A Finite-Time Ensemble Method for Mixed Layer Model Comparison

Leah Johnson[a], Baylor Fox-Kemper[a] Qing Li,[b] Hieu T. Pham,[c] Sutanu Sarkar,[c]

[a] *Earth, Environmental, and Planetary Sciences, Brown University, Providence, RI*

[b] *Earth, Ocean and Atmospheric Sciences Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China*

[c] *University of California, San Diego, San Diego, CA*

Corresponding author: Leah Johnson, leahjohn@uw.edu

Current Affiliation:Applied Physics Laboratory, University of Washington, 1013 NE 41st St, Seattle, WA 98105

ABSTRACT: This work evaluates the fidelity of various upper ocean turbulence parameterizations subject to realistic monsoon forcing and presents a finite-time ensemble vector (EV) method to better manage the design and numerical principles of these parameterizations. The EV method emphasizes the dynamics of a turbulence closure multi-model ensemble and is applied to evaluate ten different ocean surface boundary layer (OSBL) parameterizations within a single column (SC) model against two boundary layer large eddy simulations (LES). Both LES include realistic surface forcing, but one includes wind-driven shear turbulence only, while the other includes additional Stokes forcing through the wave-average equations that generates Langmuir turbulence. The finite-time EV framework focuses on what constitutes the local behavior of the mixed layer dynamical system and isolates the forcing and ocean state conditions where turbulence parameterizations most disagree. Identifying disagreement provides the potential to evaluate SC models comparatively against the LES. Observations collected during the 2018 Monsoon onset in the Bay of Bengal provide a case study to evaluate models under realistic and variable forcing conditions. The case study results highlight two regimes where models disagree a) during wind-driven deepening of the mixed layer and b) under strong diurnal forcing.

1. Introduction

The ocean surface boundary layer (OSBL) dictates the short-term heat capacity of the upper ocean and modulates the communication between the atmospheric and oceanic systems (Umlauf and Burchard 2005; Belcher et al. 2012; Li et al. 2019; Fox-Kemper et al. 2021a; Hall and Fox-Kemper 2021). Fluid motions within the OSBL are dominated by small-scale turbulence (O [1 cm to 100 m]) and so are rarely resolved and therefore parameterized in regional and global numerical models. Under realistic surface forcing, only large eddy simulations (LES) and direct numerical simulations (DNS) seek to directly simulate the important scales of boundary layer turbulence, and presently only LES can handle domains large enough to include a realistic OSBL resembling typical oceanographic conditions.

There are many approaches to approximating turbulence physics in oceanic boundary layers. LES and single column parameterization models (SC models) traditionally consider turbulence generated by wind stress and buoyancy forcing (recognized here as shear turbulence models, ST). Newer LES and SC models may also include the enhanced turbulence contribution from surface wave forcing, usually called Langmuir turbulence (LT), under the assumption that surface wave forcing can be approximated through the waves' Stokes drift in the wave-averaged equations (Leibovich 1980; Craik 1982; Holm 1996; McWilliams et al. 1997; Suzuki and Fox-Kemper 2016; D'Asaro et al. 2014; Li et al. 2019). Extensions of these equations to include stochastic waves (Holm and Hu 2021), wave breaking (Sullivan et al. 2007), and phase-dependent turbulence-wave interactions (Teixeira and Belcher 2002; Qiao et al. 2016) illuminate what is missing from the traditional wave-averaged approach. It is common to isolate the upper ocean response to atmospheric forcing in an SC modeling framework (i.e. one-dimensional (1D) models, Li et al. 2019). Validating these approaches across the wide range of ocean states and atmosphere forcing conditions or understanding the impact of an SC model on the ocean-atmosphere system is difficult due to the complexities of both the turbulence and the evolution of the OSBL. Attempts to validate modeled OSBL evolution against observations are inhibited by the difficulties in measuring turbulent motions, or confounded by other processes prevalent in the OSBL but missing inherently in the 1D framework, such as horizontal advection, fronts, and other submesoscale structures (e.g., Jaeger et al. 2020; Johnson et al. 2016). In the absence of this observational truth, OSBL SC models are compared with high-resolution LES or DNS simulations that partially resolve or resolve turbulent motions. Such simulations are computationally expensive and, except for a few examples (e.g., Rabe et al. 2015; Large et al. 2019; Pham et al. 2023; Fan et al. 2020; Whitt et al. 2022), are typically run under idealized constant forcing conditions that occupy a narrow region of the vast range of possible ocean states (estimates of regimes covered by steady-state LES are given in Li et al. 2019). Despite these many approaches, there is still a limited understanding of how well OSBL SC models work universally, under realistic conditions, or how the choice of an OSBL parameterization influences the simulated weather and climate system.

The variety of theoretical underpinnings that each turbulence parameterization is built on further complicates SC model comparison. For example, consider the common relation of turbulent motions of a variable, ϕ, to an eddy diffusivity, κ_ϕ, dependent on a velocity scale and a length scale of the turbulent motion, $\kappa_\phi = cql$, where c is a non-dimensional coefficient, q is the turbulent velocity scale and l is a typical turbulence length scale (Tennekes and Lumley 2018). While this fundamental turbulence concept is utilized by second-moment closure schemes (e.g., Rodi 1987) as well as by k-theory schemes (e.g., Large et al. 1994), each formulation's definition of length scale and turbulent velocity scale are unique to each parameterization. A unifying framework (the generic length-scale: Umlauf and Burchard 2003, 2005) was developed to connect different second-moment closure schemes. Yet, when including a broader class of SC models, key turbulent control parameters in the OSBL, such as Richardson number and turbulent velocity and length scales, are applied in widely different contexts in each specific scheme of turbulence closure. It is possible to treat each SC model as a black box and evaluate how separate SC models run under identical forcing diverge and result in different ocean states. With this method, it can be challenging to interpret diverging ocean states after a long period of time as the turbulent fluxes (and parameterizations) that define the OSBL are nonlinear, path-dependent, and exhibit hysteresis. Here, an approach is adopted to a) understand the local behavior of a non-linear dynamical system (i.e. numerical model) and b) localize approximately in time so as to quantify and evaluate the divergence across an ensemble of numerical models.

Specifically, this study presents a framework to compare models of the OSBL by evaluating the local (i.e. finite-time) behavior of the modeled OSBL subject to different turbulence physics. The goal of this work is not to identify the "best" model, but to isolate where in the state and forcing space models disagree in order to evaluate the robustness, or alternatively, the uncertainty, in the parameterized physics. Section 2 presents the mathematical foundation for understanding the modeled OSBL as a nonlinear system of equations. Leveraging dynamical systems theory, the ensemble system is first presented as a linearized one using a Taylor series expansion to highlight the distinct sources of sensitivity in the modeled OSBL system. Focusing on the sensitivity due to parameterization physics alone, a method is proposed to evaluate inter-model uncertainty.

This method is applied to a specific suite of ten OSBL SC models within the General Ocean Turbulence Model (GOTM, Burchard et al. 1999; Umlauf and Burchard 2005) compared against LES (Pham et al. 2023), and implemented in a case-study using in-situ observations of the 2018 monsoon onset collected during the ONR Oceanic Control of Monsoon Intra-Seasonal Oscillations in the Tropical Indian Ocean and the Bay of Bengal (MISO-BOB) campaign (section 3). Results are presented in section 4 and discussed in section 5. It will be shown that the finite-time ensemble method successfully isolates two regimes in the case study where models disagree a) during wind-driven deepening of the mixed layer and b) under strong diurnal forcing.

2. The Ocean Surface Boundary Layer System

Assuming horizontal homogeneity of mean fields, no mean vertical velocity, and neglecting molecular viscosity, the Boussinesq, hydrostatic, and Reynolds averaged equations for mean variables in the OSBL are:

$$\frac{\partial u}{\partial t} = fv - \frac{\partial \overline{w'u'}}{\partial z} \tag{1}$$

$$\frac{\partial v}{\partial t} = -fu - \frac{\partial \overline{w'v'}}{\partial z} \tag{2}$$

$$\frac{\partial T}{\partial t} = -\frac{\partial \overline{w'T'}}{\partial z} + \frac{\partial R}{\partial z} \tag{3}$$

$$\frac{\partial S}{\partial t} = -\frac{\partial \overline{w'S'}}{\partial z} \tag{4}$$

$$\rho = \rho(S, T, p) \tag{5}$$

with boundary conditions at the ocean-atmosphere surface (noting that here the frictional or numerical scheme sublayers that are not to be resolved, and thus the turbulent fluxes outside of the sublayers are matched by conservation to the surface fluxes):

$$\overline{w'u'} = -\tau_u(t) \text{ at } z = 0 \tag{6}$$

$$\overline{w'v'} = -\tau_v(t) \text{ at } z = 0 \tag{7}$$

$$\overline{w'T'} = F_T(t) \text{ at } z = 0 \tag{8}$$

$$\overline{w'S'} = F_S(t) \text{ at } z = 0 \tag{9}$$

The variables are given as: T is temperature [°C], S is salinity [g kg^{-1}], u is zonal velocity [m s^{-1}], v is meridional velocity [m s^{-1}], and w is vertical velocity [m s^{-1}], p is pressure [Pa or kg m^{-1} s^{-2}], R is penetrative radiative heat flux [°C m s^{-1}], ρ is density [kg m^{-3}], τ is wind input [m^2 s^{-2}], and F_T [°C m s^{-1}] and F_S [PSU m s^{-1}] are the surface heat and (virtual) salt fluxes respectively. See Fox-Kemper et al. (2021a) for a wider discussion of these equations. Primes denote turbulent properties, and overbars are the horizontal average (dropped from mean variables for clarity). All averaged variables are horizontally homogeneous but depend on vertical position z and time t.

A set of equations also predicting the flux divergence terms in Eq. (1)-(4) requires knowledge of an infinite number of higher-order moments leading to the well known turbulence closure problem. There are many avenues to turbulence closures that attempt to capture the unresolved turbulent motions in the boundary layer. Parameterizations used in this manuscript include first-order models and second-moment schemes. These models tend to utilize k-theory, where the turbulent flux of a variable ϕ is approximated by

$$\overline{w'\phi'} = -\kappa_\phi \frac{\partial \phi}{\partial z}. \tag{10}$$

First-order models have a diagnostic equation for turbulent diffusivities κ_ϕ and may include the addition of nonlocal fluxes (e.g., KPP and its implementation in CVMix: Large et al. 1994; Van Roekel et al. 2018). In second-moment schemes, prognostic equations, such as for a velocity scale and a length scale, can be used to estimate the stresses and fluxes, $\overline{w'\phi'}$ (e.g., Umlauf and Burchard 2003; Harcourt 2013). Of interest here is understanding how the choice in the closure approach impacts the trajectory of the OSBL system.

a. Understanding the OSBL as a dynamical system

A state vector $\mathbf{x}$ is taken to be all variables needed to solve the turbulence closure and Eq. (1)-(4), evaluated at all z grid points. This set is discretized in space and with a chosen time-stepping method to form a *nonlinear* diagnostic process:

$$x_j^f = \mathcal{A}_j(x_j^i; F_\mu^{i:f}; \beta) \tag{11}$$

Where $\mathcal{A}$, the system map from an initial (superscript i) to final (superscript f) time, is a nonlinear operator that depends on the initial value of all the state variables at all z locations (subscript j denotes both different variables and different locations). Due to the turbulence closure problem, a turbulence parameterization is embedded in the system $\mathcal{A}$. The nonlinear operator $\mathcal{A}$ also depends on the forcing, F, between the initial and final times through different surface conditions and radiation (subscript μ; i.e., $R, F_T, F_S, \tau_u, \tau_v$), and on time-independent model parameters β. So, given x_j^i, $F_\mu^{i:f}$ and β, the map $\mathcal{A}$ will determine the final state, x_j^f.

In many cases, the nonlinear equations are quite complex and subject to numerical concerns. As such, it can be convenient to understand the local behavior, rather than the full non-linear nature, of $\mathcal{A}$. In dynamical systems, this is done formally through a Taylor series expansion, thereby linearizing Eq. (11) around state $\mathbf{x_a}$, forcing $\mathbf{F_a}$ and parameters β_a. Bold text indicates matrices and vectors in the (approximate) linearized system, distinguishing it from the exact solution in (11). The Jacobian, gain, and parameter sensitivity matrices result from partial derivatives of $\mathcal{A}$ with respect to its arguments evaluated at the state $\mathbf{x_a}$, forcing $\mathbf{F_a}$, and parameters β_a. $\mathbf{A}|_a$ is simply the nonlinear function $\mathcal{A}$ evaluated with this standard state, forcing, and parameters. Dots indicate matrix multiplication:

$$\mathbf{x_f} = \mathbf{A}|_\mathbf{a} + \mathbf{J}|_\mathbf{a} \cdot (\mathbf{x_i} - \mathbf{x_a}) + \mathbf{G}|_\mathbf{a} \cdot (\mathbf{F_i} - \mathbf{F_a}) + \left.\frac{\partial \mathbf{A}}{\partial \beta}\right|_\mathbf{a} \cdot (\beta - \beta_a). \tag{12}$$

For the local linearization to be accurate the initial state vector $\mathbf{x_i}$ and final state vector $\mathbf{x_f}$ both must be nearby the standard state vector $\mathbf{x_a}$, and similarly the forcing and parameters must not be altered much.

For a state, $\mathbf{x_a}$, on the system map, the terms in the Taylor series expansion highlight the various aspects of a single nonlinear SC model that can impact the trajectory from $\mathbf{x_i}$ to around $\mathbf{x_a}$. This provides a useful framework for identifying sensitivities in the simulated OSBL system that are otherwise obscured by evaluating continuous simulations. Potential choices of $\mathbf{x_a}$ might arise (e.g. multi-model mean state, LES state, etc.) and the interpretation of Eq. 12 depends on this choice (see Johnson and Fox-Kemper (2023) for a more generalized discussion of $\mathbf{x_a}$).

The Jacobian, $\mathbf{J}|_\mathbf{a}$, is the evaluation at the standard state, forcing and parameters of the partial derivative of the nonlinear function $\mathcal{A}$:

$$J_{mn}(x^i;F^{i:f};\beta) = \frac{\partial \mathcal{A}_m(x^i;F^{i:f};\beta)}{\partial x_n^i} \tag{13}$$

The partial derivative captures the sensitivity of a model trajectory outcome at the final time to the initial state, but, unlike its form in the local linearization $\mathbf{J}|_\mathbf{a}$, the derivative in (13) still depends on the state, forcing, and parameters. For example, the amount of deepening of the ML by the end of an interval will be sensitive to the stratification of the ML base at the beginning of the interval.

For the surface forced OSBL, the dependence of $\mathcal{A}$ for each state variable due to infinitesimal changes in each forcing agent *over every increment of time from the initial to the final condition* can be captured by the infinite-dimensional "gain function". The gain matrix $\mathbf{G}|_\mathbf{a}$, has a nonlinear gain function form which depends on the state, forcing, and parameters:

$$G_{m\gamma}^{i:f}(x^i;F^{i:f};\beta) = \frac{\partial \mathcal{A}_m(x^i;F^{i:f};\beta)}{\partial F_\gamma^{i:f}} \tag{14}$$

It's interesting to note that the arguments to $G_{m\gamma}(x;F;\beta)$ indicate that the influence of forcing on the system is not limited to dependence on the boundary conditions necessarily, but also through parameter- and state-dependent responses to the surface fluxes. For example, SC models based on similarity theory (Monin and Obukhov 1954, hereafter MO) such as KPP are limited in the kinds of parameter- and state-dependence allowed through a small set of dimensionless relationships that may depend on surface forcing. Similarly, if the Taylor series were evaluated to higher, nonlinear order beyond (12), then the correlations between altered state and forcing would arise.

Tunable time-independent parameters, β, that appear on the right-hand side of Eq. (12) can also impact the trajectory of $\mathbf{x}$. For the discretized equations, this includes time-stepping schemes and vertical coordinates. This also includes parameters specific to each closure approach, such as Ri criteria in KPP-based formulations (Large et al. 1994; Van Roekel et al. 2018), or stability parameters in second-moment formulations (Umlauf and Burchard 2003).

The sensitivity of $\mathbf{x}$ to perturbations in the state or forcing space depends on the behavior of the OSBL system, which can be evaluated locally and formally through the eigenvalues of the Jacobian (Eq. (13)) and gain matrices (Eq. (14)). Appendix A explores this local approach for the highly simplified two-equation bulk ML model of Kraus and Turner (1967) (hereafter KT67), with results that suggest the KT67 system is stable to small perturbations in state space. While many current SC models are not tractable under the same analytical techniques, it is anticipated that they exhibit the same behavior: that the forced dissipative OSBL can be described by mean variables that evolve continuously and deterministically, and the fast timescales and stochastic, chaotic behavior (especially sensitivity to initial conditions & forcing) that characterize turbulent motions are not characteristic of the later, finite-duration SC model evolution. This is consistent with assumptions in the Reynolds averaged equations where the timescale of turbulence is less than that of the evolving BL (i.e. BL evolution is longer than the large eddy turnover timescale). BL forcing can be represented as the friction velocity, $u_* = \sqrt{\tau/\rho_o}$, and convective velocity $w_* = (B_o H)^{1/3}$. For typical values of $u_* = 0.01\ m\ s^{-1}$ and $H = 40\ m$, a timescale for the evolution of turbulence statistics can be estimated as $\tau_{eddy} \sim H/u_* \sim 1\ hr$ (Wyngaard 2010). Yet the timescale of each SC model *will differ* according to the physics and numerics employed; this work seeks to formulate a system approach to illustrate and compare these across models. The trajectory of the mean fields and turbulent fluxes beyond the turbulent eddy timescale is the focus of this system analysis.

b. The Ensemble Vector approach for intermodel comparison

As highlighted in Eq. (11), sensitivities in numerical simulations of the OSBL are defined by their physics (e.g. choice of parameterizations for unresolved processes), initial conditions, forcing conditions, as well as numerics (e.g. temporal and spatial discretization and resolution) captured in the map, $\mathcal{A}$. When different systems (i.e. SC models with different turbulence parameterizations) begin at $\mathbf{x_i} = \mathbf{x_a}$ with identical spatial resolution, time-stepping schemes and forcing ($\mathbf{F_i} = \mathbf{F_a}$), their initial trajectories will depend on the first term in the Taylor series expansion only (mirroring related approaches such as bred vectors and Lyapunov vectors). Under these conditions, two different ocean states can emerge and then diverge solely due to the choice of turbulence parameterization. While the method below can be expanded to explore different sensitivities in Eq. (12), the diverging ocean states resulting from different parameterized turbulence (i.e. across multiple models) is the focus of the rest of this manuscript. The analysis will include finite, rather than infinitesimal, duration simulations. As such, the idealized localization of Eq. (12), where model, forcing, and parameters are distinct objects for analysis, becomes increasingly poor with the duration of the analysis window. Likewise, analysis of the local objects, e.g., the eigenvalues of the matrices in Eq. (12), is not a complete description of the finite time behavior.

The impact of different systems, $\mathcal{A}^n$, on the trajectory of $\mathbf{x}$ starting at $\mathbf{x_i} = \mathbf{x_a}$ is explored here. It is helpful to establish a reference system

$$\mathbf{x_f}^{ref} = \mathcal{A}^{ref}(\mathbf{x_a}; \mathbf{F}^{i:f}; \beta) \tag{15}$$

For simplicity, we assume the system maps are deterministic, rather than stochastic, as they depend only on the behavior on timescales slower than the turbulence timescales.

From this, it is helpful to define an SC ensemble difference vector $\mathbf{y_f} = \mathbf{x_f}^n - \mathbf{x_f}^{ref}$. The trajectory of $\mathbf{y_f}$, which is the main interest of a multi-model comparison, can also be represented as a dynamical system as explored in (Johnson and Fox-Kemper 2023), which shows how linearization about a few different states and forcing conditions allow the sensitivities of the dynamical system that defines $\mathbf{y_f}$, to be compared with more commonly used methods such as Lyapunov vectors and exponents, bred vectors, and singular vectors (e.g., Wolfe and Samelson 2007; Norwood et al. 2013). While many of these approaches diagnose consequences of the Jacobian solely, SC models tend to respond as much to forcing as to initial conditions, so the gain matrix must also play a role. Yet, after an infinitesimal interval of time, the difference in trajectories between the two systems will continue to be influenced by the different gradients surrounding $\mathbf{x_a}$ between the two maps approximated by $\mathbf{J_a}$, $\mathbf{G_a}$ and $\partial \mathbf{A}^n / \partial \beta$ and like any nonlinear system, becomes increasingly challenging to evaluate.

A more computationally simple and appropriate approach evaluates the finite, nonlinear growth of error in state space between different SC models (i.e. system maps, $\mathcal{A}^n$), defined here as an ensemble vector (EV). The finite, nonlinear growth of error captured by the EV is analogous to bred vectors, commonly used for weather ensemble forecasts (Toth and Kalnay 1993, 1997). It is shown (in appendix C) that the short-time behavior of SC models converges to each model's own

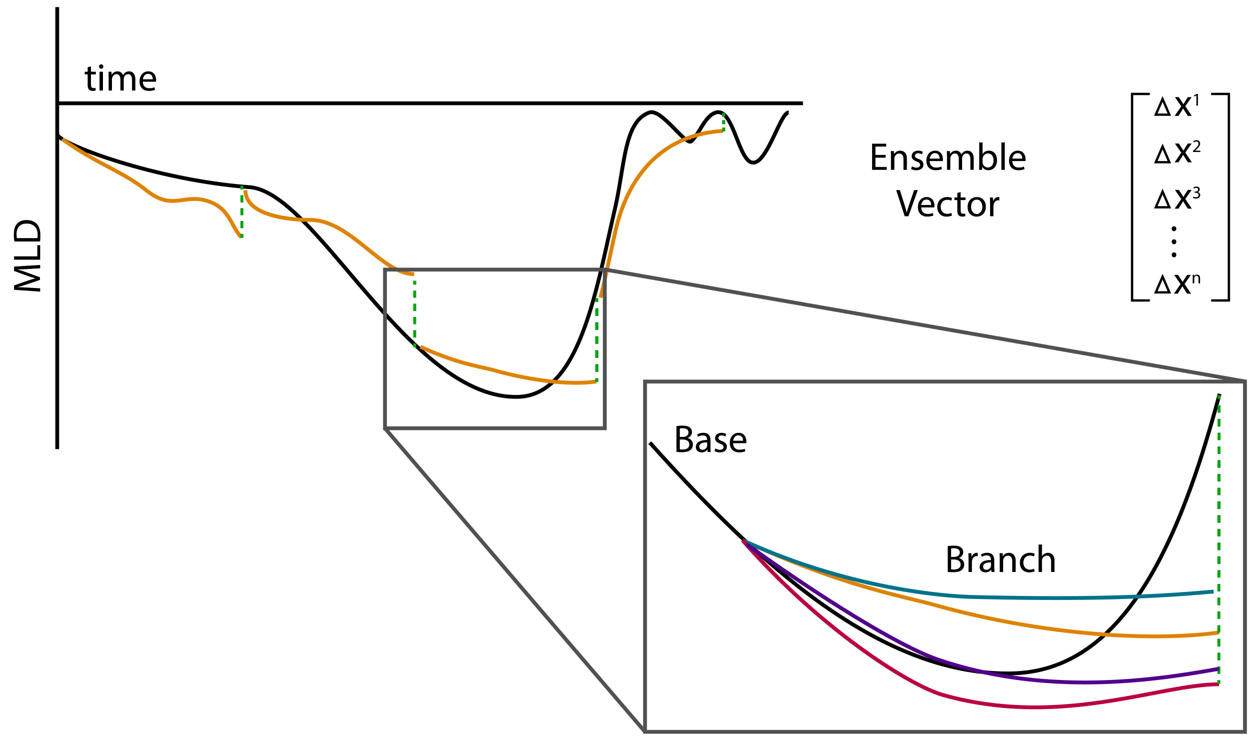


FIG. 1. Schematic of the ensemble vector method for use in inter-model comparison studies. A base run (which could be an SC model from which branches are perturbed, an SC model with reduced state space from an imperfect restart, a multi-SC-model ensemble mean, or an LES "truth") provides state variables to initialize a suite of models at different times. This example shows how the evolution of different models results in different ML depths. After a time interval (e.g. 6 hr), a difference in (non-dimensional) state space between each model, n, and the base run form the EV for that time interval, as described in section 3.

stable trajectory. Therefore, a multi-model SC EV measures the spread across an ensemble of SC models' trajectories.

For intermodel comparison, the EV is obtained by running the model $\mathcal{A}^n$ initiated with state variables from the reference (either ensemble mean or truth) run $\mathcal{A}^{ref}$ mirroring the locus of linearization $\mathbf{x_a}, \mathbf{F_a}$ in (12), referred to as a branch run. After a characteristic timescale (to be determined by the system and SC models), the difference between the modeled state and the reference state is the EV which captures deviations between the nonlinear trajectories of each system map (Fig. 1). In other words, the EV represents the fastest growing nonlinear deviations between the states (i.e. $\mathbf{y_f} = \mathbf{x_f}^n - \mathbf{x_f}^{ref}$) evolved by different turbulence parameterizations.

So far, the discussion of model comparison has been generalized, yet the execution of this method in practice will depend on the nature of the model formulations to be considered (e.g. the chosen base run and SC models and their numerical realization) and the focus of the comparison (e.g. sensitivity). The rest of this manuscript presents an example that compares a suite of ST and LT parameterizations for a case study during the 2018 monsoon onset in the Bay of Bengal. The EV method for intermodel comparison is performed using an LES as the reference base run. SC models run through GOTM (Burchard et al. 1999; Li et al. 2019) are branched from the base run to create the EV as described in section 3. The largest EVs provide a targeted examination of where and why turbulence parameterizations deviate from the LES as explored in section 4 and discussed in section 5.

3. Methodology

a. Data Processing

This analysis is motivated by the 2018 MISO-BOB field campaign that captured the upper ocean response to the onset of the monsoon intraseasonal oscillations (MISO). The details of the ocean response can be found in Shroyer et al. (2021) and are summarized here (Fig. 2). A northward propagating rain band that signaled the onset of the monsoon was associated with strong variable surface forcing (referred to as the active period). During this period, upper ocean mixing from unsteady winds and surface cooling competed with buoyancy input from strong, yet short-lived precipitation events. Later in the survey, the atmospheric forcing regime shifted to one characterized by low winds and a strong diurnal cycle (referred to as the break period). These two phases typify the oscillating wet and dry patterns that characterize the MISO and therefore provide an opportunity to evaluate the performance of upper ocean mixing parameterizations to unsteady and variable monsoon forcing.

Surface fluxes of heat, wind speed, and precipitation were collected from the meteorological system onboard the R/V *Thompson*. Surface heat fluxes and wind stresses were calculated using the COARE 3.5 algorithm and filtered in time to smooth out higher frequencies using a Butterworth filter with a cutoff frequency of one hour. Precipitation was not filtered as to capture significant rainfall events typical of the monsoon. Wave data was not collected during the survey, therefore an assumption of wind-wave alignment is made. Stokes drift profiles are computed from wind

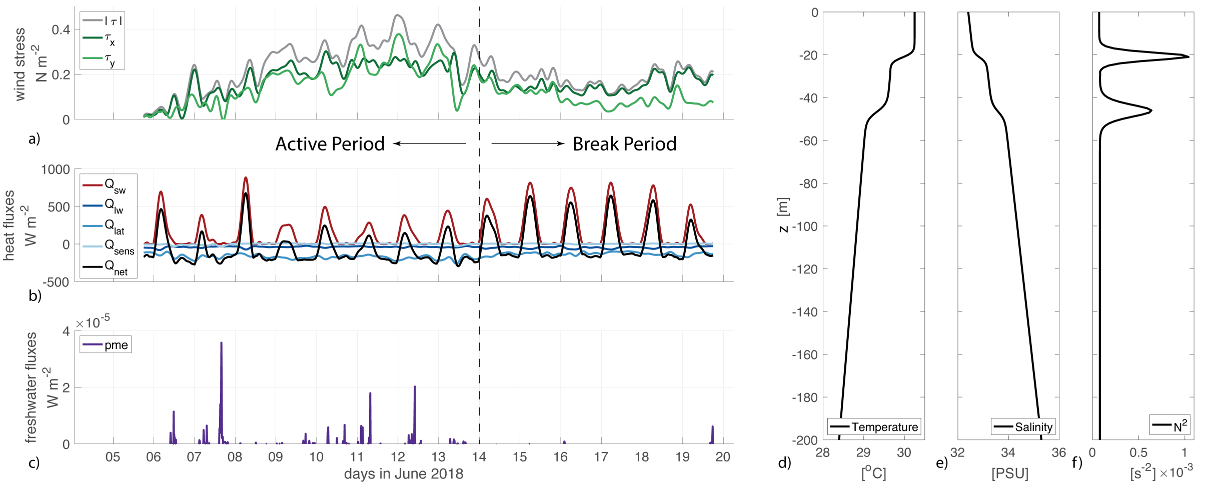


FIG. 2. Surface forcing and initial profiles motivated by observations collected during the 2018 MISO-BOB campaign in the Bay of Bengal used to drive the LES and SC models. The time series is divided into an active period which captured the monsoon onset, followed by a calm break period with strong diel forcing. a) zonal (dark green), meridional (light green), and total (black) wind stress, b) surface heat fluxes, shortwave (red - Q_sw), longwave (navy - Q_lw), latent (blue - Q_{lat}) and sensible heat (light blue Q_{sen}) , c) precipitation minus evaporation. Initial profiles of d) temperature , e) salinity , and f) stratification .

speeds at 10 m (u_{10}) using an empirical wave spectrum assuming equilibrated wind and waves (Donelan et al. 1985) similar to that described in Li and Fox-Kemper (2017). Wind-wave direction is important for LT studies, but in the absence of truth, the assumption here is appropriate for LES-SC model comparison as all LT models use the same Stokes drift profiles.

In-situ measurements collected by a fast-CTD (Pinkel et al. 2012; Lucas et al. 2016) provided motivation for idealized initial vertical profiles of temperature and salinity constructed using a tanh function (Pham et al. 2023). These surface fluxes, Stokes shear, and initial profiles were used to force a combination of LES and SC model (Table 1).

b. Large Eddy Simulation

Large-eddy simulations solve the three-dimensional grid-filtered non-hydrostatic incompressible Navier-Stokes equations under the Boussinesq approximation. Further details of the LES are in Appendix B.

Table 1. List of parameterizations used in this study.

NAME	*TYPE*	*ST*	*LT*	*references*
SMC-KEPS-ST	Second-Moment	×		Rodi (1987)
SMC-MY-ST	Second-Moment	×		Mellor and Yamada (1982)
SMC-LT	Second-Moment		×	Harcourt (2013)
KPP-CVMIX-ST	K-Profile	×		Van Roekel et al. (2018)
KPP-ROMS-ST	K-Profile	×		McWilliams et al. (2009)
KPP-ENTR-LT	K-Profile		×	Li and Fox-Kemper (2017)
KPP-EFACTOR-LT	K-Profile		×	Li et al. (2016)
KPP-R-LT	K-Profile		×	Reichl et al. (2016)
ePBL-ST	Energetic Planetary BL	×		Reichl and Hallberg (2018)
ePBL-LT	Energetic Planetary BL		×	Reichl and Li (2019)

Two LES simulations were performed (Fig. 4); one with Langmuir turbulence (LES-LT)–that is, including the Stokes vortex force, Stokes Coriolis force, and Stokes advection of the wave-averaged Boussinesq equations–and one with shear turbulence only (LES-ST). Both simulations were initialized with observationally motivated salinity and temperature profiles which consist of a 20 m OSBL on top of a 30 m remnant layer. The remnant layer is bounded by the thin layers of elevated N^2 at 20 m and 50 m depths (Fig. 3 a, e). The LES-LT model uses the same Stokes drift as the SC-LT models. Overall, the evolution of the OSBL is qualitatively similar in the two simulations. However, there are important quantitative differences between the two LES simulations due to the effects of Langmuir turbulence, for example, deeper MLs and stronger rates of turbulent mixing in the LT simulation. Detail of the differences can be found in Pham et al. (2023).

c. Single Column Models

This study explores the impact of ten different SC models on the evolution of the upper ocean using a common framework GOTM (Burchard et al. 1999; Umlauf and Burchard 2005) with the extension by Li et al. (2019) to incorporate a set of Langmuir turbulence SC models (SC-LT). Three classes of SC models used here include (1) a set of KPP variants, (2) the energetic Planetary Boundary layer (ePBL) models, and (3) a set of second-moment closure (SMC) models. Within each class, both ST and LT formulations are included. A comprehensive overview of these parameterizations can be found in Burchard et al. (1999), Umlauf and Burchard (2005), and Li

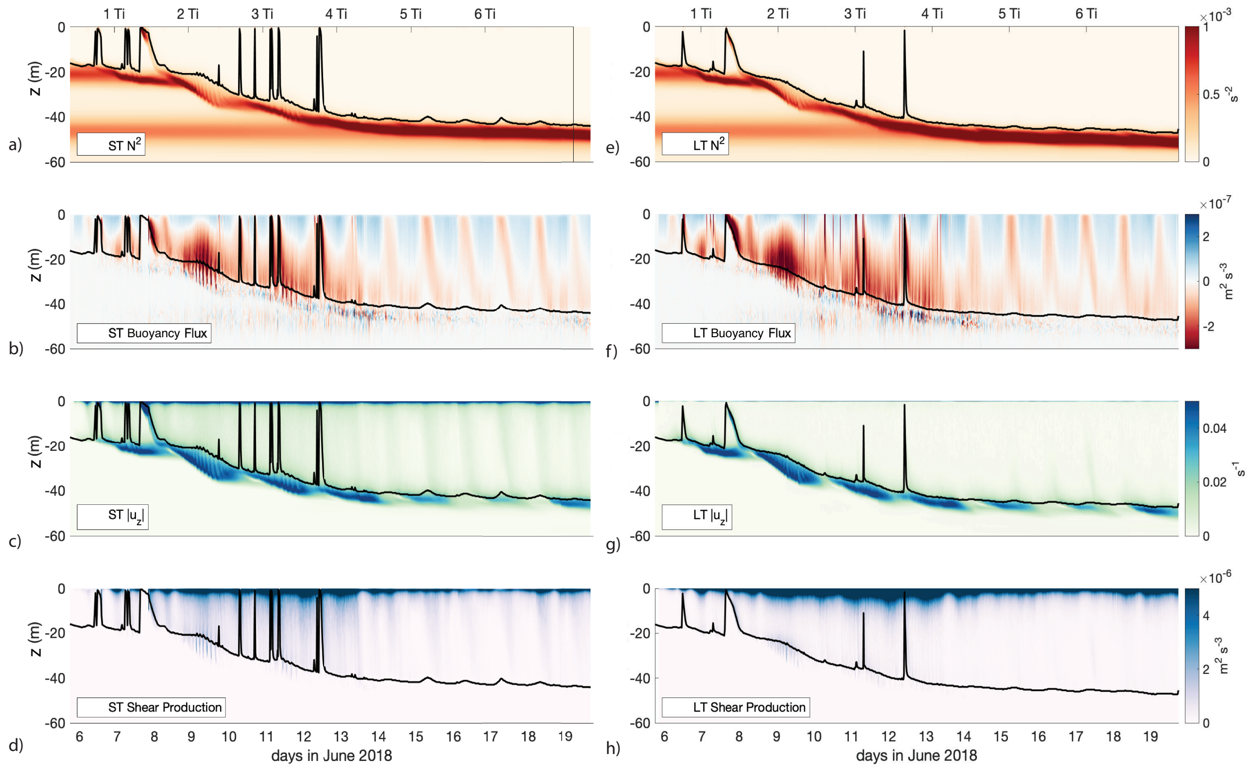


FIG. 3. Mean and turbulent fields from the LES-ST (left) and LES-LT (right) "truth" runs. During the active period, there is strong inertial shear at the ML base and ML deepening. The break period is characterized by strong diel forcing. a,e) stratification, N^2. b,f) buoyancy flux G c,g) shear $|u_z|$, d,h) shear production P.

et al. (2019). Note that SC-LT models solve Eq. (1)-(4) and do not include the Stokes vortex as in the LES. Therefore, the effect of enhanced mixing due to Langmuir turbulence is incorporated implicitly in the turbulent fluxes. The list of parameterizations used in this study and the references are summarized in Table 1.

The simulations were run with a uniform vertical grid spacing of 0.5 m, a time step of 60 s, and initialized with profiles of mean T, S, u, and v from the LES-ST and LES-LT as described in the next section. A comparison of the simulated ML depth in these SC models and LES is shown in Fig. 4.

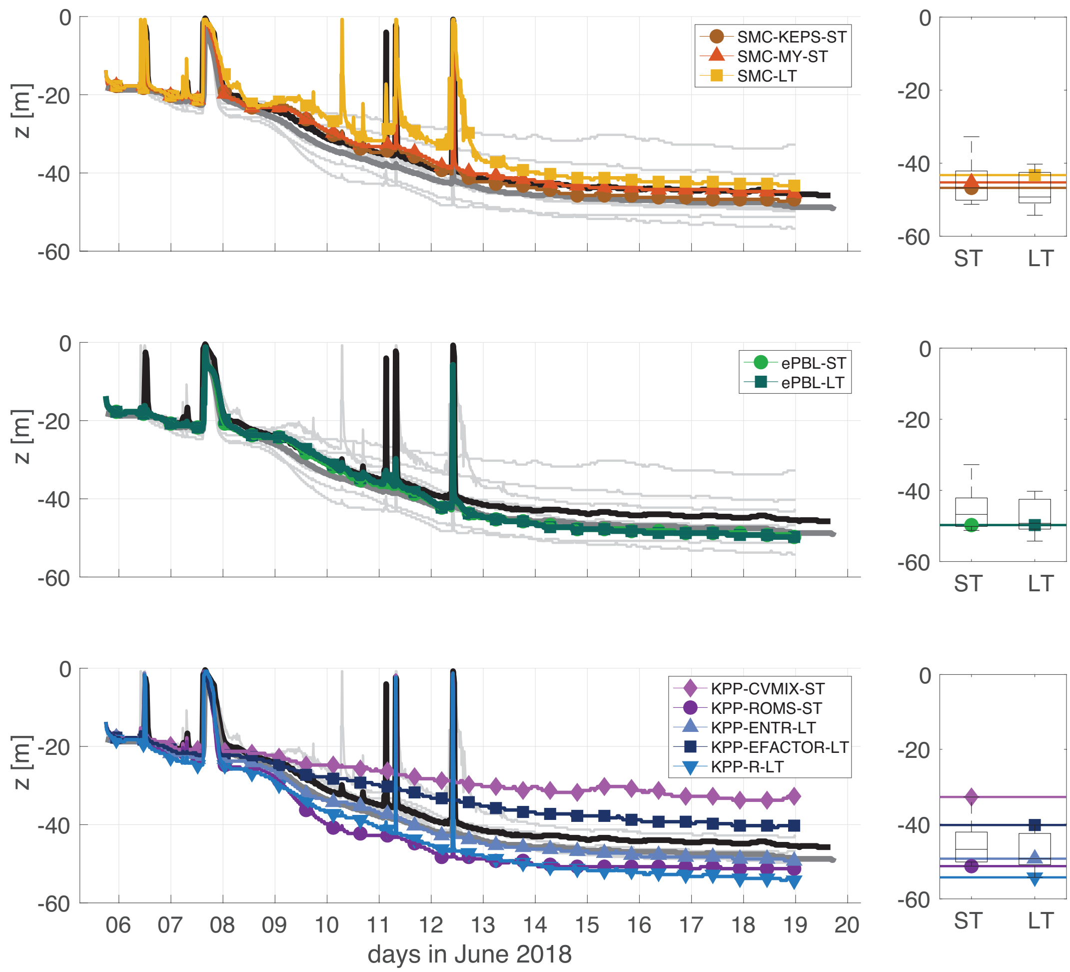


FIG. 4. Mixed layer (ML) depth from continuous model runs (not branched or restarted) (left) and box plots of the reduced-restart model spread (right). Plots are divided by single column (SC) model class for clarity. Shear turbulence (ST) LES (black), Langmuir turbulence (LT) LES (dark grey), and ensemble mean over all SC models (light grey) are the same in each plot for reference. a) second-moment closure models. b) ePBL models. c) KPP models. Box plots show the SC model ensemble spread at the end of the simulations for SC ST and SC LT models separately and the boxes are the same for each panel. Colored lines over the box plots represent ML depth at the end of each simulation.

d. Implementation

A challenge in implementing an EV method for model comparison is consolidating the many possible states that SC models rely on, each with different degrees of freedom that increase with the level of closure. Turbulence in KPP-type models relies on mean fields (to calculate a BL depth) and empirical coefficients based on surface forcing. Higher-order closures contain prognostic turbulent quantities that depend on turbulence production and dissipation. The intermodel comparison requires a reduced state space through which to compare these different maps and variables. Here, that space is reduced to the mean and turbulent fields for T, S, u, and v:

$$\mathbf{x} = \left[T, S, u, v, \overline{w'T'}, \overline{w'S'}, \overline{w'u'}, \overline{w'v'}\right]^{\mathsf{T}} \tag{16}$$

The components of the state vector in Eq. (16) are then non-dimensionalized by surface forcing, layer depth H, and the timescale of the EV interval, Δt_{EV}, such that mean variables scale as $T \sim (\mathcal{B}_o/g\alpha)\Delta t_{EV}/H$, $S \sim (\mathcal{B}_o/g\beta)\Delta t_{EV}/H$, $u, v \sim u_*^2 \Delta t_{EV}/H$, and turbulence variables scale as $\overline{w'T'} \sim \mathcal{B}_o/g\alpha$, $\overline{w'S'} \sim \mathcal{B}_o/g\beta$, and $\overline{w'u'}, \overline{w'v'} \sim u_*^2$. Here, H is defined as a mixed layer depth using a density criteria of 0.1 kg m^{-3}, $\mathcal{B}_o$ is the surface buoyancy flux, u_*^2 is the friction velocity, α is the thermal expansion coefficient and β is the haline contraction coefficient. Models are categorized into SC-ST and SC-LT to be compared with their respective LES-ST and LES-LT simulation. The EV is then defined to be a large single column vector combining the difference between the SC models and the reference LES (see Fig. 1). The reduced state space will be specific to the limitations of the SC model and experimental design. In some cases, it may be informative to look at a single variable only. For example, EV^{SST} uses a reduced state space of sea surface temperature. Specific details about the experimental setup are described in Appendix B. For implementation, SC models were branched off of LES every 3 hrs using $\Delta t = 60$ s. A 6 hr window was chosen as the EV timescale (see appendix C). Choosing a timescale of 4 and 8 hours did not significantly alter the interpretation of the results.

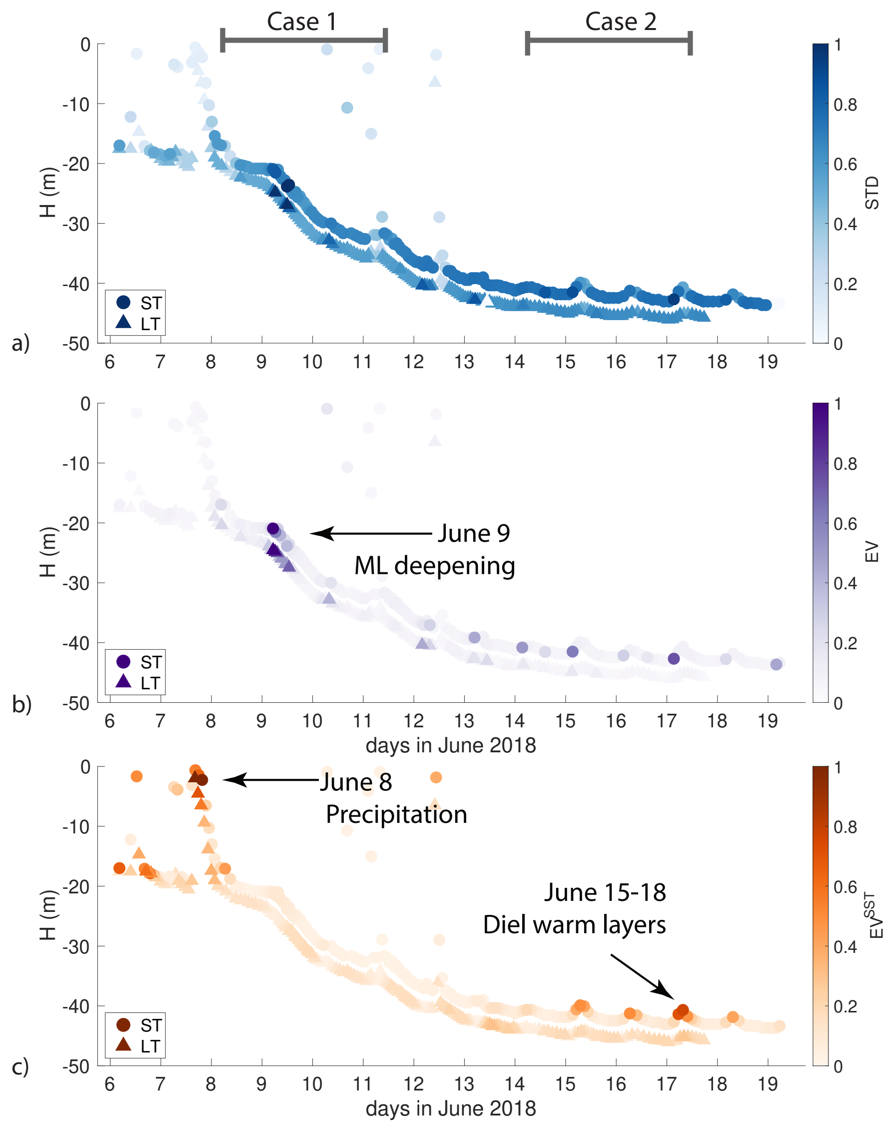
4. Results

a. Mixed Layer Evolution

A full analysis of the LES is detailed in Pham et al. (2023) and summarized here (Fig. 3). For the first 24 hours after LES initiation, shear builds up in the ML (Fig. 3) as the LES adjusts from the zero momentum initial condition. By the first inertial period ($T_i \approx 40$ hours), shear has reached the pycnocline and begins to interact with stratification at the ML base. This study will focus on ML evolution after this initial spin-up. The monsoon onset is distinguished by an increase in winds and intermittent precipitation that leads to a competition between shear production (P = $-\overline{w'u'}\partial u/\partial z - \overline{w'v'}\partial v/\partial z \approx \kappa_m((\partial u/\partial z)^2 + (\partial v/\partial z)^2)$ and buoyancy production (G = $\overline{w'b'} \approx -\kappa_s(\partial b/\partial z)$ within the active ocean surface boundary layer, where κ_m is the eddy viscosity for momentum and κ_s is the eddy diffusivity for scalars. Near-inertial oscillations develop at the local inertial frequency and are associated with enhanced shear at the ML base and rapid deepening. The injection of buoyancy by large rain events is seen as sharp streaks in G and P, yet these events are relatively short-lived and the near-surface rain pools are mixed away by the turbulence within 8 hrs. The transition from an active phase to a break phase in the monsoon occurs around June 14$^{\text{th}}$, and the remainder of forcing exhibits low winds, no precipitation, and a strong diurnal surface warming (Fig. 2).

In both the LES and SC simulations (Fig. 4), the ML deepens during the active period of high winds and cooling, then remains steady with midday shoaling during the break period of strong diurnal warming and reduced winds. The continuous ST and LT SC model runs deviate from the LES-ST and LES-LT respectively during mixed layer deepening and persists through the model run, with a spread of $\Delta H \approx 20$ m for ST and $\Delta H \approx 10$ m for LT SC models. From this example, it is impossible to isolate how the models perform under a range of forcing regimes as the ocean's states between models quickly diverge, and then subsequent behavior and sensitivity to forcing accumulates upon this initial divergence. This disagreement in MLs highlights the importance of an alternative approach to intermodel comparison as discussed in section 3c and below.

Different estimates of model error are represented in Fig. 5. The standard deviation (σ) of the full SC model ensemble difference from the LES (Fig. 4) is interpolated onto EV time intervals (Fig. 5 a) and compared with the EV and EV^{SST} (Fig. 5 b,c respectively). The full model run variance represents the model divergence over time. Alternately, the EV error highlights particular moments where BL parameterizations disagree with LES and offers an alternative depiction of the conditions in which BL parameterizations break compared to continuous runs. The EV^{SST} is also considered here to bring attention to times when SST, an essential variable for air-sea coupling, is sensitive to model physics. Two hot spots that arise provide case studies for discussion: 1) during ML deepening in the monsoon active phase as variable winds, precipitation squalls, and a damped diurnal cycle create near-inertial shear and boundary layer turbulence that erodes the pycnocline,



377 FIG. 5. MLD for LES-ST (circles) and LES-LT (triangles) colored with different measures of uncertainty, normalized such that the highest value of uncertainty for that metric is equal to 1 a) Blue; the standard deviation of the difference between continuous SC models (i.e., not branched) and the LES as seen in Fig. 4. b) Purple; the L2 norm of the EV at each branch run. c) Orange; the component of the EV containing sea surface temperature EV^{SST}. For the continuous runs (blue), the model spread increases over time. The EV (purple) highlights model disagreement during wind deepening (case study 1). The EV^{SST} (orange) is largest during diel surface warming (case study 2) and strong precipitation.

and 2) during the subsequent break period, as reduced winds and diel warming produce a strong diurnal warm layer. Exploring these two cases provide examples of how model physics influences the trajectory of the mixed layer system.

Case 1 (Fig. 6) exhibits one of the most fundamental problems in mixed layer physics, the deepening of the wind-driven mixed layer (Pollard et al. 1973), and has been a testing ground for SC model validation (Price et al. 1986; Mellor and Yamada 1982; Umlauf and Burchard 2003; Large et al. 1994). During case 1, u_* was larger than the convective velocity, $|w_*|$, the Monin-Obukhov length, L_{MO}, was more than twice H and turbulent Langmuir number ($La = [u_*/u_s]^{1/2}$), which scales the relative importance of ST to LT, was approximately 0.275. These scalings predict the dominance of wind-driven and wave-driven turbulence in the OSBL over convection (Belcher et al. 2012). Near-inertial shear reached the base of the mixed layer, resulting in enhanced shear production and buoyancy production that converted kinetic energy into potential energy. Between June 9 and 10, during wind-driven deepening (Fig. 6 d,i), buoyancy production near the ML base is not well represented by the SC models compared to LES-ST. For the ST models, KPP-CVMIX-ST produces the least vertically integrated $\overline{w'b'}$ and KPP-ROMS-ST produces the largest vertically integrated $\overline{w'T'}$ (consistent with MLDs in Fig. 4), with SMC-KEPS-ST, SMC-MY-ST, and ePBL-ST performing closer to LES. The turbulent heat flux at the base of the ML that drives entrainment is more consistent among the LT models than the ST ones. The SC-ST ensemble is closer to LES-ST in terms of velocity than the SC-LT ensemble is from LES-LT, but the SC-LT ensemble is closer to LES-LT in terms of temperature, especially near the mixed layer base where entrainment occurs.

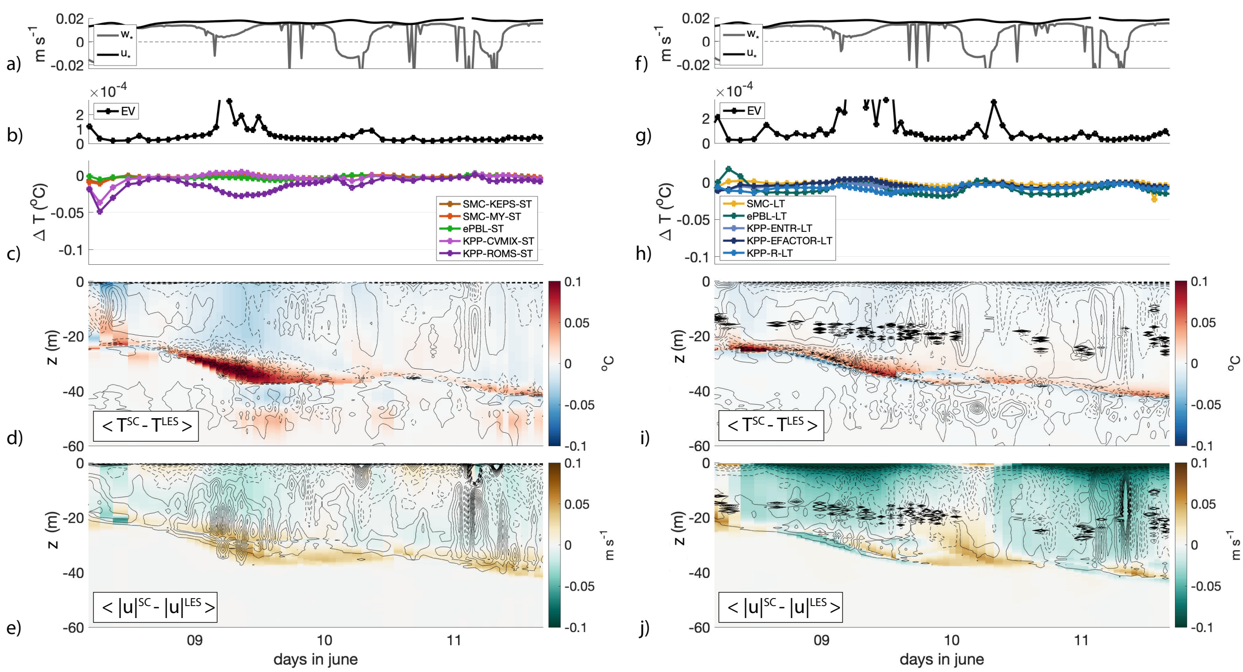


FIG. 6. Case study 1 during wind-driven deepening of the OSBL (active period) for ST (left) and LT (right). All models disagree on how to deepen the ML and induce entrainment, leading to a large EV during this time with implications for SST. a, f) - u_* and w_*. b, g) - EV, the non-dimensional L2 norm of the EV at end of each branch run. c, h) - ΔSST, the difference between SC models and LES at end of each branch run. d, i) The difference between the average of SC model temperatures and the LES temperature (colored). Contours are $\Delta|w'T'|$ with spacing $3\times10^{-6}\ ^oC\ m\ s^{-1}$ solid lines are positive and dashed lines are negative. . e, j) - The difference between the average of SC model velocity magnitudes and the LES velocity magnitude (colored). Contours are mean Eulerian $\Delta|w'u'|$ with spacing $9\times10^{-6}\ m^2\ s^{-2}$ solid lines are positive and dashed lines are negative.

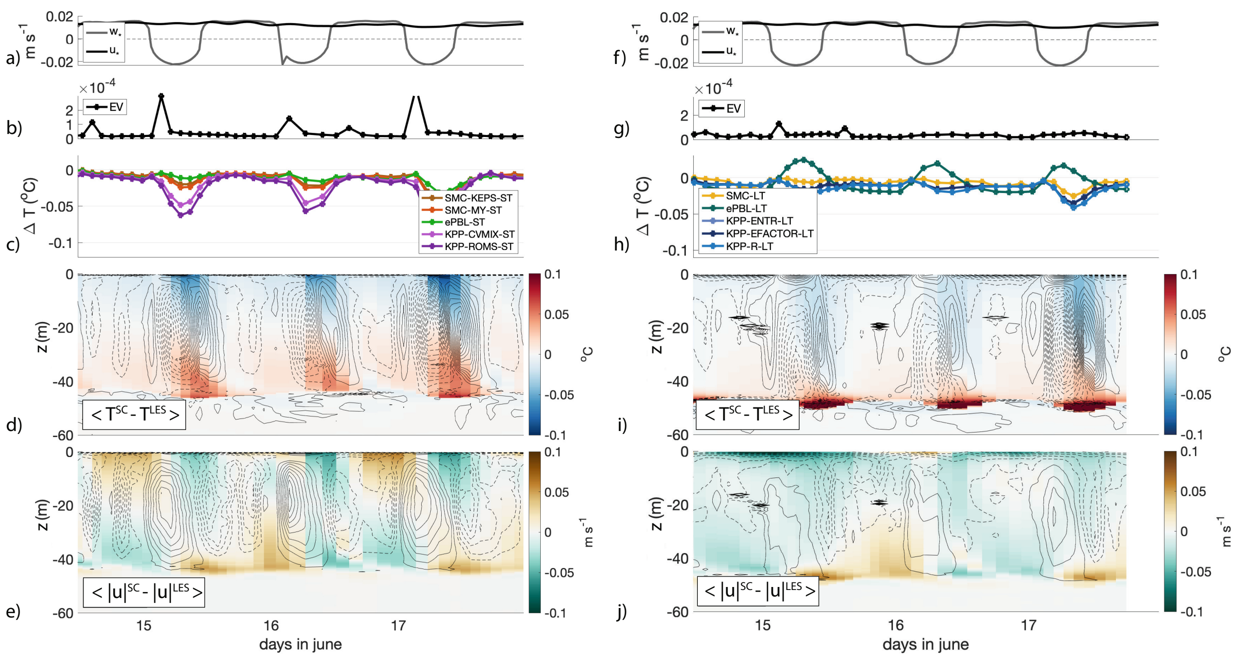


FIG. 7. Case study 2 during strong diel warming (break period) for ST (left) and LT (right). SC models tend to flux more heat away from the surface than LES, resulting in a cool SST bias. Though the L2 norm of the EV is not as large as in case 1, the localized disagreement has implications for diel SST amplitudes. a, f) - u_* and w_*. b, g) - EV, the non-dimensional L2 norm of the EV at end of each branch run. c,) - ΔSST, the difference between SC models and LES at end of each branch run. d, i) mean temperature difference between SC models and LES (colored). Contours are mean $\Delta|w'T'|$ with spacing 3×10^{-6} ^oC m s^{-1} solid lines are negative and dashed lines are positive. . e, j) - mean velocity difference between SC models and LES. Contours are mean $\Delta|w'u'|$ with spacing 9×10^{-6} m^2 s^{-2} solid lines are positive and dashed lines are negative.

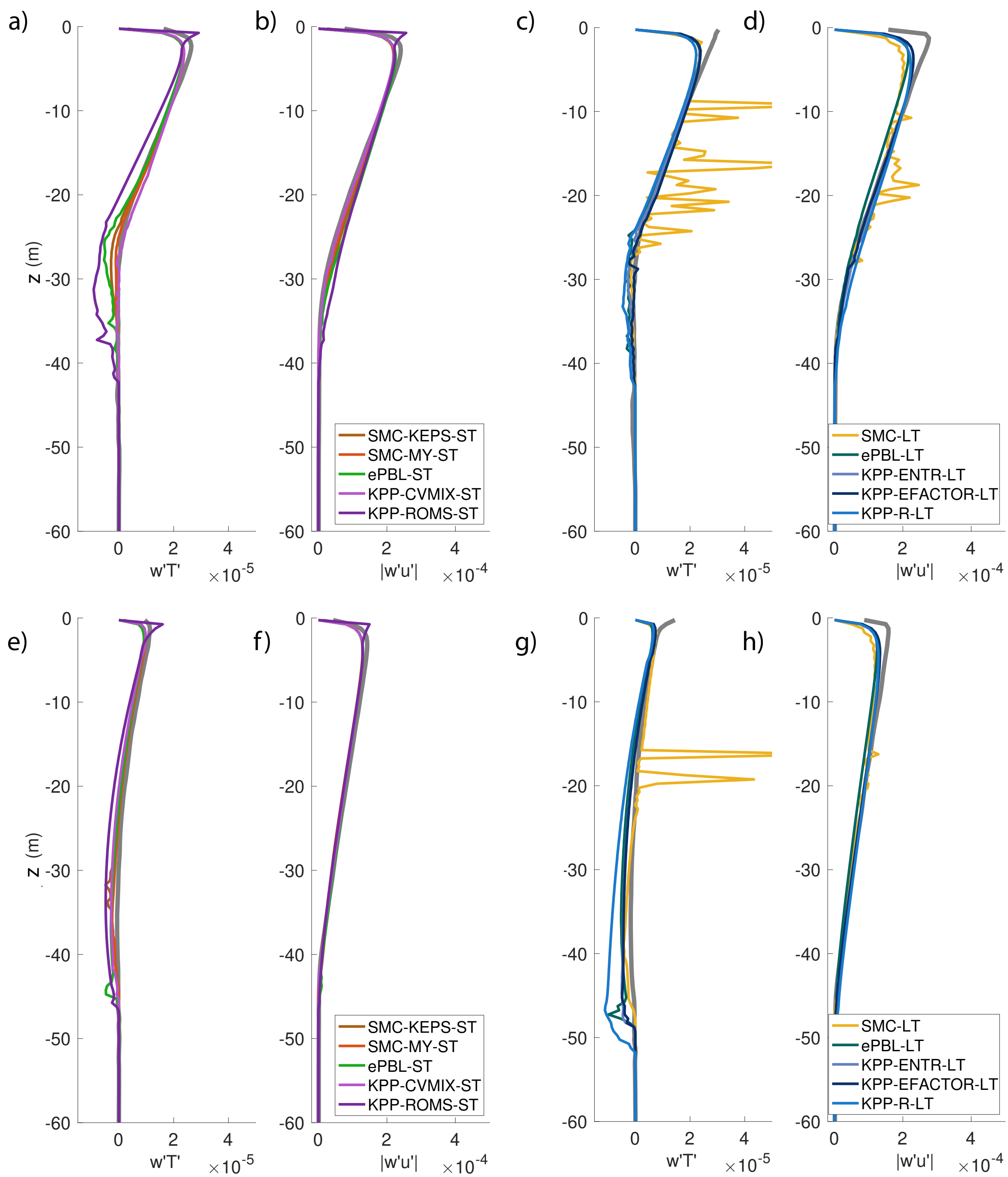
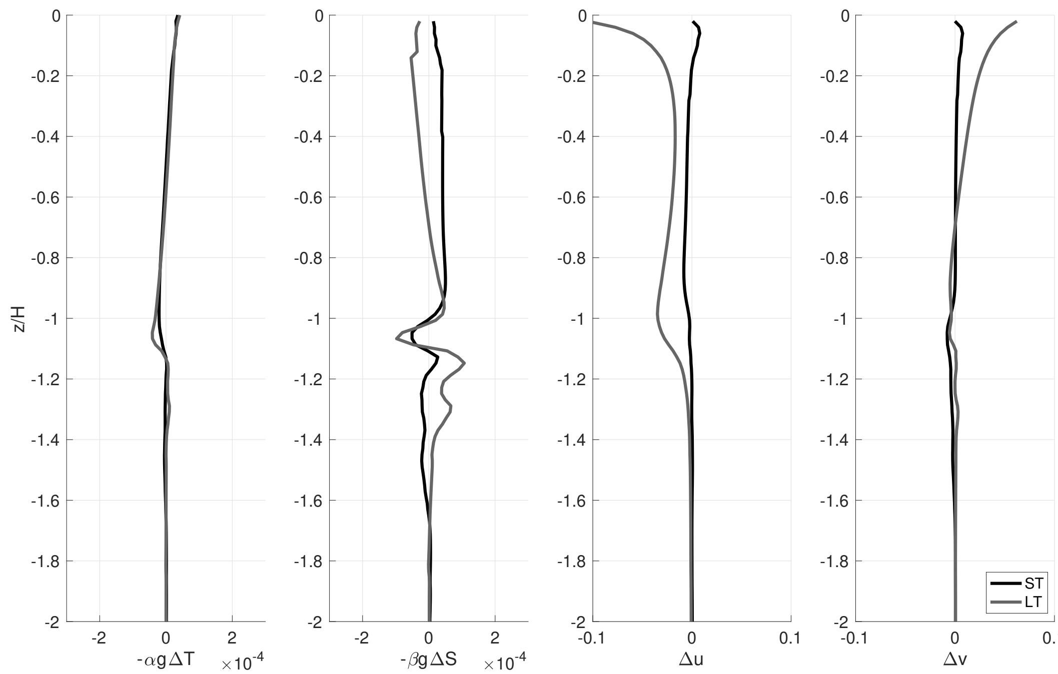


FIG. 8. Mean turbulent flux profiles for ST models (left, a, b, e, f) and LT models (right, c, d, g, h), with SC models in colors and LES in thick grey. Top (a-d): Case 1 (June 8 - 11; see Fig. 6). Bottom (e-h): Case 2 (June 14-18; see Fig. 7). a) Case 1 ST $\overline{w'T'}$, b) Case 1 ST $|\overline{w'u'}|$, c) Case 1 LT Eulerian $\overline{w'T'}$, d) Case 1 LT $|\overline{w'u'}|$, e) Case 2 ST $\overline{w'T'}$, f) Case 2 ST $|\overline{w'u'}|$, g) Case 2 LT $\overline{w'T'}$, h) Case 2 LT Eulerian $|\overline{w'u'}|$.

The initial monsoon onset is followed by a break period where deep mixed layers respond to strong daytime surface warming: case study 2 (Fig. 7). The wind speed has reduced such that u_* is smaller than the peaks in $|w_*|$ and a positive L_{MO} indicates that buoyancy forcing restratifies and acts against shear and Stokes production. The EV in the ST and LT SC simulations during this stage is less than during case 1, but their influence on SST is clear in EV^{SST}, the component of the EV reflecting sea surface temperature anomalies, as shown in Fig. 5. Both ST and LT SC models overestimate the downward turbulent heat flux and result in a damped diurnal cycle. This is consistent with the larger turbulent heat fluxed for all models between 5-20 m depth compared with LES. The temperature tendency depends on the flux divergence, and therefore the gradients in $\overline{w'T'}$. The enhanced curvature in turbulent heat flux between 5-20 m would result in more heat fluxed away from the surface (i.e. not as much warming). Note that the SC models agree with each other more than they do with LES in both LT and ST cases. The SC-LT ensemble is closer to LES-LT in terms of temperature and velocity than the SC-ST ensemble is to LES-ST (Fig. 7d-j), and has a smaller EV and EV$^{\mathrm{SST}}$ during this phase (Fig. 5c).

When averaged over the entire simulation, disagreements in T and S, and therefore ρ, between SC models and LES are largest at the ML base (Figs. 6, 7, and 9 a,b). SC models tend to be less dense above the ML and more dense below the ML indicating insufficient entrainment, with implications for stratification across the ML base and the potential energy of the water column (as discussed in section 5). Additionally, SC model velocities disagree with LES near the surface (Fig. 9 c,d), suggesting parameterized momentum flux divergences are not consistent with LES. These discrepancies in mean fields are significant for the state and energetics of the OSBL. Implications of these results are explored in the next section.



466 Fig. 9. Profiles of the difference between branched SC-ST models and LES-ST (black) and SC-LT models and LES-LT (grey) averaged over the entire study for a) buoyancy from T, b) buoyancy from S, c) u and d) v. T and S disagreements are largest at the ML base. Momentum in SC models are larger than LES near the surface.

5. Discussion

A main motivation for the EV analysis is to isolate model disagreement under different forcing conditions and ocean states to identify where parameterized physics can be improved. Using the EV method to identify when SC models disagree isolates two cases: during wind-driven deepening (case 1) and strong diel forcing (case 2). Model disagreements in the context of boundary layer theory and parameterization implementation are discussed here.

During case study 1, different variants of KPP ST formulations set the upper and lower limit of entrained turbulent heat flux during wind-driven deepening. This is consistent with the evolution of model spread in the continuous runs (Fig. 4) and suggests events such as this could kick a model state into a different trajectory over time. In this case, KPP-CVMIX-ST underestimates turbulence throughout the ML, while KPP-ROMS-ST overestimates entrainment flux. The shallow ML in KPP-CVMIX-ST is coincident with subcritical local gradient Ri at the depth of the KPP OSBL even though the bulk Richardson number criteria is met (not shown). It is common to implement a local gradient Ri number mixing criteria in KPP models for internal wave mixing, but this is effective below the OSBL depth and does not alter the results here (in GOTM).

Like many first-order mixing schemes, KPP uses a diagnostic definition for turbulence, which does not consider past turbulence statistics but instead depends on instantaneous mean variables and surface forcing. A KPP turbulence profile can, through the mean equations Eq. (1)-(2) induce an Ekman spiral and near-inertial shear, yet the translation of near-inertial energy into turbulence can only occur through the mean variables at the top and bottom of the OSBL through the bulk Ri criteria rather than through localized Ri anomalies in three dimensions as LES might. As the bulk OSBL definition in KPP-CVMIX-ST fails to deepen the mixed layer, shear builds at the ML base. This is not the case in KPP-ROMS-ST, which adopts an integral form definition for Ri (McWilliams et al. 2009). In the presence of complicated vertical shear (e.g. during times of strong wind forcing), this definition can result in a deeper OSBL depth than KPP-CVMIX-ST and therefore a different shape of κ_ϕ. In this case study, the KPP-ROMS-ST definition of bulk Ri results in significantly more mixing (as κ_ϕ in KPP is inherently linked to BL depth) than the LES and other parameterizations (e.g. Fig. 4, 6). Conversely, higher-order turbulence closure schemes tend to have stability parameters tuned to obey local gradient Ri criteria, which may be the reason why SMC-ST and KEPS-ST (and ePBL-ST which is tuned to behave like KEPS-ST) agree more with

LES during wind-driven deepening than KPP-based models do. However, in profiles (Fig. 8d, h) the local nature of second-moment closure models can produce spurious extrema.

Langmuir turbulence models are in better agreement with the LES-LT and among other SC-LT models than the shear turbulence models. KPP-based LT models (KPP-R-LT and KPP-ENTR-LT) set the upper and lower limits of the EV spread during ML deepening (Fig. 6), but the EV direction (i.e. order of model spread) is not consistent throughout case study 1. Overall, LT models agree on how to deepen the ML compared to ST models under this forcing (i.e. wind and wave-driven deepening) and state.

The active phase of the monsoon is followed by a break phase, with weak surface winds and a strong diurnal heat flux (case study 2). During this time, SC models underestimate the amplitude of diurnal sea surface temperature (Fig. 7) as a result of greater heat and momentum flux from the surface than LES during the nighttime and morning hours and less heat and momentum flux from the surface during the afternoon and evening transition. This leads to an underestimation of shear and stratification (not shown) in SC models during peak warming. This diurnal cycle of overestimation-underestimation in the turbulent fluxes does not cancel out upon averaging but results in a persistent cold SST bias in SC models compared to LES when averaged over the entire diurnal cycle. This bias is larger in ST models than LT models.

Unlike the wind-driven deepening case, turbulent heat flux profiles within SC models (both ST and LT) agree more among different parameterizations than with LES in the strong diel warming case. Because the LES and SC models use the same light attenuation curves in the temperature tendency equation, this artifact can only result from their representations of turbulence. The agreement among SC models suggests that turbulence parameterizations are built to obey similar scaling laws near the boundary. More work on near-boundary behavior is needed to understand the correct scaling and curvature of κ_ϕ during strong diurnal forcing. This challenge is a prospect for comparison between models and observations as well, as lateral effects are not expected to be important to these near-surface diel processes. The representation of these processes is likely also important for marine heatwaves (Fox-Kemper et al. 2021b).

The shape of the flux divergence determines the conversion of wind power to turbulence kinetic energy through shear and buoyancy production and turbulent transport. Since quantities $\overline{w'\phi'}$ in Eq. (1)-(4) are directly related to the turbulence kinetic energy budget, these examples confirm the importance of parameterized flux divergence on the partitioning of energy between mean and turbulent reservoirs. During the active period, buoyancy production correlates with $\tau \cdot \mathbf{u}^{z=0}/H$ (not shown), signifying the importance of the alignment of near-surface velocity and wind stress for buoyancy production (Crawford and Large 1996; Skyllingstad et al. 2000). From a turbulence energetics view, vertically averaged shear production in KPP-ROMS-ST is not different from other mixing schemes, yet buoyancy production is enhanced significantly compared to LES and other parameterizations (Fig. 10), leading to deeper mixing and more change in mean PE. The relationship between model energetics and ML depth is apparent, models with more turbulent shear and buoyancy production have deeper MLs, larger mean potential energy, and lower mean kinetic energy (Fig. 10c). A simple assumption for turbulence in the steady-state BL is that shear production and buoyancy production are balanced by dissipation, such that $P + G - \epsilon = 0$. KPP formulations don't maintain this balance as fundamentally as second-moment closures do. Instead, energetics in KPP models are expressed through the MO derived diagnostic turbulence and bulk Richardson number criteria that can result in unrealistic physical states (e.g. subcritical Ri numbers at the base of the ML). As such, energetic analysis, including EVs of energetic quantities, provide a more informative criteria for model evaluation beyond typical state variables such as ML depth and SST commonly used to discuss SC model comparison. Reichl et al. (2022) show that an energetic framework is useful even in the definition of mixed layer depth.

The ensemble vector method provides error bounds on SC model evolution that are not available when modeling any single SC model. Cases with large EV errors provide target regions for parameterization and SC model improvement. More recent work uses novel techniques such as machine learning, artificial neural networks, ensemble Kalman filters, and super-parameterizations, to constrain parameterization variables to fit LES under an array of forcing conditions (e.g., Liang et al. 2022). A commonality between parameterization fitting efforts and the ensemble error estimates presented here is acknowledging the vast array of forcing and state space that OSBL parameterizations must be able to span to accurately predict upper ocean evolution.

6. Conclusion

This work outlines an ensemble vector approach for OSBL model comparison that uses an ensemble vector methodology to isolate the nonlinear trajectories of the OSBL subject to different

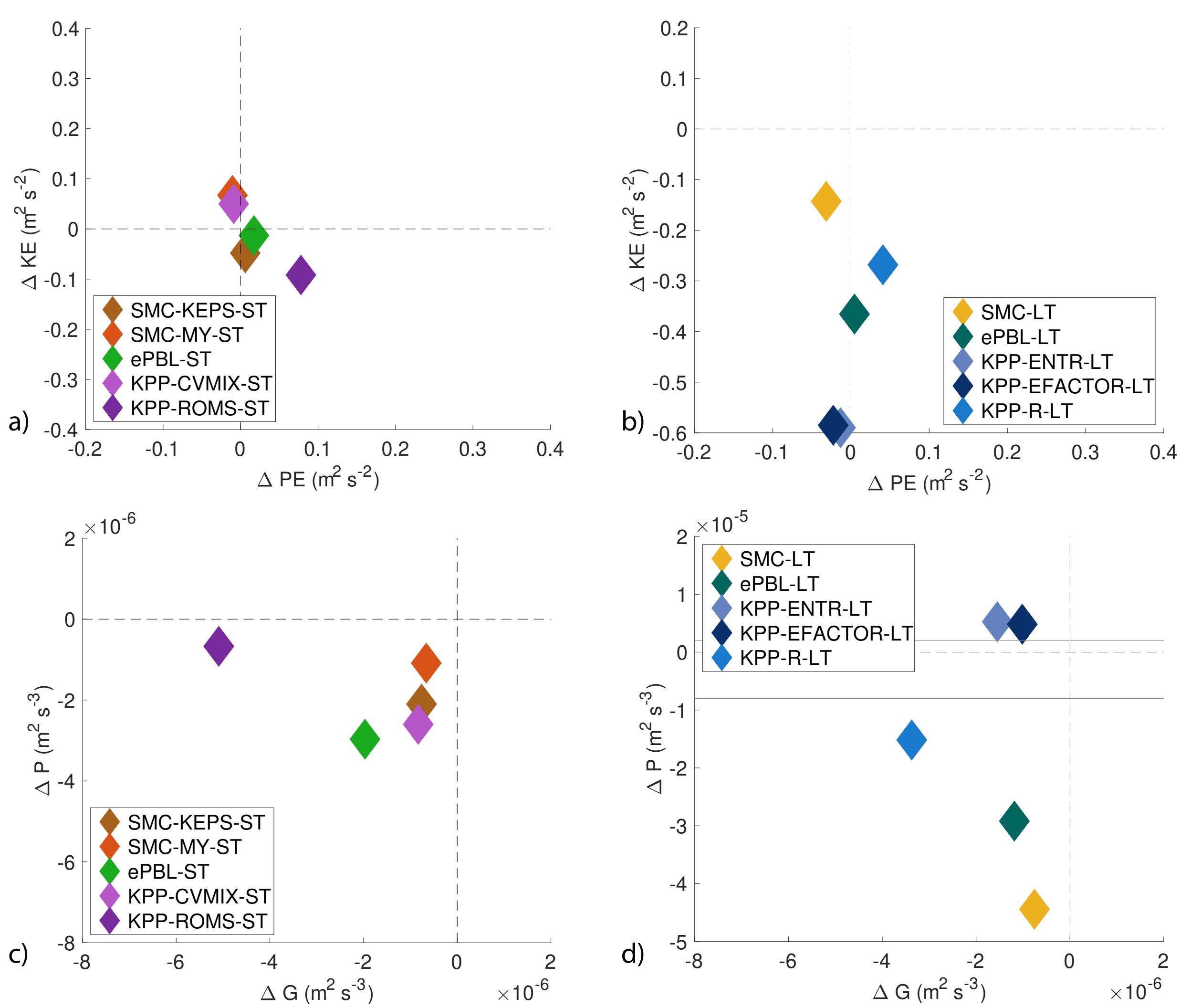


FIG. 10. Model differences in mean kinetic and potential energy (top) and turbulent kinetic and potential energy (bottom) for ST (left) and LT (right) models. a) Difference in mean kinetic and potential energy between SC-ST models and LES-ST. Models with more potential energy (deeper ML) have less kinetic energy. b) Difference in mean kinetic and potential energy between SC-LT models and LES-LT. All SC models have greater kinetic energy and mixed potential energy biases. c) Difference in shear production (P) and buoyancy production (G) between SC-ST models and LES-ST. All SC-ST models underestimate both types of production compared to LES-ST. Note the different scale for shear production in LT models which is enhanced by Stokes shear.

turbulence parameterizations. Within the ensemble vector timescale, each model exhibits initial transience, usually characterized by rapid changes in the state before returning to the state of its own base run. This initial transience hinders the application of alternative dynamical systems approaches that depend on the linearization-based analysis methods (i.e. Lyapunov vectors, singular vectors), as they often dominate the tangent linear system. The relaxation of the trajectories back to their base run as seen in Fig. C1 contrasted with the divergences of trajectories noted in the EV (Fig. 5) implies that trajectories in the OSBL are more sensitive to choice in turbulence parameterization than to perturbations in state space resulting from initial transience. In terms of the dynamical systems framework outlined in section 2, the state $\mathbf{x}$ is *more sensitive to different maps*, $\mathcal{A}^n$, than the Jacobian, Eq. (13), or gain matrix, Eq. (14), within a single map for the parameter, state and forcing space explored here.

As such, perturbed model states are not expected to diverge exponentially over time as assumed in the Lyapunov vector and bred vector approaches, but to remain diffusive as explored in the KT67 equations. Though the OSBL is a diffusive system that does not appear to exhibit chaotic behavior (i.e. appendix A and C), the non-linearity of the turbulence parameterization alters the system's trajectory so that a model's state at a given time depends on an accumulation of historical errors. This EV method identifies the nonlinear difference between stable trajectories of various maps subject to specific forcing conditions. The forcing here is key and provides a source of energy for the EV as momentum and buoyancy input at the surface are distributed differently by parameterized flux divergence formulations. The EV method highlights the key forcing when SC models diverge, unlike direct continuous simulations of transient forcing where errors build upon errors and obfuscate the interpretation of ensemble spread (Fig. 4). This work focused specifically on parameterization choice, but the Taylor series expansion in Eq. (12) sets up a framework to design other EV experiments. For example, the EV method could be adapted to explore gain matrices and evaluate sensitivity to surface forcing (e.g. uncertainty caused by reanalysis products, bulk formula or light extinction coefficients). Additionally, $\partial\mathcal{A}/\partial\beta$ could be used to evaluate sensitivities to parametric error (Souza et al. 2020), or spatial and temporal evolution (Van Roekel et al. 2018).

This case study identified windows of forcing where models deviate: 1) during wind-driven deepening and 2) under strong diurnal forcing. The isolated times of maximum EV contrast the ML spread in Fig. 4, which grows in time as model choices during the early monsoon onset are propagated throughout the continuous run. For wind-driven deepening, models disagree on how to redistribute wind power into turbulent buoyancy production, resulting in varied relationships between mean and turbulent energy in the upper ocean. Future work to improve parameterizations could consider energetic criteria to constrain mixing during these times. Under strong diel warming, SC models overestimate turbulence in the early part of the day and underestimate turbulence in the evening, with a net negative SST bias when averaged over an entire cycle. During this cycle, turbulence parameterizations agree more among each other than with LES. This suggests a need for further research on how near-surface turbulent heat flux behaves in SC models, LES, and observations.

This study did not aim to identify the best model, yet it is helpful to relate model behavior here in the context of previous studies. SMC-KEPS-ST and ePBL-ST tend to agree most with LES-ST, while ePBL-LT and KPP-ENTR-LT agree most with LES-LT in this study. These results are fairly consistent with the SC model vs. idealized LES comparisons in Li et al. (2019) where ePBL-LT and KPP-ENTR-LT were closest to the LES-LT simulations. The agreement between SMC-KEPS-ST and ePBL-ST is expected since ePBL-ST was designed to mimic SMC-KEPS-ST but under more robust numerical implementation. The two end members of the full ensemble spread in Fig. 4 are KPP-CVMIX-ST (shallowest ML) and KPP-ROMS-ST (deepest ML), again consistent with results of Li et al. (2019). This model spread originates during stage 1 as the different Ri criteria under enhanced shear due to wind-driven deepening result in drastically different OSBL depths. KPP-based models agree more during modest wind and strong diel forcing (stage two, Fig. 7). The sometimes disparate behaviors of different KPP models reinforce the importance of numerical implementation (KPP-CVMIX-ST vs. KPP-ROMS-ST in particular, which have identical theoretical foundings but different implementations), in addition to foundational aspects of OSBL theory (e.g., KPP vs. SMC. vs. ePBL vs. LES), on the trajectory of the ML system. Though this study identified two forcing regimes where models disagree, it is anticipated that the direction and magnitude of ensemble spread would shift under different forcing conditions. Therefore any statement about ”the best” model requires an EV analysis across a range of state and forcing spaces, and could be the focus of future work.

In weather forecasting, ensemble methods offer uncertainty bounds not offered by a single deterministic run (e.g., Toth and Kalnay 1997; Molteni et al. 1996). In Fig. 4, the ensemble mean (of the continuous run) is closer to LES than any single model. A rule-of-thumb that ensemble means tend to outperform individual models has long been noted in model ensembles where every model has good reason to be included (e.g., Gleckler et al. 2008), but the rule can be violated with pathological choices of models to include. Therefore, an ensemble mean of several continuous runs may provide a reliable base run along with uncertainty bounds in lieu of more computationally expensive LES. Furthermore, this suggests the potential of inter-model OSBL parameterization ensembles as a robust way to employ SC models.

The influence of turbulence parameterizations impacts upper ocean predictions during the Monsoon Intraseasonal oscillation. This work spans one active-break cycle as the onset of the northward propagating monsoon deepened the mixed layer, and the following break period reduced mixing and warmed the upper ocean. The amount of deepening predicted by the models decides the fate of air-sea interaction during the break period and the heat capacity of the upper ocean for the following monsoon period. The OSBL system, though not chaotic, is highly nonlinear and exhibits hysteresis. As such, small differences in state space identified by the EV method capture tendencies for turbulence parameterizations to set different trajectories for the OSBL system. This analysis is meant to highlight these distinctions and lead to better modeling of the OSBL and Monsoon Intraseasonal Oscillation overall.

Acknowledgments. LJ and BFK were supported by ONR N00014-17-1-2393. HTP and SS are supported by ONR N00014-17-1-2735. BFK received partial support from NSF 2148945. QL was supported by the E3SM project funded by the Office of Biological and Environmental Research in the U.S. Department of Energy's Office of Science.

Data availability statement. This manuscript used LES simulations as described in (Pham et al. 2023) and can be accesses through doi: 10.5281/zenodo.7250847. The single column simulations were run through GOTM5 with additional packages to include Langmuir turbulence (Li et al. 2019). Simulation code can be accessed at github.com/qingli411/gotm and github.com/qingli411/gotmwork. Additional code to calculate the EV can be accessed through doi:https://doi.org/10.26300/c277-dz74.

APPENDIX A

Example: the Kraus-Turner model

Understanding the simulated OSBL as a nonlinear dynamical system provides a principle framework for contextualizing the often chaotic behavior of turbulent flows. But unlike other geophysical fluid or turbulent regimes, the Reynolds averaged OSBL tends toward diffusive behavior or at least non-chaotic behavior. A simple example of OSBL behavior can be recognized by the highly simplified ML equations of KT67. Without loss of generality, the KT67 equations are written here in terms of b^T (buoyancy influenced by temperature only), the friction velocity $u_* = \sqrt{\tau/\rho_o}$, the surface buoyancy flux $\mathcal{B}_o$ and mixed layer depth, H. The variables are nondimensionalized (denoted by $\langle\,\hat{}\,\rangle$) by dividing the dimensional variable by its scale (denoted by $\langle\,\tilde{}\,\rangle$) using the following relationships $u_* \sim \tilde{u}_*\hat{u}_*$, $H \sim \tilde{H}\hat{H}$, $t \sim (\tilde{H}/\tilde{u}_*)\hat{t}$, $b^T \sim (\tilde{u}_*^2/H)\hat{b}^T$ and $\mathcal{B}_o \sim (\tilde{u}_*^3/\tilde{H})\hat{\mathcal{B}}_o$:

$$\frac{d\hat{b}^T}{d\hat{t}} = -\frac{2}{\hat{H}^2}\left[\hat{u}_*^3 + \hat{\mathcal{B}}_o\hat{H}\right] \tag{A1}$$

$$\Lambda\left(\frac{d\hat{H}}{d\hat{t}}\right)\frac{d\hat{H}}{d\hat{t}} = \left(\frac{1}{\Delta\hat{b}^T\hat{H}}\left[2\hat{u}_*^3 + \hat{\mathcal{B}}_o\hat{H}\right]\right) \tag{A2}$$

where Λ is the Heaviside step function, such that $\Lambda\left(d\hat{H}/d\hat{t}\right)$ is equal to zero when $dH/dt < 0$ (i.e. shoaling ML) and equal to one when $dH/dt > 0$ (i.e. deepening ML), and $\Delta\hat{b}^T$ is the (prescribed) buoyancy jump at the base of the ML. The state and forcing space for the KT67 are simply

$\mathbf{x} = [\hat{b}^T, \hat{H}]$ and $\mathbf{F} = [\hat{u}_*, \hat{\mathcal{B}}_o]$. The Heaviside function is an essential nonlinearity of this model, but it can be avoided by considering only shoaling or deepening conditions separately. In the KT67 equations, shoaling MLs collapse to the diagnostic relationship for ML depth, $H = -2u_*^3/\mathcal{B}_o$, which is proportional to the MO depth $L_{MO} = u_*^3/\kappa_{vk}\mathcal{B}_o$, where $\kappa_{vk} = 0.4$ is the Von Karman constant. We note that this is not a fixed point of the system, as the ML buoyancy continues to evolve under $\mathcal{B}_o$ according to Eq. (A1). For a deepening ML, the depth tendency Eq. (A2) becomes prognostic and Eq. (A1)-(A2) form a coupled system.

The eigenvalues for the Jacobian, λ^J, and gain matrix λ^G of this system are:

$$\lambda^J_{1,2} = \frac{\hat{u}_*^3}{\Delta\hat{b}^T\hat{H}^2}\left[-1 \pm \left[1 - \left(\frac{\hat{\mathcal{B}}_o\hat{H} + 2\hat{u}_*^3}{\hat{u}_*^3}\right)^2\right]^{1/2}\right] \tag{A3}$$

$$\lambda^G_{1,2} = \left(\frac{\Delta\hat{b}^T\hat{H} - 3\hat{u}_*^2\hat{H}}{\Delta\hat{b}^T\hat{H}^2}\right)\left[-1 \pm \left[1 - \frac{6\hat{u}_*^2}{\left(2\Delta\hat{b}^T\hat{H} - 6\hat{u}_*^2\hat{H}\right)^2}\right]^{1/2}\right] \tag{A4}$$

During ML deepening (i.e. when $\mathcal{B}_o\hat{H} > -2u_*^3$), $\lambda^J_{1,2}$ are negative (i.e. stable), implying nearby initial conditions will converge *eventually* rather than separate (i.e., not chaotic sensitivity). Asymptotically convergent solutions for $\lambda^J_{1,2}$ are expected due to the diffusive, non-chaotic nature of the ML equations recognized when assuming a gradient form for the flux divergence (e.g. K-theory), transforming Eq. (3) into a heat equation that would equilibrate under constant temperature boundary conditions. Eigenvalues for the gain function, $\lambda^G_{1,2}$ can be both positive or negative, and are determined by complicated relationships between u_* and the ML buoyancy jump (Δb). Unlike the Jacobian matrix, the sign of eigenvalues of the gain matrix do not indicate stability, but they do indicate sensitivity. So, surface forcing perturbations might drive neighboring trajectories together or apart, and the sign of which kind of forcing depends on the sign of $\lambda^G_{1,2}$. Therefore, small perturbations in forcing may cause diverging trajectories for specific forcing regimes. The complicated interpretation of $\lambda^G_{1,2}$ demonstrates that in a forced-dissipative system, the solution dependencies on the boundary conditions and parameters (here just $\Delta\hat{b}^T$) are critical to the interpretation of SC ensemble behavior.

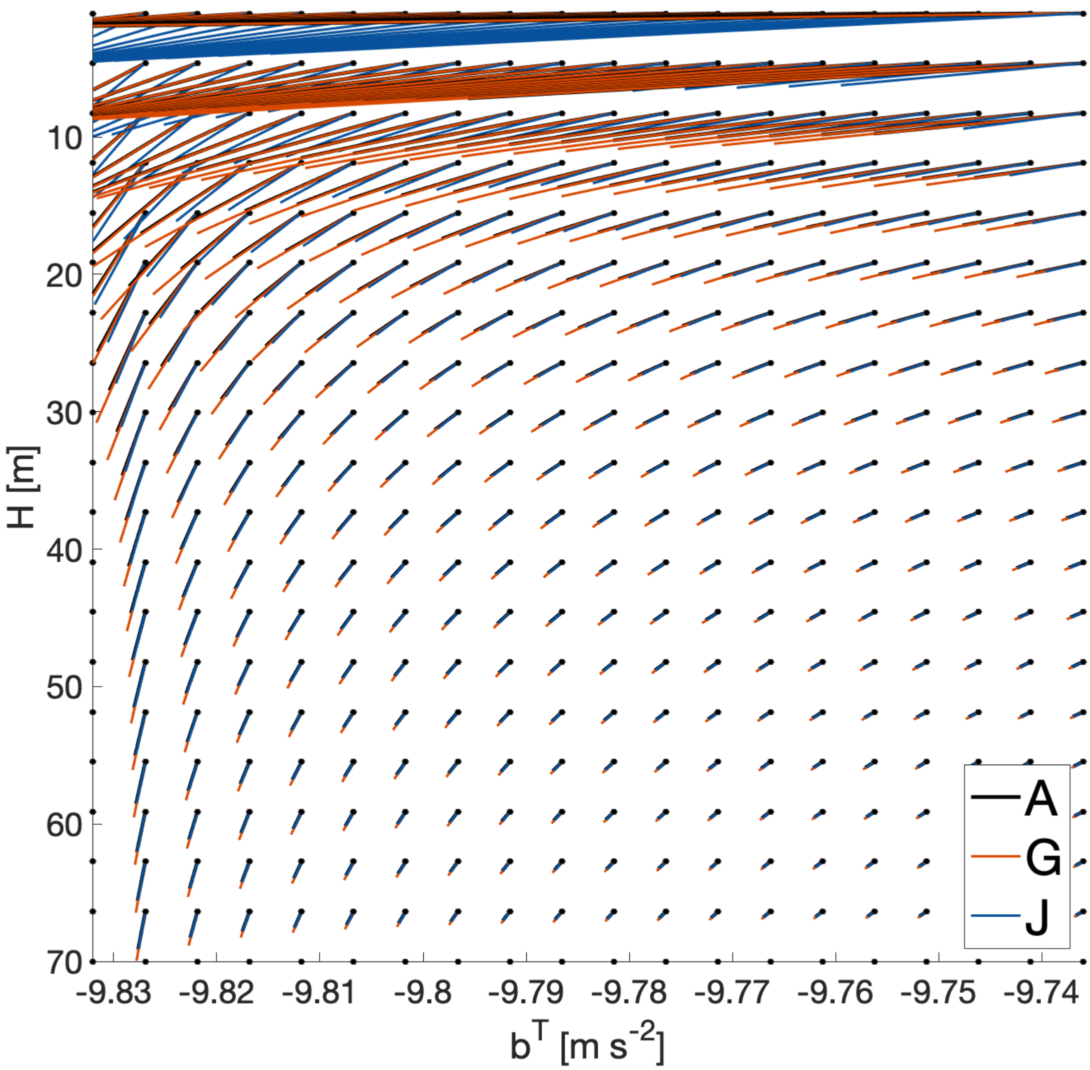


FIG. A1. A phase diagram for the Kraus-Turner (KT67) system as in Eq. (A1)-(A2) (black lines) during a case of ML deepening with $u_* = 0.013 ms^{-1}$ and $\mathcal{B}_o = 5.6\times10^{-8} m^2 s^{-3}$ (taken as the mean of the first five days of forcing in Fig. 2). For an initial condition at some point (denoted by circles), the line points to the final state $\mathbf{x}^{\mathbf{f}}$ after a single time step. Initial conditions in shallow MLDs will change more rapidly in one time step than in deeper MLDs. The linear trajectories of perturbations to state space, in terms of the Jacobian (J, blue), and to the forcing, in terms of the gain matrix (G, red), are also included.

The phase space for the deepening KT67 system (using dimensional $u_* = 0.013 ms^{-1}$ and $\mathcal{B}_o = 5.6\times10^{-8} ms^{-1}$ is demonstrated in Fig. A1) highlights the behavior of the deepening ML and sensitivity to $\Delta\hat{b}^T$ and H. The stable trajectories of small perturbations in state and forcing space are also shown. Trajectories respond to perturbations in shallow ML particularly but become less sensitive with deeper H and larger $\Delta\hat{b}^T$.

APPENDIX B

Large Eddy Simulation

Large-Eddy simulations solve the three-dimensional grid-filtered non-hydrostatic incompressible Navier-Stokes equations under the Boussinesq approximation. The wave-phase averaged equations are solved in LES to include the effects of wave-induced Stokes drift.

Subgrid momentum flux is obtained using the filtered structure function parameterization in Ducros et al. (1996). The subgrid Prandtl and Schmidt numbers are taken to be unity in the computation of subgrid heat and salinity fluxes, respectively. Further details of the numerical methods and the subgrid fluxes of the LES can be found in Pham and Sarkar (2018); VanDine et al. (2020).

The LES shown here are run on a computational domain in a rectangular box with dimensions of $192 \times 192 \times 147$ m in the zonal, meridional, and vertical directions, respectively. The horizontal grid spacing is 0.75 m while the vertical grid spacing is 0.3 m in the top 50 m and is slightly stretched in the region below. The flow is homogeneous in the horizontal directions, to arrive at the same equations as Eq. (1)-(5) after horizontal averaging, but with turbulent covariances solved for in the full 3D computation. Surface fluxes which include the wind stress, the solar and non-solar heat fluxes, and the net amount of precipitation minus evaporation as shown in Fig. 2 are applied at the top surface. The transmissive solar heat flux is parameterized using a Jerlov type I model (Paulson and Simpson 1977). A sponge region is implemented near the bottom surface to maintain constant temperature and salinity gradients in the pycnocline throughout the simulations.

APPENDIX C

Model Transience and the Ensemble Vector Timescale

The final ensemble vectors (one for SC-ST and a separate one for SC-LT) combine all SC model difference vectors $\mathbf{y_j} = \mathbf{x_j}^n - \mathbf{x_j}^{ref}$ at all depths j, with a total size determined by (the number of

SC models) × (number of depth intervals) × (length of $\mathbf{x}$). The reduced state space $\mathbf{x}$ and therefore the representation of $\mathbf{y}$ and EV, is not the full state space of all SC models. Instead, defining $\mathbf{x}$ by Eq. (16) evaluates models' ability to simulate mean and turbulent fields in relation to LES. Here the $\mathbf{x_j}^{ref}$ reference state is taken from the LES-ST or LES-LT model for the ST and LT ensemble vectors, respectively, so these are truth-informed ensemble vectors. This state space can be reduced further to focus on particular variables. For sea surface temperature, EV^{SST} is defined with $\mathbf{y}_{z=0} = \mathbf{T}^{n}_{z=0} - \mathbf{T}^{ref}_{z=0}$. Finally, the model error can be approximated as a single value through the L2 norm of the entire (dimensionless) EV.

It is also important to define the ensemble vector timescale, Δt_{EV}, which must be longer than the initial transience of each SC model, yet short enough to capture the full nonlinear response to a narrowly defined ocean state (e.g., the sampling interval of evolving surface forcing, stratification, etc). A linear EV eigenanalysis is not possible with the GOTM simulations as SC models are not initialized with each model's full state in GOTM and thus require some adjustment, particularly as higher-order schemes spin-up to statistical equilibrium. This transient behavior is evaluated by performing branch runs of each parameterization off of its own base run for a length of 24 hours, at 3 hour intervals. For example, KPP-CVMIX-ST is initialized with a state from the continuous KPP-CVMIX-ST simulation in Fig. 4c every 3 hr (as opposed to KPP-CVMIX-ST being initialized by LES as in case studies above). The L2 norms of the EVs for all branch runs highlight how initial model trajectories don't always follow the trajectory of the continuous run (Fig. C1). In other words, each SC model undergoes initial transience before it equilibrates onto its own stable trajectory (e.g. its own map $\mathcal{A}_n$). As might be expected, models with diagnostic turbulence (KPP-type and ePBL), and therefore fewer degrees of freedom and less state space reduction during restart, exhibit shorter transience than higher-order schemes (Fig. C1), with the exception of KPP-ROMS-ST that tends to deepen the ML rapidly during its transience with a long lasting imprinted effect on its EV. Models that relax back to near zero EV have initial transience that doesn't affect the ultimate trajectory. For higher-order models that do retain a perturbed state after transience, we note that this value is an order of EV magnitude less than what is shown for intermodel comparison, reinforcing that small perturbations in the model state are not the largest sources of error in these examples. However, these initial transients can constitute the fastest eigenvalues, hence the finite-time aspect of the EV method is needed.

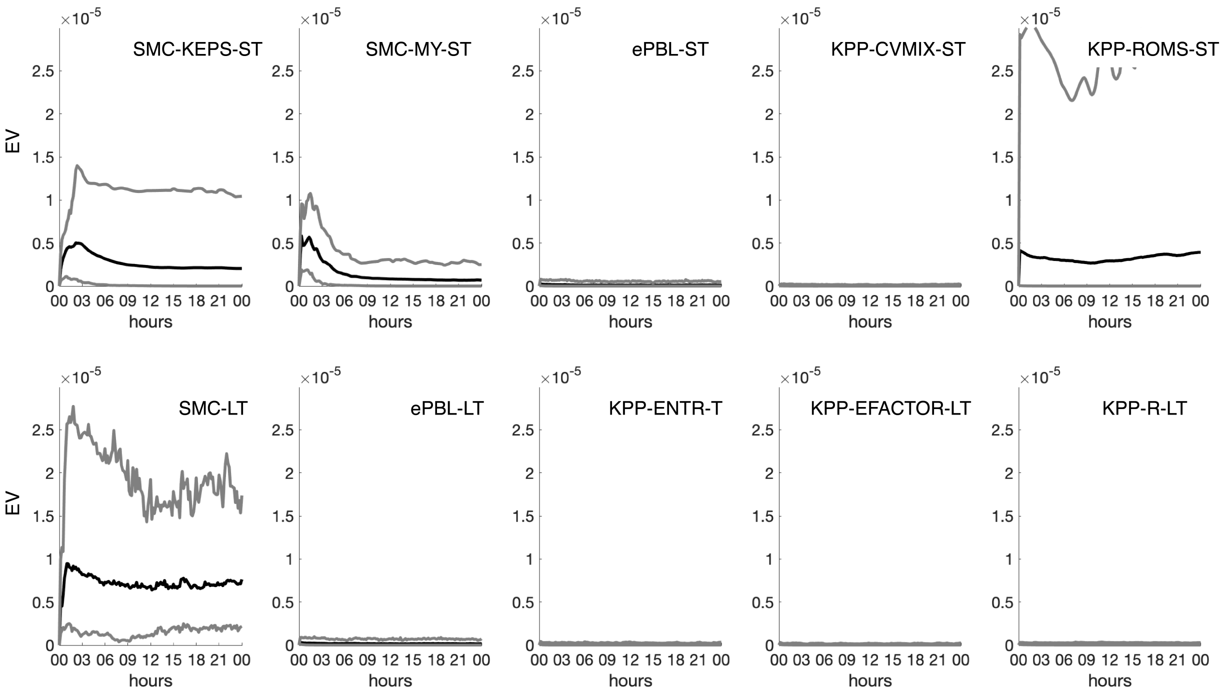


FIG. C1. The average L2 norm of the non-dimensional ensemble vector (EV) for reduced-restart branch runs initialized with each model's own base run (identical to how each SC model will be run with a chosen truth-informed or ensemble-mean base run in comparisons). Branch runs were initialized every 3 hours and run for 24 hours. Black lines are the mean for each model and the grey lines are the 5% and 95% percentiles. These represent the inherent transience in models as they reach statistical equilibrium from a set of initial conditions.

Bred vector calculations are traditionally performed in unforced systems with a chaotic divergence of nearby initial conditions and therefore require a breeding method. In this method, growing perturbations over a bred vector interval are rescaled to the initial perturbation repeatedly to find the fastest growing perturbation. The repeated rescaling identifies the direction of the largest error growth of the system and has been shown to correspond to a system's leading Lyapunov vector which can be constructed directly from the tangent linearization without repeated simulations (Kalnay et al. 2002). The forced nature of an SC model is somewhat incompatible with a breeding method because of the dual dependency on not only $\mathbf{x}$ (as in traditional breeding), but also on $\mathbf{F}$. This forcing, and the differential state-dependence sensitivity to forcing, add energy to the breeding cycle that differentiates it from traditional breeding approaches. For a given EV time interval, the growth of SC model error is explored by adding the average SC model error at the end of an ensemble vector timescale to the initial profile and re-running the simulation under the same forcing. These repeated simulations would not identify the direction of the fastest growing errors of the SC model (like traditional breeding), but instead the direction of the fastest growing errors between SC models *under a specific forcing condition*. Repeated simulations (10) tested for a six hour interval during rapid ML deepening shows that the direction of SC model spread does not evolve upon iteration. This suggests that the forced, dissipative SC model systems rapidly settle after transients onto a stable trajectory during the initial Δt_{EV} interval (Fig. C1), thereby capturing the true direction of model spread under a set of forcing and initial conditions. This is consistent with the behavior of model transience, both supporting that the largest errors between models are a result of the SC model formulation and reduced-restart issues and not a chaotic sensitivity to small perturbations in state space. Thus, the Ensemble Vector method (no rescaling and restart needed) and the Bred Vector method (rescaling and restart to identify chaotic divergence) are importantly distinct, while both seek to use finite-time simulations using the actual numerical model system to understand its nonlinear behavior.

The stabilizing tendencies in Fig. C1 also demonstrate how the different models, and therefore the EV approach, integrate statistical noise. The slow degrees of freedom within the system defined by Eq. (16) persist after the collapse of fast, transient eigenmodes. A prognostic higher moment order equation with eight or more equations to constrain turbulence would probably exhibit initially chaotic behavior (though not shown here formally, but implied by divergent transience in second-moment closure models), but as the system reaches statistical equilibrium, the mixed layer system defined by $\mathbf{x}$ in Eq. (16) represents a diffusive system captured in Fig. A1. For example, the long term behavior of k-ε does not improve by including ε into the initial conditions, suggesting its impact on the initial eigenvector (i.e. initial transience) but not the trajectory of the EV over longer timescales. Therefore it is assumed for this analysis that the transients don't importantly affect the model trajectory and that the reduced state space in Eq. (16) provides a representative subspace of the ocean surface boundary layer system suitable for initialization from restarts, LES "truth", or SC ensemble means. It is also interesting to note that the timescale of transience depends on the SC model time step, where longer time steps result in longer relaxations –this dependence reflects the fact that many of the initial transients stem from numerical spin-up techniques that depend on timestep rather than representing physical processes which are agnostic to numerical implementation. For implementation, SC models were branched off of LES every 3 hrs using a $\Delta t = 60s$, and a 6 hr window was chosen as the EV timescale. Choosing a timescale of 4 and 8 hours did not significantly alter the interpretation of the results.

References

Belcher, S. E., and Coauthors, 2012: A global perspective on langmuir turbulence in the ocean surface boundary layer. *Geophysical Research Letters*, **39 (18)**, https://doi.org/10.1029/2012GL052932, URL http://doi.wiley.com/10.1029/2012GL052932.

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

Craik, A., 1982: The generalized lagrangian-mean equations and hydrodynamic stability. *Journal of Fluid Mechanics*, **125**, 27–35, https://doi.org/10.1017/S002211208200322X.

Crawford, G. B., and W. G. Large, 1996: A numerical investigation of resonant inertial response of the ocean to wind forcing. *Journal of Physical Oceanography*, **26 (6)**, 873–891, https://doi.org/10.1175/1520-0485(1996)026⟨0873:ANIORI⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/10.1175/1520-0485(1996)026{\%}3C0873:ANIORI{\%}3E2.0.CO;2.

D'Asaro, E. A., J. Thomson, A. Y. Shcherbina, R. R. Harcourt, M. F. Cronin, M. A. Hemer, and B. Fox-Kemper, 2014: Quantifying upper ocean turbulence driven by surface waves. *Geophysical Research Letters*, **41 (1)**, 102–107, https://doi.org/10.1002/2013GL058193, URL http://doi.wiley.com/10.1002/2013GL058193.

Donelan, M. A., J. Hamilton, and W. H. Hui, 1985: Directional spectra of wind-generated waves. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **315**, 509–562, https://doi.org/10.1098/rsta.1979.0079.

Ducros, F., P. Comte, and M. Lesieur, 1996: Large-eddy simulation of transition to turbulence in a boundary layer developing spatially over a flat plate. *Journal of Fluid Mechanics*, **326**, 1–36.

Fan, Y., and Coauthors, 2020: The effect of langmuir turbulence under complex real oceanic and meteorological forcing. *Ocean Modelling*, **149**, 101 601.

Fox-Kemper, B., L. Johnson, and F. Qiao, 2021a: *Ocean Mixing*, chap. Ocean Near-Surface Layers. Elsevier, URL http://www.geo.brown.edu/research/Fox-Kemper/pubs/pdfs/Fox-KemperJohnson21.pdf, in press.

Fox-Kemper, B., and Coauthors, 2021b: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Ocean, Cryosphere and Sea Level Change. Cambridge University Press, URL https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_09.pdf, in press.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, **113 (D6)**, https://doi.org/10.1029/2007JD008972.

Hall, G., and B. Fox-Kemper, 2021: Regional mixed layer depth as a climate diagnostic and emergent constraint. In preparation.

Harcourt, R. R., 2013: A second-moment closure model of langmuir turbulence. *Journal of Physical Oceanography*, **43 (4)**, 673–697.

Holm, D. D., 1996: The ideal Craik-Leibovich equations. *Physica D: Nonlinear Phenomena*, **98 (2-4)**, 415–441, https://doi.org/10.1016/0167-2789(96)00105-4.

Holm, D. D., and R. Hu, 2021: Stochastic effects of waves on currents in the ocean mixed layer. *Journal of Mathematical Physics*, **62 (7)**, 073 102, https://doi.org/10.1063/5.0045010.

Jaeger, G. S., J. MacKinnon, A. Lucas, E. Shroyer, J. Nash, A. Tandon, J. Farrar, and A. Mahadevan, 2020: How spice is stirred in the bay of bengal. *Journal of Physical Oceanography*, **50 (9)**, 2669–2688, https://doi.org/10.1175/JPO-D-19-0077.1.

Johnson, L., and B. Fox-Kemper, 2023: Example scripts, notes, and data for 'a dynamical systems approach to mixed layer model comparison. *Brown University Open Data Collection. Brown Digital Repository. Brown University Library.*, https://doi.org/https://doi.org/10.26300/c277-dz74.

Johnson, L., C. M. Lee, and E. A. D'Asaro, 2016: Global estimates of lateral springtime restratification. *Journal of Physical Oceanography*, **46 (5)**, 1555–1573, https://doi.org/10.1175/JPO-D-15-0163.1.

Kalnay, E., M. Corazza, and M. Cai, 2002: Are bred vectors the same as lyapunov vectors? *EGS general assembly conference abstracts*, 6820.

Kraus, E. B., and J. S. Turner, 1967: A one-dimensional model of the seasonal thermocline ii. the general theory and its consequences. *Tellus*, **19 (1)**, 98–106, https://doi.org/10.3402/tellusa.v19i1.9753, URL https://doi.org/10.3402/tellusa.v19i1.9753https://www.tandfonline.com/doi/full/10.3402/tellusa.v19i1.9753.

Large, W. G., J. C. McWilliams, and S. C. Doney, 1994: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, **32 (4)**, 363, https://doi.org/10.1029/94RG01872, URL https://doi.org/10.1029/94RG01872http://doi.wiley.com/10.1029/94RG01872.

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

Leibovich, S., 1980: On wave-current interaction theories of langmuir circulations. *Journal of Fluid Mechanics*, **99 (4)**, 715–724, https://doi.org/10.1017/S0022112080000857.

Li, Q., and B. Fox-Kemper, 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, https://doi.org/10.1175/JPO-D-17-0085.1, URL https://doi.org/10.1175/JPO-D-17-0085.1.

Li, Q., A. Webb, B. Fox-Kemper, A. Craig, G. Danabasoglu, W. G. Large, and M. Vertenstein, 2016: Langmuir mixing effects on global climate: Wavewatch iii in cesm. *Ocean Modelling*, **103**, 145–160.

Li, Q., and Coauthors, 2019: Comparing ocean surface boundary vertical mixing schemes including langmuir turbulence. *Journal of Advances in Modeling Earth Systems*, **11 (11)**, 3545–3592, https://doi.org/10.1029/2019MS001810.

Liang, J.-H., J. Yuan, X. Wan, J. Liu, B. Liu, H. Jang, and M. Tyagi, 2022: Exploring the use of machine learning to parameterize vertical mixing in the ocean surface boundary layer. *Ocean Modelling*, **176**, 102 059.

Lucas, A. J., and Coauthors, 2016: Adrift upon a salinity-stratified sea: A view of upper-ocean processes in the bay of bengal during the southwest monsoon. *Oceanography*, **29 (2)**, 134–145.

McWilliams, J. C., E. Huckle, and A. F. Shchepetkin, 2009: Buoyancy effects in a stratified ekman layer. *Journal of Physical Oceanography*, **39 (10)**, 2581–2599.

McWilliams, J. C., P. P. Sullivan, and C.-H. Moeng, 1997: Langmuir turbulence in the ocean. *J. Fluid Mech.*, **334**, 1–30.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, **20 (4)**, 851, https://doi.org/10.1029/RG020i004p00851, URL http://doi.wiley.com/10.1029/RG020i004p00851.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ecmwf ensemble prediction system: Methodology and validation. *Quarterly journal of the royal meteorological society*, **122 (529)**, 73–119.

Monin, A., and A. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, **151 (163)**, e187.

Norwood, A., E. Kalnay, K. Ide, S.-C. Yang, and C. Wolfe, 2013: Lyapunov, singular and bred vectors in a multi-scale system: an empirical exploration of vectors related to instabilities. *Journal of Physics A: Mathematical and Theoretical*, **46 (25)**, 254 021.

Paulson, C. A., and J. J. Simpson, 1977: Irradiance measurements in the upper ocean. *Journal of Physical Oceanography*, **7 (6)**, 952–956.

Pham, H. T., and S. Sarkar, 2018: Ageostrophic secondary circulation at a submesoscale front and the formation of gravity currents. *Journal of Physical Oceanography*, **48 (10)**, 2507–2529.

Pham, H. T., S. Sarkar, L. Johnson, B. Fox-Kemper, P. P. Sullivan, and Q. Li, 2023: Multi-scale temporal variability of turbulent mixing during a Monsoon Intraseasonal Oscillation in the Bay of Bengal: An LES study. *Journal of Geophysical Research: Oceans*, **128**, e2022JC018 959, https://doi.org/10.1029/2022JC018959.

Pinkel, R., M. Buijsman, and J. M. Klymak, 2012: Breaking topographic lee waves in a tidal channel in luzon strait. *Oceanography*, **25 (2)**, 160–165.

Pollard, R. T., P. B. Rhines, and R. O. R. Y. Thompson, 1973: The deepening of the wind-mixed layer. *Geophysical Fluid Dynamics*, **4 (4)**, 381–404, https://doi.org/10.1080/03091927208236105, URL https://doi.org/10.1080/03091927208236105.

Price, J. F., R. A. Weller, and R. Pinkel, 1986: Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing. *Journal of Geophysical Research*, **91 (C7)**, 8411, https://doi.org/10.1029/JC091iC07p08411, URL https://doi.org/10.1029/JC091iC07p08411http://doi.wiley.com/10.1029/JC091iC07p08411.

Qiao, F., Y. Yuan, J. Deng, D. Dai, and Z. Song, 2016: Wave–turbulence interaction-induced vertical mixing and its effects in ocean and climate models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374 (2065)**, 20150 201, https://doi.org/10.1098/rsta.2015.0201.

Rabe, T. J., T. Kukulka, I. Ginis, T. Hara, B. G. Reichl, E. A. D'Asaro, R. R. Harcourt, and P. P. Sullivan, 2015: Langmuir turbulence under hurricane gustav (2008). *Journal of Physical Oceanography*, **45 (3)**, 657–677, https://doi.org/10.1175/JPO-D-14-0030.1.

Reichl, B. G., A. Adcroft, S. M. Griffies, and R. Hallberg, 2022: A potential energy analysis of ocean surface mixed layers. *JGR Oceans*, submitted.

Reichl, B. G., and R. Hallberg, 2018: A simplified energetics based planetary boundary layer (epbl) approach for ocean climate simulations. *Ocean Modelling*, **132**, 112–129, https://doi.org/10.1016/j.ocemod.2018.10.004, URL https://linkinghub.elsevier.com/retrieve/pii/S1463500318301069.

Reichl, B. G., and Q. Li, 2019: A parameterization with a constrained potential energy conversion rate of vertical mixing due to langmuir turbulence. *Journal of Physical Oceanography*, **49 (11)**, 2935–2959, https://doi.org/10.1175/JPO-D-18-0258.1, URL https://journals.ametsoc.org/view/journals/phoc/49/11/jpo-d-18-0258.1.xml.

Reichl, B. G., D. Wang, T. Hara, I. Ginis, and T. Kukulka, 2016: Langmuir turbulence parameterization in tropical cyclone conditions. *Journal of Physical Oceanography*, **46 (3)**, 863–886.

Rodi, W., 1987: Examples of calculation methods for flow and mixing in stratified fluids. *Journal of Geophysical Research*, **92 (C5)**, 5305, https://doi.org/10.1029/JC092iC05p05305, URL http://doi.wiley.com/10.1029/JC092iC05p05305.

Shroyer, E., and Coauthors, 2021: Bay of bengal intraseasonal oscillations and the 2018 monsoon onset. *Bulletin of the American Meteorological Society*, **102 (10)**, E1936–E1951.

Skyllingstad, E. D., W. D. Smyth, and G. B. Crawford, 2000: Resonant wind-driven mixing in the ocean boundary layer. *Journal of Physical Oceanography*, **30 (8)**, 1866–1890, https://doi.org/10.1175/1520-0485(2000)030⟨1866:RWDMIT⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/10.1175/1520-0485(2000)030{\%}3C1866:RWDMIT{\%}3E2.0.CO;2.

Souza, A. N., and Coauthors, 2020: Uncertainty quantification of ocean parameterizations: Application to the k-profile-parameterization for penetrative convection. *Journal of Advances in Modeling Earth Systems*, **12 (12)**, e2020MS002 108.

Sullivan, P. P., J. C. McWILLIAMS, and W. K. Melville, 2007: Surface gravity wave effects in the oceanic boundary layer: Large-eddy simulation with vortex force and stochastic breakers. *Journal of Fluid Mechanics*, **593**, 405–452.

Suzuki, N., and B. Fox-Kemper, 2016: Understanding Stokes forces in the wave-averaged equations. *Journal of Geophysical Research–Oceans*, **121**, 1–18, https://doi.org/10.1002/2015JC011566, URL http://dx.doi.org/10.1002/2015JC011566.

Teixeira, M., and S. Belcher, 2002: On the distortion of turbulence by a progressive surface wave. *Journal of Fluid Mechanics*, **458**, 229–267, https://doi.org/10.1017/S0022112002007838.

Tennekes, H., and J. L. Lumley, 2018: *A first course in turbulence*. MIT press.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at nmc: The generation of perturbations. *Bulletin of the american meteorological society*, **74 (12)**, 2317–2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at ncep and the breeding method. *Monthly Weather Review*, **125 (12)**, 3297–3319, https://doi.org/10.1175/1520-0493(1997)

125⟨3297:EFANAT⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/10.1175/1520-0493(1997)125{\%}3C3297:EFANAT{\%}3E2.0.CO;2.

Umlauf, L., and H. Burchard, 2003: A generic length-scale equation for geophysical turbulence models. *Journal of Marine Research*, **61 (2)**, 235–265, https://doi.org/10.1357/002224003322005087, URL http://www.ingentaselect.com/rpsv/cgi-bin/cgi?ini=xref{\&}body=linker{\&}reqdoi=10.1357/002224003322005087.

Umlauf, L., and H. Burchard, 2005: Second-order turbulence closure models for geophysical boundary layers. a review of recent work. *Continental Shelf Research*, **25 (7-8)**, 795–827, https://doi.org/10.1016/j.csr.2004.08.004, URL https://linkinghub.elsevier.com/retrieve/pii/S0278434304003152.

Van Roekel, L., and Coauthors, 2018: The kpp boundary layer scheme for the ocean: Revisiting its formulation and benchmarking one-dimensional simulations relative to les. *Journal of Advances in Modeling Earth Systems*, **10 (11)**, 2647–2685, https://doi.org/10.1029/2018MS001336, arXiv: 1710.02558v2.

VanDine, A., H. T. Pham, and S. Sarkar, 2020: Investigation of les models for a stratified shear layer. *Computers & Fluids*, **198**, 104 405.

Whitt, D., D. Cherian, R. Holmes, S. Bachman, R.-C. Lien, W. Large, and J. Moum, 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.

Wolfe, C. L., and R. M. Samelson, 2007: An efficient method for recovering lyapunov vectors from singular vectors. *Tellus A: Dynamic Meteorology and Oceanography*, **59 (3)**, 355–366.

Wyngaard, J., 2010: Turbulence in the atmosphere.