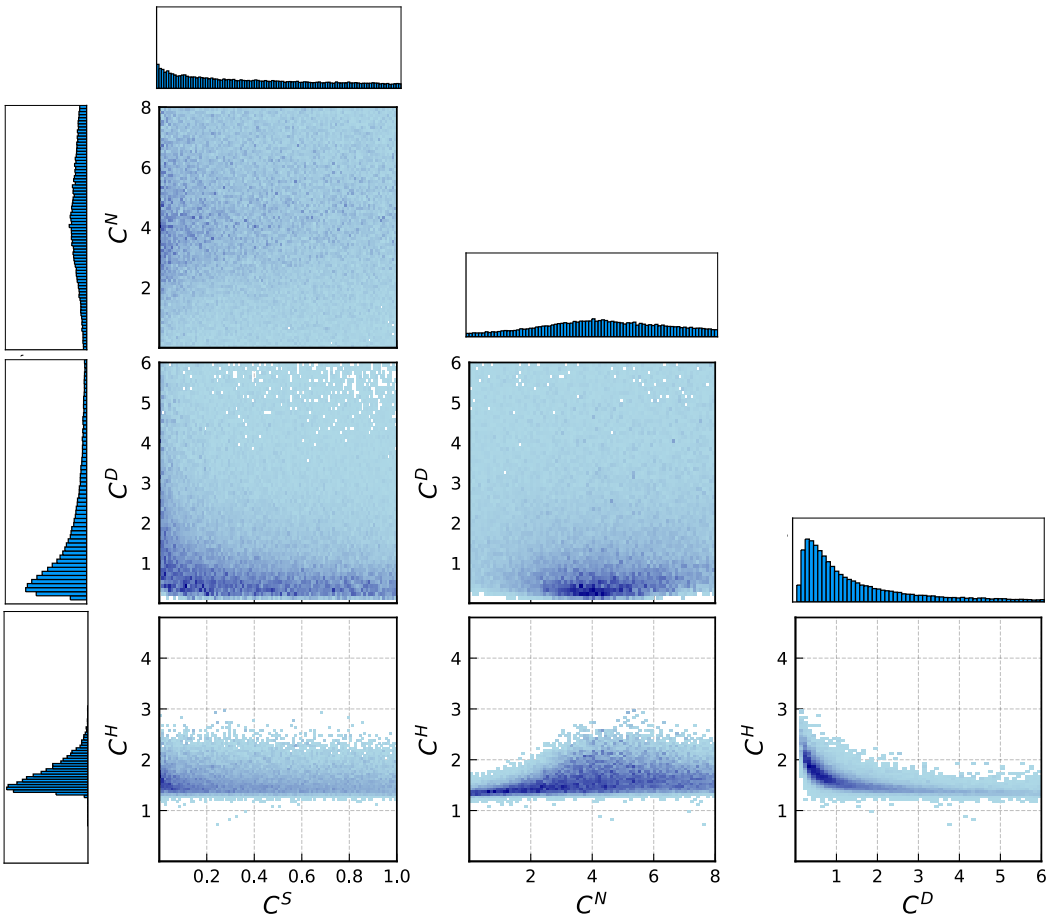

Figure 5. Marginal Distributions for KPP Parameters. The dark blue regions correspond to regions of high probability and the light blue regions are regions of low probability. The white space corresponds to regions that the RW-MCMC algorithm never visited. The corresponding 1D marginal distributions (corresponding to integrals of the 2D marginal distributions) are displayed on the left and on top of the plots for reference.

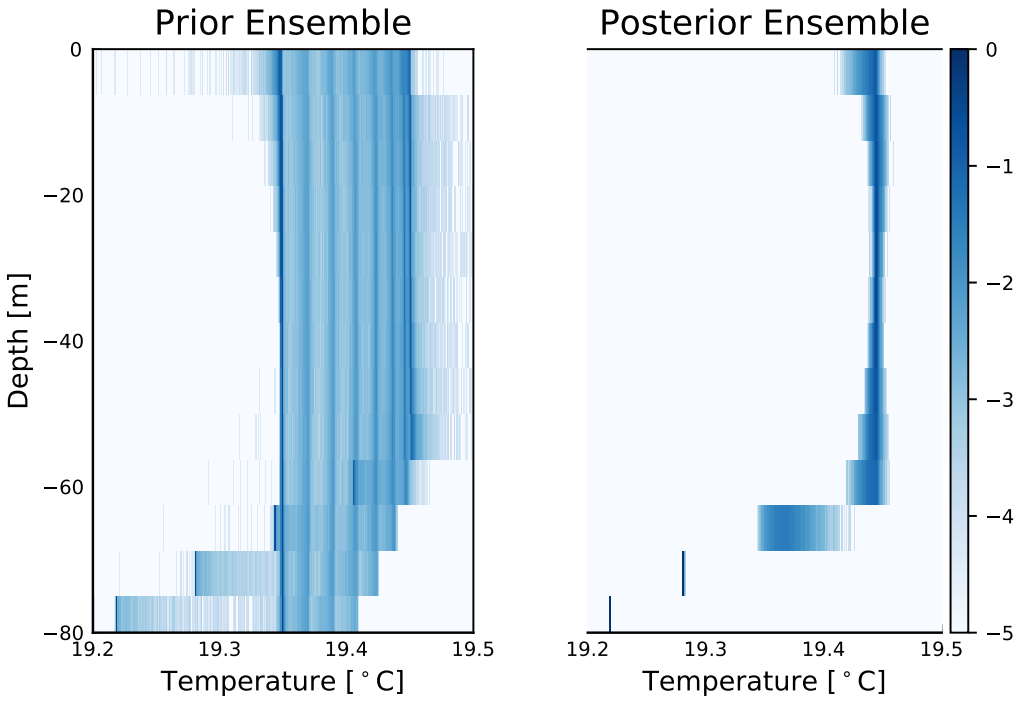

Figure 6. Uncertainty propagation of the temperature profile with respect to the prior and posterior probability distributions. The use of probability distributions for parameters has the consequence that the temperature field is no longer a point estimate, but rather a probability distribution at each moment in space and time. By sampling from the parameter probability distributions and evolving the parameterization forward in time, we obtain a succinct representation of what it means to "fiddle" with parameters. The legend on the right shows what the colors correspond to in terms of the base 10 logarithm of the probability distributions.

Now that we have applied the Bayesian methodology to one LES simulation and explored its implications, we are ready to apply the method to multiple LES simulations covering different regimes in the following section.

3.2 Calibration of KPP parameters from multiple LES simulations

We now use our Bayesian framework to explore possible sources of bias in the KPP model. To this end we investigate what happens when we change the initial stratification in penetrative convection simulations. This is motivated by recent work on boundary layer depth biases in the Southern Ocean (DuVivier et al., 2018; Large et al., 2019). In those studies, KPP failed to simulate deep boundary layers in winter when the subsurface summer stratification was strong.

We perform 32 large eddy simulations and calculate parameter distributions for each case. In the previous section we saw that C^H is the most sensitive parameter. Thus our focus now will be on the optimization and uncertainty quantification of C^H. In the background, however, we are estimating *all* parameters. We keep the surface cooling constant at 100 W/m^2 for all regimes, and only vary the initial stratification. The integration time was stopped when the boundary layer depth filled about 70% of the domain in each simulation. We used 128^3 grid points in the LES, ≈ 0.8 meter resolution in each direction[5]. We use a lower resolution for the LES in these trend studies as compared to those in the previous section, but results were not sensitive to this change. In the Bayesian inference, each one of the probability distributions were calculated 10^5 iterations of RW-MCMC, leading to an effective sample size on the order of 10^3. The stratifications ranged from $N^2 \approx 1 \times 10^{-6}$ to $N^2 \approx 3.3 \times 10^{-5} s^{-2}$.

We find, as visualized in figure 7, that C^H is not constant but depends on the background stratification, N^2. The blue dots are the median values of the probability distributions and the stars are the modes (minimum of the loss function). The error bars correspond to 90% probability intervals, meaning that 90% of parameter values fall between the error bars. The large discrepancy between the median and the mode is due to the mode being the optimal value of the entire four dimensional distribution whereas the median only corresponds to the marginal distribution. The reference KPP value is plotted as a dashed line.

The median values and optimal values increase monotonically with the initial stratification revealing a systematic bias. Furthermore, it exposes *where* the systematic bias comes from: no single value of C^H, equation 15, can correctly reproduce the deepening of the boundary layer for all initial stratifications. This suggests that the scaling law for the boundary layer depth criteria is incommensurate with the LES data.

The failure of equation 15 can be understood by going back to the buoyancy budget in equation 7. Using the KPP estimate for the buoyancy jump across the entrainment layer,

$$\Delta b \equiv \frac{1}{C^S h} \int_{-C^S h}^{0} B(z) \mathrm{d}z - B(-h), \tag{25}$$

and introducing $N_h^2 \equiv \partial_z B(-h)$ for the stratification at the base of the entrainment layer to distinguish it from the interior stratification N^2, we find that the boundary layer depth criterion, equation 15, implies,

$$h \Delta b \simeq C^H h^{4/3} \left(Q_b\right)^{1/3} N_h. \tag{26}$$

[5] Although the parameter estimates will vary upon using less LES resolution, the qualitative trends are expected to be robust.

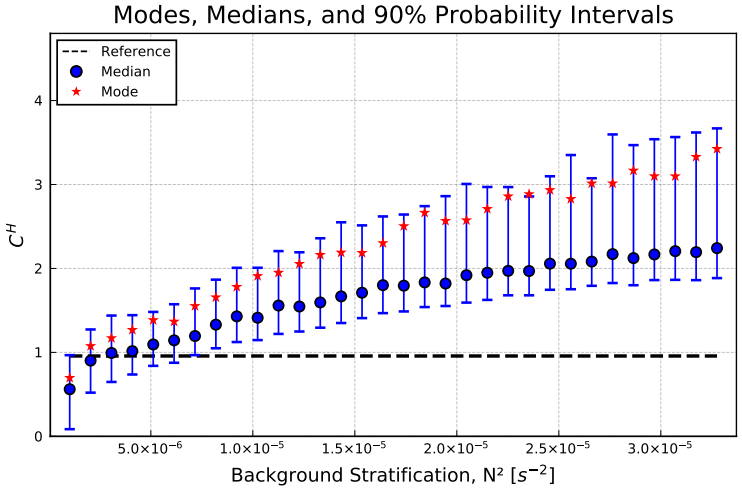

Figure 7. Mixing depth parameter optimized across various background stratification. The dots are the median values, the stars are the mode, and the error bars correspond to 90% probability intervals. The horizontal dashed line is the default value of the mixing depth parameter for reference. Here one can see that the mixing depth parameter when estimated across various regimes produces different results. This is a signature of a systematic bias in the parameterization.

Substituting this expression in the buoyancy budget, equation 7, one obtains an implicit equation for the evolution of the boundary layer depth h,

$$\left(\frac{1}{2}N^2 - C^H \left(Q_b\right)^{1/3} h^{-2/3} N_h\right) h^2 \simeq Q_b t. \tag{27}$$

The LES simulation described in section 2.1, and many previous studies of penetrative convection, e.g. (Van Roekel et al., 2018; Deardorff et al., 1980), show that the boundary layer depth grows as $\sqrt{t}$. To be consistent, N_h would have to scale as $h^{2/3}$, but this is not observed in the LES simulations nor supported by theory. This suggests that we must modify the formulation of boundary layer depth, as we now describe.

3.3 Modification of the KPP parameterization to reduce biases

From the multi-regime study of the previous section we found that there is no optimal KPP mixing depth parameter C^H that works for arbitrary initial stratification. This prompted us to look for an alternative formulation of the depth criterion which satisfies the well known empirical result that the boundary layer depth deepens at a rate,

$$h \simeq \sqrt{c \frac{Q_b}{N^2} t}, \tag{28}$$

where c is a dimensionless constant found to be close to 3.0 with the LES simulation in section 2.1. Furthermore, c was found to be close to 3.0 across all the numerical experiments from section 3.2. Substituting this expression into the buoyancy budget, equation 7, we find that,

$$\frac{\Delta b}{h N^2} \simeq \left(\frac{1}{2} - \frac{1}{c}\right). \tag{29}$$

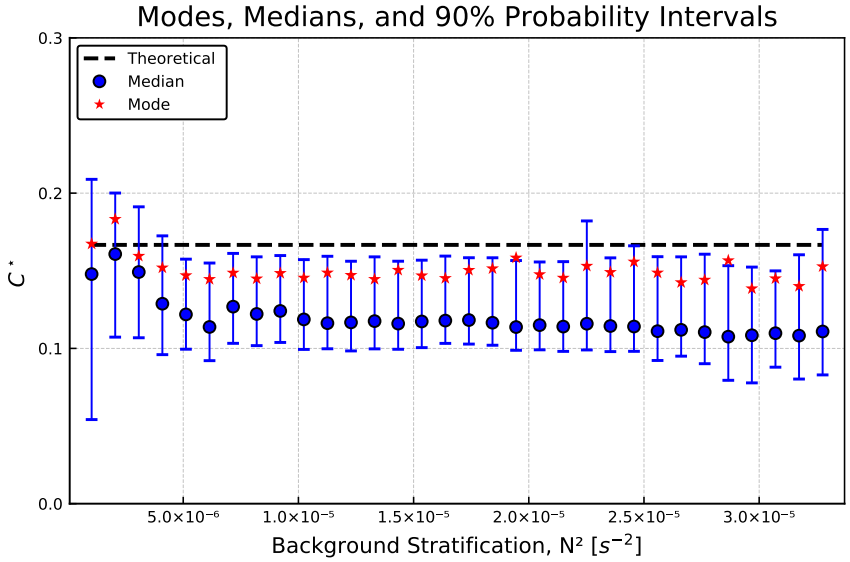

Figure 8. The modified mixing depth parameter optimized across various background stratification. The dots are the median values, the stars are the mode, and the error bars correspond to 90% probability intervals. The dashed line corresponds to 1/6, the theoretical expectation based on equation 31. This is similar to figure 7, but using the modification from section 3.3. Here one can see that there mixing depth parameter when estimated across various regimes produces similar results. This is a desirable feature in a parameterization.

This expression can then be used as a new boundary layer depth criterion to replace equation 15,

$$C^\star = \frac{h\left(\frac{1}{C^S h}\int_{-C^S h}^{0} B(z)dz - B(-h)\right)}{N^2 h^2 + 10^{-11}\mathrm{m^2 s^{-2}}}, \tag{30}$$

where $C^\star$ replaces C^H as the dimensionless parameter whose value sets the boundary layer depth. The value of N^2 here is the background stratication. Based on equation 29 and our LES data, we expect

$$C^\star \simeq \left(\frac{1}{2} - \frac{1}{c}\right) \simeq \frac{1}{6}. \tag{31}$$

Equation 30 is an implicit equation for h which guarantees that equation 28 holds.

We now repeat the model calibration in section 3.2 but using this new boundary layer depth criterion to test whether there is an optimal value of $C^\star$ that is independent of initial stratification. We estimate all KPP parameters and show the new mixing depth parameter for simulations with different initial stratifications in figure 8. Encouragingly there is no obvious trend in the optimal values of $C^\star$ and the error bars overlap for all cases. This supports the new criterion in the sense that parameters estimated in different regimes are now consistent with one another. The uncertainties in $C^\star$ translate into an uncertainty in boundary layer depth prediction. In particular, values between $0.05 \leq C^\star \leq 0.2$ imply a boundary layer depth growth in the range $\sqrt{2.22tQ_b/N^2} \leq h \leq \sqrt{3.33tQ_b/N^2}$.

Additionally, one can check if the constants estimated following the methodology of section 3 are consistent with an *independent* measure directly from the diagnosed LES

simulation. In particular the LES simulations suggest that $C^\star \simeq 1/6$ as per equation 31. From figure 8 we see that the optimal $C^\star$ is smaller than $1/6 = 0.167$ (the dashed black line) and the value $1/6$ is not within the confidence intervals for many of the simulations. There are several potential reasons for the discrepancy, e.g., the neglect of curvature in the buoyancy budget (since we assumed a piece-wise linear buoyancy profile) or the finite resolution of the parameterization. Perhaps the most likely explanation is the difference in how the boundary layer depth was diagnosed in the LES, which need not have the same meaning as the one in KPP. A different definition in the LES simulation, such as the depth of maximum stratification, would yield a different scaling law, but still proportional to $\sqrt{t}$. Whatever the choice, the Bayesian parameter estimation bypasses these ambiguities/inconsistencies by direct comparison with the entire horizontally average temperature profile from the LES.

We do not explore other modifications to the boundary layer depth criterion as this would greatly expand the scope of this article. Furthermore, biases in KPP are not limited to the cases explored here, see Van Roekel et al. (2018) for discussions and remedies. The criterion described in this section assumes a constant initial stratification and a constant surface heat loss, which leads to the $\sqrt{t}$ growth of the boundary layer depth. It would be interesting to extend the criterion to arbitrary initial stratification, variable surface heat fluxes, not to mention the interaction with wind-driven mixing. The goal here is not to derive a new parameterization, but rather to illustrate and argue for a Bayesian methodology in the refinement and assessment of parameterizations.

4 Discussion

We presented a Bayesian approach to assess the skill of the K-Profile Parameterization (KPP) for turbulent convection triggered by surface cooling in an initially stably stratified ocean. The KPP model for this physical setting consists of a one dimensional model with an algebraic constraint for the mixing-layer depth together with four non-dimensional parameters. These parameters consisted of an algebraic reorganization of the original KPP parameters so that terms in the equations could be associated with choices of parameters. Parameters were estimated by reducing the mismatch between the vertical buoyancy profile predicted by KPP and the area-averaged buoyancy profile simulated with a three dimensional LES code for the same initial conditions and surface forcing. Using Bayes' formula we further estimated the full joint probability distribution of the four parameters. Furthermore, the probability distribution was used to quantify inter-dependencies among parameters and their uncertainty around the optimal values.

Repeating the Bayesian parameter optimization and uncertainty quantification for different initial stratifications, we found that no unique set of parameters could capture the deepening of convection in all cases. This implied that the KPP formulation does not capture the dependence of convection on the initial stratification in the simple test case considered here: constant surface cooling, constant initial stratification, no wind, and no background flow. The parameter that required re-tuning for each case was the one associated with the boundary layer depth criterion, thereby suggesting that this criterion has the wrong functional dependence on stratification. We thus reformulated the boundary layer depth criterion to capture the semi-analytical result, supported by the LES simulations, that the boundary layer depth deepens as the square root of time when the initial stratification is constant. The validity of the new formulation was vindicated because the Bayesian approach was able to find a set of parameters which captured the evolution of the boundary layer, as compared to the LES simulations, for all initial stratifications. In this way, the Bayesian methodology allowed us identify and remove a bias in KPP formulation.

The methodology outlined here could be as easily applied to other parameterizations of boundary layer turbulence, such as those reviewed in CVMix (Griffies et al., 2015), Pacanowski and Philander (1981), Mellor and Yamada (1982), Price et al. (1986), and Kantha and Clayson (1994) . It is expected that the inclusion of additional physics, such as wind-driven mixing and its interaction with convection, would also be amenable to the techniques described in this manuscript. Our experience is that progress is faster if one starts with simple idealized setups, like the ones considered here, and then move to progressively more realistic ones which accounted for variable stratification and surface heat fluxes, wind-stress forcing, background shear, surface waves, etcetera. The Bayesian method would then provide a rigorous evaluation of parameter uncertainty, parameter dependencies, and biases in the formulation of the parameterization.

Ultimately, our hope is that parameter probability distributions estimated in local regimes will serve as useful prior information for calibration/tuning of Earth System Models (ESMs). Local simulations of turbulence must be carefully designed and incorporate suites of subgrid-scale processes that have leading order impact in global ocean dynamics: surface and bottom boundary layer turbulence, surface wave effects, deep convection, mesoscale and submesoscale turbulence, and so forth. Bayesian calibration of parameterization for each subgrid-scale process will then result in probability distributions for all the nondimensional parameters associated with the parameterizations. These distributions can then be used as prior information for what is a reasonable range of values that each parameter can take, when the parameterizations are implemented in an ESMs.

With regards to calibration of ESMs, the parameterizations of different subgrid-scale processes may nonlinearly interact with each other and with the resolved physics. Additional calibration is then required for the full ESM. Presently this is achieved by perturbing the parameters within plausible ranges (Mauritsen et al., 2012; Schmidt et al., 2017). The Bayesian approach provides an objective approach to determine a plausible range. The same algorithm cannot be used to calibrate the ESM, because the methodologies described here are not computationally feasible when applied to larger systems. Promising approaches to address this challenge through the use of surrogate models are described in Sraj et al. (2016) and Urrego-Blanco et al. (2016). Such models bring internal sources of uncertainty and it is not clear to what extent one can trust a surrogate of a full ESM. One potential way to address this additional challenge is the Calibrate, Emulate, and Sample (CES) approach outlined in Cleary et al. (2020). There the surrogate model's uncertainty is estimated through the use of Gaussian processes and included as part of a consistent Bayesian calibration procedure.

Should the global problem still exhibit significant biases, even when all available prior information about parameterizations and about global data are leveraged utilizing emulators or traditional methods of tuning, then one would have to conclude that there is a fundamental deficiency in our understanding of how the different components of the climate system interact with one another, or that perhaps the models do not include some key process. For example, Rye et al. (2020) argue that glacial melt might be one such missing process which is not currently represented in ESMs. The advantage of the systematic calibration approach outlined here is that it allows us to quantify uncertainty in ESM projections and identify the sources of such uncertainty.

Appendix A Oceananigans.jl

Oceananigans.jl (Ramadhan et al., 2020) is open source software for ocean process studies written in the Julia programming language (Bezanson et al., 2017; Besard et al., 2019). For the large eddy simulations (LESs) reported in this paper, Oceananigans.jl is configured to solve the spatially-filtered, incompressible Boussinesq equations with a temperature tracer. Letting $\boldsymbol{u} = (u, v, w)$ be the three-dimensional, spatially-filtered ve-

locity field, θ be the conservative temperature, p be the kinematic pressure, f be the Coriolis parameter, and $\boldsymbol{\tau}$ and $\boldsymbol{q}$ be the stress tensor and temperature flux due to subfilter turbulent diffusion, the equations of motion are A1–A3,

$$\partial_t \boldsymbol{u} + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} + f \hat{z} \times \boldsymbol{u} + \nabla p = b\hat{z} - \nabla \cdot \boldsymbol{\tau}, \tag{A1}$$

$$\partial_t \theta + \boldsymbol{u} \cdot \nabla \theta = -\nabla \cdot \boldsymbol{q}, \tag{A2}$$

$$\nabla \cdot \boldsymbol{u} = 0. \tag{A3}$$

The buoyancy b appearing in A1 is related to conservative temperature by a linear equation of state,

$$b = \alpha g\, (\theta_0 - \theta)\,, \tag{A4}$$

where $\theta_0 = 20^\circ\,\mathrm{C}$ is a reference temperature, $\alpha = 2 \times 10^{-4}\,(^\circ\mathrm{C})^{-1}$ is the thermal expansion coefficient, and $g = 9.81\,\mathrm{m^2\,s^{-1}}$ is gravitational acceleration at the Earth's surface.

A1 Subfilter stress and temperature flux

The subfilter stress and momentum fluxes are modeled with downgradient closures, such that

$$\tau_{ij} = -2\nu_e \Sigma_{ij} \qquad \text{and} \qquad \boldsymbol{q} = -\kappa_e \nabla \theta, \tag{A5}$$

where $\Sigma_{ij} \equiv \frac{1}{2}\left(\partial_i u_j + \partial_j u_i\right)$ is the strain rate tensor, and ν_e and κ_e are the eddy viscosity and eddy diffusivity of conservative temperature. The eddy viscosity ν_e and eddy diffusivity κ_e in equation A5 are modeled with the anisotropic minimum dissipation (AMD) formalism introduced by (Rozema et al., 2015) and (Abkar et al., 2016), refined by (Verstappen, 2018), and validated and described in detail for ocean-relevant scenarios by (Vreugdenhil & Taylor, 2018). AMD is simple to implement, accurate on anisotropic grids (Vreugdenhil & Taylor, 2018), and relatively insensitive to resolution (Abkar et al., 2016).

A2 Numerical methods

To solve equations A1–A3 with the subfilter model in equation A5 we use the software package '`Oceananigans.jl`' written in the high-level Julia programming language to run on Graphics Processing Units, also called 'GPUs' (Bezanson et al., 2017; Besard et al., 2019; Besard et al., 2019). `Oceananigans.jl` uses a staggered C-grid finite volume spatial discretization (Arakawa & Lamb, 1977) with centered second-order differences to compute the advection and diffusion terms in equation A1 and equation A2, a pressure projection method to ensure the incompressibility of $\boldsymbol{u}$, a fast, Fourier-transform-based eigenfunction expansion of the discrete second-order Poisson operator to solve the discrete pressure Poisson equation on a regular grid (Schumann & Sweet, 1988), and second-order explicit Adams-Bashforth time-stepping. For more information about the staggered C-grid discretization and second-order Adams Bashforth time-stepping, see section 3 in (Marshall et al., 1997) and references therein. The code and documentation are available for perusal at `https://github.com/CliMA/Oceananigans.jl`.

Appendix B Parcel Theory Derivation for the KPP Boundary Layer Depth Criterion

Here we summarise the derivation of the KPP boundary layer depth criterion for penetrative convection, because we could not find a succinct description in the published literature. Following (Deardorff et al., 1980) we consider the vertical momentum equation for a parcel punching through the entrainment layer,

$$w' \frac{dw'}{dz} \simeq -(b' - \bar{b}) \tag{B1}$$

where b' is the buoyancy of the parcel, assumed to be equal to the mixed layer value, and $\bar{b}$ is the area mean buoyancy profile in the entrainment layer. This equation holds if the area occupied by sinking plumes is small compared to the total area so that $\bar{b}$ is a good proxy for the buoyancy in the environment around the plumes and $b'-\bar{b}$ represents the buoyancy force experienced by the parcel. The parcel velocity decelerates from $w' \equiv w_e$ at the mixed layer depth ($z = -h + \Delta h$) to zero at the boundary layer depth ($z = -h$) where turbulence vanishes. Assuming that the background stratification N_e^2 is approximately constant in the entrainment layer we also have $b' - \bar{b} = N_e^2(-h + \Delta h) - N_e^2 z$. The momentum equation can then be integrated from $z = -h + \Delta h$ to $z = -h$,

$$(w_e)^2 \simeq N_e^2 \Delta h^2, \tag{B2}$$

assuming that the background stratification N_e^2 is constant in the entrainment layer. Introducing Δb as the difference between the environment buoyancy in the mixed layer and that at the base of the entrainment layer, we have $\Delta b = N_e^2 \Delta h$, and hence,

$$\Delta b \propto w^* N_e, \tag{B3}$$

and (Deardorff et al., 1980) assumes that $w_e \propto w^*(-h + \Delta h)$. The criterion for diagnosing the boundary layer depth follows from this relationship; h is defined as the first depth z below the ocean surface where,

$$\frac{\Delta b(-h)}{w^*(-h) N_e(-h)} = C^H, \tag{B4}$$

for some universal constant C^H. In the main text we show this scaling fails to predict the rate of deepening of the boundary layer depth in LES simulations. Further analysis, not reported here, show that this failure stems from relationship (B3) which is not supported by the simulations.

Equation (B4) is often referred to as a critical Richardson number criterion which may seem odd given that no Richardson number appears in the expression. This is best understood if one extends the criterion to the case when there is a momentum shear in the boundary layer, typically induced by mechanical stresses, such that in addition to a density jump $\Delta b(z)$ there is also a momentum jump $\Delta u(z)$ across the entrainment layer. The entrainment layer base is then found where the Richardson number matches a critical value Ri_c,

$$\mathrm{Ri}_c = \frac{\Delta b(-h)}{(\Delta u(-h))^2 + \frac{C^H}{\mathrm{Ri}_c} w^*(-h) N_e(-h)}. \tag{B5}$$

The rationale behind this extended criterion can be found in (Large et al., 1994). For the purely convective limit $\Delta u(-h) = 0$ and the dependence on Ri_c drops out.

Appendix C Relationship between the model in section 2.2 and Large et al. (1994)'s formulation of KPP

The formulation of KPP in Section 2.2 represents an algebraic reorganization of the formulation proposed by Large et al. (1994). The two formulations are mathematically equivalent. In this appendix, we discuss in detail how the four free parameters C^H, C^S, C^D, and C^N are algebraically related to the free parameters proposed by Large et al. (1994).

Large et al. (1994)'s formulation of KPP for the case of penetrative convection with no horizontal shear introduces six nondimensional parameters: the Von Karman constant $\kappa = 0.4$, the ratio of the entrainment flux to the surface flux $\beta_T = 0.2$, a constant that sets the amplitude of the non-local flux $C^* = 10$, a constant that ensures the continuity of the buoyancy flux profile $c_s = 98.96$, the surface layer fraction $\epsilon = 0.1$, and a

parameter that controls the smoothing of the buoyancy profile at the base of the boundary layer depth C_v. Large et al. (1994) argue that C_v can take any value between 1 and 2. We set the reference value $C_v = 1.7$, which corresponds to the strong stratification limit in the model proposed by Danabasoglu et al. (2006) and given by equation (8.184) in Griffies et al. (2015).

In our formulation we introduce four parameters which are related to the original Large et al. (1994) parameters as follows,

$$C^H = \frac{C_v(\beta_T)^{1/2}}{(c_s\kappa^4\epsilon)^{1/6}}, \quad C^S = \epsilon, \quad C^D = (c_s\kappa^4)^{1/3}, \quad \text{and} \quad C^N = C^*(c_s\kappa^4\epsilon)^{1/3}. \tag{C1}$$

We are able to reduce the number of parameters from six $(\epsilon, c_s, C_V, \beta_T, \kappa, C^*)$ to four (C^H, C^S, C^D, C^N), because in the case of penetrative convection the two combinations $C_v(\beta_T)^{1/2}$ and $c_s\kappa^4$ always appear together.

Using the reference KPP parameter values reported above, our parameters take the values:

$$C^H = 0.956, \quad C^S = 0.1, \quad C^D = 1.36, \quad C^N = 6.3275. \tag{C2}$$

We refer to these as the reference parameters.

It is worth commenting why the critical Richardson number, the focus of much literature on KPP, does not appear when considering penetrative convection. The boundary layer depth is determined implicitly through equations (21) and (23) in Large et al. (1994),

$$\mathrm{Ri}_b(z) = \frac{(B_r - B(z))(-z)}{|V_r - V(z)|^2 + V_t^2(z)} \quad \text{and} \quad V_t^2(z) = \frac{C_v(\beta_T)^{1/2}}{Ri_c\kappa^2}(c_s\epsilon)^{-1/2}(-z)Nw_s, \tag{C3}$$

where B is buoyancy and B_r is the average of B between the surface and the depth ϵz. The boundary layer depth is defined as the depth $z = -h$ where $\mathrm{Ri}_b(-h) = \mathrm{Ri}_c$. For convection without shear, the case considered in this paper, $|V_r - V(z)|^2 = 0$ and $w_s = w^*(c_s\epsilon)^{1/3}\kappa^{4/3}$. The two equations can therefore be combined together:

$$\frac{C_v(\beta_T)^{1/2}}{\kappa^{2/3}}(c_s\epsilon)^{-1/6} = \frac{(B_r - B(-h))h}{hNw^*}. \tag{C4}$$

and the critical Richardson number drops out from the expression. This expression further supports our decision to introduce the single parameter C^H in favor of the combination of original parameters appearing on the left hand side of (C4). In penetrative convection it is the parameter C^H that controls the boundary layer depth rather than the critical Richardson number.

The optimal parameters and probability distributions for (C^H, C^S, C^D, C^N) can be mapped onto $(\epsilon, C_v(\beta_T)^{1/2}, c_s\kappa^4, C^*)$ using the inverse transformation,

$$\epsilon = C^S, \; c_s\kappa^4 = (C^D)^3, \; C^* = \frac{C^N}{C^D(C^S)^{1/3}}, \text{ and } C_v(\beta_T)^{1/2} = C^H(C^D)^{1/2}(C^S)^{1/6}. \tag{C5}$$

Appendix D A Primer on Uncertainty Quantification

The probability distribution of the parameters in a parameterization must quantify the likelihood that the parameters take on values other than those that minimize the loss function $\mathcal{L}$. To achieve this the probability distribution must satisfy two key properties:

1. In the limit of no uncertainty, the probability distribution should collapse to a delta function centered at the optimal parameter values that minimize the loss function.

553 2. The uncertainty of a parameter value $\boldsymbol{C}$ should increase proportionally to the value
554 of $\mathcal{L}(\boldsymbol{C})$.

There are many probability distributions that satisfy the above properties. We choose the following:

$$\rho(\boldsymbol{C}) \propto \rho^0(\boldsymbol{C}) \exp\left(-\mathcal{L}(\boldsymbol{C})/\mathcal{L}_0\right), \tag{D1}$$

555 where ρ^0 is a uniform prior distribution, $\mathcal{L}$ is a loss function, and $\mathcal{L}_0$ is a hyperparam-
556 eter.

557 The hyperparameter $\mathcal{L}_0$ sets the shape of the likelihood function $\exp\left(-\mathcal{L}(\boldsymbol{C})/\mathcal{L}_0\right)$
558 and its associated uncertainty quantification. The limit $\mathcal{L}_0 \to 0$ corresponds to no un-
559 certainty, because the likelihood function and the probability distribution collapse to a
560 delta function peaked at the optimal parameter values that minimize the loss function.
561 The limit $\mathcal{L}_0 \to \infty$ instead corresponds to a likelihood function that adds no informa-
562 tion to reduce the uncertainty and the posterior distribution $\rho(\boldsymbol{C})$ is equal to the prior
563 one $\rho_0(\boldsymbol{C})$. Thus $\mathcal{L}_0$ must take finite values between zero and infinity, if the likelihood
564 function is to add useful information.

For any finite value of $\mathcal{L}_0$, the probability distribution has its mode (maximum) at the optimal parameters, if the prior distribution is uniform. This can be easily demonstrated. Let $\mathbf{C}^*$ denote the parameter values for which the loss function has its global minimum and $\mathbf{C}$ denote any other set of parameter values. It is then the case that $\rho(\mathbf{C})$ is smaller than $\rho(\mathbf{C}^*)$ for any $\mathbf{C}$,

$$\mathcal{L}(\mathbf{C}^*) \leq \mathcal{L}(\mathbf{C}) \;\Rightarrow\; \exp\left(-\mathcal{L}(\mathbf{C})/\mathcal{L}_0\right) \leq \exp\left(-\mathcal{L}(\mathbf{C}^*)/\mathcal{L}_0\right) \Rightarrow\; \rho(\mathbf{C}) \leq \rho(\mathbf{C}^*). \tag{D2}$$

565 Hence the most probable value of the probability distribution is achieved at the mini-
566 mum of the loss function independent of $\mathcal{L}_0$ for a uniform prior distribution.

567 As mentioned in section 3, it is convenient to set the hyperparameter $\mathcal{L}_0$ to be equal
568 to the minimum of the loss function $\mathcal{L}(\mathbf{C}^*)$. This choice satisfies two key requirements.
569 First, the uncertainties of parameters should be independent of the units of the loss func-
570 tion. Second, the hyperparameter $\mathcal{L}_0$ should be larger the larger the loss function $\mathcal{L}(\mathbf{C}^*)$,
571 because the latter is a measure of the parameterization bias and the former should be
572 larger if there is more uncertainty about acceptable parameter values.

573 In practice it is seldom possible to find the global minimum of $\mathcal{L}$ and instead we
574 adopt a "best guess" of the optimal parameters $\tilde{\mathbf{C}}$ and set $\tilde{\mathcal{L}}_0 = \mathcal{L}(\tilde{\mathbf{C}})$. Since $\mathcal{L}(\mathbf{C}^*) \leq$
575 $\mathcal{L}(\tilde{\mathbf{C}})$, our choice is conservative because a larger $\mathcal{L}_0$ corresponds to *more* uncertainty.

576 Appendix E Random Walk Markov Chain Monte Carlo

577 We use the Random Walk Markov Chain Monte Carlo Method (RW-MCMC) in-
578 troduced by Metropolis et al. (1953) to sample values from the probability distribution.
579 While other more efficient algorithms exist, our parameter space is only four dimensional
580 and computational cost is not an issue. The RW-MCMC samples the probability func-
581 tion by taking a random walk through parameter space. The algorithm generates a se-
582 quence of sample parameter values $\boldsymbol{C}_i$ in such a way that, as more and more sample val-
583 ues are produced, the distribution of values more closely approximates the joint param-
584 eter probability distribution of the parameters. At each iteration, the algorithm picks
585 a candidate parameter set for the next sample value based on the current sample value.
586 Then, with some probability, the candidate parameter set is either accepted (in which
587 case the candidate value is used in the next iteration) or rejected (in which case the can-
588 didate value is discarded, and current values reused in the next iteration). The criterion
589 for acceptance and its relation to the probability distribution is best described by sketch-
590 ing the algorithm:

1. Choose a set of initial parameter values $\boldsymbol{C}_0$. We pick our best guess at the set of values that minimize the log-likelihood function as estimated from standard minimization techniques.
2. Choose a new set of candidate parameters by adding a Gaussian random variable with mean zero and covariance matrix Σ to the inital set, $\tilde{\boldsymbol{C}}_1 = \boldsymbol{C}_0 + \mathcal{N}(0, \Sigma)$. The algorithm is guaranteed to work *independently* of the choice of Σ as long as it is nonzero and does not vary throughout the random walk. However suitable choices can speed up convergence and will be discussed below.
3. Calculate $\Delta\ell = \ell(\boldsymbol{C}_0) - \ell(\tilde{\boldsymbol{C}}_1)$. This is a measure of how much more likely $\tilde{\boldsymbol{C}}_1$ is relative to $\boldsymbol{C}_0$.
4. Draw a random variable from the interval $[0, 1]$, e.g, calculate $u = \mathcal{U}(0, 1)$. If $\log(u) < \Delta\ell$ accept the new parameter values and set $\boldsymbol{C}_1 = \tilde{\boldsymbol{C}}_1$. Otherwise reject the new parameter values $\boldsymbol{C}_1 = \boldsymbol{C}_0$. This is the "accept / reject" step. Note that if $\Delta\ell > 0$, i.e. if the proposed parameter produces a smaller output in the negative log-likelihood function, the proposal is always accepted.
5. Repeat steps 2-4 , replacing $\boldsymbol{C}_0 \to \boldsymbol{C}_i$ and $\boldsymbol{C}_1 \to \boldsymbol{C}_{i+1}$, to generate a sequence for $\boldsymbol{C}_i$ of parameter values.

The sequence of parameter values generated by the algorithm can then be used to construct any statistics of the probability distribution 18, including empirical distributions, marginal distributions, and joint distributions. In the context of KPP it can generate the uncertainty of the temperature value at any depth and time as well as the uncertainty of the boundary layer depth at a given time.

To guide the choice of an appropriate value for Σ, one diagnoses the "number of independent samples" by using approximations of the correlation length as described by Sokal (1997). If Σ is too small then the acceptance rate is too large since each candidate parameter is barely any different from the original one. Too large a Σ yields too low acceptance rates. To find an appropriate compromise we perform a preliminary random walk and estimate the covariance matrix of the resulting distribution. We then set Σ equal to this covariance matrix.

Last, in order to sample parameters within a finite domain, we artificially make the parameter space periodic and the random walk is therefore guaranteed to never leave the desired domain.

Acknowledgments

The authors would like to thank Carl Wunsch, Tapio Schneider, Andrew Stuart, and William Large for numerous illuminating discussions. We would also like to thank the reviewers for their helpful suggestions with the manuscript. Our work is supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program, and by the National Science Foundation under grant AGS-6939393. Code and data may be found via figshare at (Souza, 2020).

References

Abkar, M., Bae, H. J., & Moin, P. (2016). Minimum-dissipation scalar transport model for large-eddy simulation of turbulent flows. *Physical Review Fluids*, *1*(4), 041701. doi: 10.1103/PhysRevFluids.1.041701

Arakawa, A., & Lamb, V. R. (1977). Computational design of the basic dynamical processes of the ucla general circulation model. In J. Chang (Ed.), *General circulation models of the atmosphere* (Vol. 17, p. 173 - 265). Elsevier. doi: 10.1016/B978-0-12-460817-7.50009-4

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, *53*, 370–418.

Besard, T., Churavy, V., Edelman, A., & Sutter, B. D. (2019). Rapid software prototyping for heterogeneous and distributed platforms. *Advances in Engineering Software*, *132*, 29 - 46. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0965997818310123` doi: https://doi.org/10.1016/j.advengsoft.2019.02.002

Besard, T., Foket, C., & De Sutter, B. (2019, April). Effective extensible programming: Unleashing julia on gpus. *IEEE Transactions on Parallel and Distributed Systems*, *30*(4), 827-841. doi: 10.1109/TPDS.2018.2872064

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. doi: 10/f9wkpj

Cenedese, C., Marshall, J., & Whitehead, J. A. (2004, 03). A Laboratory Model of Thermocline Depth and Exchange Fluxes across Circumpolar Fronts*. *Journal of Physical Oceanography*, *34*(3), 656-667. Retrieved from `https://doi.org/10.1175/2508.1` doi: 10.1175/2508.1

Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., & Stuart, A. M. (2020). Calibrate, emulate, sample. *arXiv:2001.03689 [stat.CO]*. Retrieved from `https://arxiv.org/abs/2001.03689`

Danabasoglu, G., Large, W. G., Tribbia, J. J., Gent, P. R., Briegleb, B. P., & McWilliams, J. C. (2006). Diurnal coupling in the tropical oceans of ccsm3. *Journal of climate*, *19*(11), 2347–2365.

Deardorff, J. W., Willis, G. E., & Stockton, B. H. (1980). Laboratory studies of the entrainment zone of a convectively mixed layer. *Journal of Fluid Mechanics*, *100*(1), 41–64. doi: 10.1017/S0022112080001000

DuVivier, A. K., Large, W. G., & Small, R. J. (2018). Argo observations of the deep mixing band in the southern ocean: A salinity modeling challenge. *Journal of Geophysical Research: Oceans*, *123*(10), 7599-7617. doi: 10.1029/2018JC014275

Frisch, U. (1995). *Turbulence: The legacy of a. n. kolmogorov*. Cambridge University Press. doi: 10.1017/CBO9781139170666

Griffies, S., Levy, M., Adcroft, A., Danabasoglu, G., Hallberg, R., Jacobsen, D., ... Ringler, T. (2015). *Theory and numerics of the community ocean vertical mixing (cvmix) project* (Draft from March 9, 2015).

Harcourt, R. R. (2015). An improved second-moment closure model of langmuir turbulence. *Journal of Physical Oceanography*, *45*(1), 84-103. Retrieved from `https://doi.org/10.1175/JPO-D-14-0046.1` doi: 10.1175/JPO-D-14-0046.1

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589-602. doi: 10.1175/BAMS-D-15-00135.1

Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). Cambridge University Press. doi: 10.1017/CBO9780511790423

Kantha, L. H., & Clayson, C. A. (1994). An improved mixed layer model for geophysical applications. *Journal of Geophysical Research: Oceans*, *99*(C12), 25235-25266. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC02257` doi: 10.1029/94JC02257

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680. doi: 10.1126/science.220.4598.671

Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, *32*(4), 363-403. doi: 10.1029/94RG01872

Large, W. G., Patton, E. G., DuVivier, A. K., Sullivan, P. P., & Romero, L. (2019). Similarity theory in the surface layer of large-eddy simulations of the wind-, wave-, and buoyancy-forced southern ocean. *Journal of Physical Oceanography*,

49(8), 2165-2187. doi: 10.1175/JPO-D-18-0066.1

Li, Q., & Fox-Kemper, B. (2017). Assessing the effects of langmuir turbulence on the entrainment buoyancy flux in the ocean surface boundary layer. *Journal of Physical Oceanography*, *47*(12), 2863-2886. Retrieved from https://doi.org/10.1175/JPO-D-17-0085.1 doi: 10.1175/JPO-D-17-0085.1

Marshall, J., Adcroft, A., Hill, C., Perelman, L., & Heisey, C. (1997). A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research: Oceans*, *102*(C3), 5753-5766. doi: 10.1029/96JC02775

Marshall, J., & Schott, F. (1999). Open-ocean convection: Observations, theory, and models. *Reviews of Geophysics*, *37*(1), 1-64. doi: 10.1029/98RG02739

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., ... Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*(3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012MS000154 doi: 10.1029/2012MS000154

Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, *20*(4), 851-875. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG020i004p00851 doi: 10.1029/RG020i004p00851

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087-1092. doi: 10.1063/1.1699114

Morrison, H., van Lier-Walqui, M., Kumjian, M. R., & Prat, O. P. (2020). A bayesian approach for statistical–physical bulk parameterization of rain microphysics. part i: Scheme description. *Journal of the Atmospheric Sciences*, *77*(3), 1019-1041. Retrieved from https://doi.org/10.1175/JAS-D-19-0070.1 doi: 10.1175/JAS-D-19-0070.1

Nadiga, B., Jiang, C., & Livescu, D. (2019). Leveraging bayesian analysis to improve accuracy of approximate models. *Journal of Computational Physics*, *394*, 280 - 297. Retrieved from http://www.sciencedirect.com/science/article/pii/S0021999119303407 doi: https://doi.org/10.1016/j.jcp.2019.05.015

Pacanowski, R. C., & Philander, S. G. H. (1981, 11). Parameterization of Vertical Mixing in Numerical Models of Tropical Oceans. *Journal of Physical Oceanography*, *11*(11), 1443-1451. Retrieved from https://doi.org/10.1175/1520-0485(1981)011<1443:POVMIN>2.0.CO;2 doi: 10.1175/1520-0485(1981)011⟨1443:POVMIN⟩2.0.CO;2

Price, J. F., Weller, R. A., & Pinkel, R. (1986, July). Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing. , *91*(C7), 8411-8427. doi: 10.1029/JC091iC07p08411

Ramadhan, A., Wagner, G. L., Hill, C., Campin, J.-M., Churavy, V., Besard, T., ... Marshall, J. (2020). Oceananigans.jl: Fast and friendly geophysical fluid dynamics on GPUs. *Journal of Open Source Software*, *5*(53), 2018. Retrieved from https://doi.org/10.21105/joss.02018 doi: 10.21105/joss.02018

Reichl, B. G., Wang, D., Hara, T., Ginis, I., & Kukulka, T. (2016). Langmuir turbulence parameterization in tropical cyclone conditions. *Journal of Physical Oceanography*, *46*(3), 863-886. Retrieved from https://doi.org/10.1175/JPO-D-15-0106.1 doi: 10.1175/JPO-D-15-0106.1

Rozema, W., Bae, H. J., Moin, P., & Verstappen, R. (2015). Minimum-dissipation models for large-eddy simulation. *Physics of Fluids*, *27*(8), 085107. doi: 10.1063/1.4928700

Rye, C. D., Marshall, J., Kelley, M., Russell, G., Nazarenko, L. S., Kostov, Y., ... Hansen, J. (2020). Antarctic glacial melt as a driver of recent southern ocean climate trends. *Geophys. Res. Lett.*, *47*(11), e2019GL086892. doi: 10.1029/2019GL086892

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., ... Saha, S. (2017). Practice and philosophy of climate model tuning across six us modeling centers. *Geoscientific Model Development*, *10*(9), 3207–3223. Retrieved from https://gmd.copernicus.org/articles/10/3207/2017/ doi: 10.5194/gmd-10-3207-2017

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*(24), 12,396-12,417. doi: 10.1002/2017GL076101

Schneider, T., Teixeira, J., Bretherton, C., Brient, F., Pressel, K., Schär, C., & Siebesma, A. (2017, 01). Climate goals and computing the future of clouds. *Nature Climate Change*, *7*, 3-5. doi: 10.1038/nclimate3190

Schumann, U., & Sweet, R. A. (1988). Fast fourier transforms for direct solution of poisson's equation with staggered boundary conditions. *Journal of Computational Physics*, *75*(1), 123 - 137. doi: https://doi.org/10.1016/0021-9991(88)90102-7

Sokal, A. (1997). Monte carlo methods in statistical mechanics: Foundations and new algorithms. In C. DeWitt-Morette, P. Cartier, & A. Folacci (Eds.), *Functional integration: Basics and applications* (pp. 131–192). Boston, MA: Springer US. doi: 10.1007/978-1-4899-0319-8_6

Souza, A. (2020, Jul). *Uncertainty quantification of ocean parameterizations: Code and data.zip.* figshare. Retrieved from https://figshare.com/articles/software/Uncertainty_quantification_of_ocean_parameterizations_Code_and_Data_zip/12743834/1 doi: 10.6084/m9.figshare.12743834.v1

Sraj, I., Iskandarani, M., Thacker, W. C., Srinivasan, A., & Knio, O. M. (2014). Drag parameter estimation using gradients and hessian from a polynomial chaos model surrogate. *Monthly Weather Review*, *142*(2), 933-941. Retrieved from https://doi.org/10.1175/MWR-D-13-00087.1 doi: 10.1175/MWR-D-13-00087.1

Sraj, I., Zedler, S. E., Knio, O. M., Jackson, C. S., & Hoteit, I. (2016). Polynomial chaos–based bayesian inference of k-profile parameterization in a general circulation model of the tropical pacific. *Monthly Weather Review*, *144*(12), 4621-4640. Retrieved from https://doi.org/10.1175/MWR-D-15-0394.1 doi: 10.1175/MWR-D-15-0394.1

Sullivan, P. P., & Patton, E. G. (2011, 10). The Effect of Mesh Resolution on Convective Boundary Layer Statistics and Structures Generated by Large-Eddy Simulation. *Journal of the Atmospheric Sciences*, *68*(10), 2395-2415. Retrieved from https://doi.org/10.1175/JAS-D-10-05010.1 doi: 10.1175/JAS-D-10-05010.1

Thacker, W. C. (1989). The role of the hessian matrix in fitting models to measurements. *Journal of Geophysical Research: Oceans*, *94*(C5), 6177-6196. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC094iC05p06177 doi: 10.1029/JC094iC05p06177

Urrego-Blanco, J. R., Urban, N. M., Hunke, E. C., Turner, A. K., & Jeffery, N. (2016). Uncertainty quantification and global sensitivity analysis of the los alamos sea ice model. *Journal of Geophysical Research: Oceans*, *121*(4), 2709-2732. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JC011558 doi: 10.1002/2015JC011558

van Lier-Walqui, M., Vukicevic, T., & Posselt, D. J. (2012). Quantification of cloud microphysical parameterization uncertainty using radar reflectivity. *Monthly Weather Review*, *140*(11), 3442-3466. Retrieved from https://doi.org/10.1175/MWR-D-11-00216.1 doi: 10.1175/MWR-D-11-00216.1

Van Roekel, L., Adcroft, A. J., Danabasoglu, G., Griffies, S. M., Kauffman, B., Large, W., ... Schmidt, M. (2018). The kpp boundary layer scheme for the ocean: Revisiting its formulation and benchmarking one-dimensional simula-

tions relative to les. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2647-2685. doi: 10.1029/2018MS001336

Verstappen, R. (2018). How much eddy dissipation is needed to counterbalance the nonlinear production of small, unresolved scales in a large-eddy simulation of turbulence? *Computers & Fluids*, *176*, 276–284. doi: 10.1016/j.compfluid.2016.12.016

Vreugdenhil, C. A., & Taylor, J. R. (2018). Large-eddy simulations of stratified plane Couette flow using the anisotropic minimum-dissipation model. *Physics of Fluids*, *30*(8), 085104. doi: 10.1063/1.5037039

Wang, D., Large, W. G., & McWilliams, J. C. (1996). Large-eddy simulation of the equatorial ocean boundary layer: Diurnal cycling, eddy viscosity, and horizontal rotation. *Journal of Geophysical Research: Oceans*, *101*(C2), 3649-3662. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95JC03441` doi: 10.1029/95JC03441

Wunsch, C. (2006). *Discrete inverse and state estimation problems: With geophysical fluid applications.* Cambridge University Press. doi: 10.1017/CBO9780511535949

Zedler, S. E., Kanschat, G., Korty, R., & Hoteit, I. (2012). A new approach for the determination of the drag coefficient from the upper ocean response to a tropical cyclone: a feasibility study. *Journal of Oceanography*, *68*(2), 227–241. Retrieved from `https://doi.org/10.1007/s10872-011-0092-6` doi: 10.1007/s10872-011-0092-6