

PhD thesis

# Boundary Layer Turbulence

Exploring Strategies to Improve Mixing Parameterizations in Ocean Models

Marta Agnieszka Mrozowska

Advisors: Prof. Markus Jochum and Associate Prof. James Avery

Submitted: November 30, 2024

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

Marta Agnieszka Mrozowska: *“Boundary Layer Turbulence — Exploring Strategies to Improve Mixing Parameterizations in Ocean Models”*

This PhD project has been carried out at

**NIELS BOHR INSTITUTE, UNIVERSITY OF COPENHAGEN**

under the supervision of

**PROF. MARKUS JOCHUM**

and

**ASSOCIATE PROF. JAMES AVERY**

(Niels Bohr Institute, University of Copenhagen)



## ABSTRACT — ENGLISH

The oceanic boundary layer is the barrier connecting the ocean interior to other climate components. Simultaneously, it is the region in the sea with the highest turbulence density. In global ocean simulations, turbulent fluxes are too small to resolve and must therefore be parameterized. Vertical mixing schemes are thus integral for accurate representation of upper ocean dynamics in climate models. However, persistent surface biases continue to limit the accuracy of numerical climate simulations. In this thesis, two strategies to improve oceanic vertical mixing schemes are explored. The first study investigates whether a range of parameterizations can reproduce the observed near-inertial wave-induced mixing at two sites in the Tropical Atlantic. Shipboard turbulence observations are compared to two forced, eddy-rich ocean model simulations. The observed mixing is not reproduced in any of the models, but near-inertial wave amplitude is found to be sensitive to parameterization choice. In the second study, Bayesian optimization as a method for automated tuning of ocean models is proposed. The Turbulent Kinetic Energy (TKE) scheme is tuned to minimize mixed layer depth (MLD) biases in the ocean model Veros. The default TKE parameter values fall within the parameter space region for which MLD bias is minimized.

## ABSTRACT — DANSK

Det oceaniske grænselag er barrieren, der forbinder havets indre med andre klimakomponenter. Samtidig er dette område i havet med den højeste turbulensstæthed. I globale havsimuleringer er turbulente fluxer for små til at blive løst direkte og skal derfor parameteriseres. Vertikale blandings-skemaer er således afgørende for en nøjagtig repræsentation af dynamikken i det øvre hav i klimamodeller. Dog vedbliver vedvarende overfladebiaser med at begrænse nøjagtigheden af numeriske klimamodeller. I denne afhandling efterforskes to strategier til forbedring af vertikale blandings-skemaer i havet. Det første studie undersøger, om en række parameteriseringer kan genskabe den observerede blanding forårsaget af nær-inertielle bølger på to lokaliteter i det Tropiske Atlanterhav. Skibsbårne turbulensobservationer sammenlignes med to forcerede, eddy-rige havmodelsimuleringer. Den observerede blanding bliver ikke genskabt i nogen af modellerne, men amplituden af de nær-inertielle bølger viser sig at være følsom over for valget af parameterisering. I det andet studie foreslås Bayesiansk optimering som en metode til automatisk justering af havmodeller. Turbulent Kinetisk Energi (TKE)-skemaet finjusteres for at minimere fejl i blandingslagsdybden (MLD) i havmodellen Veros. De standardmæssige TKE-parameterværdier falder inden for det parameterområde, hvor MLD-bias minimeres.

“ *Women waste what would otherwise,  
in sinister reflections of universes  
or worrisome projections of worlds forgotten,  
be called precious time.  
To think it spent, to think it lost  
to wonder if it was ever there at all;  
the woes of a worried woman.*

*Men, we exhibit macabre maneuvers,  
fishing, trawling, dragging waters.  
No more to that shallow beast.*

”

—J. M. Søndermølle, *sir mister esq.*

## ACKNOWLEDGMENTS

The title image of this thesis is the eddy-rich mixed layer depth field from the Cycle 2 Hackathon of the nextGEMS project. It was an honor and a pleasure to make a contribution to the joint effort of this huge community of researchers, with 26 partners all across Europe. I have learned immensely much from the brilliant, creative people involved in the project. The members of the Storms & Oceans work group have served as continuous inspiration and support throughout the wild journey that the process of writing this thesis has been. I'd like to in particular thank the co-authors of the first article, including Swantje Bastin, Rebecca Hummels, Marcus Dengler and Tim Fischer, as well as Johann Jungclauss and Sergey Danilov, for the support throughout the process.

One of the most amazing experiences of my time as a PhD was the possibility to join the M189 research cruise lead by Marcus Dengler. This experience will forever stay in my memory, and I hope it results in friendships lasting for a lifetime. I've been fortunate to meet many friends along the way, inspiring early career scientists that make me excited for the future of climate science. I was also fortunate to start my PhD alongside a group of other Doctorate students at the Physics of Ice, Climate and Earth section. Always knowing that we could share the ups and the downs, the complaints and the joys, has brought me peace on days when the going got tough. Having been at PICE since my Bachelor, I shared many worries and woes with my colleagues during lunches and coffee breaks.

I want to thank the TeamOcean group and the co-authors of my second paper, namely Aster Stoustrup and Carl-Johannes Johnsen for all the work they have put in to make the manuscript possible. I also want to thank Dion Häfner for his work on Veros, JAX and this beautiful document class. Dion has been a continuous inspiration on how to do science not only well, but also beautifully.

I would like to thank my host at the Hereon institute, Jeff Carpenter, for providing the space for me to work on the VerOpt project and for his advice.

I shared many laughs and great moments with my two co-supervisors, Markus Jochum and James Avery, and Roman Nuterman, who could always provide a solution to my technical problems and advice on how to navigate the PhD process.

James has been an adventuring partner and an inspiration on how to tackle the academic lifestyle. I hope to become a teacher like him one day. Markus has taught me to love the ocean. He has an incredible skill in steering me in the right direction, while letting me come into my own as a physicist and an oceanographer. I want to thank them both for enabling me to complete this work in a way I am proud of.

Finally, I want to thank my family (and friends that feel like family), Kiss, Oliver, Aragorn, my mom Agnieszka, my dad Marcin, and my fantastic siblings, Ala and Mateusz. Thank you for your patience in this time, and I promise to see you more often now.

One person deserves a co-authorship of this thesis due to the sheer amount of support he provided, the love, and the complete and unbreaking belief in me in the moments when I lacked it. This work could never be completed without my partner of over a decade, Jonas Søndermølle. Thank you for everything.

# Contents

- o Introduction 1
- 1 Navigating the Turbulent Seas 3
  - 1.1 Ocean Mixing and Climate 3
  - 1.2 Ocean Turbulence Theory 7
  - 1.3 Turbulence Closure Schemes 11
- 2 Mixing in Eddy-Rich Models 20
  - 2.1 Near-Inertial Waves 20
  - 2.2 Article I – Using NIW Observations to Assess Mixed Layer Parameterizations: A Case Study in the Tropical Atlantic 22
- 3 Automated tuning 42
  - 3.1 Gaussian Emulators 42
  - 3.2 Article II – Fast and Efficient: Bayesian Optimization with GPU Acceleration for Ocean Models 47
- 4 Uncharted Depths 75
  - 4.1 MLD for Model Validation 75
  - 4.2 Extended Optimization 76
  - 4.3 Turbulence Closures of Tomorrow 76
- References 80

# Introduction



Ocean flows are turbulent, and the chaotic motions span over length scales from hundreds of kilometers (mesoscale) to millimeters (microscale). Observations indicate that ocean turbulence is integral to nearly all fundamental physical processes affecting weather and climate. Upper ocean turbulence modulates surface fluxes of geochemical tracers and heat, and sets global water mass properties. Internal turbulence drives the meridional overturning circulation, which impacts climate on millennial time scales.

Reynolds decomposition provides a method to theoretically describe how turbulence affects the mean flows and physical state of the sea. The Kolmogorov spectrum describes the energy cascade, where it induces increasingly finer eddy structures until it is dissipated into heat. The length scale at which this irreversible process occurs varies from centimeter to micrometer scales. Global ocean simulations at these scales are unfeasible. Turbulent fluxes must therefore be approximated by vertical mixing schemes.

Parameterizing boundary layer turbulence has been named as one of the most prominent challenges in ocean modeling Fox-Kemper et al. (2019). In this PhD thesis, various strategies for coping with this challenge are explored.

The following chapters provide an overview of ocean turbulence in observations, theory and numerical models. The role of upper ocean mixing in climate is explored, followed by insights into the limits of theoretical understanding and a summary of how turbulence is represented in Ocean General Circulation Models (OGCMs). Various vertical mixing schemes are introduced, and the methods for assessing their accuracy summarized.





## CHAPTER CONTENTS

- 1 Navigating the Turbulent Seas 3
  - 1.1 Ocean Mixing and Climate 3
    - 1.1.1 Surface Mixed Layer 4
  - 1.2 Ocean Turbulence Theory 7
    - 1.2.1 Notation and Governing Equations 7
    - 1.2.2 Energy Cascade 9
  - 1.3 Turbulence Closure Schemes 11
    - 1.3.1 Turbulent Kinetic Energy Closure 12
    - 1.3.2 The Choice of TKE Parameters 14
    - 1.3.3 K-profile Parameterization 15
    - 1.3.4 Vertical Mixing Schemes in Contemporary Ocean Models 16

# Navigating the Turbulent Seas



## 1.1 OCEAN MIXING AND CLIMATE

The vertical structure of the ocean can be broken down into two major parts: the surface, which is characterized by strong stratification, vigorous currents and interactions with the overlying atmosphere, and the slow, weakly stratified interior. *Diapycnal mixing* is integral to both, and plays a key role in processes which connect them. The following section summarizes the ways diapycnal mixing shapes the climate, with the focus on upper ocean turbulence. It was inspired by the excellent book *Ocean Mixing: Drivers, Mechanisms and Impacts* by Meredith and Garabato (2021).

The ocean is mostly adiabatic, which means that in the absence of external forcing, water masses in its interior will retain their physical properties. The ocean density varies on the order of tens of kilograms per cubic meter, a small fraction of the mean density of seawater  $\rho_0 \approx 1024 \text{ kg m}^{-3}$ . Even so, the ocean is *stably stratified* in most places, with lighter water overlaying the denser. *Buoyancy frequency*  $N$  is used to quantify the strength of stratification in the ocean:

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \rho}{\partial z}, \quad (1.1)$$

BUOYANCY FREQUENCY

where  $z$  symbolizes the vertical coordinate,  $\rho$  is local density and  $g$  is the magnitude of the gravitational acceleration. Mixing across the constant density surfaces - the isopycnals - requires energy to overcome the stratification. Diapycnal mixing away from basin boundaries is mainly driven by internal wave breaking.

The water mass properties are thus attained at the surface, set by the interaction with the overlying atmosphere<sup>1</sup>. Flow velocities in the atmosphere are typically 1 – 2 orders of magnitude larger than in the ocean. Laws of fluid dynamics, however, assert that the velocities must match at the interface. This leads to the exchange of momentum at the sea surface and creates *the atmospheric and oceanic boundary layers*, where the strong vertical velocity gradient induces turbulent mixing. It is the oceanic *mixed layer* that transfers heat and momentum from the surface into the stratified layers below. This surface barrier, which can at times become thinner than 10 meters<sup>2</sup>, plays a central role in setting the global climate.

1. Sea surface boundary conditions are also locally set by sea ice and land hydrosphere.

2. Amounting to approximately 0.25% of the total average depth of the ocean.

Water is highly efficient at storing heat, as opposed to air. Because of this, “the heat capacity of the top 2.5 m of the ocean equals that of the whole column of air above it” (Garrett, 1996). In fact, over 90% of the heat excess resulting from

anthropogenic climate change has been absorbed by the ocean (Levitus et al., 2012). The rate of the surface heat uptake depends on stratification (Marshall and Zanna, 2014). On annual average, heat enters the ocean in the tropics and is released at higher latitudes, which necessarily involves heat transfer across isopycnal layers. Diapycnal mixing is therefore important to both the mean rate of ocean heat uptake, as well as the geographic distribution of surface heat fluxes which affect local weather.

Mixed layer processes impact the exchange of gases between the ocean and atmosphere. The rate of surface uptake of carbon dioxide and its sequestration into the deep ocean, where water masses can remain with no atmospheric contact for hundreds of years, impacts climate on millennial time scales. The ocean's long term carbon storage properties are the reason for its key role in theories explaining the CO<sub>2</sub> and global temperature correlation across glacial-interglacial transitions (e.g., Galbraith and Eggleston, 2017; Peacock, Lane, and Restrepo, 2006). Central to these theories is the sluggish meridional overturning circulation (MOC), which sinks the carbon-rich surface waters into the ocean interior. The MOC itself is driven partially by diapycnal mixing of abyssal waters (Kuhlbrodt et al., 2007).

The surface ocean is home to a vast ecosystem, which relies on the oxygen and nutrients supplied by diapycnal mixing. One of the major challenges of contemporary oceanography is the observed growth of oxygen minimum zones in the warming climate, which threaten marine life (Breitburg et al., 2018; Schmidtko, Stramma, and Visbeck, 2017). Proposed pathways to explain this trend include the reduction in surface ocean ventilation due to stronger stratification (Brandt et al., 2015; Oschlies et al., 2018).

Diapycnal mixing impacts climate on time scales ranging from millennial to inter-annual. Surface mixed layer sets the physical and geochemical properties of the global water masses, thereby connecting the surface ocean to the abyss. The fluxes of oxygen and carbon dioxide across the air-sea interface and isopycnal layers have direct consequences for aquatic life. Understanding and accurate modeling of diapycnal mixing in the ocean is therefore of utmost importance.

### 1.1.1 Surface Mixed Layer

Figure 1.1 shows the main properties of the oceanic boundary layer (OBL). An important distinction must be made between the *mixing layer* and the *mixed layer* (ML). The former is a region of active mixing where the turbulent kinetic dissipation rate  $\epsilon$  is elevated. In Fig. 1.1, the mixing layer has the depth of  $h_\epsilon$ , at the point where  $\log(\epsilon)$  exceeds its background value. The mixed layer forms as a consequence of mixing. The two depths may be, but often are not equal.

The mixed layer is characterized by an approximately constant vertical density profile  $\rho(z)$ . It overlays the thermocline, the stratified layer extending from ML base down to the abyss. Beyond the thermocline, the stratification weakens. Away from the atmospheric forcing, the mean currents become more sluggish, but the diapycnal mixing is still active in the presence of internal wave breaking. High rates of turbulent mixing appear again at the bottom of the ocean, driven by the friction between the fluid and the inert sediment. It is the surface mixed layer, however, which contains the most energy-dense turbulence in the ocean.

The sea surface is well mixed, because the energy supplied by the atmosphere at the boundary results in a strong vertical gradient of horizontal velocity. This shear stress is a source of turbulent kinetic energy, of which a portion is dissipated into internal energy, and a portion is used to lift the underlying denser fluid. As a result, stratification becomes unstable and the lifted fluid element gains potential energy. When it sinks, it accelerates, eventually forming the characteristic vortex-like structure of turbulence. The energy from the atmosphere is continuously supplied, so the mixing maintains the constant density profile at the sea surface.

*Shear-driven mixing* as described above is one of the two main processes which maintain the surface ML. *Buoyancy flux*, i.e. the exchange of heat and freshwater with the atmosphere, can lead to *convective mixing* at the air-sea interface. The heating of the ML base by the penetrating shortwave radiation can result in convective mixing as well.

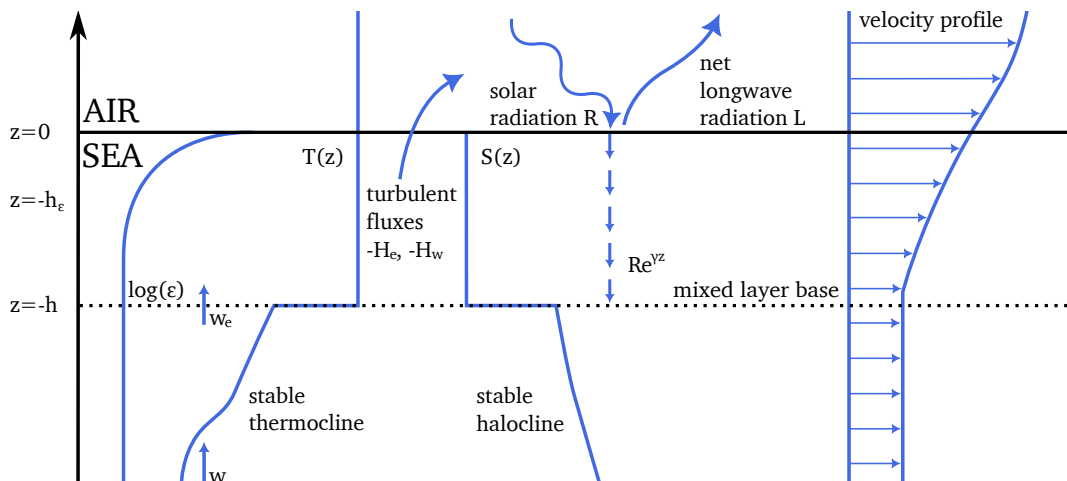


Figure 1.1: Oceanic boundary layer schematic adapted from Meredith and Garabato (2021). The profile of  $\log(\epsilon)$  sets the mixing layer depth  $h_e$ , while the temperature and salinity profiles  $T(z)$  and  $S(z)$  set the mixed layer depth  $h$ . Sensible  $H_e$  and latent  $H_w$  heat fluxes, and the net longwave radiation  $L$  are confined to the surface, while the shortwave solar radiation  $R$  penetrates the water column. Upwelling  $w$  squeezes the temperature profile in the thermocline. Entrainment fluxes  $w_e$  originate from shear instability at the ML base.

Due to its dependence on buoyancy fluxes and wind stress, mixed layer depth (MLD) varies significantly in space and time. During the day, the heat from the sun stratifies and thins the mixed layer, while nocturnal cooling destabilizes and deepens it. The annual MLD cycle is driven by surface wind strength and heat flux variations (Figure 1.2<sup>3</sup>).

The thinnest mixed layers coincide with peak heat flux into the ocean and lowest wind speeds. At 11.5°N, which is located below the Tropic of Cancer, heat flux peaks at the fall equinox. For mid- and high-latitude regions, the mixed layer is thinnest at summer solstice. At winter solstice, the rate of heat flux leaving the ocean is highest, but MLD is not yet at its deepest. Between winter solstice and spring equinox, continued heat loss drives buoyancy loss, while the strong winds provide mechanical energy to deepen the ML. The mixed layer does not respond to atmospheric forcing *instantaneously*, and thus is characterized by a degree of lag behind active mixing processes.

Considerable effort has been committed to finding the best possible criterion for computing the depth of the mixed layer in order to validate it in ocean models (e.g., de Boyer Montégut et al., 2004; Holte and Talley, 2009; Kara, Rochford, and Hurlburt, 2000). The most common is the threshold method, where MLD is defined as the depth for which the density (or temperature) changes by some quantity  $\delta\rho$  relative to some reference depth  $z_{\text{ref}}$ :

$$\text{DENSITY THRESHOLD} \quad \rho(\text{MLD}) = \rho(z_{\text{ref}}) + \delta\rho. \quad (1.2)$$

The difficulty lies in the realization that MLD is highly nonlinear, and thus not commutable (de Boyer Montégut et al., 2004). MLD computed from mean density profiles is not the same as the average of MLDs computed from individual density observations. A criterion for MLD computation must therefore be effective for the full range of MLD values in the ocean: from a few meters in the tropical summer to hundreds of meters during the polar winter. No single threshold criterion can fulfill this task. Holte and Talley (2009) developed a promising alternative, where MLD is computed by a complex algorithm, which utilizes density, temperature and salinity gradients. However, the many steps involved in the calculation make it impractical to include in ocean models, thus threshold methods are often preferred.

Despite these difficulties, MLD is an important variable used to validate ocean and climate models (e.g., Huang, Qiao, and Dai, 2014; Treguier et al., 2023) and vertical mixing parameterizations (e.g., Damerell et al., 2020; Pottapinjara and Joseph, 2022). Over the past 25 years, the Argo programme contributed over two million profiles of drifter observations (Wong et al., 2020), enabling the construction of global 1° MLD climatologies (e.g., de Boyer Montégut, 2022; Holte et al., 2017). Remote sensing of turbulence observations is more difficult, and although the global mixing dataset of Waterhouse et al. (2014) is an invaluable tool for modelers, MLD climatology provides a more complete global reference for upper ocean turbulence.

3. Data source of SST and wind speed: Prediction and Research Moored Array in the Tropical Atlantic (PIRATA, Bourlès et al., 2019); of MLD and Q: enhanced PIRATA (Foltz, Schmid, and Lumpkin, 2018).

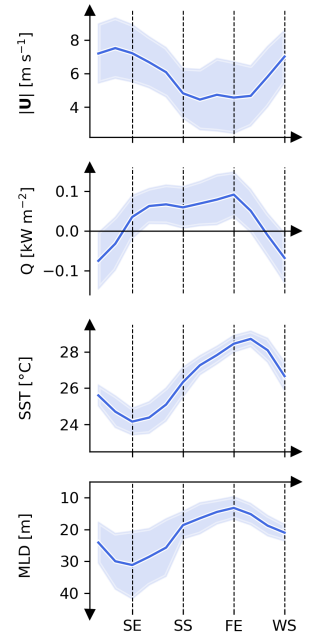


Figure 1.2: The mean seasonal cycle of wind speed ( $|U|$ ), total heat flux ( $Q$ ), sea surface temperature (SST) and mixed layer depth (MLD) at the PIRATA mooring site 11.5°N, 23°W. The vertical lines indicate the spring equinox (SE), the summer solstice (SS), the fall equinox (FE) and the winter solstice (WS).

## 1.2 OCEAN TURBULENCE THEORY

What is the basis for the current understanding and prediction of upper ocean turbulence? The answer to this question requires a deep dive into the theory of geophysical fluid dynamics (GFD). The following sections are a synthesis of numerous GFD texts, in particular Kundu et al. (2016), Thorpe (2007) and Umlauf and Burchard (2020). The aim is to address the three points:

- ▷ How is ocean turbulence described theoretically?
- ▷ How is ocean turbulence modeled and why is it difficult?
- ▷ What are the limitations of turbulence parameterizations?

### 1.2.1 Notation and Governing Equations

Continuum mechanics relies on the so-called *continuum hypothesis*, where the motion of matter in a fixed volume can be described by its macroscopic properties, such as the mean flow and density. Second order tensors are used for a complete description of internal forces acting on fluid particles. While the second order tensor is similar to a matrix, it represents physical quantities invariant under coordinate transformations, meaning the underlying physical relationships remain the same regardless of the coordinate system used. For this reason, the equations that describe fluid motion are often expressed in Einstein notation, which serves as a shorthand for summation<sup>4</sup>:

$$\text{EINSTEIN NOTATION} \quad \mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + x_3 \mathbf{e}_3 = x_i \mathbf{e}_i. \quad (1.3)$$

Here, the three basis vectors  $\mathbf{e}_1$ ,  $\mathbf{e}_2$  and  $\mathbf{e}_3$  span the three-dimensional Cartesian space, and the three scalar coordinates  $\{x_1, x_2, x_3\}$  define a point  $\mathbf{x}$  in the space. The left hand side shows an example of Einstein notation, where the repeated index  $i \in \{1, 2, 3\}$  indicates summation.

Fluid motion is governed by the Navier-Stokes (N-S) equations. Due to their complexity, analytical solutions exist only for highly idealized flows. Numerical solutions in ocean models are most often based on the Boussinesq approximation for incompressible fluids, where it is assumed that density fluctuations in the fluid are small:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} - \nu \nabla^2 \mathbf{u} = -\frac{\nabla p}{\rho_0} + \frac{\rho}{\rho_0} \mathbf{g}, \quad (1.4)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (1.5)$$

Here,  $\mathbf{u}$  is the three-dimensional velocity vector,  $p$  is pressure,  $\rho$  is density and  $\nu$  is the kinematic viscosity. The reference density  $\rho_0$  is the mean density of seawater.  $\boldsymbol{\Omega} = (0, \Omega \cos \phi, \Omega \sin \phi)$  is the Earth rotation vector and

4. Consider, for example, tensor contraction:

$$\begin{aligned} \mathbf{A} : \mathbf{B} &= A_{ij} B_{kl} (\mathbf{e}_i \otimes \mathbf{e}_j) : (\mathbf{e}_l \otimes \mathbf{e}_k) \\ &= A_{ij} B_{kl} \delta_{ik} \delta_{jl} \\ &= A_{ij} B_{ji}, \end{aligned}$$

where  $A_{ij}$  and  $B_{kl}$  are elements of the second order tensors  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, and  $\delta$  is the Kronecker delta. The expression  $A_{ij} B_{ji}$  preserves the brevity of the vector notation, but is easier to parse.

NAVIER-STOKES MOMENTUM EQUATIONS  
CONTINUITY EQUATION

$\mathbf{g} = (0, 0, -g)$  is the gravitational acceleration vector. Using Einstein notation<sup>5</sup>, the equations become:

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} + 2\epsilon_{ijk}\Omega_j u_k - \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} = -\frac{1}{\rho_0} \frac{\partial p}{\partial x_i} + \frac{\rho}{\rho_0} g_i, \quad (1.6)$$

$$\frac{\partial u_i}{\partial x_i} = 0. \quad (1.7)$$

The complete description of the physical state of sea water includes (the linear approximation of) the *equation of state*:

$$\rho(\theta, S, p) = \rho_0(-\alpha(\theta - \theta_0) + \beta(S - S_0)), \quad (1.8)$$

5. In oceanographic literature, it is a custom to use  $\mathbf{x} = (x, y, z)$  as geographical coordinates and  $\mathbf{u} = (u, v, w)$  as the velocity vector. Hence,  $(x, y, z) = (x_1, x_2, x_3)$  and  $(u, v, w) = (u_1, u_2, u_3)$  will be used interchangeably.

EQUATION OF STATE

where  $\alpha$  is the thermal expansion coefficient,  $\beta$  is the haline contraction coefficient, while  $\theta_0$  and  $S_0$  are reference temperature and salinity, respectively. The transport equations for potential temperature  $\theta$  and salinity  $S$ :

$$\frac{\partial \theta}{\partial t} + u_j \frac{\partial \theta}{\partial x_j} = \nu^\theta \frac{\partial^2 \theta}{\partial x_j \partial x_j} \quad \text{and} \quad \frac{\partial S}{\partial t} + u_j \frac{\partial S}{\partial x_j} = \nu^S \frac{\partial^2 S}{\partial x_j \partial x_j}, \quad (1.9)$$

TRANSPORT EQUATIONS OF TEMPERATURE AND SALINITY

where  $\nu^\theta$  and  $\nu^S$  are molecular viscosities, complete the seven equations (1.6)-(1.9) for sea water dynamics.

Fluid motion is by nature oscillatory. As will be discussed in the following text, it is useful to *decompose* velocity and other fluid properties into mean and fluctuating components, where the mean of the fluctuations is by definition equal to zero:

$$u = \bar{u} + u', \quad \bar{u}' = 0. \quad (1.10)$$

REYNOLDS DECOMPOSITION

The decomposition formalism is borrowed from the theory of *stochastic processes*, where  $u$  is assumed to be a random variable. Independent, repeated measurements of  $u$  under identical conditions make up an ensemble. The *expected value* of  $\bar{u}^m$  is:

$$\langle u^m(\mathbf{x}, t) \rangle = \lim_{N \rightarrow \infty} \overline{u^m(\mathbf{x}, t)} \equiv \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N u^m(\mathbf{x}, t : n), \quad (1.11)$$

EXPECTED VALUE OF A RANDOM VARIABLE ENSEMBLE

where  $u(\mathbf{x}, t : n)$  is the  $n^{\text{th}}$  set of measurements of  $u(\mathbf{x}, t)$ . In practice, repeated conditions are never identical, so the choice of the spatial and temporal span of the measurements needs to be defined based on the individual problem. For example, consider the annual mean cycles of MLD in Fig. 1.2. The climatology is based on daily measurements over 28 years, i.e. each monthly ensemble is composed of  $N \approx 840$  measurements. The ensemble means approximate

the expected values of MLD for each month of the year. Any sub-monthly departures from the mean are considered as random fluctuations.

The exponent  $m$  symbolizes the  $m^{\text{th}}$  *statistical moment* of  $u$ . As a direct consequence of Eq. (1.11), the product of statistical moments is not commutable. For  $m > 1$  and for two random variables  $u$  and  $v$ :

$$\overline{u^m} \neq \bar{u}^m \quad \text{and} \quad \overline{uv} \neq \bar{u}\bar{v}. \quad (1.12)$$

PRODUCT OF STATISTICAL MOMENTS

This simple fact turns out to be fundamentally tied to the energetics of turbulent flows.

### 1.2.2 Energy Cascade

The laws of fluid motion were first described over two centuries ago, but the formalism to study turbulence followed only 72 years later in a seminal work by Osborne Reynolds (1895). In his paper *On the Dynamical Theory of Incompressible Viscous Fluids and the Determination of the Criterion*, Reynolds introduced the method to decompose the Navier-Stokes equations into fluctuating and mean components (Eq. (1.10)). The formalism was used to derive a theoretical foundation for the empirically observed limit at which fluid flow becomes turbulent:

$$\frac{\mathcal{L}\mathcal{U}}{\nu}, \quad (1.13)$$

REYNOLDS NUMBER

where  $\mathcal{L}$  is the length scale of the flow and  $\mathcal{U}$  is the relevant flow velocity. Today, this ratio of *advective to viscous forces* in the fluid is known as the *Reynolds number*,  $Re$ . At the time, it was empirically determined that the fluid flow became unstable when  $Re \approx 2000$  or higher. Today, the generally accepted value is  $Re = \mathcal{O}(10^3)$  (Thorpe, 2007). Typical values of  $Re$  for the ocean<sup>6</sup> exceed  $10^4$ .

Reynolds' original aim to demonstrate the theoretical basis for the limit of flow stability was far surpassed in his work, because *Reynolds decomposition* became a fundament for the modern understanding of turbulence. N-S momentum equations (Eq. (1.6)) can be re-written to describe the evolution of the mean and fluctuating components of the flow separately. It turns out that *Reynolds stress*

$$\overline{u'_i u'_j} \quad (1.14)$$

6. With  $\nu \approx 10^{-6} \text{ m}^2 \text{ s}^{-1}$ , some oceanic  $Re$  values can be estimated. On the high end, there is the turbulent Gulf Stream with width  $\mathcal{L} \approx 10^5 \text{ m}$  and surface velocity  $\mathcal{U} \approx 1 \text{ m s}^{-1}$ , giving  $Re = \mathcal{O}(10^{11})$ . On the low end, e.g. in coastal waters,  $\mathcal{L} \approx 10 \text{ m}$ ,  $\mathcal{U} \approx 0.1 \text{ m s}^{-1}$  and  $Re = \mathcal{O}(10^6)$ .

REYNOLDS STRESS

appears in both. Through Reynolds stress, kinetic energy is removed from the mean flow and becomes a driver of chaotic motions of the fluid.

Turbulent kinetic energy (tke) is defined as:

$$\bar{e} = \frac{1}{2} \overline{u'_i u'_i}. \quad (1.15)$$

TURBULENT KINETIC ENERGY



The transport equation for  $\bar{e}$  can be derived by contracting the N-S equations with  $u'_i$  and averaging. The resulting budget is:

$$\begin{aligned} \frac{\partial \bar{e}}{\partial t} + \bar{u}_j \frac{\partial \bar{e}}{\partial x_j} = \frac{\partial}{\partial x_j} & \left( -\frac{1}{2} \overline{u'_i u'_i u'_j} + 2\nu \overline{u'_i s'_{ij}} - \frac{\overline{p' u'_j}}{\rho_0} \right) \\ & + \overline{u'_i u'_j} \frac{\partial \bar{u}_i}{\partial x_j} - \frac{g}{\rho_0} \overline{u'_3 \rho'} - 2\nu \overline{s'_{ij} s'_{ij}}. \end{aligned} \quad (1.16)$$

TURBULENT KINETIC ENERGY EQUATION

The first three terms on the right hand side of the equation represent the transport of tke due to turbulent stresses, viscous diffusion and turbulent pressure fluctuations, respectively. The last three terms represent the sources and sinks of tke: *shear production*, *buoyancy flux* and the *tke dissipation*  $\epsilon$ :

$$\epsilon = 2\nu \overline{s'_{ij} s'_{ij}}. \quad (1.17)$$

TURBULENT KINETIC ENERGY DISSIPATION

Buoyancy *flux* is a source of tke when the density profile is unstable, and a sink in stably stratified conditions. Shear production is a source of tke, and it is equal and opposite to the energy removed from the mean flow via Reynolds stress. The dissipation term  $\epsilon$  represents the irreversible conversion of turbulent kinetic energy into internal energy.

Typical values of  $\epsilon$  in the ocean are between  $10^{-9}$  and  $10^{-10}$  W kg $^{-1}$ , but in turbulent patches such as the ocean surface, they can reach  $10^{-1}$  W kg $^{-1}$  (Thorpe, 2007). Following dimensional analysis, the length scale at which the term makes a significant contribution to the tke budget is:

$$l_k = \left( \frac{\nu^3}{\epsilon} \right)^{1/4}. \quad (1.18)$$

KOLMOGOROV LENGTH SCALE

The order of magnitude of kinematic viscosity  $\nu$  in sea water is  $10^{-6}$  m $^2$  s $^{-1}$ , resulting in the range  $l_k \in [10^{-2}$  m,  $6 \cdot 10^{-5}$  m]. Let us consider this scale in relation to the dimensions of ocean flows. Turbulent kinetic energy is supplied at mesoscales of  $\mathcal{O}(100)$  km to geostrophic, two-dimensional turbulent motions. Largest scales supporting three-dimensional turbulence associated with diapycnal mixing are of the order of  $\mathcal{O}(100)$  m, much larger than the dissipation scales. What happens in between?

The idea of the *energy cascade* is credited to Lewis Fry Richardson (1922)<sup>7</sup>. However, it was the Russian mathematician Andrey Kolmogorov who described it mathematically, contributing a monumental step in the development of the theoretical description of turbulence (Kolmogorov, 1941/1991).

In his 1941 paper, Kolmogorov formalized the idea of the *inertial subrange* at which turbulent motions are not dissipated nor forced, but transfer turbulent kinetic energy from low to high wave numbers. In this description, the kinetic

7. In this work, Richardson paraphrased Augustus de Morgan to formulate the famous rhyme: “We realise thus that: big whirls have little whirls that feed on their velocity, and little whirls have lesser whirls and so on to viscosity - in the molecular sense.”

energy from large eddies is transferred to smaller eddies, until eventually the dissipation scale is reached where tke is irreversibly converted into heat.

Turbulent kinetic energy and its dissipation can be expressed in spectral form:

$$\bar{e} = \int_0^\infty E(k)dk \quad \text{and} \quad \epsilon = \int_0^\infty D(k)dk, \quad (1.19)$$

where the wave number  $k = 2\pi l^{-1}$  is inversely proportional to the eddy diameter  $l$ . Kolmogorov postulated that at high  $Re$  numbers, turbulence is locally isotropic. Under the assumption that  $\partial E(k)/\partial t = 0$ , this indicates that there exists a range of eddy wave numbers for which the tke dissipation is balanced by the spectral energy flux<sup>8</sup>. In this *universal equilibrium range*,  $E(k)$  is only a function of  $k$ ,  $\epsilon$  and  $\nu$ . Eventually, eddies become small enough for molecular viscosity to take effect. This threshold is defined as the Kolmogorov wave number  $k_k = 2\pi l_k^{-1}$ , with  $l_k$  from Eq. (1.18). In the inertial subrange  $k_k \ll k$ , the tke spectrum is dominated by the energy cascade from large to small eddies:

$$\text{KOLMOGOROV'S } k^{-5/3} \text{ LAW} \quad E(k) = k_0 \epsilon^{2/3} k^{-5/3}, \quad (1.20)$$

where  $k_0$  is a constant and  $k_k \ll k \ll K_{Oz}$ . The lower bound for  $k$  is the Ozmidov length scale,  $K_{Oz}$ . At lower wave numbers, tke is generated by the mean flow. The Kolmogorov spectrum is illustrated on Figure 1.3.

The results of Kolmogorov and Reynolds explain the profound difficulty of modeling large scale flows:

- ▷ Turbulence removes energy from the mean flow, and thus needs to be accounted for in the discretized momentum equations which constitute the physical basis of ocean models and
- ▷ The resolution required to simulate the full spectrum of turbulence down to the dissipation scale is unfeasible.

When the fluid motion equations are discretized, the energy spectrum is truncated. This inevitably results in error, which propagates from small to large scale flows. One of the fundamental questions of computational oceanography is: *can this error be rectified and if so, how?*

### 1.3 TURBULENCE CLOSURE SCHEMES

Reynolds averaged Navier-Stokes (RANS) equations offer valuable insights into the physical properties of turbulence. However, in the resulting formulation, variables outnumber the equations, leading to the infamous *turbulence*

#### TURBULENT KINETIC ENERGY SPECTRA

8. For homogeneous, isotropic turbulence,

$$\frac{\partial E}{\partial t} + \frac{\partial \mathcal{T}}{\partial k} = -2\nu k^2 E(k) = D(k),$$

where  $\mathcal{T}$  is the spectral energy flux.

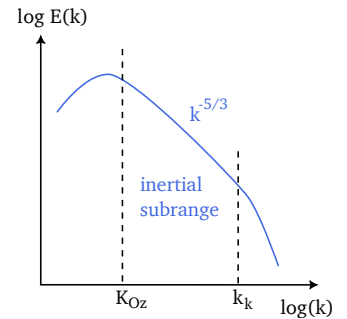


Figure 1.3: Kolmogorov spectrum (adapted from Meredith and Garabato (2021)).

*closure problem*. Numerical models based on RANS equations involve expressions that represent turbulent fluxes in terms of the properties of the mean flow, the so-called *turbulence closure schemes*.

In ocean general circulation models (OGCMs), where the horizontal grid spacing is much larger than the vertical, diapycnal mixing budget is assumed to be dominated by the vertical gradient terms<sup>9</sup>. Horizontal isopycnal mixing is treated separately by other parameterizations. Vertical turbulent closures are formulated for one-dimensional columns and approximate the turbulent fluxes of momentum and tracers across the isopycnal layers.

Two main types of vertical turbulence closures are used in oceanography: statistical and empirical<sup>10</sup>. The former estimate the mixing rate for the whole vertical column based on prognostic turbulent transport equations of second moment quantities such as  $\text{tke}$  in Eq. (1.16). The latter typically consider the boundary layer and ocean interior separately, and estimate turbulent fluxes by directly modeling specific geophysical mixing processes.

The schemes discussed in this PhD thesis are the Turbulent Kinetic Energy closure (TKE, Blanke and Delecluse, 1993; Gaspar, Grégoris, and Lefevre, 1990) and the K-profile parameterization (KPP, Large, McWilliams, and Doney, 1994). Both are widely used in contemporary ocean models and each represents one of the two approaches: KPP is an empirically based closure, while TKE is a statistical scheme. The following sections outline their structure, as well as their performance across models of varying complexity.

### 1.3.1 Turbulent Kinetic Energy Closure

The scheme proposed by Gaspar, Grégoris, and Lefevre (GGL90, 1990) is a simplified version of the Mellor and Yamada (MY82, 1982) level 2.5 model, one of the classic second moment closures (SMC), where two prognostic equations for  $\bar{e}$  and  $\bar{e}l$  are used to estimate vertical turbulent fluxes. The version of the scheme used in OGCMs in Articles I and II is further altered variant from Blanke and Delecluse (BD93, 1993). Although lengthy, the steps between the formulations of MY82 and BD93 help to explain the motivation for default parameter values in modern uses of the scheme, and the results of TKE optimization in Article II. They will be therefore summarized in this section.

Contemporary turbulence closures are based on the *turbulent viscosity* and *diffusion hypotheses*. The full Reynolds stress is parameterized by:

$$\overline{u'_i u'_j} = \frac{2}{3} \bar{e} \delta_{ij} - K_m \left( \frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right), \quad (1.21)$$

a tensor of the same form as the stress tensor for a Newtonian fluid. Here, the  $\text{tke}$  acts as turbulent pressure. Due to the boundary layer and the hydro-

9. This is known as the *boundary layer approximation*.

10. A third approach has emerged in recent years, namely machine learning (ML)-based parameterizations. The results of this work are focused on evaluating the classical approaches, but ML-based methods are briefly discussed in Section 4.3.

static approximations, only two terms of the full Reynolds stress need to be parameterized:

$$\overline{u'w'} = -K_m \frac{\partial \bar{u}}{\partial z} \quad \text{and} \quad \overline{v'w'} = -K_m \frac{\partial \bar{v}}{\partial z}. \quad (1.22)$$

The turbulent flux of tracer values such as potential temperature is approximated by analogy to Fick's law of molecular diffusion:

$$\overline{u'_i \theta'} = -K_h \frac{\partial \bar{\theta}}{\partial x_i}. \quad (1.23)$$

TURBULENT DIFFUSION  
HYPOTHESIS

*Eddy viscosity*  $K_m$  and *eddy diffusivity*  $K_h$  are then determined using a method dependent on the type of closure. Dimensional analysis suggests that they should be a product of a characteristic *length scale* and *velocity scale*. A natural choice for the velocity scale is the turbulent kinetic energy, so that:

$$K_m = c_\mu l \bar{e}^{1/2} \quad \text{and} \quad K_h = c'_\mu l \bar{e}^{1/2}, \quad (1.24)$$

SMC EDDY DIFFUSIVITY

where  $l$  is the mixing length scale, while  $c_\mu$  and  $c'_\mu$  are the stability functions of squared mean shear  $Sh^2$  and buoyancy frequency  $N^2$ .  $K_m$  and  $K_h$  are related by the turbulent Prandtl number  $P_{rt}$ , such that  $K_m = P_{rt}^{-1} K_h$ . To parameterize the diffusive fluxes,  $\bar{e}$ ,  $l$ ,  $c_\mu$  and  $c'_\mu$  have to be determined.

SMCs typically involve prognostic equations for the turbulent kinetic energy and another second moment quantity, such as the dissipation ( $k - \epsilon$ , Rodi, 1987)<sup>11</sup>, mixing length scale ( $k - kl$ , Mellor and Yamada, 1982) or mixing time scale ( $k - \omega$ , Umlauf and Burchard, 2003; Wilcox, 1988). The two prognostic equations are used to derive all other relevant quantities. The GGL90 scheme inherits the prognostic tke equation from MY82 in the following form:

11. In this thesis, I adapt the notation from GGL90 where tke is symbolized by  $\bar{e}$ . However, in SMC literature, tke is typically symbolized by  $k$ .

$$\frac{\partial \bar{e}}{\partial t} - \frac{\partial}{\partial z} \left( K_e \frac{\partial \bar{e}}{\partial z} \right) = K_m Sh^2 - K_h N^2 - c_\epsilon \frac{\bar{e}^{3/2}}{l_\epsilon}. \quad (1.25)$$

PROGNOSTIC TKE EQUATION

This is a considerable simplification of the full tke equation in Eq. (1.16). The most dramatic changes originate from the boundary layer approximation, which removes all horizontal gradient terms. Horizontal advection by the mean flow is explicitly modeled in the OGCM, and thus not included.

The two dominant turbulent transport terms are approximated with a diffusion model:

$$-\frac{\partial}{\partial z} \left( \overline{ew'} + \frac{\overline{p'w'}}{\rho_0} \right) = K_e \frac{\partial \bar{e}}{\partial z}, \quad (1.26)$$

VERTICAL TKE DIFFUSION

where  $K_e = \alpha_{tke} K_m$  is the diffusivity. Shear production and buoyancy flux are derived using Eq. (1.21) and Eq. (1.23):

$$\begin{aligned} \overline{uw} \frac{\partial \bar{u}}{\partial z} + \overline{vw} \frac{\partial \bar{v}}{\partial z} &= -K_m \left( \frac{\partial u}{\partial z} \right)^2 - K_m \left( \frac{\partial v}{\partial z} \right)^2 \\ \frac{g}{\rho_0} \overline{w' \rho'} &= -K_h \frac{g}{\rho_0} \frac{\partial \bar{\rho}}{\partial z} \end{aligned} \quad (1.27) \quad \text{SHEAR AND BUOYANCY PRODUCTION}$$

Finally,  $\epsilon$  is parameterized following the Kolmogorov spectrum in Eq. (1.19):

$$\epsilon = c_\epsilon \frac{\bar{e}^{3/2}}{l_\epsilon} \propto \bar{e}^{3/2} k. \quad (1.28) \quad \text{TKE DISSIPATION PARAMETERIZATION}$$

### 1.3.2 The Choice of TKE Parameters

In MY82, all second moment quantities  $\overline{u_i u_j}$  are parameterized in terms of model variables from Equations (1.25) - (1.28) and used to derive the stability functions  $c_\mu$  and  $c'_\mu$ . The free parameters of the model all scale the master length  $l$ . The parameter values are set so the model fulfills the logarithmic law of the wall near solid surfaces.

It is the complex formulation of  $l$  which motivated GGL90 to edit the MY82 scheme and introduce a simplified version where only two length scales are considered: the dissipation length  $l_\epsilon$  and the mixing length  $l_k$ . The latter replaces  $l$  in Equations (1.24):

$$K_m = c_k l_k P_{rt} \bar{e}^{-1/2} \quad \text{and} \quad K_h = c_k l_k \bar{e}^{-1/2}, \quad (1.29) \quad \text{GGL90 EDDY DIFFUSIVITY}$$

where  $c_k$  is now a constant model parameter. The values of  $l_k$  and  $l_\epsilon$  are computed from the integrals:

$$\begin{aligned} \frac{g}{\rho_0} \int_z^{z+l_u} [\bar{\rho}(z) - \bar{\rho}(z')] dz &= \bar{e}(z), \\ \frac{g}{\rho_0} \int_z^{z-l_d} [\bar{\rho}(z) - \bar{\rho}(z')] dz &= \bar{e}(z), \end{aligned} \quad (1.30) \quad \text{GGL90 LENGTH SCALES}$$

where  $l_\epsilon = (l_u l_d)^{1/2}$  and  $l_k = \min(l_u, l_d)$ .

The length scale computation in GGL90's version of the closure requires an iterative algorithm, which is the most expensive part of the scheme. Contrary to GGL90, BD93 test the scheme in a three-dimensional ocean model and find the GGL90 length scale computation to be a bottleneck. BD93 implement instead:

$$l_d = l_u = l_k = l_\epsilon = 2^{1/2} \bar{e}^{-1/2} N^{-1}, \quad (1.31) \quad \text{BD93 LENGTH SCALES}$$

which is an approximation of Eq. (1.30) obtained by assuming constant stratification.

In MY82 and GGL90, the constant<sup>12</sup> scaling the vertical diffusive flux of the  $\alpha_{tke} = 1$ . BD93 find that this value sometimes causes numerical instabilities due to coarse vertical resolution of the three-dimensional model, and therefore recommend  $\alpha_{tke} = 30$  for OGCMs<sup>13</sup>.

GGL90 use  $P_{rt} = 1$ , similar to MY82's  $P_{rt} = 0.8$ . BD93 use a value which is dependent on gradient Richardson number  $Ri$  instead:

$$P_{rt} = 6.6Ri, \quad P_{rt} \in [1, 10]. \quad (1.32)$$

This choice is motivated by observations of  $P_{rt}$  dependence on stratification. As a result of this relation,  $Ri_c \propto c_k c_\epsilon^{-1}$  in the BD93 model<sup>14</sup>.

The quasi-equilibrium is the state for which the shear and buoyancy production terms  $P_s + P_b$  equal the dissipation. As pointed out in Umlauf and Burchard (2003), this state can support a range of gradient Richardson numbers, of which  $Ri_c$  is the maximum. They find that it is in fact this *steady-state* gradient Richardson number  $Ri_{st}(c_\mu, c'_\mu)$  that determines the depth of the mixed layer in the column SMC models. In BD93, the stability functions are constant, which leads to a constant steady-state gradient Richardson number equal to  $Ri_c$ . Mrozowska et al. (n.d.) demonstrates that similarly to the one-dimensional SMC cases, it is the  $Ri_c$  which sets the global MLD.

Simplifications implemented by GGL90 and BD93 are motivated by computational efficiency and numerical stability considerations. The changes implemented do not tackle any of the typical SMC limitations, such as the lack of parameterization of nonlocal mixing and internal wave breaking.

### 1.3.3 K-profile Parameterization

The KPP scheme addresses these issues by including a nonlocal transport term  $\gamma_\psi$  in the closure:

$$\overline{w'\psi'} = -K_\psi \frac{\partial \bar{\psi}}{\partial z} - K_\psi \gamma_\psi, \quad (1.33)$$

where  $\overline{w'\psi'}$  represents the turbulent flux of a quantity  $\psi$ , such as momentum or temperature. Just as in TKE, turbulent diffusivity and viscosity hypotheses are applied, but the  $\gamma_\psi$  term simulates the mixing which occurs when the vertical  $\bar{\psi}$  gradient is weak. Non-zero values of  $\gamma_\psi$  are only prescribed to tracer quantities in unstable surface buoyancy forcing.

Eddy diffusivity in KPP is a product of three terms:

$$K_\psi = h \bar{w}_\psi(\sigma) G(\sigma), \quad (1.34)$$

12. And the reciprocal of the Schmidt number  $\sigma_k$ .

13. In Article II, we find that the  $\alpha_{tke}$  value does not impact model stability when 60 vertical layers are used.

BD93 PRANDTL NUMBER

14. For realistic  $Ri_c$  numbers. See Section 3.2 for the discussion about  $Ri_c > 1.5$ .

KPP VERTICAL TURBULENT FLUXES

KPP EDDY DIFFUSIVITY

where  $h$  is the depth of the ocean surface boundary layer (OSBL),  $G(\sigma)$  is the vertical polynomial shape function and  $\sigma$  is the dimensionless vertical coordinate  $\sigma \in [0, 1]$ , which extends from the ocean surface to the OSBL base.

The depth of the boundary layer is prescribed based on the value of the bulk Richardson number:

$$Ri_b(d) = \frac{(\bar{B}_r - \bar{B}(d))d}{|\bar{\mathbf{v}}_r - \bar{\mathbf{v}}(d)|^2 + v_t^2(d)}, \quad (1.35)$$

BULK RICHARDSON NUMBER

where  $\bar{B}$  is buoyancy,  $\bar{\mathbf{v}}$  is the horizontal velocity vector,  $r$  symbolises near-surface values,  $d = -z + \eta$ ,  $\eta$  is the position of the dynamic ocean free surface and  $v_t$  parameterizes the unresolved vertical shear. Boundary layer depth  $h$  is the shallowest depth layer for which  $Ri_b > Ri_c$ . Below the boundary layer, eddy diffusivity is set to a constant minimum value to simulate unresolved internal wave breaking. Gradient Richardson number  $Ri$  exceeding a critical value may increase the interior eddy diffusivity above the background value.

#### 1.3.4 Vertical Mixing Schemes in Contemporary Ocean Models

The KPP and TKE schemes summarized in previous sections are widely used in the community, but many more proposed closures exist.<sup>15</sup> The schemes undergo continued development due to the changes in OGCM structure and resolution. For example, Van Roekel et al. (2018) test the KPP scheme against Large Eddy Simulations (LES). The original parameterization for unresolved shear by Large, McWilliams, and Doney (1994) is sensitive to vertical resolution. Motivated by this, new formulation of  $v_t$  is proposed.

Additional expressions to simulate specific physical processes may also be implemented. For example, Jochum et al. (2013) extend the KPP scheme to include a parameterization of NIW-induced mixing. Li et al. (2019) test the impact of simulating the effects of Langmuir circulation across various models, and Brüggemann et al. (2024) show that the inclusion of the tidal mixing parameterization IDEMIX (Nielsen et al., 2018; Olbers and Eden, 2013) in three OGCMs can lead to a better representation of diffusivity in the ocean interior.

Despite a plethora of choices, no scheme has yet emerged as generally favored. Partially, this is due to the nature of model development. Implementing new model components is time consuming and cumbersome, thus modelers often prefer to continue working on the schemes that are already there. Modeling aims also determine which mixing scheme is used. For example, the TKE scheme can be implemented in an ocean model which favors energetic consistency (Eden, Czeschel, and Olbers, 2014).

15. For example, the Pacanowski and Philander model (PP, Pacanowski and Philander, 1981), NASA Goddard Institute for Space Studies scheme (GISS, Canuto et al., 2002), the Planetary Boundary Layer model (ePBL, Reichl and Hallberg, 2018), and many more.

The volume of literature comparing vertical mixing scheme performance is gradually expanding. The task carries with itself a number of challenges, such as computational cost of sensitivity studies in OGCMs and coupled models; the contamination of biases originating from forcing and model structure; the scheme sensitivity to resolution, and the impact of numerical mixing. Nevertheless, the persistent near-surface OGCM biases motivate the continued research into strategies to improve the schemes.

The coarse and eddy-permitting OMIP<sup>16</sup> models exhibit shallow summer and deep winter MLD biases (Treguier et al., 2023). Despite identical forcing and relatively small differences in SST across the OGCMs, MLD biases have a significantly larger spread. Increasing horizontal resolution does not lead to universal improvements. The eddying models cannot rectify the MLD biases in the Antarctic Circumpolar Current (ACC) and subpolar regions seen in their coarser counterparts. MLD is generally shallower in the high resolution models, likely due to the presence of mesoscale eddies which restratify the water column (Gaube, J. McGillicuddy Jr., and Moulin, 2019).

Approaches to testing the model sensitivity to mixing scheme choice include column, OGCM and full GCM studies. For example, Damerell et al. (2020) compare the ability of five mixing schemes, TKE and KPP among them, to reproduce point observations from the North Atlantic Ocean. ML is too deep in all models, and none of the mixing schemes performs systematically better than others. Pottapinjara and Joseph (2022) use three different mixing schemes, including KPP, to simulate the climate of the tropical Indian Ocean. They find that ML is biased deep across all the tested schemes. Once again, none of them is found to be systematically better.

Another approach is to compare the schemes in full GCMs. Gutjahr et al. (2021) investigate the four parameterizations in this context. The response to TKE and KPP is similar, with warm SST biases in the Atlantic ocean. The authors propose a pathway, where the enhanced Atlantic MOC strengthens the heat transport to high northern latitudes, resulting in SST increase.

Mixed layer depth biases persist across different models and schemes, which shows that turbulence parameterizations are still in need of improvement. Validation and sensitivity studies can be done using column models, which allow for a cleaner comparison, but cannot quantify the sensitivity of large scale circulation to the parameterization. For this purpose, OGCMs can be used. The downside is that OGCM biases result from the combined model deficiencies originating from resolution, geometry, forcing, initial conditions and other parameterizations. Isolating the vertical mixing scheme impact on the biases is not possible. On top of this, OGCMs do not include ocean interactions with the atmosphere and other climate components. Mixing schemes may induce feedbacks in coupled models, which are not captured in OGCMs. Coupled model studies are the most expensive option for analyzing climate sensitivity to vertical mixing parameterizations. The feedback pathways resulting from changing the scheme used, parameter values, or scheme

16. Ocean Model Intercomparison Project (Griffies et al., 2016). OMIP mostly includes models with variations of the TKE and KPP schemes.



components are difficult to determine due to the complexity of full GCMs and ESMs.

How can we move forward? Although OGCM and full GCM biases in physical properties of the upper ocean exhibit persistent biases, the modeling community is not short on ideas for alternative approaches. In the following chapters, two novel strategies for assessing oceanic vertical mixing schemes are explored.

## CHAPTER CONTENTS

- 2 Mixing in Eddy-Rich Models 20
  - 2.1 Near-Inertial Waves 20
  - 2.2 Article I – Using NIW Observations to Assess Mixed Layer Parameterizations: A Case Study in the Tropical Atlantic 22
    - 2.2.1 A Note on Insufficient Resolution 23

## Mixing in Eddy-Rich Models

# 2

Structural OGCM and GCM biases fundamentally stem from the inability to simulate the full energy spectrum of the atmospheric and oceanic flows. In coarse resolution models, where the horizontal grid spacing is on the order of  $1^\circ$ , geostrophic turbulence is not simulated and must be parameterized. As a result, transient climate processes that induce vertical mixing are either absent or only approximated.

A natural solution to this problem is to increase the model resolution. Unfortunately, reducing GCM grid spacing by a factor of 10 is approximately  $10^3$  times more computationally expensive (Hewitt et al., 2022). Nevertheless, the development of global eddy-rich models is one of the major current focuses of the climate community.<sup>1</sup>

Simulated atmosphere benefits significantly from increased resolution due to the better representation of deep convection, top-of-the-atmosphere radiative balance and rainfall. The hope for km-scale ESMs is that biases in upper ocean dynamics are reduced due to improved atmospheric forcing, and due to more realistic statistical representation of oceanic flows.

Even down to kilometer scale, the Kolmogorov spectrum informs that vertical turbulence schemes continue to be a necessity. Contrary to their coarse counterparts, eddy-rich ESMs have not been subject to decades of tests and tuning. One universal impact of increasing ocean model resolution is ML shallowing due to enhanced levels of baroclinic instability (e.g., Lévy et al., 2010; Oschlies, 2002; Treguier et al., 2023), suggesting that vertical mixing parameterizations in eddy-rich OGCMs may require adjusting.

### 2.1 NEAR-INERTIAL WAVES

ML deepening can occur due to entrainment of water below its base. One of the main drivers of this phenomenon are near-inertial waves (NIWs), which get their name from their frequency:

$$f = 2\Omega \sin\phi, \quad (2.1)$$

where  $\phi$  is latitude,  $\Omega$  is Earth's angular velocity and  $f$  is the Coriolis or *inertial* frequency. The near-inertial currents are excited by winds and sustained by resonance.

1. The rate of the development is impressive. In the first intercomparison of global storm-resolving models, Stevens et al. (2019) demonstrate how the 40-day DYAMOND simulations at less than 5 km resolution pass the so-called *Palmer-Turing test* (Palmer, 2016), where snapshots of the atmospheric cloud fields are indistinguishable from observations. Recently, in association with the nextGEMS project, Rackow et al. (2024) report the results of global multi-year IFS-FESOM simulations at 2.8 km resolution.

NIWs propagate out of the mixed layer with the group velocity:

$$c_{g,z} \approx -\frac{N^2 |\mathbf{k}_h|^2}{f m^3}, \quad (2.2)$$

VERTICAL GROUP VELOCITY OF  
NIWS

where  $\mathbf{k}_h$  is the horizontal wavenumber and  $m$  is the vertical wavenumber. They move equatorward, and therefore are most often observed oscillating with frequency slightly higher than the local inertial frequency. NIWs are believed to contribute significantly to internal diapycnal mixing, but most of the wind-driven NIW energy has been shown to be dissipated at the surface (Furuichi, Hibiya, and Niwa, 2008; Zhai et al., 2009).

Observations of sea surface temperature (SST) along hurricane tracks indicate that near-inertial motions induced by resonant winds contribute significantly to surface ocean cooling. Horizontal NIW currents can reach speeds of  $1 \text{ m s}^{-1}$ , inducing strong shear at the ML base and mixing via entrainment. Jochum et al. (2013) show that by parameterizing near-inertial motions in the Community Earth System Model, surface SST and rainfall biases were reduced. The interest to estimate surface NIW dissipation is therefore two-fold: 1) it allows to determine the fraction of near-inertial wind energy available for mixing in the ocean interior and 2) it provides insight into the impact of NIW-induced mixing on the upper ocean.

NIW dissipation rates have been shown to vary significantly depending on the model used. The first NIW model was proposed by Pollard and Millard (1970), where the near-inertial currents are driven by wind stress in a one-dimensional slab mixed layer. The unrealistic linear damping of this model significantly overestimates NIW amplitude (Plueddemann and Farrar, 2006). Efforts to simulate NIWs in three-dimensional models show that high temporal resolution of at least 6 hours is necessary to resolve the NIW forcing mechanism (Rimac et al., 2013). Moreover, NIW propagation into ocean interior is enhanced in the presence of eddies (Zhai, Greatbatch, and Zhao, 2005; Zhai, Greatbatch, and Eden, 2007). This implies that studying the impact of NIWs on surface ocean on global scale requires realistic, eddy-rich ocean models.

## 2.2 ARTICLE I — USING NIW OBSERVATIONS TO ASSESS MIXED LAYER PARAMETERIZATIONS: A CASE STUDY IN THE TROPICAL ATLANTIC

Article I: “Using NIW Observations to Assess Mixed Layer Parameterizations: A Case Study in the Tropical Atlantic” (Mrozowska et al., 2024) has been published in the Journal of Geophysical Research: Oceans.

The research in Article I is motivated by the question: *can local mixing be used as a metric for the performance of turbulence closures in eddy-rich OGCMs?*<sup>2</sup>

The main aims of the study are:

1. To investigate how well eddy-rich OGCMs reproduce the observed NIW characteristics and associated mixing;
2. To determine whether the simulated NIW characteristics are sensitive to the vertical mixing parameterization.

The two schemes tested are KPP and TKE, although model sensitivity to parameter values is investigated only for the latter<sup>3</sup>. The analysis is focused at two PIRATA sites: 11.5°N and 15°N.

The comparison of modeled and observed NIW characteristics is complicated by slight biases of the atmospheric forcing. Modeled winds in the near-inertial band are too weak at both sites, with significant bias of 27% at 15°N. As a result, although the near-inertial currents at 15°N are in phase with observations, they are too weak in all models.

Due to shallow ML bias at the 11.5°N site, some TKE models simulate stronger than observed near-inertial currents despite the weaker forcing. The simulated NIW amplitude is smaller than observed at the time of the microstructure observations, and thus the NIW-induced mixing is not simulated. In contrast, enhanced eddy diffusivity can be seen at the 15°N site after storm winds substantially deepen the mixing layer.

TKE models with lower  $c_k$  values generally simulate stronger near-inertial currents at the 11.5°N site, but the sensitivity is reduced at 15°N. This may be due to the generally higher stratification and shallower ML at 11.5°N.

The answer to the question posed at the beginning is thus mixed: on one hand, the models simulate the NIWs and the comparison with microstructure observations was possible. On the other, the combination of surface ocean biases and forcing biases makes it difficult to isolate the impact of the vertical mixing schemes on NIW characteristics.

In association with the nextGEMS project, a mixing database was developed which contains shipboard observations of local mixing processes, NIWs

2. The nextGEMS project begun shortly after the publication of Hummels et al. (2020), which provided a set of NIW turbulence observations for model validation. The availability of the data and the considerations outlined in the previous section amounted to NIWs becoming the research focus.

3. Based on communication with Nils Brüggemann, Johann Jungclaus and Sergey Danilov, it was hypothesized that the TKE parameter  $c_k$  was likely to have the largest impact on the upper tropical ocean. This turns out to be true, see Article II in Section 3.2

among them (Dengler et al., 2023; Fischer, Dengler, and Melzer, 2024). Investigating NIW mixing across the whole Tropical Atlantic could provide a more robust basis for testing the sensitivity of NIW characteristics to the vertical mixing schemes.

### 2.2.1 A Note on Insufficient Resolution

One alternative reason for the absence of NIW-induced mixing is not discussed in Article I. Namely, that higher modes of NIWs are not resolved in the models, and thus the modeled shear is not sufficient to overcome the stratification and induce mixing.<sup>4</sup>

As shown in Eq. (2.2), lower NIW modes propagate out of the mixed layer the quickest due to the inverse proportionality between  $m$  and  $c_{g,z}$ . It is thus the higher modes that are associated with upper ocean dissipation. The contribution to the near-inertial kinetic energy and shear of each mode can be computed via modal decomposition. However, the full profile of the model velocities in the study was not saved due to limited disk space, and therefore this approach was not possible.

The reasoning is plausible, however, considering that RAJA2022 find that in global HYCOM runs with  $1/25^\circ$  horizontal and 8 m surface vertical resolution, only the 5 lowest NIW modes are resolved. In Article I, it is evident that shear below the mixed layer is too weak in all models (see Figure 5 in the article), but the observed shear is a product of the whole unresolved internal wave field. The near-inertial kinetic energy is also generally too weak in the models (Fig. 3), although this may result from the wind forcing bias.

Moreover, the models do simulate enhanced eddy diffusivities associated with NIWs at the  $15^\circ\text{N}$  site, which suggests that some degree of mixing can be resolved at the present resolution. The Fischer, Dengler, and Melzer (2024) database provides an observational basis for a broader assessment of NIW characteristics in the models.

4. Lars Umlauf provided this idea through personal correspondence.

# JGR Oceans

## RESEARCH ARTICLE

10.1029/2024JC020985

### Key Points:

- Observations of inertial oscillations are used to evaluate the performance of two vertical mixing schemes in two high-resolution models
- Both the K-profile parameterization and the Turbulent Kinetic Energy closure underestimate the NIW-induced mixing
- Near-inertial kinetic energy biases are sensitive to the vertical mixing parameterization

### Correspondence to:

M. A. Mrozowska,  
marta.mrozowska@nbi.ku.dk

### Citation:







Mrozowska, M. A., Jochum, M., Bastin, S., Hummels, R., Koldunov, A., Dengler, M., et al. (2024). Using NIW observations to assess mixed layer parameterizations: A case study in the tropical Atlantic. *Journal of Geophysical Research: Oceans*, 129, e2024JC020985. <https://doi.org/10.1029/2024JC020985>

Received 30 JAN 2024  
Accepted 20 APR 2024

### Author Contributions:

**Conceptualization:** M. Jochum  
**Data curation:** S. Bastin, A. Koldunov, M. Dengler, T. Fischer, R. Nuterman  
**Formal analysis:** M. A. Mrozowska, R. Hummels  
**Funding acquisition:** M. Jochum  
**Investigation:** M. A. Mrozowska, M. Jochum, R. R. Hansen  
**Methodology:** M. A. Mrozowska, M. Jochum, R. Hummels  
**Project administration:** M. Jochum  
**Resources:** M. Jochum, S. Bastin, R. Hummels, A. Koldunov, M. Dengler, T. Fischer, R. Nuterman  
**Software:** M. A. Mrozowska, S. Bastin, A. Koldunov, R. Nuterman  
**Supervision:** M. Jochum  
**Visualization:** M. A. Mrozowska, R. R. Hansen  
**Writing – original draft:** M. A. Mrozowska  
**Writing – review & editing:** M. A. Mrozowska, M. Jochum, S. Bastin, R. Hummels, M. Dengler, T. Fischer, R. Nuterman

## Using NIW Observations to Assess Mixed Layer Parameterizations: A Case Study in the Tropical Atlantic

M. A. Mrozowska<sup>1</sup> , M. Jochum<sup>1</sup> , S. Bastin<sup>2</sup>, R. Hummels<sup>3</sup>, A. Koldunov<sup>4</sup> , M. Dengler<sup>3</sup> , T. Fischer<sup>3</sup> , R. Nuterman<sup>1</sup> , and R. R. Hansen<sup>1</sup>

<sup>1</sup>Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark, <sup>2</sup>Max Planck Institute for Meteorology, Hamburg, Germany, <sup>3</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany, <sup>4</sup>Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

**Abstract** Tropical sea surface temperature (SST) biases can cause atmospheric biases on global scales, hence SST needs to be represented well in climate models. A major source of uncertainties is the representation of turbulent mixing in the oceanic boundary layer, or mixed layer (ML). In the present study we focus on near-inertial wave (NIW) induced mixing. The performance of two mixing schemes, Turbulent Kinetic Energy and K-profile parameterization (KPP), is assessed at two sites (11.5°N, 23°W and 15°N, 38°W) in the tropical Atlantic. At 11.5°N, turbulence observations (eddy diffusivities, shear and stratification) are available for comparison. We find that the schemes differ in their representation of NIWs, but both under-represent the observed enhanced diffusivities below the observed ML. However, we find that the models do mix below the ML at 15°N when a storm passes nearby. The near-inertial oscillations remain below the ML for the following 10 days. Near-inertial kinetic energy (NIKE) biases in the models are not directly correlated with the wind speed, the MLD biases, or the stratification at the ML base. Instead, NIKE biases are sensitive to the vertical mixing scheme parameterization. NIKE biases are lowest when the KPP scheme is used.

**Plain Language Summary** The surface temperature of the ocean is highly dependent on the depth of the mixed layer (ML), the uppermost layer in the water column, where density, temperature and salinity are approximately constant. In climate models, the vertical mixing processes cannot be resolved, and instead they are computed with the use of vertical mixing schemes. We assess how well two of such schemes can represent the mixing induced by a specific type of ocean waves, near-inertial waves (NIWs). We compare recent observations of turbulent mixing induced by NIWs in the tropical Atlantic with numerical simulations that resolve storms. Our results show that the models are able to reproduce the observed NIWs, but underestimate their mixing and amplitude. Our analysis also shows that NIWs are a driver of mixing below the uppermost ocean layer in the models. The strength of the near-inertial currents is sensitive to the vertical mixing parameterization.

## 1. Introduction

The tropical ocean is one of the main energy sources for the tropical and extra-tropical atmosphere and even small sea surface temperature (SST) anomalies there can lead to a restructuring of the global climate (Barsugli & Sardeshmukh, 2002; Jochum & Potemra, 2008). Weather forecasting and climate prediction therefore rely on accurate representation of the tropical SST. A large source of bias could be the vertical mixing (Foltz et al., 2003).

Diapycnal mixing in the ocean has been studied extensively, as it is theorized to be a significant driver of the meridional overturning circulation (Munk & Wunsch, 1998). Vertical turbulent mixing in the ocean interior is mainly driven by internal wave breaking powered by winds and tides (Wunsch & Ferrari, 2004). Near-inertial waves (NIWs) are a dominant mode of high-frequency variability in the ocean (Fu, 1981; Garrett & Munk, 1975). They can be generated by a range of processes, such as oceanic Lee waves and wave-wave interactions (Alford et al., 2016), but their most notable and well-recorded driver is the surface wind stress (e.g., D'Asaro, 1985). Due to their ubiquity in the ocean and the potential for driving diapycnal mixing, a multitude of estimates for the fraction of the wind power input (WPI) transferred into deep ocean by NIWs exist. While the estimates vary in magnitude, a common conclusion is that the majority of the NIW energy is dissipated in the surface ocean. Zhai et al. (2009), for example, show in a model study that almost 70% of the near-inertial WPI is dissipated in the top 200 m.

It is well established that NIWs are important in deepening of the surface mixed layer (ML) in the tropical ocean. Strong vertical shear across the ML base induced by near-inertial currents leads to elevated mixing and vertical diffusive heat flux that cools and subsequently deepens the ML (e.g., Large & Crawford, 1995). Jochum et al. (2013) find that by accounting for near-inertial currents in the coarse resolution CCSM4 model, the tropical rainfall bias is reduced through deepening of the ML and changing the SST gradients. NIW-induced mixing is also the dominant process which contributes to SST cooling during and after the passing of tropical cyclones (e.g., Greatbatch, 1983; Price, 1981; Vincent et al., 2012).

The database of direct measurements of NIW-induced mixing has been expanding over the past two decades (e.g., Alford & Gregg, 2001; Brizuela et al., 2023; Hummels et al., 2020), but is still limited. Large scale studies of NIW-induced mixing are therefore based on numerical models of varying complexity (e.g., Kuwano-Yoshida et al., 2017; Raja et al., 2022; Rimac et al., 2016). In climate models, the strength of NIWs is often underestimated because their main generation mechanism is a rapid change of wind speed across the fronts of a storm system, something that is often not properly resolved in ESMs (Rimac et al., 2013). The amplitude of wind-driven NIWs is highly dependent on the depth of the mixed layer (MLD) (Pollard & Millard, 1970), and their rate of dissipation is primarily dependent on ocean stratification (Pollard, 1970).

The biases in surface ocean MLD remain an unsolved problem in ocean modeling (Huang et al., 2014). One issue is that the depth of the ML is a highly spatially and temporally variable field (de Boyer Montégut et al., 2004). While ocean models are able to simulate realistic MLD values, it remains a question whether local mixing processes are accurately represented. Our study attempts to bridge this gap by evaluating the ability of two commonly used mixing schemes, the K-profile parameterization (KPP, Large et al., 1994) and the turbulent kinetic energy closure scheme (TKE, Gaspar et al., 1990), to reproduce NIW-induced mixing in the tropical Atlantic. We use two forced, eddy ocean models: the Ocean component of the ICOSahedral and Nonhydrostatic weather and climate model (ICON-O, Korn et al., 2022) and the Finite Element/volumE Sea ice-Ocean Model (FESOM, Danilov et al., 2017; Scholz et al., 2019). Furthermore, we assess the sensitivity of the NIW-induced mixing to the TKE scheme parameterization by conducting five TKE sensitivity experiments. We present the results of this evaluation at two sites in the tropical Atlantic.

The text is structured as follows: we motivate the choice of sites (Section 2.1), specify the setup of the sensitivity studies (Section 2.2) and the data selected to evaluate the model performance (Section 2.3); the atmospheric conditions at the sites are compared to model forcing (Section 2.4). We present an overview of the MLD bias in the tropical Atlantic (Section 3.1), modeled NIW characteristics (Section 3.2), their mixing signature (Section 3.3), and the sensitivity of the near-inertial kinetic energy (NIKE) to the vertical mixing schemes (Section 3.4). Following a summary and discussion (Section 4), the main conclusions of the study are listed in Section 5.

## 2. Methods

### 2.1. Choice of Study Period and Locations

A study by Hummels et al. (2020) (hereafter referred to as HDRF20) documents a strong NIW observed during a research cruise in the tropical Atlantic. The vessel-mounted Acoustic Doppler Current Profiler (vmADCP) measurements show a clear signal of the wave at 11°N, 21°W between thirteenth and fifteenth of September 2015, as the tropical storm Ida develops (Cangialosi, 2015). The shear induced by the wave cooled the ML at the rate of 244 Wm<sup>-2</sup>. The observations provide a complete set of parameters required to investigate the NIW-induced mixing in the models, therefore we pick the years 2014/2015 as the period of the ocean simulations.

The Prediction and Research Moored Array in the Tropical Atlantic (PIRATA, Bourlès et al., 2019) offers a range of long-term measurements of ocean currents and state. In the Atlantic, the near-inertial velocity variance is highest between 7.5°S and 12.5°N (Elipot & Lumpkin, 2008). In the vicinity of this latitude band, high frequency current and co-located wind observations are available in 2014/15 at two PIRATA moorings off the equator: 11.5°N, 23°W; 15°N, 38°W. Coincidentally, Ida also passes across this second mooring, and the two periods of NIW-induced mixing are the focus of the present study. Throughout the article, they are referred to as the 11.5°N and the 15°N site, respectively.



## 2.2. Setup of Sensitivity Experiments

We use two models for the study: the Ocean component of the ICOSahedral and Nonhydrostatic weather and climate model (ICON-O, Korn et al., 2022) and the Finite Element/volumeE Sea ice-Ocean Model (FESOM, Danilov et al., 2017; Scholz et al., 2019). Both of them are forced with hourly ECMWF Reanalysis product (ERA5, Hersbach et al., 2020). FESOM spin-up is initialized from the Polar Science Hydrographic winter Climatology (PHC3.0, updated from Steele et al., 2001) in 2009. Both the base run and the sensitivity runs start from rest. ICON-O spin-up is initialized with Ocean Reanalysis System 5 (ORAS5, Zuo et al., 2019) in 2010. The simulation branches out into control and sensitivity experiments in 2014. The models run for 2 simulated years after the spin-up with a temporal output frequency of 3 h.

The models run on a triangular mesh grid, and the output is interpolated on a regular  $0.1^\circ$  grid in post-processing. There are 40 vertical levels in the top 200 m, with monotonously increasing distance between the layers: 2 m at the surface and 10.7 m at the bottom. Due to constraints on available output storage space, the simulation domain covers the whole global ocean down to the sediment, but only the upper 200 m are saved. FESOM runs on a B-grid of 13 km resolution in the tropical Atlantic and 50 km resolution in the rest of the ocean. ICON-O runs on a C-grid of approximately 10 km horizontal resolution globally. We use two widely employed vertical mixing schemes for the analysis: the turbulent kinetic energy (TKE, Gaspar et al., 1990) and the KPP (Large et al., 1994).

TKE is a parameterization where the turbulent vertical fluxes are determined from the vertical property gradients. Eddy diffusivities are dependent on TKE:

$$K_m = c_k l_k \bar{\epsilon}^{1/2}, \quad K_h = K_s = K_m / Pr_t, \quad (1)$$

where  $l_k$  is a mixing length,  $c_k$  is a constant and  $\bar{\epsilon}$  is the TKE. The momentum diffusivity,  $K_m$ , is related to the heat and salinity diffusivities,  $K_h$  and  $K_s$ , through the Prandtl number. The constant  $c_k$  scales the contribution of turbulent kinetic energy to mixing. The TKE itself is computed by a prognostic equation in the model. It can be enhanced by external forcing, such as the winds, or internal waves and currents producing shear. Aside from the TKE dissipation term, TKE sinks include vertical TKE diffusion and stable stratification.

KPP is a non-local closure parameterization. Like TKE, KPP utilizes the concept of eddy diffusivities for determining the turbulent vertical fluxes. The diffusivities no longer depend on TKE, but have a vertical profile dependent on the depth of the boundary layer  $h$ , which is determined using the bulk Richardson number:

$$Ri_b(d) = \frac{(B_r - B(d)) d}{|\mathbf{v}_r - \mathbf{v}(d)|^2 + v_t^2(d)}, \quad (2)$$

where  $B$  is buoyancy,  $\mathbf{v} = (u, v, w)$  is the mean velocity, and the subscript  $r$  denotes near-surface values;  $v_t$  is an additional term, the turbulent velocity shear, which ensures that entrainment is independent of stratification  $N$ . The depth of the boundary layer  $h$  is then the smallest  $d$  for which Equation 2 is less than a critical bulk Richardson number,  $Ri_c$ . Below the boundary layer, the ocean vertical mixing in KPP is parameterized as a superposition of three effects: instability due to resolved shear, internal wave breaking, and double diffusion. The former is based on the local gradient Richardson number, and the diffusivities arising from internal wave breaking are set to constant values (typically  $10^{-4} \text{ m}^2 \text{ s}^{-2}$  for momentum and  $10^{-5} \text{ m}^2 \text{ s}^{-2}$  for tracer diffusivity).

The vertical mixing schemes are implemented in the models with the Community Ocean Vertical Mixing package (CVMix, Griffies et al., 2015; Van Roekel et al., 2018). The evaluation of the schemes can be found in Korn et al. (2022), Gutjahr et al. (2021) (ICON-O), and Scholz et al. (2022) (FESOM). In the control runs, the vertical mixing is parameterized with the TKE scheme. The minimum background TKE is  $10^{-6} \text{ J kg}^{-1}$ , with no minimum diffusivity imposed, and  $c_k = 0.2$ . The sensitivity experiments in both of the models include TKE runs with  $c_k = 0.1$  and  $c_k = 0.3$ , as well as a simulation done with KPP. During initial runs, we noticed a low diffusivity band developing at the bottom of the ML, leading to introducing two additional sensitivity runs: one with minimum background TKE of  $10^{-5} \text{ J kg}^{-1}$ ; and another with a minimum background  $K_m$  of  $10^{-4} \text{ m}^2 \text{ s}^{-1}$  and a minimum background  $K_h$  of  $10^{-5} \text{ m}^2 \text{ s}^{-1}$ . The two KPP experiments – ICON\_kpp and FESOM\_kpp – are both run with the critical Richardson number  $Ri_c = 0.3$ . The minimum background  $K_h$  and  $K_m$  values are set as in ICON\_kappa. In ICON\_kpp, the mixing scheme below the ocean boundary layer is PP (Pacanowski & Philander, 1981), while in

FESOM\_kpp, it is KPP at both the surface and in the interior of the ocean. Table 1 outlines the specifications of the experiments.

### 2.3. Observations Used for Model Evaluation

The modeled mixing in the tropical Atlantic is assessed both by comparison to MLD climatology and by investigating the local effect of NIW mixing at the two selected sites. The latter requires model forcing evaluation to ensure that major discrepancies between modeled and observed NIW properties originate from the mixing parameterizations, rather than biased surface boundary conditions. We use three main data sources for these purposes: ship based climatology, the PIRATA array, and observations from the R/V Meteor cruise M119.

For the MLD, we use the IFREMER climatology (de Boyer Montégut et al., 2004), which is a global data set computed from profiles collected between 1941 and 2008. The MLD is calculated using the criterion  $\Delta T = \pm 0.2^\circ\text{C}$  from the in-situ temperature at 10 m below the surface. The annual climatology set contains spatially averaged median MLD values, computed from profiles grouped into  $2^\circ \times 2^\circ$  grid boxes. Additionally, we use the enhanced PIRATA data set (Foltz et al., 2018) to compare the time series of the modeled and observed MLD at the two PIRATA sites. In ePIRATA, Argo data is used to construct vertical temperature and salinity profiles from the mooring observations; the MLD time series is then calculated from the criterion of  $\Delta\rho = 0.12\text{ kg m}^{-3}$  relative to 1 m depth, which is approximately equivalent to  $\Delta T = -0.35^\circ\text{C}$ .

A range of PIRATA data is used to validate the forcing and the NIW response. We note that the data from the PIRATA moorings are assimilated into the ERA5 product as a part of the DRIBU data set (Hersbach et al., 2020). In the forcing comparison, we use the following: wind vector data measured at 4 m above the sea surface (accuracy of magnitude  $\pm 2\%$  and direction  $\pm 3.4^\circ$ ); longwave and shortwave radiation data measured at 3.5 m above the sea surface ( $\pm 1\%$  and  $\pm 2\%$ , respectively); and total precipitation data ( $\pm 0.4\text{ mm hr}^{-1}$ ). The wind vector at 10 m is computed following the logarithmic wind profile assuming neutral stability (Oke, 2002):

$$\mathbf{U}_{10} = \mathbf{U}_4 \frac{\ln(10/z_0)}{\ln(4/z_0)}, \quad (3)$$

where  $z_0 = 2 \times 10^{-4}\text{ m}$  is the roughness length and  $\mathbf{U} = (U, V)$  is the wind vector. For NIW response validation, we use high frequency ocean current ( $\pm 0.5\text{ cm s}^{-1}$ ) and temperature data ( $\pm 0.002^\circ\text{C}$ ).

Finally, for the comparison of the mixing parameters, we use the microstructure and vmADCP measurements from the M119 cruise (Brandt et al., 2017; Fischer, 2020). The NIW observations published by HDRF20 were recorded between the thirteenth and fifteenth of September. The shear ( $Sh^2 = (\partial u/\partial z)^2 + (\partial v/\partial z)^2$ ) is computed from the vmADCP current measurements, which span the depths 17–800 m, with 8 m spatial resolution. The shear values are averaged between the thirteenth and fifteenth of September (with cutoff at midnight). The 25 microstructure profiles were recorded between the thirteenth of September, 10:00 UTC and the fourteenth of September, 06:00 UTC, and have the vertical resolution of about 0.5 m. The 95% confidence intervals of  $N^2$  and  $Sh^2$  are estimated using the bias-corrected and accelerated bootstrapping algorithm. Eddy diffusivity values across the 25 profiles are collected in 10 m depth bins and averaged. The 95% confidence intervals of  $K_\rho$  are estimated using error propagation from Schafstall et al. (2010).

### 2.4. Comparison of Forcing to PIRATA Data

The significant factors for NIW generation and dissipation are the wind stress, ML depth, and stratification. The relevant local forcing parameters to evaluate are therefore wind velocities, radiation, and rainfall. The agreement between the ERA5 forcing and the local measured PIRATA climate is evaluated and presented in Figure 1. The ML depth is evaluated in Section 3.1.

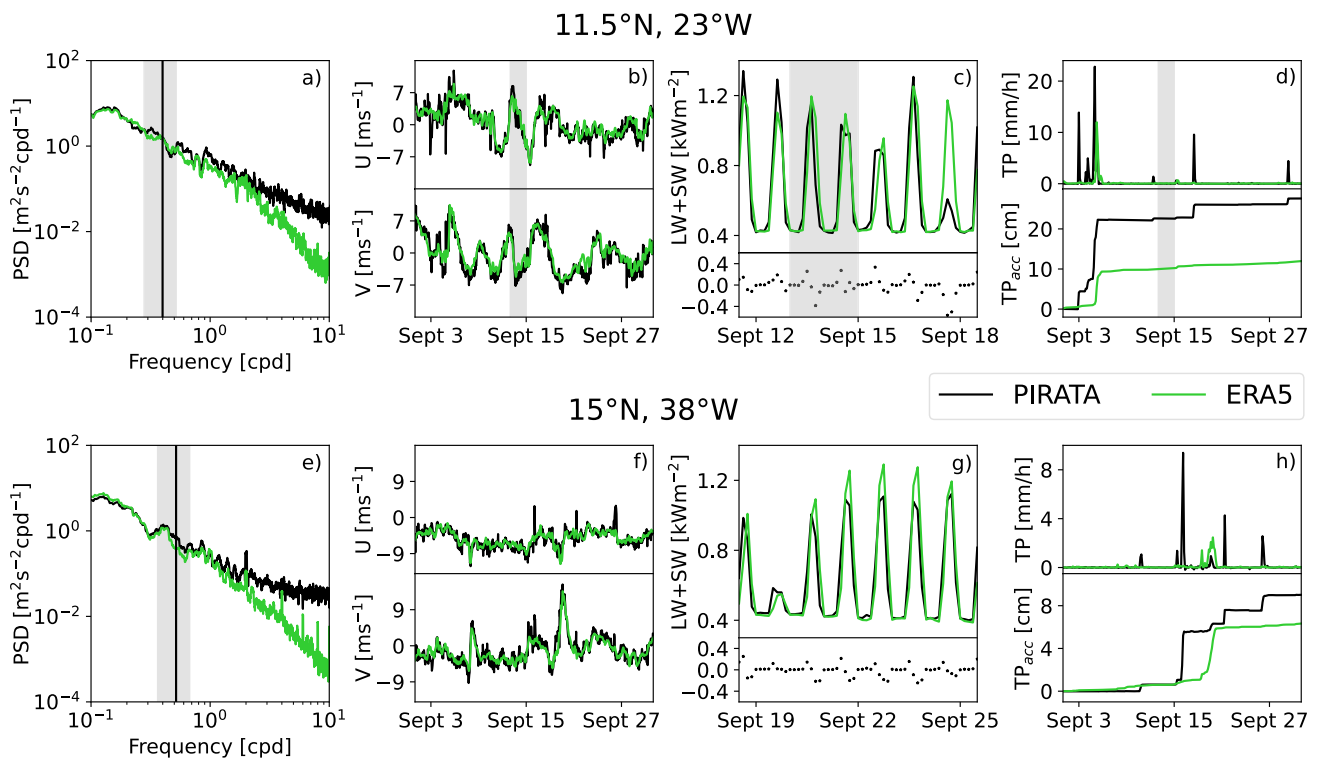
NIWs are forced by the wind stress. In comparison to the direct PIRATA observations, the power spectral density (PSD) of the ERA5 wind speed at the two sites in 2015 shows a clear difference at periods of hours or less (Figures 1a and 1e). This difference is expected, because at scales close to grid scale viscosity is designed to remove energy to suppress numerical instability (Jochum et al., 2008). Within the near-inertial band, the wind energy is 10% lower at the  $11.5^\circ\text{N}$  site and 27% lower at the  $15^\circ\text{N}$  site in ERA5 relative to PIRATA. The wind

**Table 1**  
Overview of the Experiments

Name	Mixing scheme	$c_k$ value	Minimum TKE	Minimum $\kappa_H$	Minimum $\kappa_M$	$Ri_c$
ICON_ck1	TKE	0.1	$10^{-6}$	—	—	—
ICON_ck2	TKE	0.2	$10^{-6}$	—	—	—
ICON_ck3	TKE	0.3	$10^{-6}$	—	—	—
ICON_kpp	KPP	—	—	$10^{-5}$	$10^{-4}$	0.3
ICON_mintke	TKE	0.2	$10^{-5}$	—	—	—
ICON_kappa	TKE	0.2	$10^{-6}$	$10^{-5}$	$10^{-4}$	—
FESOM_ck1	TKE	0.1	$10^{-6}$	—	—	—
FESOM_ck2	TKE	0.2	$10^{-6}$	—	—	—
FESOM_ck3	TKE	0.3	$10^{-6}$	—	—	—
FESOM_kpp	KPP	—	—	$10^{-5}$	$10^{-4}$	0.3

velocity time series in Figures 1b and 1f confirm the coherence of the wind forcing. We can therefore expect coherence in the observed and simulated NIW signal, but lower amplitude due to the discrepancy in WPI.

The shortwave and longwave series reveal a diurnal cycle bias of 200–300  $\text{Wm}^{-2}$  (Figures 1c and 1g). This bias originates from the shortwave radiation data. ERA5 tends to show higher daily radiation values compared to PIRATA, especially at the 15°N site. Considering that the analyzed NIWs were associated with the passing of a storm, these biases likely arise from discrepancies between in-situ and modeled cloud cover. Additionally, Foltz et al. (2013) report that some of the PIRATA moorings can show a bias in SW measurements during boreal



**Figure 1.** The comparison of the ERA5 forcing (green) to PIRATA mooring data (black) at 11.5°N and 15°N in September 2015. Subplots (a) and (e) show the power spectral density of the wind speed time series spanning from January–October 2015, the black vertical lines indicate the local Coriolis frequency  $f$ , and the shaded region spans from  $0.7f$  to  $1.3f$ ; (b) and (f) show the zonal and meridional wind velocities ( $U$ ,  $V$ ); (c) and (g) show the sum of the long- and shortwave radiation (LW + SW) at the time of the near-inertial wave events; the plots below indicate the residual (PIRATA-ERA5); (d) and (h) show the total precipitation and accumulated precipitation ( $TP_{acc}$ ). In panels (b), (c), and (d), the shaded region indicates the duration of HDRF20 observations.

summer and fall due to aeolian dust accumulation. The magnitude of the bias can be as large as  $50 \text{ Wm}^{-2}$ , which could account for 25% of the radiation offset between ERA5 and PIRATA. The contribution could be significantly higher during the day, as the reported biases are based on daily averages, including zero values at night.

While the precipitation biases are relatively low at the times the NIWs are measured, long-term precipitation differences can introduce persistent salinity offsets. Therefore, we show the hourly precipitation and the accumulated precipitation over the entirety of September 2015 (Figures 1d and 1h). Overall, PIRATA registers higher rates of rainfall, but ERA5 overestimates the precipitation at the  $15^\circ\text{N}$  site when the storm passes by the mooring on the nineteenth of September.

Surface salinity fluxes are not consistently impacted by the systematic underestimation of rainfall in ERA5. Table 2 shows the mean salinity and potential density in the ICON-O models and observations. The differences across the ML are computed by subtracting the surface layer values from the values at a depth layer  $d$  below the ML:  $d = 20 \text{ m}$  at  $11.5^\circ\text{N}$ ;  $d = 40 \text{ m}$  at  $15^\circ\text{N}$ . The depths are selected based on PIRATA data availability. The means are calculated over the course of September 2015 at both sites. At  $11.5^\circ\text{N}$ , we additionally report the mean biases during HDRF20 observations. “TKE average” is an average of all TKE sensitivity experiments: the temporal means are computed for each run, and an average of those is presented in the table.

Surface salinity is not systematically higher, nor consistently biased at the two sites as a result of reduced rainfall in ERA5. In the tropics, a high salinity layer develops below the fresher ML in the summer. The reduced rainfall in ERA5 would therefore serve to reduce the density gradient across the ML. Instead, the density gradient is generally too high in the models. The surface ocean is too dense regardless of the magnitude of surface salinity biases. There is therefore no clear indication that the models are significantly affected by the underestimated precipitation in ERA5 forcing.

In summary, the WPI within the near-inertial band is lower by 10% at  $11.5^\circ\text{N}$  and by 27% at  $15^\circ\text{N}$  in ERA5 compared to PIRATA. Despite the biases in radiative forcing and precipitation, the surface layer salinity and potential density means in the models are all within 1.3% of the observed PIRATA values. We conclude that the ERA5 reanalysis product is sufficiently close to the PIRATA data to force the observed NIWs and their response in the models.

### 3. Results

#### 3.1. Mixed Layer Depth in the Models

The modeled tropical Atlantic MLD is calculated from the temporally averaged output using the criterion of  $\Delta T = -0.2^\circ\text{C}$  relative to temperature at 10 m depth. The mean is taken over the summer, defined as first of July to the first of October. The map of median values in  $2^\circ \times 2^\circ$  boxes is computed for all the model runs. In Figure 2, the climatological MLD distributions are shown, together with comparison to the control FESOM run, FESOM\_ck2, and the simulation which resembles observations most, FESOM\_kpp. The zonal averages of MLD distributions are presented in Figure 2g for the general overview of the model performances. Local daily MLD time series at  $11.5^\circ\text{N}$  and  $15^\circ\text{N}$  are shown in Figures 2c and 2f, respectively, and compared to the ePIRATA reconstructions. The model values are calculated from  $\Delta T = -0.35^\circ\text{C}$  relative to 1 m depth.

We note that 2015 marked the beginning of the strongest Pacific El Niño event of the 21st century (Santoso et al., 2017). Motivated by this, we check for wind anomalies in the Atlantic. The winds do not diverge significantly from the climatology, but the El Niño event caused reduction in hurricane activity in the tropical Atlantic in 2015 (Stewart, 2016).

The two maps in Figures 2d and 2e show the shallow MLD bias which spans nearly the entire basin and is present in all experiments. The exception is a band north of the equator where the zonal average of ICON\_kpp MLD is about 1.5 times deeper than the IFREMER climatology (Figure 2g). North of  $10^\circ\text{N}$ , all but ICON\_kpp modeled MLD profiles are too shallow and fall outside of the IFREMER confidence bound. The ICON-O TKE runs are generally more biased than FESOM, despite identical forcing and ML scheme. The TKE sensitivity experiments do not affect the zonal MLD average significantly; instead we see considerable differences between the two models, and the two mixing schemes. The KPP MLD values are systematically deeper than the TKE based ones.

At the  $11.5^\circ\text{N}$  site, the MLD is generally too shallow in all TKE runs, and too deep in the KPP runs, as indicated by the comparison to the ePIRATA data (Figure 2c). During the HDRF20 observations, ICON\_kpp simulates

**Table 2**

Mean Surface Salinity ( $S$ ), Salinity Difference Across the Mixed Layer (ML) ( $\Delta S$  Across the ML), Surface Potential Density ( $\sigma_\theta$ ) and Potential Density Difference Across the ML ( $\Delta\sigma_\theta$  Across the ML) in the ICON-O Model Runs and PIRATA Mooring Data

Site	11.5°N	11.5°N	11.5°N	11.5°N	15°N	15°N
Time period	HDRF20	HDRF20	Sept 2015	Sept 2015	Sept 2015	Sept 2015
Surface salinity ( $S$ ) (psu)						
PIRATA	35.63	–	35.37	–	36.48	–
TKE average	35.37	<b>–0.26</b>	35.49	<b>0.12</b>	36.46	<b>–0.02</b>
KPP	35.27	<b>–0.36</b>	35.32	<b>–0.05</b>	36.57	<b>0.09</b>
$\Delta S$ across the ML (psu)						
PIRATA	0.16	–	0.32	–	0.03	–
TKE average	0.31	<b>0.15</b>	0.35	<b>0.03</b>	0.11	<b>0.08</b>
KPP	0.15	<b>–0.01</b>	0.17	<b>–0.15</b>	0.09	<b>0.06</b>
Surface $\sigma_\theta$ ( $\text{kg m}^{-3}$ )						
PIRATA	22.77	–	22.55	–	23.52	–
TKE average	22.95	<b>0.18</b>	22.81	<b>0.26</b>	23.63	<b>0.11</b>
KPP	22.88	<b>0.11</b>	22.82	<b>0.27</b>	23.82	<b>0.30</b>
$\Delta\sigma_\theta$ across the ML ( $\text{kg m}^{-3}$ )						
PIRATA	0.51	–	0.83	–	0.73	–
TKE average	1.70	<b>1.19</b>	1.79	<b>0.96</b>	1.24	<b>0.51</b>
KPP	0.49	<b>–0.02</b>	0.55	<b>–0.28</b>	1.08	<b>0.35</b>

Note. The means are computed over two time periods (duration of HDRF20 observations, thirteenth to fifteenth of September 2015; and September 2015) at the 11.5° site and over September 2015 at the 15°N site. "Across the ML" refers to salinity and potential density values at the surface subtracted from values at a depth  $d$  below the mixed layer. At 11.5°N,  $d = 20$  m; at 15°N,  $d = 40$  m. "TKE average" stands for the average of all the TKE sensitivity runs. In bold, model biases are highlighted. Positive bias indicates the modeled values are higher compared to PIRATA data.

MLD closest to observed. The TKE and KPP runs are not as easily distinguishable at 15°N (Figure 2d). Here, the MLD before the passing of the tropical storm on the nineteenth of September varies between 15 and 30 m; afterward, the ML is restratified.

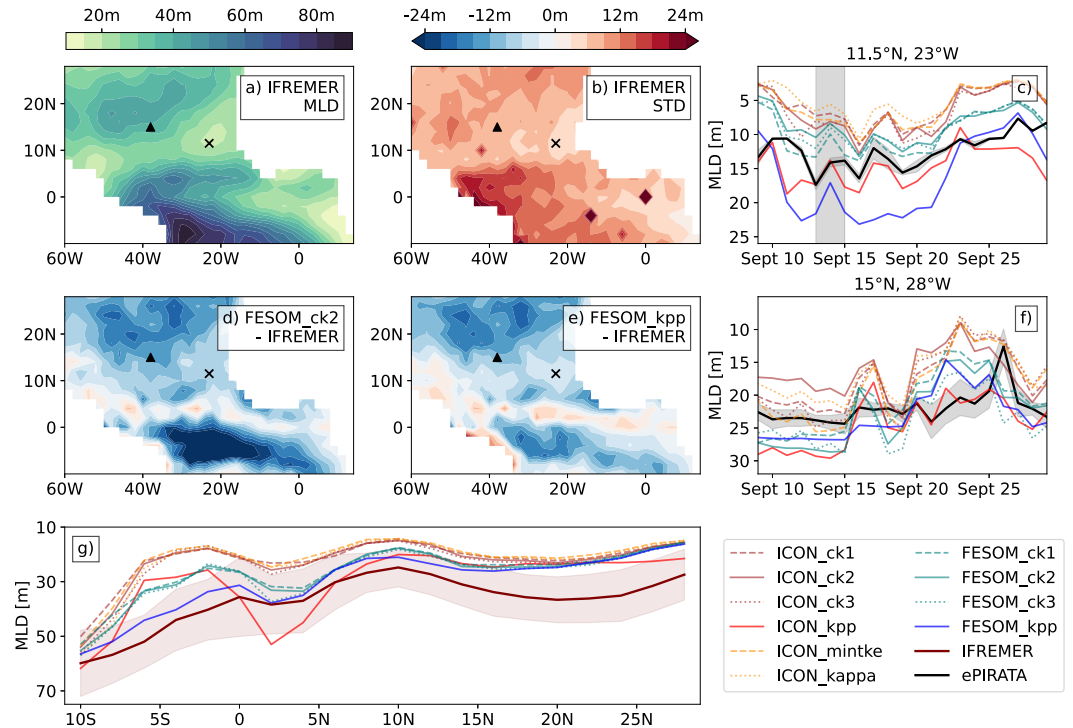
### 3.2. Near-Inertial Wave Characteristics

The time series of the zonal and meridional velocities at 10 m depth is shown in Figures 3a and 3b. A bandpass filter is used to extract the near-inertial signal from the raw data (Figures 3c and 3d). The Gaussian window envelops frequencies between  $0.7f$  and  $1.3f$ . The filtered velocities,  $u_i$  and  $v_i$ , are then used to compute the NIKE:

$$NIKE = \frac{1}{2}\rho_0(u_i^2 + v_i^2). \quad (4)$$

The NIKE values averaged over September 2015 and the three days of the HDRF20 observations are listed in Table 3. We choose to present the two NIKE diagnostics for the 11.5°N site, as the duration of the HDRF20 observations does not span over a full inertial period. The temporal distribution of NIKE is presented in Figures 3e and 3f.

The models generally underestimate the strength of the near-inertial currents at both sites. At 15°N, the negative bias is higher. About 27% of the lacking NIW energy at this site can be associated with the underestimated WPI within the near-inertial band in ERA5 (Section 2.4). At 11.5°N, despite 10% weaker winds, the mean September NIKE is higher than observed in Icon\_ck1, Icon\_ck2 and Icon\_ck3. In the period of HDRF20 observations, NIWs are too weak in all experiments.



**Figure 2.** The mixed layer depth in the model runs compared to the IFREMER climatology in the tropical Atlantic. The IFREMER data in panel (a) and the corresponding std in panel (b) show the summer (JAS) average between the years 1941 and 2008, while the model output is the average of summer 2015. The difference maps (d) FESOM\_ck2-IFREMER and (e) FESOM\_kpp-IFREMER show the shallow bias which is consistent in all model runs. The 11.5°N site is marked on the maps with an x, and the 15°N site - with a triangle. In panel (g), the zonal averages of all experimental MLD profiles are compared to the climatology. The shading indicates the median std from the raw IFREMER MLD data. (c) and (f) show the modeled September MLD time series at 11.5°N and 15°N, respectively; the ePIRATA (black) daily MLD values are presented with the estimated error (shading). In panel (c), the gray vertical bar indicates the duration of the HDRF20 observations.

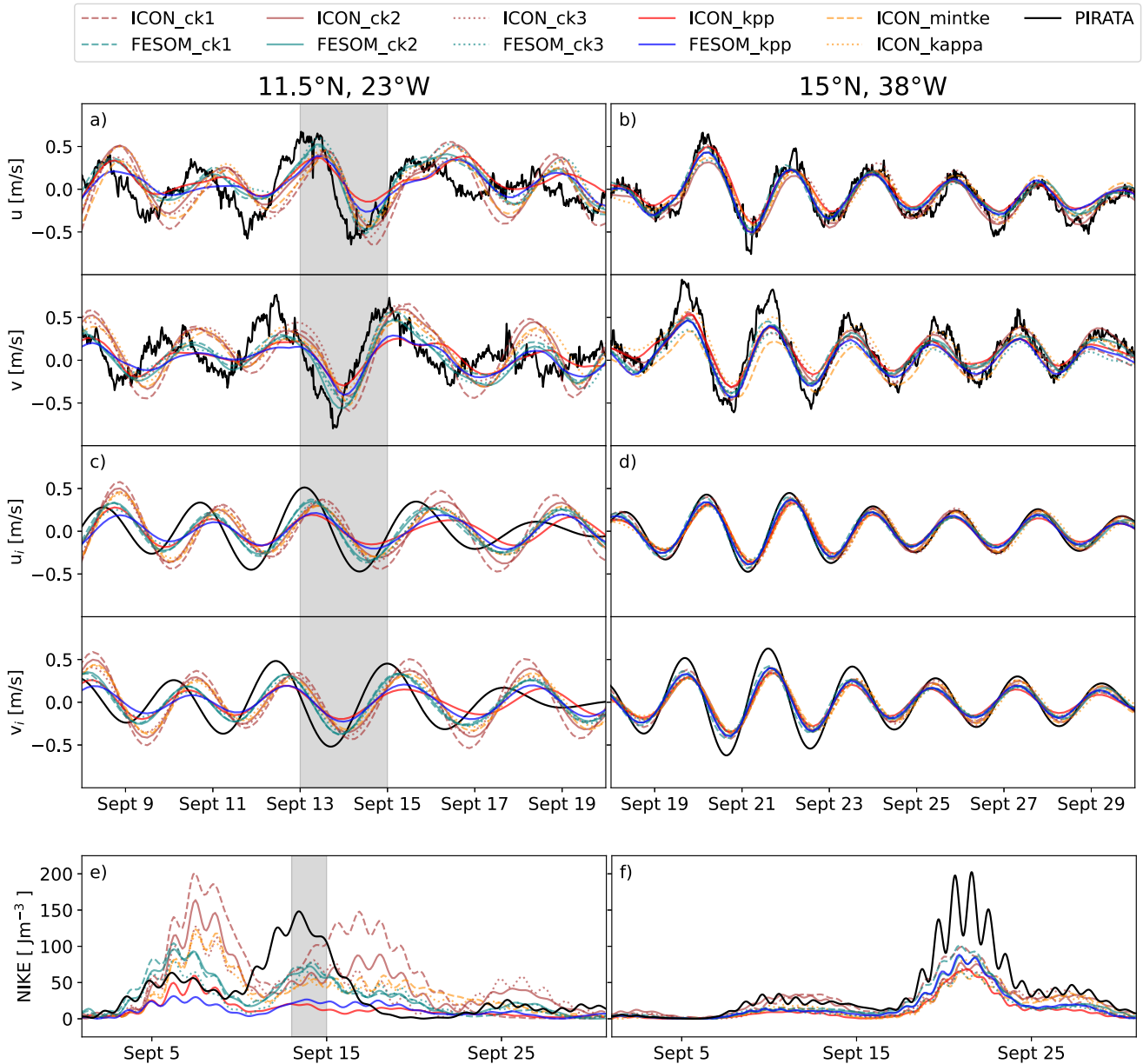
There are key differences in how the NIWs are modeled at the two sites. The waves are in phase at 15°N and out of phase at 11.5°N, and the strength of the near-inertial currents is more sensitive to the vertical mixing scheme at 11.5°N. The difference most likely stems from stratification: at 11.5°N, the potential density gradient across the ML is more than two times higher in the ICON-O TKE models compared to PIRATA (Table 2); at 15°N, the density gradient bias is lower. The magnitudes of the relative MLD biases at the two sites are similar (Figures 2c and 2f), but the ML is thinner at 11.5°N. NIW amplitude is inversely proportional to the depth of the ML (D'Asaro, 1985), hence shallow MLD biases at 11.5°N potentially have a larger impact on the strength of NIWs. At both sites, the mean NIKE is sensitive to the amount of TKE which is used for mixing in the TKE scheme: generally, low  $c_k$  values lead to more energetic NIWs. The sensitivity of NIKE bias to the vertical mixing scheme parameterization is further discussed in Section 3.4.

### 3.3. The Mixing Signature of NIWs

Detailed turbulence observations are only available for the 11.5°N site. We therefore use the PIRATA temperature data to assess the vertical mixing of heat in the models at both locations (Section 3.3.1). The simulated eddy diffusivities, stratification, and shear are evaluated at the 11.5°N site in Section 3.3.2. We then present the deep mixing in the models after the storm at 15°N in Section 3.3.3.

#### 3.3.1. Temperature

The measured and modeled temperatures at the two sites are compared in Figure 4. In Figures 4f and 4g, the mean September vertical temperature profiles show that in most models, not enough heat penetrates below the ML. The standard deviation from the PIRATA monthly means is indicated by the black horizontal bars. The temperature



**Figure 3.** In panels (a) and (b), the raw PIRATA mooring velocity data in comparison to modeled velocity output at 11.5°N and 15°N at 10 m depth are presented. The bandpassed velocities  $u_i$  and  $v_i$  are plotted in panels (c) and (d). The September near-inertial kinetic energy time series for both sites is shown in panels (e) and (f). The gray shading indicates the duration of the HDRF20 observations.

oscillations below the ML have a near-inertial period, which is reproduced by the models at the 15°N site (demonstrated in Figures 4k and 4l), but not at 11.5°N (Figure 4e). At the surface, the diurnal cycle is evident, and it is reproduced by the models. Within the ML, the temperature is approximately constant; the diurnal cycle signal does not reach the bottom of the ML, and this is consistent between most of the models and the PIRATA observations.

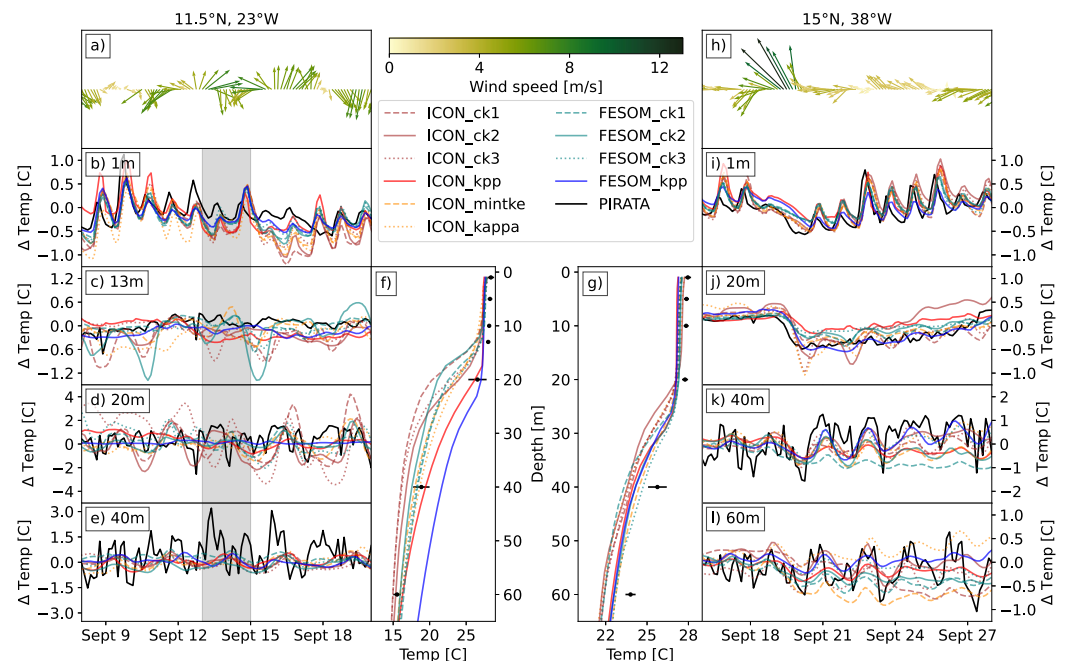
ICON\_kpp exhibits the best agreement with the observed September mean temperature profile at 11.5°N. The additional ICON TKE runs, ICON\_mintke and ICON\_kappa, show an improved vertical temperature profile compared to the base run ICON\_ck2. The modeled vertical temperature profile is more consistent across the sensitivity studies at the 15°N site. The systematic overestimation of the temperature gradient steepness could be a

**Table 3**  
Overview of the Mean Near-Inertial Kinetic Energy

Site	11.5°N	11.5°N	11.5°N	11.5°N	15°N	15°N
Time period	HDRF20 ( $\text{Jm}^{-3}$ )	HDRF20 (%)	Sept 2015 ( $\text{Jm}^{-3}$ )	Sept 2015 (%)	Sept 2015 ( $\text{Jm}^{-3}$ )	Sept 2015 (%)
PIRATA	116.6	—	35.6	—	36.3	—
ICON_ck1	80.8	<b>−30.7</b>	59.2	<b>66.3</b>	25.1	<b>−30.9</b>
ICON_ck2	48.6	<b>−58.3</b>	46.0	<b>29.2</b>	22.7	<b>−37.5</b>
ICON_ck3	76.0	<b>−34.8</b>	42.3	<b>18.8</b>	18.8	<b>−48.2</b>
ICON_kpp	13.2	<b>−88.7</b>	12.9	<b>−63.8</b>	12.5	<b>−65.6</b>
ICON_mintke	40.7	<b>−65.1</b>	35.6	<b>0.0</b>	16.3	<b>−55.1</b>
ICON_kappa	42.9	<b>−63.1</b>	32.8	<b>−7.9</b>	18.2	<b>−49.9</b>
FESOM_ck1	72.6	<b>−37.7</b>	32.1	<b>−9.6</b>	20.6	<b>−43.3</b>
FESOM_ck2	68.0	<b>−41.7</b>	31.6	<b>−11.2</b>	18.1	<b>−50.1</b>
FESOM_ck3	61.6	<b>−47.2</b>	27.3	<b>−23.3</b>	13.9	<b>−61.7</b>
FESOM_kpp	24.1	<b>−79.2</b>	12.3	<b>−65.4</b>	17.5	<b>−51.8</b>

Note. At 11.5°N, the mean is taken both over the entire September 2015 and over the duration of the HDRF20 observations (thirteenth to fifteenth of September). The relative difference between the model results and the PIRATA data is highlighted in bold text. Negative values indicate that modeled NIKE is weaker than observed.

result of the lacking wind energy in ERA5 forcing. The rapid cooling of the surface occurs when the storm passes by the mooring. All models reproduce the rapid ML cooling and subsequent warming after the storm. Below the ML, the temperature oscillations are present, but have reduced amplitude compared to the observations (Figures 4k and 4l). Contrary to the 11.5°N site, none of the vertical mixing schemes can be singled out as best performing.



**Figure 4.** The comparison of measured and modeled temperatures at the two sites. (a) and (h) show the PIRATA wind vectors. The mean September vertical temperature profiles are given in panels (f) and (g); the PIRATA temperature averages for each available depth are indicated with the black points, and the horizontal bars indicate the standard deviation. The time series show departures from September mean temperature at (b) 1 m, (c) 13 m, (d) 20 m, and (e) 40 m depths for the 11.5°N site; and (i) 1 m, (j) 20 m, (k) 40 m, and (l) 60 m depths for the 15°N site. The gray shaded area in panels (b), (c), (d), and (e) indicates the duration of the HDRF20 observations.



### 3.3.2. Comparison to Microstructure Measurements at 11.5°N

The snapshots of the mixing parameters at the 11.5°N site are examined in the model sensitivity runs and compared to the HDRF20 data. The temporal locations of the peaks in diffusivities and shear coincide with maximum NIW velocities, which differ across the models due to the NIW phase shift. The snapshots have a vertical structure which persists over time, and therefore we choose to average the profiles to highlight the similarities between the modeled and observed mixing structures. These are presented in Figure 5.

Stratification is well represented in the models beyond the ML. The KPP runs simulate the observed  $N^2$  profile the best, and all models can reproduce the observed stratification below the ML. Above 10 m,  $N^2$  is significantly overestimated in all the runs. Shear is systematically underestimated by the models despite coarser vertical resolution of observed horizontal currents. We suspect unresolved internal waves are the reason for this discrepancy (Large et al., 1994).

All models exhibit enhanced mixing above approximately the same depth of 20 m. The diffusivity values from the microstructure profiler are only valid below the ML, which is not constant across the 25 measurements. The top bin of microstructure diffusivity data in Figure 5 is set to depth below 18 m, which is the average observed MLD, while the modeled MLD is shallower in the TKE runs. Hence, the TKE simulations exhibit the observed enhanced diffusivities below the simulated ML, but the effect does not reach the same depth due to the thinner ML.

The enhanced diffusivity below 20 m is not reproduced by the models. The TKE model runs with no restrictions to background diffusivities match the interior  $K_p$  values below 50 m depth well. Although the vertical resolution of the models is higher than the observations, the interior shear is lower than observed in all of the runs. ICON\_ck3 is an exception, coming close to the observed shear at 40 m. Despite this local consistency with both the  $N^2$  and  $Sh^2$  values, the run does not exhibit enhanced diffusivity at this depth.

### 3.3.3. Vertical Mixing Profile

The lack of deep mixing response to NIWs at 11.5°N puts to question whether such process occurs in the models at all. Due to the proximity and strength of the storm at the 15°N site, we can verify whether mixing below the ML can be simulated.

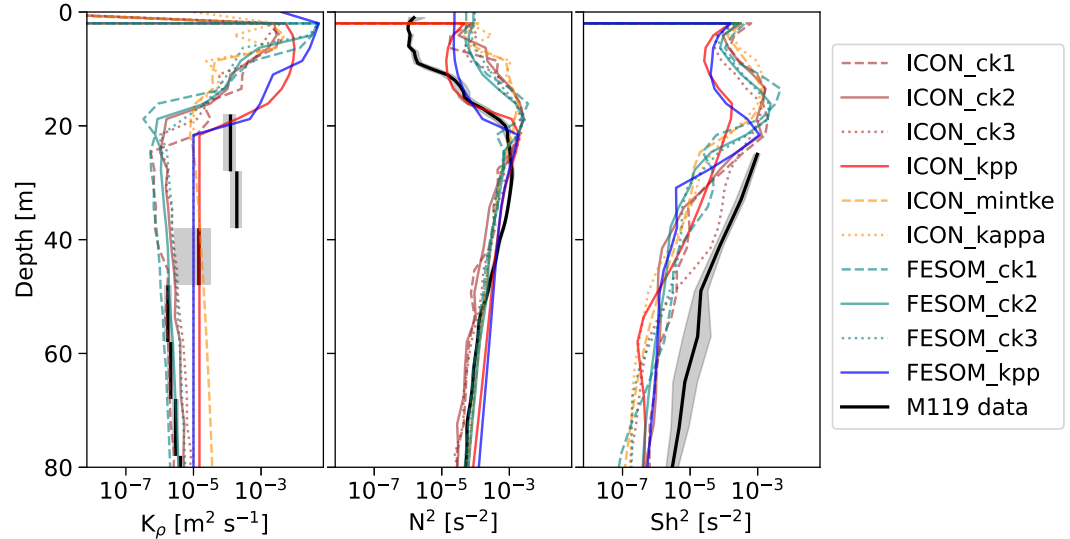
Snapshots of  $K_p$  in the model runs reveal that indeed, the enhanced mixing reaches below the recorded ML after the storm passes on the nineteenth of September (Figure 6). Diffusivities of  $10^{-3} \text{ m}^2 \text{ s}^{-1}$  reach 40 m depth in most of the models. Oscillations at near-inertial frequencies can be seen in the TKE runs; the oscillations persist for over 10 days after the storm passes. The PSD of diffusivities at 53.6 and 58 m show clear peaks at the near-inertial frequency.

## 3.4. NIKE Sensitivity to Mixing Scheme Parameterization

There are a few key differences in how NIWs are simulated at the two PIRATA sites: at 11.5°N, the mean NIKE bias is positive for some models, the NIWs are out of phase, and the vertical temperature gradient is sensitive to the mixing scheme choice; at 15°N, the mean NIKE bias is strictly negative, the NIWs are in phase, and the vertical temperature gradient is not highly sensitive to the mixing scheme choice. We now consider the NIKE biases at the two sites throughout the entire simulation period to illustrate what might be the cause of these differences.

We compare the probability density function (PDF) of the wind vector components and the wind speed, as well as the PDF of the relative bias between ERA5 output and PIRATA measurements. Throughout 2014 and 2015, the wind speed is systematically underestimated by ERA5. The magnitude of the bias is 7.88% at 11.5°N and 2.92% at 15°N. The PIRATA wind speed sensors have the precision of 2%–3%. We therefore believe that the agreement between PIRATA and ERA5 simulated winds is good enough for NIW analysis in 2014 and 2015.

We define a "NIW event" in the time series of the bandpassed PIRATA current signal as the period in which NIKE is continuously above  $25 \text{ Jm}^{-3}$  for more than one inertial period. All 2014 and 2015 NIW events take place during boreal summer and fall, and in total there are nine of such events at 11.5°N and only one at 15°N. The probability distributions of relative NIKE biases are calculated based on data from periods when NIWs are present at the sites: between June and November at 11.5°N, and between August and October at 15°N. While the PIRATA



**Figure 5.** Eddy diffusivities ( $K_\rho$ ), the buoyancy frequency ( $N^2$ ) and shear ( $Sh^2$ ) mean vertical profiles in the models compared to microstructure and vADCP current data from the M119 R/V Meteor cruise. The gray shading indicates 95% confidence bounds. The top bin in the diffusivity data is set to 18 m, which is the mean depth of the mixed layer (ML) at the time of the HDRF20 measurements. Modeled diffusivities within the ML (above 18 m) are available and shown.

observations only show one NIW event at  $15^\circ\text{N}$  in 2015, the data from 2014 is also included, as in some of the TKE simulations NIKE is higher than  $25 \text{ Jm}^{-3}$  during that period.

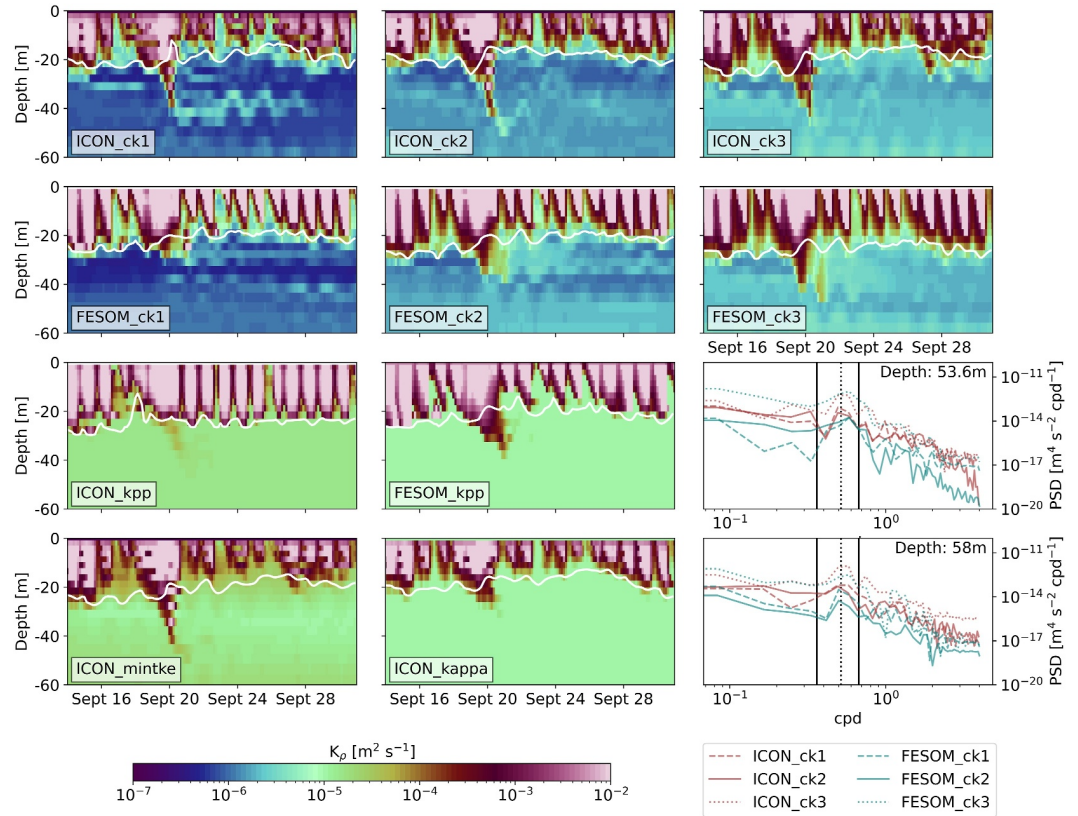
Figure 7 shows that NIKE bias is sensitive to the mixing scheme choice at  $11.5^\circ\text{N}$ . The sensitivity is significantly reduced at the  $15^\circ\text{N}$  site (Figure 8). The relative bias is defined as:

$$\text{Relative bias} = \frac{\text{NIKE}_{sim} - \text{NIKE}_{obs}}{\text{NIKE}_{obs} + c} \cdot 100\%, \quad (5)$$

where  $\text{NIKE}_{obs}$  is the observed (PIRATA) NIKE,  $\text{NIKE}_{sim}$  is the simulated NIKE, and the constant  $c = 25 \text{ Jm}^{-3}$  is added to handle NIKE values that are close to zero. On the figure, the distribution of the relative bias is shown, together with the mean bias and its standard deviation. At  $11.5^\circ\text{N}$ , all TKE models overestimate the observed NIKE, while KPP underestimates it. The bias is sensitive to the TKE parameterization: when a higher percentage of TKE is used for mixing, the NIKE bias is reduced. The mean NIKE bias is smallest in the KPP simulation.

The distribution of the relative NIKE bias is also sensitive to the vertical mixing scheme at  $15^\circ\text{N}$ , but the sensitivity is not consistent with  $11.5^\circ\text{N}$ . Similarly to  $15^\circ\text{N}$ , Icon\_ck1 simulates the largest positive bias and largest bias overall, and Icon\_kpp is the only simulation which systematically underestimates NIKE. However, apart from Icon\_ck1 and Icon\_mintke, the average NIKE bias at  $15^\circ\text{N}$  is less than 3.5% across the sensitivity experiments. The  $15^\circ\text{N}$  site is therefore fundamentally different than  $11.5^\circ\text{N}$  during the simulation period: at  $15^\circ\text{N}$ , the major source of near-inertial energy is the tropical storm Ida. The NIWs at  $11.5^\circ\text{N}$  are forced by weaker but more frequent bursts of rotating winds, leading to generally higher near-inertial energy.

We found no statistically significant correlation of the NIKE bias with the wind speed, the MLD bias, or the stratification at the base of the ML, the latter defined as the maximum  $N^2$  value along the vertical coordinate. Hence, the contrasting characteristics of the modeled NIWs most likely stem from the combination of the physical conditions at the sites. At  $11.5^\circ\text{N}$ , the high density gradient across the ML (Table 2) produces strong near-inertial currents. NIW amplitude is inversely proportional to the depth of the ML (D'Asaro, 1985). The strength of the near-inertial currents is therefore more sensitive to MLD biases when the ML is thin. At  $15^\circ\text{N}$ , the density gradient is lower and the ML deeper, hence the inertial current sensitivity to MLD biases is reduced.



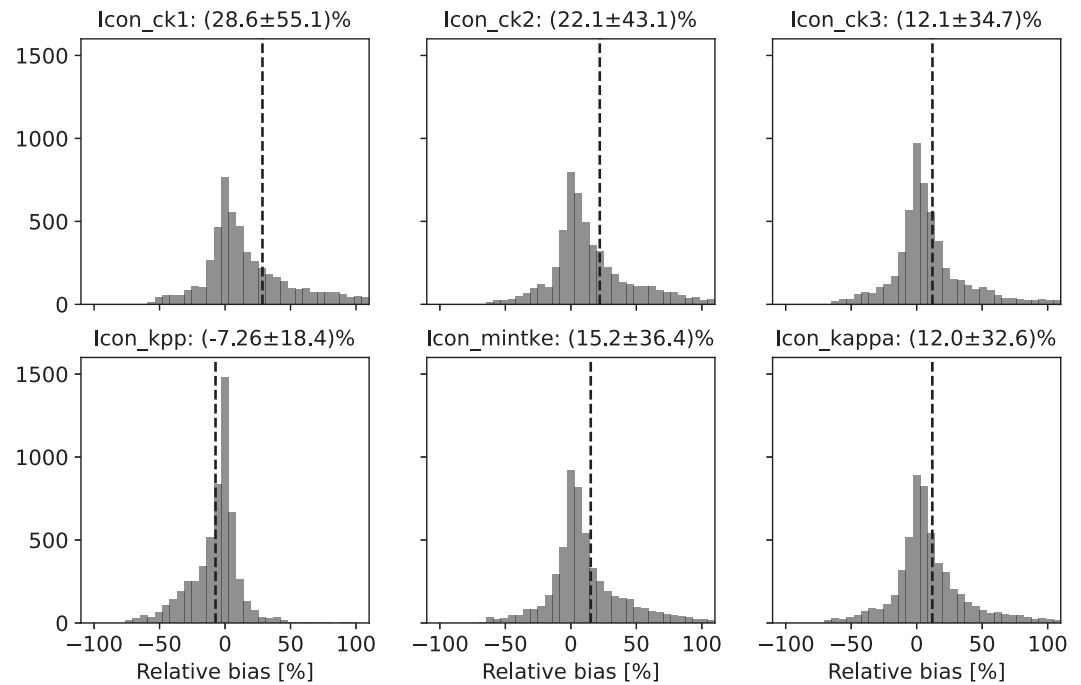
**Figure 6.** Snapshots of eddy diffusivities at the 15°N site in the models. The snapshots reveal the deep mixing caused by the storm, and the subsequent near-inertial diffusivity oscillations below the mixed layer (ML). The depth of the ML is indicated with the white line. The spectra of diffusivity time series between September 19th and 31st, 2015, at depths 53.6 and 58 m confirm the presence of diffusivity oscillations at the near-inertial frequencies. The dotted lines mark  $f$ , and the two black solid lines envelop the range from  $0.7f$  to  $1.3f$ .

#### 4. Discussion

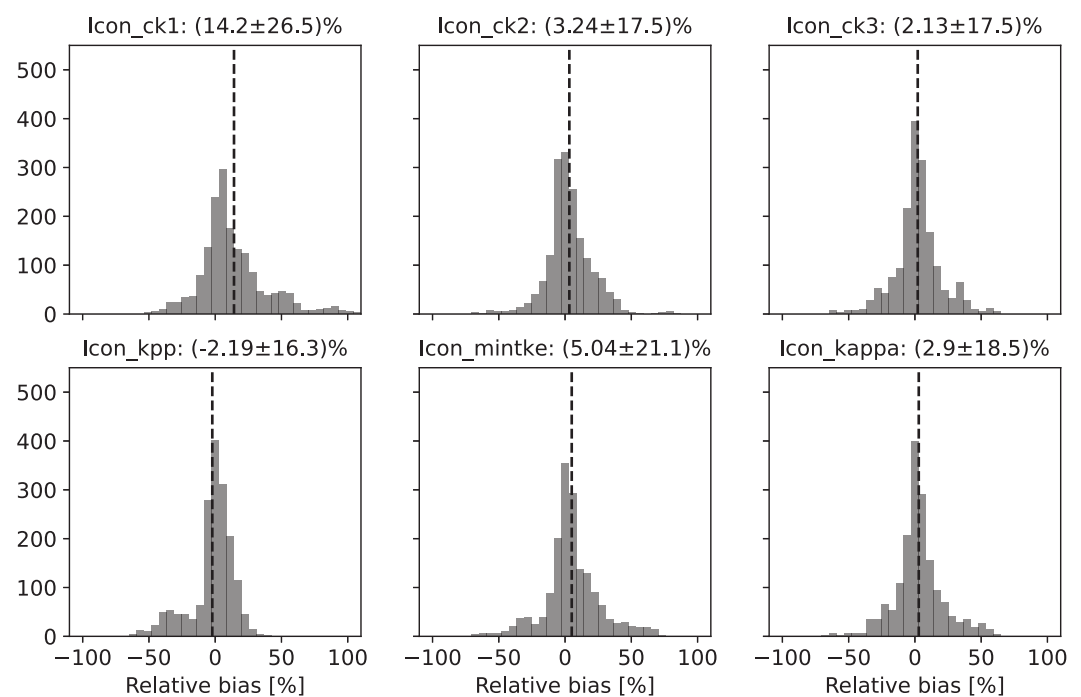
We present an evaluation of NIW-induced mixing at two sites in the tropical Atlantic in a range of sensitivity experiments using two different vertical mixing schemes: the Turbulent Kinetic Energy closure and the K-profile parameterization. We demonstrate that the eddy-resolving ocean models can simulate NIWs in the tropics to an extent, but the modeled near-inertial currents are too weak and the NIW-induced mixing underestimated. Out of the two parameterizations tested, KPP can model the observed mixing profiles better. At the 11.5°N site, ICON\_kpp excels especially at simulating the depth of the ML (Figure 2c) and mean vertical temperature profile (Figure 4f). Both KPP runs model the vertical stratification profile well (Figure 5). The analysis of NIKE biases at 11.5°N also shows that KPP is least biased in simulating the magnitude of the near-inertial currents (Figure 7).

The models do not reproduce the enhanced diffusivities below 20 m at the 11.5°N site (Figure 5). However, at the 15°N site, mixing below the ML occurs (Figure 6). The modeled and observed NIWs are in phase at the surface. In the models, the NIWs radiate down the water column and produce shear which induces mixing below the ML. The near-inertial oscillations in diffusivity persist below the ML for 10 days after the storm passes by the mooring. The local winds are weak during those remaining days of September (Figure 4h), suggesting that the interior near-inertial motions originate from the storm forcing. Hence, NIW-induced mixing could be a significant contributor to the local vertical turbulent cooling (see Foltz et al. (2020) for a detailed discussion about the seasonal cycle of the ML heat budget at this site).

The TKE scheme across all sensitivity runs overestimates the stratification within the ML, which traps the NIWs within it and suppresses the mixing below it. ICON\_ck3 is the only run which shows enhanced shear below the ML at 11.5°N. The enhanced TKE allows for vertical turbulent flux of momentum, while the thin ML produces



**Figure 7.** Probability distribution of near-inertial kinetic energy (NIKE) relative bias in the ICON-O sensitivity studies between June and November in the years 2014–2015 at 11.5°N. The black dashed line indicates the mean bias, and the standard deviation of the bias is given in figure titles. Positive bias means that simulated NIKE is higher than observed. For example, in Icon\_ck1, the systematic NIKE bias is 28.6%. In this simulation, 68% of the NIKE values in the time series are at minimum 26.5% too weak and at maximum 83.7% too strong compared to observations.



**Figure 8.** Probability distribution of near-inertial kinetic energy (NIKE) relative bias in the ICON-O sensitivity studies between August and October of the years 2014 and 2015 at 15°N. The black dashed line indicates the mean bias, and the standard deviation of the bias is given in figure titles. Positive bias means that simulated NIKE is higher than observed. At this site, only one near-inertial wave event is detected. The time series is therefore dominated by near-zero NIKE values.

relatively strong NIWs, but the shear remains too low to overcome the stratification and induce mixing (Figure 5). While ICON-O and FESOM differ significantly across the tested parameters, they consistently show the same discrepancies between the mixing schemes: TKE simulates shallower ML, lower ML diffusivities, and higher ML  $N^2$  and  $Sh^2$  compared to KPP and observations. Mean NIKE bias at 11.5°N is sensitive to the TKE scheme parameter  $c_k$ , which determines the fraction of TKE allocated for vertical mixing. In the prognostic TKE equation, the eddy viscosity  $K_m$  also scales the vertical flux of shear and TKE. Hence, lower  $c_k$  values result in stronger stratification, which increase the inertial current velocity.

The TKE scheme is based on the 2.5 level turbulence closure model by Mellor and Yamada (1982). In order to simplify the scheme, they neglect the material derivative and the dissipation of the temperature variance. It is justified for stable and slightly unstable flows, but is likely to introduce numerical errors in highly unstable conditions. Fan et al. (2021) use the 2.5 level Mellor-Yamada scheme in the Navy Coastal Ocean Model (NYCOM) and find that NIW mixing is under-represented when compared to large eddy simulation results. They speculate that the highly unstable conditions induced by strong cooling of the ML are the dominant cause for the scheme's inability to reproduce the NIW-induced mixing. Our results are consistent with theirs, indicating that the TKE scheme might be subject to the same limitation.

In this work, our assumption is that the NIW-induced mixing is not properly simulated because of the weaknesses of the mixing schemes. Other sources of biases that were not considered in this study might contribute to misrepresentation of NIWs in the models. Fan et al. (2021) find that turbulent mixing is better represented in the NYCOM model when the resolution at the ML base is as high as in the surface layers. While this does not directly affect the NIW-induced mixing in their case, a finer resolution at the ML base in our experiments could have countered biases in stratification (Table 2) and thereby indirectly countering the biases in NIW characteristics. Horizontal resolution might also play a role, as is explored in a study by Lévy et al. (2010). In an idealized basin, they find that increasing horizontal resolution in the NEMO model decreases the depth of the ML. Aside from the resolution, non-local model biases such as advected offsets in temperature and salinity might have contributed to the suppression of NIW-induced mixing.

We demonstrate that NIW observations can be used to identify limitations in the performance of vertical mixing parameterizations. The presence of deep near-inertial oscillations following the storm at 15°N confirm the potential of high resolution simulations for exploring the dynamics of NIW-induced mixing in the ocean interior.

## 5. Conclusions

Two vertical mixing schemes (KPP and TKE) in two different models (ICON-O and FESOM) are evaluated on their ability to reproduce NIW characteristics and NIW-induced mixing at two PIRATA mooring locations (11.5°N and 15°N) in the tropical Atlantic. The main findings of our study are:

- Microstructure measurements show enhanced diffusivities below the ML after the passing of a near-inertial wave at 11.5°N. The enhanced diffusivities are absent in the models regardless of the mixing scheme used.
- NIKE at both sites is sensitive to the mixing scheme choice. The sensitivity is higher at 11.5°N due to high stratification and shallow ML. KPP systematically underestimates the observed NIKE, while TKE systematically overestimates it.
- The KPP scheme simulates the vertical stratification profile and the depth of the ML at 11.5°N more realistically than TKE. NIKE bias at 11.5°N is also generally smallest in the KPP runs.
- The models simulate NIW-induced mixing following a storm at 15°N. The near-inertial oscillations remain below the ML for 10 days after the storm passes.

## Data Availability Statement

The M119 cruise data are available at <https://doi.pangaea.de/10.1594/PANGAEA.877375> (vmADCP; Brandt et al., 2017) and <https://doi.pangaea.de/10.1594/PANGAEA.920592> (microstructure; Fischer, 2020). The PIRATA data can be accessed at <https://www.pmel.noaa.gov/tao/drupal/disdell/>. The ePIRATA data set is available at <https://www.aoml.noaa.gov/phod/epirata/> (Foltz et al., 2018). The ERA5 reanalysis product can be found at <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview> (Hersbach et al., 2020). The IFREMER MLD climatology is available at <https://cerweb.ifremer.fr/deboyer/mlld/home>.

php (de Boyer Montégut et al., 2004). Model output from ICON-O and FESOM used in this study is available at [https://sid.erd.dk/cgi-sid/lis.py?share\\_id=DDGK0EDcTu](https://sid.erd.dk/cgi-sid/lis.py?share_id=DDGK0EDcTu).

### Acknowledgments

We would like to thank Sergey Danilov and Nikolay Koldunov (Alfred Wegener Institute for Polar and Marine Research), and Johann Jungclaus (Max Planck Institute for Meteorology) for their advice and guidance on model details. The model output was generated for the purposes of nextGEMS, a European H2020 funded project under the Grant agreement No. 101003470. The PIRATA mooring array is an open source of ocean data managed by GTMBA Project Office of NOAA/PMEL. We would like to thank the many people involved in the projects for making the model data and observations available for use in our research. We would also like to thank the two anonymous reviewers whose valuable comments inspired additional analysis, as well as the editor Lars Umlauf for guiding us through the review process.

### References

- Alford, M. H., & Gregg, M. C. (2001). Near-inertial mixing: Modulation of shear, strain and microstructure at low latitude. *Journal of Geophysical Research*, *106*(C8), 16947–16968. <https://doi.org/10.1029/2000JC000370>
- Alford, M. H., MacKinnon, J. A., Simmons, H. L., & Nash, J. D. (2016). Near-inertial internal gravity waves in the ocean. *Annual Review of Marine Science*, *8*(1), 95–123. <https://doi.org/10.1146/annurev-marine-010814-015746>
- Barsugli, J. J., & Sardeshmukh, P. D. (2002). Global atmospheric sensitivity to tropical sst anomalies throughout the indo-pacific basin. *Journal of Climate*, *15*(23), 3427–3442. [https://doi.org/10.1175/1520-0442\(2002\)015<3427:GASTTS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3427:GASTTS>2.0.CO;2)
- Bourlès, B., Araujo, M., McPhaden, M. J., Brandt, P., Foltz, G. R., Lumpkin, R., et al. (2019). Pirata: A sustained observing system for tropical Atlantic climate research and forecasting. *Earth and Space Science*, *6*(4), 577–616. <https://doi.org/10.1029/2018EA000428>
- Brandt, P., Czeschel, R., Schütte, F., Kopte, R., & Krahnmann, G. (2017). ADCP current measurements (38 and 75 kHz) during METEOR cruise M119 [Dataset]. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.877375>
- Brizuela, N. G., Alford, M. H., Xie, S.-P., Sprintall, J., Voet, G., Warner, S. J., et al. (2023). Prolonged thermocline warming by near-inertial internal waves in the wakes of tropical cyclones. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(26), e2301664120. <https://doi.org/10.1073/pnas.2301664120>
- Cangialosi, J. P. (2015). *Tropical storm ida*. (Tech. Rep.). National Oceanic and Atmospheric Administration. Retrieved from [https://www.nhc.noaa.gov/data/tcr/AL102015\\_Ida.pdf](https://www.nhc.noaa.gov/data/tcr/AL102015_Ida.pdf)
- Danilov, S., Sidorenko, D., Wang, Q., & Jung, T. (2017). The finite-volume sea ice–ocean model (FESOM2). *Geoscientific Model Development*, *10*(2), 765–789. <https://doi.org/10.5194/gmd-10-765-2017>
- D’Asaro, E. A. (1985). The energy flux from the wind to near-inertial motions in the surface mixed layer. *Journal of Physical Oceanography*, *15*(8), 1043–1059. [https://doi.org/10.1175/1520-0485\(1985\)015<1043:TEFFTW>2.0.CO;2](https://doi.org/10.1175/1520-0485(1985)015<1043:TEFFTW>2.0.CO;2)
- de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology [Dataset]. *Journal of Geophysical Research*, *109*(C12). <https://doi.org/10.1029/2004JC002378>
- Eliot, S., & Lumpkin, R. (2008). Spectral description of oceanic near-surface variability. *Geophysical Research Letters*, *35*(5). <https://doi.org/10.1029/2007GL032874>
- Fan, Y., Yu, Z., & Jacobs, G. (2021). Vertical coordinate and resolution dependence of the second moment turbulent closure models and their limitations. *Ocean Modelling*, *168*, 101896. <https://doi.org/10.1016/j.ocemod.2021.101896>
- Fischer, T. (2020). Microstructure measurements during meteor cruise m119 [Dataset]. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.920592>
- Foltz, G. R., Evan, A. T., Freitag, H. P., Brown, S., & McPhaden, M. J. (2013). Dust accumulation biases in PIRATA shortwave radiation records. *Journal of Atmospheric and Oceanic Technology*, *30*(7), 1414–1432. <https://doi.org/10.1175/JTECH-D-12-00169.1>
- Foltz, G. R., Grodsky, S. A., Carton, J. A., & McPhaden, M. J. (2003). Seasonal mixed layer heat budget of the tropical Atlantic Ocean. *Journal of Geophysical Research*, *108*(C5). <https://doi.org/10.1029/2002JC001584>
- Foltz, G. R., Hummels, R., Dengler, M., Perez, R. C., & Araujo, M. (2020). Vertical turbulent cooling of the mixed layer in the Atlantic ITCZ and trade wind regions. *Journal of Geophysical Research: Oceans*, *125*(2), e2019JC015529. <https://doi.org/10.1029/2019JC015529>
- Foltz, G. R., Schmid, C., & Lumpkin, R. (2018). An enhanced PIRATA dataset for tropical Atlantic Ocean–atmosphere research [Dataset]. *Journal of Climate*, *31*(4), 1499–1524. <https://doi.org/10.1175/JCLI-D-16-0816.1>
- Fu, L.-L. (1981). Observations and models of inertial waves in the deep ocean. *Reviews of Geophysics*, *19*(1), 141–170. <https://doi.org/10.1029/RG019i001p00141>
- Garrett, C., & Munk, W. (1975). Space-time scales of internal waves: A progress report. *Journal of Geophysical Research*, *80*(3), 291–297. <https://doi.org/10.1029/JC080i003p00291>
- Gaspar, P., Grégoris, Y., & Lefevre, J.-M. (1990). A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station papa and long-term upper ocean study site. *Journal of Geophysical Research*, *95*(C9), 16179–16193. <https://doi.org/10.1029/JC095iC09p16179>
- Greatbatch, R. J. (1983). On the response of the ocean to a moving storm: The nonlinear dynamics. *Journal of Physical Oceanography*, *13*(3), 357–367. [https://doi.org/10.1175/1520-0485\(1983\)013<0357:OTROTO>2.0.CO;2](https://doi.org/10.1175/1520-0485(1983)013<0357:OTROTO>2.0.CO;2)
- Griffies, S. M., Levy, M., Adcroft, A. J., Danabasoglu, G., Hallberg, R. W., Jacobsen, D., et al. (2015). Theory and numerics of the community ocean vertical mixing (CVMIX) project (Tech. Rep.). <https://climatemodeling.science.energy.gov/technical-highlights/community-ocean-vertical-mixing-cvmix-project>
- Gutjahr, O., Brüggemann, N., Haak, H., Jungclaus, J. H., Putrasahan, D. A., Lohmann, K., & von Storch, J.-S. (2021). Comparison of ocean vertical mixing schemes in the Max Planck Institute Earth System Model (MPI-ESM1. 2). *Geoscientific Model Development*, *14*(5), 2317–2349. <https://doi.org/10.5194/gmd-14-2317-2021>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis [Dataset]. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Huang, C. J., Qiao, F., & Dai, D. (2014). Evaluating CMIP5 simulations of mixed layer depth during summer. *Journal of Geophysical Research: Oceans*, *119*(4), 2568–2582. <https://doi.org/10.1002/2013JC009535>
- Hummels, R., Dengler, M., Rath, W., Foltz, G. R., Schütte, F., Fischer, T., & Brandt, P. (2020). Surface cooling caused by rare but intense near-inertial wave induced mixing in the tropical Atlantic. *Nature Communications*, *11*(1), 3829. <https://doi.org/10.1038/s41467-020-17601-x>
- Jochum, M., Briegleb, B. P., Danabasoglu, G., Large, W. G., Norton, N. J., Jayne, S. R., et al. (2013). The impact of oceanic near-inertial waves on climate. *Journal of Climate*, *26*(9), 2833–2844. <https://doi.org/10.1175/JCLI-D-12-00181.1>
- Jochum, M., Danabasoglu, G., Holland, M., Kwon, Y.-O., & Large, W. G. (2008). Ocean viscosity and climate. *Journal of Geophysical Research*, *113*(C6). <https://doi.org/10.1029/2007JC004515>
- Jochum, M., & Potemra, J. (2008). Sensitivity of tropical rainfall to Banda Sea diffusivity in the community Climate System Model. *Journal of Climate*, *21*(23), 6445–6454. <https://doi.org/10.1175/2008JCLI2230.1>
- Korn, P., Brüggemann, N., Jungclaus, J. H., Lorenz, S. J., Gutjahr, O., Haak, H., et al. (2022). Icon-o: The ocean component of the icon earth system model—Global simulation characteristics and local telescoping capability. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2021MS002952. <https://doi.org/10.1029/2021MS002952>
- Kuwano-Yoshida, A., Sasaki, H., & Sasai, Y. (2017). Impact of explosive cyclones on the deep ocean in the north pacific using an eddy-resolving ocean general circulation model. *Geophysical Research Letters*, *44*(1), 320–329. <https://doi.org/10.1002/2016GL071367>

- Large, W. G., & Crawford, B. G. (1995). Observations and simulations of upper-ocean response to wind events during the ocean storms experiment. *Journal of Physical Oceanography*, 25(11), 2831–2852. [https://doi.org/10.1175/1520-0485\(1995\)025<2831:OASOUO>2.0.CO;2](https://doi.org/10.1175/1520-0485(1995)025<2831:OASOUO>2.0.CO;2)
- Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, 32(4), 363–403. <https://doi.org/10.1029/94RG01872>
- Lévy, M., Klein, P., Tréguier, A.-M., Iovino, D., Madec, G., Masson, S., & Takahashi, K. (2010). Modifications of gyre circulation by sub-mesoscale physics. *Ocean Modelling*, 34(1–2), 1–15. <https://doi.org/10.1016/j.ocemod.2010.04.001>
- Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, 20(4), 851–875. <https://doi.org/10.1029/RG020i004p00851>
- Munk, W., & Wunsch, C. (1998). Abyssal recipes ii: Energetics of tidal and wind mixing. *Deep Sea Research Part I: Oceanographic Research Papers*, 45(12), 1977–2010. [https://doi.org/10.1016/S0967-0637\(98\)00070-3](https://doi.org/10.1016/S0967-0637(98)00070-3)
- Oke, T. R. (2002). *Boundary layer climates*. Routledge.
- Pacanowski, R. C., & Philander, S. G. H. (1981). Parameterization of vertical mixing in numerical models of tropical oceans. *Journal of Physical Oceanography*, 11(11), 1443–1451. [https://doi.org/10.1175/1520-0485\(1981\)011<1443:POVMIN>2.0.CO;2](https://doi.org/10.1175/1520-0485(1981)011<1443:POVMIN>2.0.CO;2)
- Pollard, R. T. (1970). On the generation by winds of inertial waves in the ocean. *Deep-Sea Research and Oceanographic Abstracts*, 17(4), 795–812. [https://doi.org/10.1016/0011-7471\(70\)90042-2](https://doi.org/10.1016/0011-7471(70)90042-2)
- Pollard, R. T., & Millard, R. C. (1970). Comparison between observed and simulated wind-generated inertial oscillations. *Deep-Sea Research and Oceanographic Abstracts*, 17(4), 813–821. [https://doi.org/10.1016/0011-7471\(70\)90043-4](https://doi.org/10.1016/0011-7471(70)90043-4)
- Price, J. F. (1981). Upper ocean response to a hurricane. *Journal of Physical Oceanography*, 11(2), 153–175. [https://doi.org/10.1175/1520-0485\(1981\)011<0153:UORTAH>2.0.CO;2](https://doi.org/10.1175/1520-0485(1981)011<0153:UORTAH>2.0.CO;2)
- Raja, K. J., Buijsman, M. C., Shriver, J. F., Arbic, B. K., & Siyanbola, O. (2022). Near-inertial wave energetics modulated by background flows in a global model simulation. *Journal of Physical Oceanography*, 52(5), 823–840. <https://doi.org/10.1175/JPO-D-21-0130.1>
- Rimac, A., Von Storch, J.-S., & Eden, C. (2016). The total energy flux leaving the ocean's mixed layer. *Journal of Physical Oceanography*, 46(6), 1885–1900. <https://doi.org/10.1175/JPO-D-15-0115.1>
- Rimac, A., von Storch, J.-S., Eden, C., & Haak, H. (2013). The influence of high-resolution wind stress field on the power input to near-inertial motions in the ocean. *Geophysical Research Letters*, 40(18), 4882–4886. <https://doi.org/10.1002/grl.50929>
- Santoso, A., McPhaden, M. J., & Cai, W. (2017). The defining characteristics of ENSO extremes and the strong 2015/2016 El Niño. *Reviews of Geophysics*, 55(4), 1079–1129. <https://doi.org/10.1002/2017RG000560>
- Schafstall, J., Dengler, M., Brandt, P., & Bange, H. (2010). Tidal-induced mixing and diapycnal nutrient fluxes in the Mauritanian upwelling region. *Journal of Geophysical Research*, 115(C10). <https://doi.org/10.1029/2009JC005940>
- Scholz, P., Sidorenko, D., Danilov, S., Wang, Q., Koldunov, N., Sein, D., & Jung, T. (2022). Assessment of the finite-volume sea ice–ocean model (FESOM2.0)—Part 2: Partial bottom cells, embedded sea ice and vertical mixing library CVMIX. *Geoscientific Model Development*, 15(2), 335–363. <https://doi.org/10.5194/gmd-15-335-2022>
- Scholz, P., Sidorenko, D., Gurses, O., Danilov, S., Koldunov, N., Wang, Q., et al. (2019). Assessment of the finite-volume sea ice–ocean model (FESOM2.0)—Part 1: Description of selected key model elements and comparison to its predecessor version. *Geoscientific Model Development*, 12(11), 4875–4899. <https://doi.org/10.5194/gmd-12-4875-2019>
- Steele, M., Morley, R., & Ermold, W. (2001). Phc: A global ocean hydrography with a high-quality arctic ocean. *Journal of Climate*, 14(9), 2079–2087. [https://doi.org/10.1175/1520-0442\(2001\)014<2079:PAGOHW>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<2079:PAGOHW>2.0.CO;2)
- Stewart, S. R. (2016). *2015 atlantic hurricane season*. (Tech. Rep.). National Oceanic and Atmospheric Administration. Retrieved from [https://www.nhc.noaa.gov/data/tcr/summary\\_atlc\\_2015.pdf](https://www.nhc.noaa.gov/data/tcr/summary_atlc_2015.pdf)
- Van Roekel, L., Adcroft, A. J., Danabasoglu, G., Griffies, S. M., Kauffman, B., Large, W. G., et al. (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>
- Vincent, E. M., Lengaigne, M., Madec, G., Vialard, J., Samson, G., Jourdain, N. C., et al. (2012). Processes setting the characteristics of sea surface cooling induced by tropical cyclones. *Journal of Geophysical Research*, 117(C2). <https://doi.org/10.1029/2011JC007396>
- Wunsch, C., & Ferrari, R. (2004). Vertical mixing, energy, and the general circulation of the oceans. *Annual Review of Fluid Mechanics*, 36(1), 281–314. <https://doi.org/10.1146/annurev.fluid.36.050802.122121>
- Zhai, X., Greatbatch, R. J., Eden, C., & Hibiya, T. (2009). On the loss of wind-induced near-inertial energy to turbulent mixing in the upper ocean. *Journal of Physical Oceanography*, 39(11), 3040–3045. <https://doi.org/10.1175/2009JPO4259.1>
- Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, 15(3), 779–808. <https://doi.org/10.5194/os-15-779-2019>

## CHAPTER CONTENTS

- 3 Automated tuning 42
  - 3.1 Gaussian Emulators 42
    - 3.1.1 The Linear Model 43
    - 3.1.2 The Nonlinear Model 45
  - 3.2 Article II – Fast and Efficient: Bayesian Optimization with GPU Acceleration for Ocean Models 47
    - 3.2.1 The Critical Richardson number 48



## Automated tuning

Automated calibration methods are increasingly used to test model response to changing parameter values or, in other words, to *quantify the model parametric uncertainty*. For example, Souza et al. (2020) calibrate the KPP scheme to reproduce Large Eddy Simulation (LES) convective mixing. Their method requires  $10^6$  iterations to sample the parameter probability distributions, and thus is limited to single column models. Williamson, Blaker, and Sinha (2017) are able to reduce the number of samples needed by using a Gaussian emulator. In their work, they search the 24-dimensional parameter space of an OGCM to find ranges of parameters which best simulate the observed temperature and salinity profiles. The method, first introduced in Williamson et al. (2015), is called *history matching*, and uses three waves of 400 simulations.

The use of emulators greatly reduces the number of simulations needed for model calibration. The Calibrate-Emulate-Sample (CAS) method introduced by Cleary et al. (2021) proposes emulating the parameter-to-data map of an expensive inverse problem, such that a Gaussian emulator can be sampled for calibration instead.

Bayesian optimization (BO), in contrast, samples the parameter space iteratively. The surrogate model is constructed at each step to maximize the use of information gained from new simulations. This can be highly beneficial in OGCM and full GCM studies, especially when simulations have exceedingly high computational cost. In this chapter, the mechanism behind Gaussian emulators is summarised<sup>1</sup>. The second article in Section 3.2 demonstrates how BO can be used to tune the TKE scheme in a coarse-resolution OGCM.

### 3.1 GAUSSIAN EMULATORS

A *Gaussian process* is a “a collection of random variables, any finite number of which have a joint Gaussian distribution” (Williams and Rasmussen, 2006):

$$g(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (3.1)$$

In other words, the *probability distribution* of  $g(\mathbf{x})$  at any  $\mathbf{x} \in \mathcal{X}^d$  is Gaussian with mean  $m(\mathbf{x})$  and variance  $k(\mathbf{x}, \mathbf{x}')$ . The variance is determined by the *kernel* function, which is a measure of *correlation* between any two points

1. A more extensive and general discussion about the modern uses of Gaussian processes can be found in Williams and Rasmussen (2006). The summary in this PhD thesis is inspired by their Chapter 2, as well as Stoustrup (2021).

$\mathbf{x}, \mathbf{x}' \in \mathcal{X}^d$ . The function  $g(\mathbf{x})$  is a  $\mathcal{GP}$  sample. It shows what a function obeying the probability distribution defined by  $\mathcal{GP}$  may look like.

Essentially, Gaussian emulators use Gaussian processes to interpolate between the (usually few) known data points, with the additional advantage of providing the uncertainty estimate of the model. In the following text, the linear regression model is derived and generalized for non-linear problems.

### 3.1.1 The Linear Model

Given a set of  $n$  datapoints  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ , where  $x, y \in \mathbb{R}$ , the simplest regression model is a linear function:

$$f(x) = wx, \quad y = f(x) + \epsilon, \quad (3.2)$$

ONE-DIMENSIONAL LINEAR MODEL

where  $w$  is the slope of the line and  $\epsilon$  is noise, which is assumed to be Gaussian with zero mean and standard deviation  $\sigma_n$ :  $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ . The noise represents the observational uncertainty. The aim now is to find  $w$  which can best describe the data. In the Bayesian formalism, this can be achieved by optimizing the *posterior*, or the *probability of the model parameters given the data*:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \quad : \quad p(w|\mathbf{x}, y) = \frac{p(y|\mathbf{x}, w)p(w)}{p(y|\mathbf{x})}. \quad (3.3)$$

BAYES RULE

The *prior* over the model parameters reflects the first guess of  $w$ . It is assumed that  $w \sim \mathcal{N}(0, \sigma_p^2)$ , i.e. that the slope of the line in the model is drawn from a Gaussian distribution with 0 mean and  $\sigma_p^2$  variance<sup>2</sup>:

$$\text{PRIOR} \quad p(w) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{w^2}{2\sigma_p^2}\right), \quad (3.4)$$

2. This is equivalent to presupposing that there is *no correlation* in the data. Broadly speaking, this is a good first guess when we have no idea what  $f(x)$  may look like.

The *marginal likelihood* is independent of the model parameters and serves as a normalizing constant:

$$p(y|\mathbf{x}) = \int p(y|\mathbf{x}, w)p(w)dw. \quad (3.5)$$

MARGINAL LIKELIHOOD

We assume that the set of inputs is independent, and thus can calculate the *likelihood*, i.e. the *probability distribution of the observations given the model*:

$$\begin{aligned}
p(\mathbf{y}|\mathbf{x}, \mathbf{w}) &= \prod_{i=1}^n p(y_i|x_i, \mathbf{w}) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(y_i - \mathbf{w}x_i)^2}{2\sigma_n^2}\right) \\
&= \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left(-\frac{|\mathbf{y} - \mathbf{w}\mathbf{x}|^2}{2\sigma_n^2}\right) \\
&= \mathcal{N}(\mathbf{w}\mathbf{x}, \sigma_n^2 I),
\end{aligned} \tag{3.6} \quad \text{LIKELIHOOD}$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are vectors containing all datapoints and  $|\cdot|$  symbolizes the Euclidean distance. The result is another Gaussian distribution with mean  $\mathbf{w}\mathbf{x}$  and a covariance matrix  $\sigma_n^2 I$ .

The posterior can be computed from Equations (3.3)-(3.6), isolating the part dependent on model parameters:

$$\begin{aligned}
p(\mathbf{w}|\mathbf{x}, \mathbf{y}) &\propto p(\mathbf{y}|\mathbf{x}, \mathbf{w})p(\mathbf{w}) \\
&\propto \exp\left(-\frac{|\mathbf{y} - \mathbf{w}\mathbf{x}|^2}{2\sigma_n^2} - \frac{\mathbf{w}^2}{2\sigma_p^2}\right) \\
&\propto \exp\left(-\frac{a}{2}(\mathbf{w} - a^{-1}\sigma_n^{-2}\mathbf{x} \cdot \mathbf{y})^2\right),
\end{aligned} \tag{3.7} \quad \text{POSTERIOR}$$

where  $a = \sigma_p^{-2} + \sigma_n^{-2}|\mathbf{x}|^2$ . Thus, the posterior is a Gaussian distribution with mean  $\bar{\mathbf{w}} = a^{-1}\sigma_n^{-2}\mathbf{x} \cdot \mathbf{y}$  and variance  $a^{-1}$ . Note that this expression only informs about the most likely value of  $\mathbf{w}$ , and the degree of certainty about it. The posterior is normalized, so the information about how well  $\mathbf{w}$  describes the data in relation to all other possible slope values is contained in the marginal likelihood in Eq. (3.5).

Finally, the *predictive distribution* for  $f_* \triangleq f(x_*)$  at a test point  $x_*$  defines the regression model. Since  $f(x_*)$  is linear in  $\mathbf{w}$ :

$$\begin{aligned}
\mathbb{E}[f(x_*)] &= \mathbb{E}[x_*\mathbf{w}] = x_*\mathbb{E}[\mathbf{w}] = x_*\bar{\mathbf{w}}, \\
\text{Var}[f(x_*)] &= \text{Var}[x_*\mathbf{w}] = x_*^2\text{Var}[\mathbf{w}] = x_*^2 a^{-1}.
\end{aligned} \tag{3.8} \quad \text{EXPECTATION AND VARIANCE AT A TEST POINT}$$

Thus,  $f(x_*)$  is normally distributed with mean  $x_*\bar{\mathbf{w}}$  and variance  $x_*^2 a^{-1}$ . The regression model is complete with a mean, a standard deviation, and the estimation of confidence of the fit given by the marginal likelihood.

This simple example provides a blueprint for how an emulator is constructed. The following two steps take us from a restrictive, one-dimensional linear model to a regression model that can predict a vast number of functions  $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ :

1. Expanding the linear model to allow multi-dimensional input and
2. Projecting the input vector  $\mathbf{x}$  into a feature space spanned by a set of basis functions  $\boldsymbol{\phi}(\mathbf{x})$  to allow nonlinearity.

### 3.1.2 The Nonlinear Model

We now have a set of  $n$  datapoints  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ , with  $\mathbf{x} \in \mathbb{R}^d$  and  $y \in \mathbb{R}$ . The linear model becomes:

$$f(x) = \mathbf{x}^\top \mathbf{w}, \quad y = f(\mathbf{x}) + \epsilon, \quad (3.9) \quad \text{1. MULTI-DIMENSIONAL INPUT}$$

where  $\mathbf{w}$  is a  $d$ -dimensional vector of weights, which is drawn from the normal distribution  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p)$ .  $\Sigma_p$  is the covariance matrix on the weights. To consider all datapoints simultaneously, the vector  $\mathbf{y}$  contains all outputs and the matrix  $X$  is a stack of vector inputs  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ . Likelihood, posterior and the predictive distribution follow from the one-dimensional example, and the extended formulas are listed in Table 3.1.

We can define a set of basis functions  $\{\phi_k(\mathbf{x}) | k \in 1, \dots, N\}$  to construct a vector function  $\boldsymbol{\phi} : \mathbb{R}^d \rightarrow \mathbb{R}^N$ , which maps the  $d$ -dimensional vector  $\mathbf{x}$  into an  $N$ -dimensional feature space with a new weight vector  $\boldsymbol{\omega}$ :

$$f(\mathbf{x}) = \sum_k^N \phi_k(\mathbf{x}) \omega_k = \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\omega}. \quad (3.10) \quad \text{2. NONLINEAR MODEL}$$

Table 3.1: GP regression equations.

Equation	Multi-dimensional input	Nonlinear models
Dataset	$\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$	$\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$
Model	$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{w}$	$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\top \boldsymbol{\omega}$
Likelihood	$\mathcal{N}(X^\top \mathbf{w}, \sigma_n^2 I)$	$\mathcal{N}(\boldsymbol{\Phi}^\top \boldsymbol{\omega}, \sigma_n^2 I)$
Posterior covariance	$A^{-1} = (\Sigma_p^{-1} + \sigma_n^{-2} X X^\top)^{-1}$	$\mathcal{A}^{-1} = (\Sigma_p^{-1} + \sigma_n^{-2} \boldsymbol{\Phi} \boldsymbol{\Phi}^\top)^{-1}$
Posterior mean	$\bar{\mathbf{w}} = \sigma_n^{-2} A^{-1} X \mathbf{y}$	$\bar{\boldsymbol{\omega}} = \sigma_n^{-2} \mathcal{A}^{-1} \boldsymbol{\Phi} \mathbf{y}$
Predictive distribution	$\mathcal{N}(\mathbf{x}_*^\top \bar{\mathbf{w}}, \mathbf{x}_*^\top A^{-1} \mathbf{x}_*)$	$\mathcal{N}(\boldsymbol{\phi}(\mathbf{x}_*)^\top \bar{\boldsymbol{\omega}}, \boldsymbol{\phi}(\mathbf{x}_*)^\top \mathcal{A}^{-1} \boldsymbol{\phi}(\mathbf{x}_*))$

Given that the basis functions are independent of  $\omega$ ,  $f(\mathbf{x})$  remains a linear function of the weights, but its flexibility is considerably expanded.<sup>3</sup>

In the general case, we can define  $\Phi$  as the  $n \times N$  matrix of projected input points  $[\phi(\mathbf{x}_1) \cdots \phi(\mathbf{x}_n)]$ . The equations for the likelihood  $p(\mathbf{y}|\Phi, \omega)$ , the posterior  $p(\omega|\Phi, \mathbf{y})$  mean and variance, and the predictive distribution at a test point are listed in Table 3.1.

The ingenuity of the method emerges with the realization that the predictive distribution of the nonlinear model can be re-written in terms of  $K = \Phi^\top \Sigma_p \Phi$ :

$$p(f_*|\mathbf{x}_*, X, \mathbf{y}) = \mathcal{N}(\phi_*^\top \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \mathbf{y}, \phi_*^\top \Sigma_p \phi_* - \phi_*^\top \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \Phi^\top \Sigma_p \phi_*). \quad (3.11)$$

The notation eliminates the need to compute  $\phi(\mathbf{x})$ , because Eq. (3.11) can be expressed *solely* as a function of the *kernel*:  $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \Sigma_p \phi(\mathbf{x}')$ . This powerful formulation is called *the kernel trick*.

The predictive distribution in Eq. 3.11 can be expanded to include a set of test points stacked in a matrix  $X_*$  with the corresponding vector  $\mathbf{f}_* \triangleq \mathbf{f}(X_*)$ :

$$\begin{aligned} p(\mathbf{f}_*(X_*)|X_*, \mathbf{y}) &= \mathcal{N}(\bar{\mathbf{f}}_*, \text{Cov}(\mathbf{f})), \\ \bar{\mathbf{f}}_* &= K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} \mathbf{y}, \\ \text{Cov}(\mathbf{f}) &= K(X_*, X_*) - K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*). \end{aligned} \quad (3.12)$$

where the matrices  $K(X, X')$  contain the covariances  $k(\mathbf{x}, \mathbf{x}')$ . When  $\sigma_n = 0$ , i.e. when the output is assumed to be noise-free, the formulation is identical to Equation (2) in Article II.

3. For example, consider the basis set  $\{x^{k-1} | k \in \{1, 2, 3, 4\}\}$  defining the mapping  $\phi(x) = [1, x, x^2, x^3]^\top$ . The Gaussian regression model becomes:

$$f(x) = \omega_1 + \omega_2 x + \omega_3 x^2 + \omega_4 x^3.$$

The model  $f(x)$  is now able to fit a 3<sup>rd</sup> order polynomial to input points  $x$ , but the predictive distribution at a test point  $x_*$  can still be easily found by computing the linear transformation of the multivariate normal posterior  $p(\omega|\Phi, \mathbf{y})$ , just like in Eq. 3.8:

$$\begin{aligned} \mathbb{E}[\phi(x_*)^\top \omega] &= \phi(x_*)^\top \bar{\omega} \\ \text{Cov}[\phi(x_*)^\top \omega] &= \phi(x_*)^\top \mathcal{A}^{-1} \phi(x_*). \end{aligned}$$

### 3.2 ARTICLE II — FAST AND EFFICIENT: BAYESIAN OPTIMIZATION WITH GPU ACCELERATION FOR OCEAN MODELS

Article II: “*Fast and Efficient: Bayesian Optimization with GPU Acceleration for Ocean Models*” (Mrozowska et al., n.d.) has been submitted to Journal of Geophysical Research: Machine Learning and Computation.

In the second article, the tested method for improving vertical mixing parameterization is automated calibration. Bayesian optimization is used to tune the TKE scheme in the Versatile Ocean Simulator (Veros, Häfner et al., 2018). The main aims of the study are:

1. To determine whether there exists a set of TKE parameter values which can reduce MLD biases in Veros;
2. To demonstrate the robustness of VerOpt and the potential of Bayesian optimization for OGCM tuning.

The answer to the first point turns out to be *no*. In the optimization experiment where the target is MLD climatology, the default parameter values are among those which give the best results. Unfortunately, *the best* in this context means MLD bias of 43%. This is not satisfactory, but Figure 6 in Article II shows that TKE parameter values do not impact the global MLD distribution enough to improve the biases globally. The shallow bias in the Southern Hemisphere can be rectified by increasing the ratio of  $c_k c_\epsilon^{-1}$ , but this simultaneously worsens MLD everywhere else.

Consistent with SMC column studies (e.g. Burchard and Bolding, 2001; Umlauf and Burchard, 2003), the vertical extent of mixing in the model is solely<sup>4</sup> dependent on  $Ri_c \propto c_k c_\epsilon^{-1}$ . The geographical MLD distribution, however, is not affected. The TKE parameterization can therefore only scale the local rate of mixing supplied by the atmospheric forcing.

The second point is achieved in the TWIN experiment, a supervised optimization where the target MLD is a field simulated using Veros. VerOpt can identify TKE parameter values which reproduce the target MLD up to 1.18% accuracy within 180 model simulations. In fact, parameter values that reproduce the target MLD with a comparably low error are already found after 60 simulations.

An unexpected yet significant outcome of the study is the quantification of energy efficiency gain with the use of Veros JAX backend (Häfner, Nuterman, and Jochum, 2021). Even the most efficient calibration techniques require a number of ocean model simulations which carry a high computational cost. We find that the use of Lumi GPUs for Veros simulations can cost up to 17 times less energy compared to CPU computation. Of course, computational

4. In SMCs, this would be the steady-state  $Ri_{st}$  dependent on the stability functions, but in the simplified TKE scheme, the stability functions are constant and  $Ri_{st} = Ri_c$ .

resources vary across research groups and can therefore be highly subjective<sup>5</sup>. It should be stressed, however, that the adaptability of the model can significantly reduced the resources needed for calibration.

5. In Article II, the accessible GPU resources are considerably newer than the accessible CPU resources.

### 3.2.1 The Critical Richardson number

Due to the ranges chosen for the parameters  $c_k$  and  $c_\epsilon$ , some tested  $Ri_c$  values are much larger than the widely accepted  $Ri_c < 1$ . In the submitted version of the manuscript, a detail about the calculation of the critical Richardson number is omitted, which results in an overestimation of  $Ri_c$  values in the range  $Ri_c > 1.5$ .

In Article II, we use the quasi-equilibrium state  $P_s + P_b = \epsilon$  to derive  $Ri_c$  as a function of the TKE parameters. BD<sup>6</sup> arrive at the same relation by setting the vertical tke transport to zero:

6. Blanke and Delecluse, 1993

$$\frac{\partial \bar{e}}{\partial t} = \sqrt{2}c_k N \left( \frac{1}{Ri} - \frac{2c_k + P_{rt}c_\epsilon}{2c_k P_{rt}} \right) \bar{e}. \quad (3.13)$$

This expression implies that tke is in the quasi-equilibrium state when:

$$Ri = Ri_c = \frac{P_{rt}}{1 + 2^{-1}P_{rt}c_\epsilon c_k^{-1}}, \quad (3.14)$$

which can be re-written as a polynomial function of  $Ri$  using the definition  $P_{rt} = 6.6Ri$ :

$$f(Ri) = \frac{6.6c_\epsilon}{2c_k} Ri^2 - 5.6Ri. \quad (3.15)$$

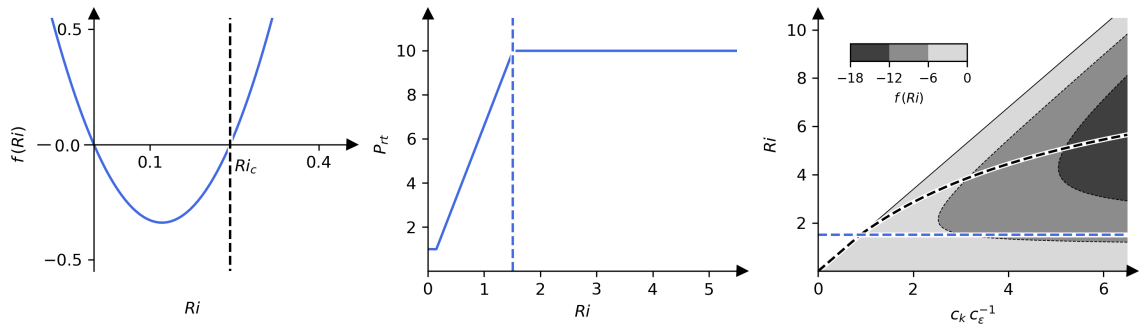


Figure 3.1: The left panel shows the polynomial  $f(Ri)$  with the two roots 0 and  $Ri_c \approx 0.23$ . The Prandtl as a function of  $Ri_c$  is plotted in the middle panel. For values larger than  $Ri \approx 1.5$  (dashed blue line),  $P_{rt}$  is constant. The critical Richardson number is a function of  $P_{rt}$ : as a result, for values greater than 1.5,  $f(Ri)$  overestimates  $Ri_c$  (right panel). Black dashed line on the right panel shows  $Ri_c$  as a function of  $c_k c_\epsilon^{-1}$ .  $Ri_c$  defined as a root of  $f(Ri)$  can be compared to the corrected  $Ri_c$  computed with constant  $P_{rt}$  for  $Ri > 1.5$ .

The polynomial  $f(Ri)$  is plotted in Fig. 3.1 for the default TKE parameterization<sup>7</sup> and as a function of  $Ri$  and  $c_k c_\epsilon^{-1}$ . The two roots of  $f(Ri)$  are zero and  $Ri_c$ .  $P_{rt} = 6.6Ri$  is only true in the model when  $1 < P_{rt} < 10$ , i.e.  $f(Ri)$  only determines  $Ri_c$  in the range  $6.6 > Ri^{-1} > 0.66$ <sup>8</sup>. Outside of this range,  $Ri_c$  can be determined using Eq. (3.14) with constant Prandtl number. For  $P_{rt} = 1$ , Eq. (3.14) is approximately equal to the root of  $f(Ri)$ . For  $P_{rt} = 10$ , Eq. 3.3 increases slower than  $1.7c_k c_\epsilon^{-1}$ , which means that the  $Ri_c$  values in the range  $Ri_c > 1.5$  are smaller than reported in Article II. This does not change the main scientific results of the work, but affects the discussion on how sensitive the model is to  $Ri_c > 1.5$ . The highest  $Ri_c$  value in the OBS experiment is 9.9, which corresponds to  $Ri_c \approx 5.38$  when computed with constant  $P_{rt} = 10$ .

7.  $c_k = 0.1$  and  $c_\epsilon = 0.7$ .

8. Approximately  $0.15 < Ri < 1.5$ .

In the approximation of Eq. (3.14),

$$\bar{e} \propto \exp \left[ (Ri^{-1} - Ri_c^{-1}) t \right], \quad (3.16)$$

which illustrates the local exponential growth of tke for  $Ri < Ri_c$  and decay for  $Ri > Ri_c$ . In Article II, we find that the effect of ML deepening with increasing  $Ri_c$  is strongest at high latitudes where stratification is weakest. Eq. (3.5) shows that increasing  $Ri_c$  will disproportionately affect regions with typical  $Ri$  values much smaller than  $Ri_c$ .



# Fast and efficient: Bayesian optimization with GPU acceleration for ocean models

Marta Agnieszka Mrozowska<sup>1</sup>, James Avery<sup>2,3</sup>, Aster Stoustrup<sup>1</sup>,  
Roman Nuterman<sup>1</sup>, Carl-Johannes Johnsen<sup>3</sup>, Markus Jochum<sup>1</sup>

<sup>1</sup>Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup>Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark

<sup>3</sup>Department of Computer Science DIKU, University of Copenhagen, Copenhagen, Denmark

## Key Points:

- We use Bayesian optimization to find the TKE scheme parameter values that minimize the mixed layer depth bias in the ocean model Veros.
- GPU computation with the Veros JAX backend consumes up to 17 times less energy than CPU computation.
- The MLD bias is smallest when  $Ri_c < 1$ , and the default parameterization falls within this range.

---

Corresponding author: Marta Agnieszka Mrozowska, [marta.mrozowska@nbi.ku.dk](mailto:marta.mrozowska@nbi.ku.dk)

**Abstract**

Ocean general circulation models (OGCMs) contain numerous parameterizations of sub-grid scale processes. The parameter tuning procedure is rarely reported and often done by hand. We present an automated alternative: Bayesian optimization, a method which has recently emerged as a frontier in expensive black box optimization. VerOpt, a Python package for the ocean model Veros, adapts Bayesian optimization to climate model tuning. We use VerOpt to identify a set of parameter values of the Turbulent Kinetic Energy (TKE) closure scheme that minimize mixed layer depth (MLD) bias in Veros. We present the results of two optimization procedures: TWIN and OBS. The goal is to minimize modeled MLD error relative to a target map. In TWIN, the target is MLD simulated using Veros with a known parameterization. The ratio of two TKE parameters  $c_k c_\epsilon^{-1}$ , proportional to the critical Richardson number  $Ri_c$ , is the dominant factor in setting the global MLD. After 180 model simulations, the lowest error in the TWIN experiment is 1.18%. In OBS, the target is MLD climatology. The MLD bias is smallest when  $Ri_c < 1$ , and the default TKE parameterization falls within this range. We find, however, that altering the TKE parameterization is not sufficient to reduce the significant MLD bias of 42.62%. The OBS experiment results indicate that the TKE scheme parameters are not the dominant source of MLD bias in Veros. We discuss other possible sources of MLD bias, as well as the potential of extending of the optimization procedure to other parameterizations.

**Plain Language Summary**

In the ocean, the mixing of temperature and salinity across layers of equal density happens on centimeter scales. In comparison, the spacing between the depth layers in ocean models is on the order of meters. Vertical mixing schemes approximate the effect of unresolved turbulence, but they require parameter tuning to accurately represent it. We use Bayesian optimization (BO) to determine which parameter values of the Turbulent Kinetic Energy (TKE) mixing scheme minimize the mixed layer depth (MLD) biases in the ocean model Veros. Veros Optimizer (VerOpt) adapts BO to climate modeling problems. We present the results of two optimization sequences: TWIN and OBS. In the TWIN experiment, the optimizer is tasked with finding a set of parameters which reproduce MLD simulated with Veros. VerOpt identifies a range of parameter values which reproduce the target MLD up to 1.18% accuracy. In the OBS experiment, the target is the observed MLD. We find that the default values are among the parameters that simulate MLD closest to observed. Furthermore, significant MLD biases in Veros cannot be rectified by tuning the TKE scheme. We discuss other possible sources of bias, such as structural limitations of the scheme and other model parameterizations.

**1 Introduction**

The representation of vertical mixing in ocean models remains a major challenge. The horizontal resolution of ocean general circulation models (OGCMs) has increased to order of kilometers in the recent decade in an effort to eliminate the need for parameterizing mesoscale physics and the associated mixing (e.g. Small et al., 2014; Korn et al., 2022). The same tactic cannot be applied to diapycnal mixing, because it happens on centimeter to meter scales, too small to be resolved in the foreseeable future. Therefore, the vertical turbulent fluxes of momentum, temperature and salinity have to be computed by turbulence closure schemes.

Accurate mixed layer depth (MLD) representation is necessary for realistic energy exchange between the ocean and atmospheric components in coupled models, as biases in MLD lead to sea surface temperature (SST) anomalies. In the tropical ocean, SST anomalies can lead to restructuring of the global climate (Jochum & Potemra, 2008). A significant challenge in the modeling of vertical mixing in the ocean are the limited ob-

65 servations. Global datasets of diffusivity profiles have been emerging over the past decade,  
 66 but the data coverage is orders of magnitude lower than that of density profiles (Waterhouse  
 67 et al., 2014). MLD climatology is thus the best indication for the performance of ver-  
 68 tical mixing schemes. Despite the community effort spanning multiple decades, MLD bi-  
 69 ases still persist in OGCMs (Huang et al., 2014; Treguier et al., 2023).

70 Only some of the parameters in vertical mixing schemes have a solid physical ba-  
 71 sis. For example, in the Turbulent Kinetic Energy (TKE) closure in Gaspar et al. (1990),  
 72 the parameters  $c_k$  and  $c_\epsilon$  are picked based on the empirical value of the mixing coeffi-  
 73 cient  $\gamma_{R_f}$ , which is highly uncertain (Gregg et al., 2018). The OGCM parameters which  
 74 need to be selected by the modeler are often tuned by hand, and the process is rarely  
 75 reported (Mauritsen et al., 2012). Here, *tuning* refers to changing the model parameters  
 76 slightly such that the model output resembles observations as closely as possible. This  
 77 process carries with itself a number of risks, such as over-fitting (Williamson et al., 2017).  
 78 There’s therefore a need for methods which systematize and automate OGCM param-  
 79 eter tuning.

80 In this paper, we demonstrate such a method by tuning the TKE parameter val-  
 81 ues in the ocean model Veros (Versatile Ocean Simulator, Häfner et al., 2018, 2021). We  
 82 use the Python package VerOpt (Veros Optimizer, Stoustrup, 2021), which is designed  
 83 to adapt Bayesian optimization (BO) to climate modeling problems. In recent decades,  
 84 BO has become a popular method to handle expensive black box optimization and has  
 85 been successfully applied in disciplines such as robotics, environmental sensing, drug de-  
 86 sign, as well as in machine learning for hyper-parameter optimization (Shahriari et al.,  
 87 2015; Wang et al., 2023). The method’s popularity has catalyzed its development, as well  
 88 as software availability. In Python, BO models can be constructed using BoTorch (Balandat  
 89 et al., 2020). VerOpt is built in the PyTorch ecosystem (Ansel et al., 2024) and imple-  
 90 ments BoTorch optimization routines and functions.

91 We present the results of two optimization experiments: TWIN and OBS. In each  
 92 experiment, a set of TKE parameter values that can best simulate a target MLD map  
 93 is identified by VerOpt. The optimization process and the fundamentals of BO are in-  
 94 troduced in section 2. The setup of the two optimization procedures is described in sec-  
 95 tion 3, including Veros model specifications, the description of the TKE scheme, the choice  
 96 of MLD climatology (section 3.1), the optimizer specifications (section 3.2) and the com-  
 97 putational resources (section 3.3). The results of the TWIN experiment are reported in  
 98 section 4.1, and of the OBS experiment - in section 4.2. The discussion in section 5 ex-  
 99 plores possible improvements and adaptations of the optimization procedure, followed  
 100 by conclusions in section 6.

## 101 2 Bayesian optimization with VerOpt

102 Bayesian optimization is a black box optimization method, i.e. it locates the global  
 103 minimum of an unknown scalar objective function  $f(\mathbf{x}): \mathcal{X}^d \rightarrow \mathbb{R}$ . Two aspects of the  
 104 method make it especially well suited for climate science problems. Firstly, the optimiza-  
 105 tion does not rely on the gradient of the objective function. A full simulation spanning  
 106 millions of iterations is difficult to differentiate, and for most existing earth system mod-  
 107 els it is completely out of reach. Furthermore, the gradients may contain singularities  
 108 or bifurcations, which break the optimization sequence. Secondly, BO requires relatively  
 109 few objective function evaluations. Optimization algorithms which do not rely on gra-  
 110 dient information often depend on tens of thousands to millions of input points. Run-  
 111 ning such vast numbers of OGCM simulations is unfeasible. Efficient sampling is there-  
 112 fore a top priority in OGCM calibration.

113 The data efficiency of BO stems from utilizing Gaussian process (GP) regression  
 114 to construct a model of  $f(\mathbf{x})$  over  $\mathcal{X}$ . The so-called *surrogate* of the objective function

115 is then used to compute the *acquisition* function, which highlights the regions in the pa-  
 116 rameter space that are most likely to contain the global objective function minimum:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{X}^d} f(\mathbf{x}).^1 \quad (1)$$

117 The objective function is sampled iteratively following the three key steps of the opti-  
 118 mization procedure:

- 119 1. the construction of the surrogate model, which emulates the distribution of possi-  
 120 ble objective functions given our existing knowledge,
- 121 2. the calculation of the acquisition function from the mean and uncertainty of the  
 122 surrogate model, and
- 123 3. the minimization of the acquisition function to determine the new objective func-  
 124 tion coordinates to evaluate.

125 The steps involved in optimizing an unknown function  $f: \mathcal{X} \rightarrow \mathbb{R}$  with VerOpt  
 126 are summarized in Figure 1. The following description of the method is highly condensed;  
 127 more in-depth explanations can be found in Stoustrup (2021) and Williams and Rasmussen  
 128 (2006).

129 Let us consider a set of objective function evaluations:  $\{(\mathbf{x}_i, f_i) | i = 1, \dots, n\}$ . Here,  
 130  $\mathbf{x}_i$  is an input point in the parameter space  $\mathcal{X}^d$  with  $d$  parameters. The  $d \times n$  matrix  
 131  $X$  contains all  $n$  input points. The corresponding objective function values  $f_i(\mathbf{x}_i)$   
 132 are contained in the vector  $\mathbf{f}(X) \in \mathbb{R}^n$ . GP regression allows us to construct a predictive  
 133 distribution of the surrogate model  $\mathbf{f}_*$  at a set of test points  $X_*$ :

$$p(\mathbf{f}_*(X_*) | X_*, \mathbf{f}(X)) = \mathcal{N}(K(X_*, X)K(X, X)^{-1}\mathbf{f}, \quad (2)$$

$$K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)).$$

134  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$  symbolizes the normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ . We  
 135 assume noise-free input  $y_i = f_i$ . The Gram matrices  $K(X, X')$  contain covariances be-  
 136 tween any two set of points  $X$  and  $X'$ , such that each element of the matrix can be de-  
 137 scribed as  $k_{ij} = k(\mathbf{x}_i, \mathbf{x}'_j)$ . For example, an element of the covariance matrix between  
 138 the test and input points,  $K(X_*, X)$ , is equal to  $k(\mathbf{x}_{*i}, \mathbf{x}_j)$ .

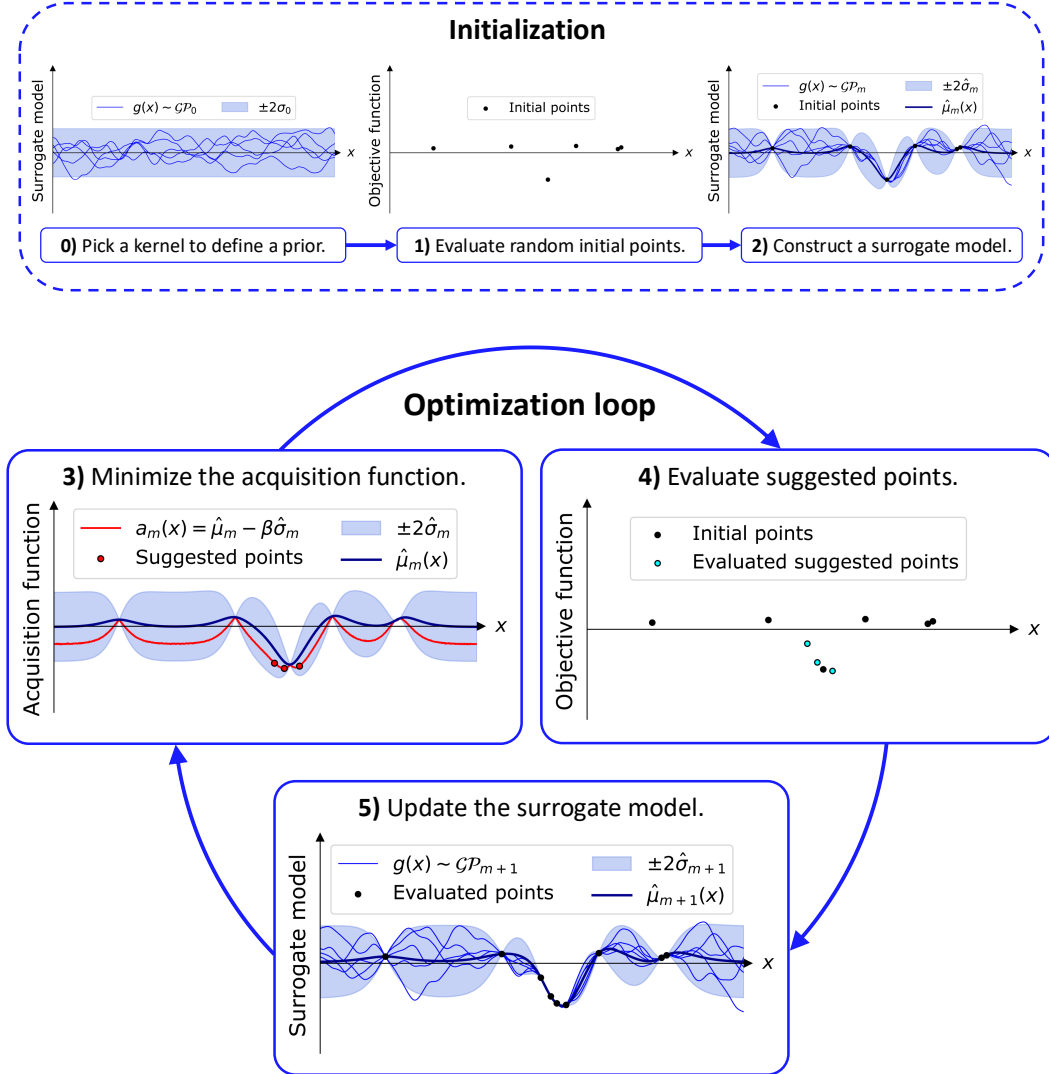
139 The *kernel*  $k(\mathbf{x}, \mathbf{x}')$  determines the shape and behavior of the surrogate model; it  
 140 contains the *prior* information about the objective function. The Matérn kernel is the  
 141 default VerOpt option:

$$k(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \sqrt{2\nu} \frac{\|\mathbf{x} - \mathbf{x}'\|}{l} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{\|\mathbf{x} - \mathbf{x}'\|}{l} \right). \quad (3)$$

142  $\Gamma(\cdot)$  is the gamma function,  $K_\nu$  is the modified Bessel function of the order  $\nu$ , and  $l$  is  
 143 the characteristic length scale. The Matérn kernel has high utility across a wide range  
 144 of black box optimization problems because:

- 145 1. It is  $\lceil \nu \rceil - 1$  differentiable. Climate model output, just as most physical systems,  
 146 is not expected to be infinitely-differentiable and smooth, so the hyper-parameter  
 147  $\nu$  provides control over an important aspect of the objective function surrogate.
- 148 2. Aside from  $\nu$ , the kernel contains only one other hyper-parameter - the charac-  
 149 teristic length scale  $l$ , which can be determined by maximizing the marginal like-  
 150 lihood in Eq. (4).

<sup>1</sup> VerOpt follows the Bayesian optimization literature custom where the objective function is *maximized* rather than minimized. However, since the aim of the optimization in this paper is to minimize model biases, we present it as such for the sake of readability and consistency.



**Figure 1.** Bayesian optimization with VerOpt demonstrated using a one-dimensional example. The prior is a Gaussian process  $\mathcal{GP}_0(0, k(x, x'))$ , where  $k(x, x')$  is the Matérn kernel. Five sample functions  $g_{j \in \{1, \dots, 5\}}(x) \sim \mathcal{GP}_0$  illustrate our assumption about the shape of the objective function. The shaded region indicates the 95% confidence bound. In panel 1), the optimization is initialized by evaluating  $n_{init} = 6$  random objective function coordinates. In the first optimization round  $m = 1$ , the surrogate model is constructed based on the initial points (panel 2). For  $m \in \{1, \dots, n_{rounds}\}$ , the acquisition function  $a_m(x)$  with  $\beta = 1$  is calculated from the surrogate model mean  $\hat{\mu}_m$  and uncertainty  $\hat{\sigma}_m$  (panel 3). The acquisition function minima are suggested to evaluate. Panel 4 shows the updated set of objective function values: the basis for the updated surrogate model  $\mathcal{GP}_{m+1}$  in panel 5. Each optimization round consists of the steps shown in panels 3-5.

151 In VerOpt, the default choice for  $\nu$  is 2.5, making the kernel twice-differentiable. The  
 152 optimal length scale is determined by maximizing the marginal log likelihood (MLL) of  
 153 the surrogate model:

$$\log(p(\mathbf{f}(X)|\boldsymbol{\theta})) = -\frac{1}{2}\mathbf{f}^T K(X, X)\mathbf{f} - \frac{1}{2}\log|K(X, X)| - \frac{n}{2}\log(2\pi), \quad (4)$$

154 where  $\boldsymbol{\theta}$  is the hyper-parameter vector. In our case,  $\boldsymbol{\theta} = \mathbf{l} = [l_1, \dots, l_d]$ . The anisotropic  
 155 Matérn kernel allows the length scale parameter  $l$  to be different for each dimension in  
 156  $\mathcal{X}$ . The radial distance  $r = \frac{\|\mathbf{x}-\mathbf{x}'\|}{l}$  in Eq. (3) becomes  $r = \left\| \frac{\mathbf{x}-\mathbf{x}'}{\mathbf{l}} \right\|$ .

157 The three terms in the MLL equation represent the model fit, the model complex-  
 158 ity, and the normalization factor, respectively. The term  $-\frac{1}{2}\mathbf{f}^T K(X, X)\mathbf{f}$  measures how  
 159 well the surrogate model covariance fits with the variance of the objective function val-  
 160 ues. At large length scales, the GP model becomes approximately constant, while low  
 161 length scales produce highly nonlinear models. The punishment term for model complex-  
 162 ity  $\frac{1}{2}\log|K(X, X)|$  is therefore high at low length scales where the surrogate is flexible  
 163 and can fit the evaluated points well. The bad fit punishment  $\frac{1}{2}\mathbf{f}^T K(X, X)\mathbf{f}$  grows ex-  
 164 ponentially as  $l$  increases. This makes the MLL convex and the maximum easy to find.

165 After the optimal hyper-parameter vector  $\hat{\boldsymbol{\theta}}$  is determined, the predictive distri-  
 166 bution of the surrogate model  $p(\mathbf{f}_*)$  is used to compute the acquisition function. The de-  
 167 fault VerOpt acquisition function is the modified Lower Confidence Bound (LCB, Sriniva-  
 168 vas et al., 2009):

$$a(\mathbf{x}) = \hat{\mu}(\mathbf{x}) - \hat{\sigma}(\mathbf{x})\beta - r\hat{\sigma}(\mathbf{x})\beta\gamma, \quad (5)$$

169 where  $\beta$  and  $\gamma$  are positive acquisition function hyper-parameters,  $\hat{\mu}$  and  $\hat{\sigma}$  are the mean  
 170 and standard deviation of the optimal surrogate model, and  $r$  is a random number drawn  
 171 from  $\mathcal{N}(0, 1)$ . The standard form of LCB does not include the  $\gamma$  term. The stochastic  
 172 noise is added to improve optimizer performance when the surrogate model is nearly con-  
 173 stant along a parameter. The value of  $\gamma = 0.01$  is sufficiently small to only affect the  
 174 optimization in that case.

175 The acquisition function is optimized using the L-BFGS-B algorithm with multi-  
 176 ple restarts (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Box constraints,  
 177 Byrd et al., 1995), as implemented in the scikit-learn package (Pedregosa et al., 2011).  
 178 The new set of the optimal coordinates is evaluated by the objective function, i.e. by run-  
 179 ning a set of simulations with parameterizations  $\{\mathbf{x}_1, \dots, \mathbf{x}_{n_{evals}}\}$ . In total, the opti-  
 180 mization produces  $n$  objective function evaluations, where  $n = n_{init} + n_{Bayes}$ . The  $n_{init}$   
 181 initial points are randomly picked from a uniform distribution over the finite bounds of  
 182 the parameter space  $\mathcal{X}$ . The  $n_{Bayes}$  points are the minima of the acquisition function.  
 183 At each round of optimization,  $n_{evals}$  points are evaluated. The total number of opti-  
 184 mization rounds is therefore  $n_{rounds} = n_{Bayes}/n_{evals}$ . The evaluated set of parameter  
 185 values for which the objective function is minimized is denoted by  $\boldsymbol{\chi}^*$ .

186 In the optimization sequence in Figure 1, the task is to determine the parameter  
 187  $x$  for which  $f(x)$  is minimized. Panel 0 shows 5 random vector draws from the GP prior  
 188  $\mathcal{GP}_0(0, k_0(x, x'))$ , where  $k_0(x, x')$  is the Matérn kernel with  $\nu = 2.5$  and  $l_0 = 0.25$  (Eq. (3)).  
 189 The shaded region indicates  $\mu_0(x) \pm 2\sigma_0(x)$ . The optimization is initialized by evalu-  
 190 ating  $n_{init}$  randomly selected initial points (panel 1). The GP prior distribution is condi-  
 191 tioned on the evaluated points (Eq. (2)) to construct the first surrogate model at op-  
 192 timization round  $m = 1$ . The optimal length scale  $\hat{l}_1$  is found by locating the MLL max-  
 193 imum from Eq. (3). The draws from  $\mathcal{GP}_1$  show the possible shape of the objective func-  
 194 tion given the information gained from evaluating the initial points. In panel 3, the ac-  
 195 quisition function  $a(x)$  (Eq. (5)) computed using the mean and variance from the sur-  
 196 rogate model is plotted. The result of the acquisition function optimization is the new  
 197 set of suggested points: the optimizer’s best guesses for  $\mathbf{x}^*$ . In panel 4, the new evalu-  
 198 ated points are plotted alongside the initial points. The full set is used to update the sur-  
 199 rogate model to  $\mathcal{GP}_{m+1}$  shown in panel 5. Each optimization round  $m$  produces:

- 200 • the surrogate model  $\mathcal{GP}_m(\hat{\mu}_m(x), k_m(x, x'))$  with the kernel defined using the op-
- 201 timal length scale  $\hat{l}_m$ ,
- 202 • the surrogate model mean  $\hat{\mu}_m(x)$  and standard deviation  $\hat{\sigma}_m(x)$ ,
- 203 • the acquisition function  $a_m(x)$  and
- 204 • the set of points to evaluate  $\{x_{m,1}, \dots, x_{m,n\_evals}\}$ .

205 Panels 3-5 summarise the steps of the optimization procedure repeated for  $m \in \{1, \dots, n_{rounds}\}$ .

### 206 3 Experimental setup

207 The parameterization to optimize is the Turbulent Kinetic Energy (TKE, Gaspar  
208 et al., 1990) closure scheme in Veros. The optimal TKE parameters are found for two  
209 setups: the coarse setup with  $4^\circ \times 4^\circ$  horizontal resolution (Veros  $4^\circ \times 4^\circ$ ), and the stan-  
210 dard setup with  $1^\circ \times 1^\circ$  horizontal resolution (Veros  $1^\circ \times 1^\circ$ ). The model employs a reg-  
211 ular, three-dimensional staggered Arakawa-C grid. Both setups have 60 vertical layers  
212 with spacing monotonously increasing towards the bottom. In Veros  $1^\circ \times 1^\circ$ , the surface  
213 vertical resolution is 2 m, while in Veros  $4^\circ \times 4^\circ$  it is 4 m. The models are forced with  
214 monthly ERA-Interim (ECMWF Reanalysis v4, Dee et al., 2011) climatology. Further  
215 description of Veros components can be found in Häfner et al. (2018).

216 The duration of all simulations presented in this paper is 30 model years, which  
217 is long enough for the tropical and mid-latitude mixed layer to spin up (Liu et al., 1994).  
218 The integration begins from initial conditions based on climatological temperature and  
219 salinity profiles from World Ocean Atlas 2005 (WOA05, Locarnini et al., 2006; Antonov  
220 et al., 2006). The average of the last two years is used for analysis.

221 The TKE scheme is a 1.5-order turbulence closure commonly used in ocean mod-  
222 els, such as Nucleus for European Modelling of the Ocean (NEMO, Madec & the NEMO team,  
223 2016) included in the Ocean Model Intercomparison Project (OMIP, Griffies et al., 2016).  
224 Eddy viscosity  $K_m$  is parameterized as:

$$K_m = c_k l_k \bar{e}^{1/2}, \quad (6)$$

225 where  $l_k$  is the mixing length scale,  $c_k$  is the mixing coefficient and  $\bar{e} = \frac{1}{2}(\overline{u'^2} + \overline{v'^2} + \overline{w'^2})$   
226 is the turbulent kinetic energy. The zonal, meridional and vertical velocities are sym-  
227 bolized by  $u$ ,  $v$  and  $w$ . The overbar and apostrophe indicate the mean and fluctuating  
228 components in Reynolds decomposition, respectively:  $u = \bar{u} + u'$ . The prognostic equa-  
229 tion for  $\bar{e}$  is:

$$\frac{\partial \bar{e}}{\partial t} = \frac{\partial}{\partial z} \left( K_e \frac{\partial \bar{e}}{\partial z} \right) - c_\epsilon \frac{\bar{e}^{3/2}}{l_\epsilon} - K_h N^2 + K_m S h^2, \quad (7)$$

230 where  $K_e = \alpha_{tke} K_m$  is the vertical diffusive flux of  $\bar{e}$ ,  $K_h = K_m P_{rt}^{-1}$  is the eddy dif-  
231 fusivity with Prandtl number  $P_{rt}$ ,  $N$  is the buoyancy frequency,  $S h^2 = (\frac{\partial \bar{u}}{\partial z})^2 + (\frac{\partial \bar{v}}{\partial z})^2$   
232 is the vertical squared shear, and  $l_\epsilon$  is the dissipation length scale. Aside from the three  
233 parameters  $\alpha_{tke}$ ,  $c_k$  and  $c_\epsilon$ , the model allows for setting a limit for minimum eddy vis-  
234 cosity and diffusivity:  $min_{K_m}$  and  $min_{K_h}$ , respectively.

235 The computation of the length scales  $l_k$  and  $l_\epsilon$  in Veros follows Blanke and Delecluse  
236 (1993):

$$l_k = l_\epsilon = 2^{1/2} \bar{e}^{1/2} N^{-1}. \quad (8)$$

237 Furthermore, Prandtl number is dependent on the Richardson number  $Ri$ :  $P_{rt} = 6.6 Ri$ ,  
238  $P_{rt} \in [1, 10]$ . Assuming homogeneous and stationary turbulence, it is possible to de-  
239 rive the critical Richardson number  $Ri_c$  as a function of the TKE parameters. In this  
240 regime, Equation (7) becomes a balance between the buoyancy flux, shear production  
241 and dissipation:

$$K_m S h^2 - K_h N^2 - c_\epsilon \frac{\bar{e}^{3/2}}{l_\epsilon} = 0. \quad (9)$$

**Table 1.** The default values of the TKE parameters and their bounds.

parameter	default value	bounds
$c_k$	0.1	[0.05, 1.0]
$c_\epsilon$	0.7	[0.05, 1.0]
$\alpha_{tke}$	30.0	[0.5, 50]
$\log_{10}(\min_{K_m} [\text{m}^2 \text{ s}^{-1}])$	$\log_{10}(2 \cdot 10^{-4})$	[-7.0, -3.5]

242 After rearranging the terms, Eq. (9) becomes:

$$\frac{P_{rt}}{P_{rt} - Ri_c} Ri_c = \frac{6.6}{6.6 - 1} Ri_c = c_\epsilon^{-1} \bar{\epsilon}^{3/2} l_\epsilon N^2 K_m \quad (10)$$

243 Recalling that  $K_h = K_m P_{rt}^{-1}$  and using the length scale definition from Eq. (8), we get  
244  $K_m = c_k 2^{1/2} \bar{\epsilon} N^{-1}$ . We end up with  $Ri_c$  as a function of  $c_k$  and  $c_\epsilon$ :

$$Ri_c \approx 1.7 c_k c_\epsilon^{-1}. \quad (11)$$

245 Tuning the  $c_k c_\epsilon^{-1}$  ratio is therefore equivalent to picking the critical Richardson number  
246 of the model.

247 The default parameter values of the TKE scheme in Veros are reported in Table 1;  
248 the parameterization is used in the control simulations Default Veros  $1^\circ \times 1^\circ$  and De-  
249 fault Veros  $4^\circ \times 4^\circ$ . For the optimization, parameter bounds are necessary to define the  
250 finite search space. The choices are motivated as follows:

- 251 • The default  $c_k$  and  $c_\epsilon$  values are taken from Gaspar et al. (1990). Since the two  
252 parameters scale  $\bar{\epsilon}$  production and dissipation, a natural pick for the scale for both  
253 parameters is  $[0, 1]$ . However, at  $c_k$  values close to zero, MLD becomes noisy - likely  
254 due to numerical diffusivity dominating the model diffusivity. With no sink for  
255  $\bar{\epsilon}$ , Eq. (7) might become unstable. The minimum values for  $c_k$  and  $c_\epsilon$  should there-  
256 fore be larger than zero. We pick 0.05 as the minimum bound for both.
- 257 • The default value for  $\alpha_{tke}$  in Gaspar et al. (1990) is 1, but has since been changed  
258 to 30 in modern TKE implementations (e.g., Williamson et al., 2017; Brüggemann  
259 et al., 2024). For the bounds, we adapt the range used in Williamson et al. (2017).
- 260 • The Prandtl number is set to 10 in the ocean interior, i.e.  $\min_{K_h} = 0.1 \min_{K_m}$ .  
261 The default  $\min_{K_m} = 2 \cdot 10^{-4}$  is picked based on the mean observed diffusiv-  
262 ity:  $K_h \approx 10^{-5} \text{ m}^2 \text{ s}^{-1}$  (Ledwell et al., 1998). The lowest values of  $K_h$  in mi-  
263 crostructure profiles are typically on the order of  $10^{-7} \text{ m}^2 \text{ s}^{-1}$ . When  $\min_{K_h}$  is  
264 below this value, it effectively acts as a limit for numerical stability and does not  
265 have a significant effect on the diffusivity profile. At the higher end of the range,  
266  $\min_{K_h}$  imposes a constant background diffusivity below the mixed layer.

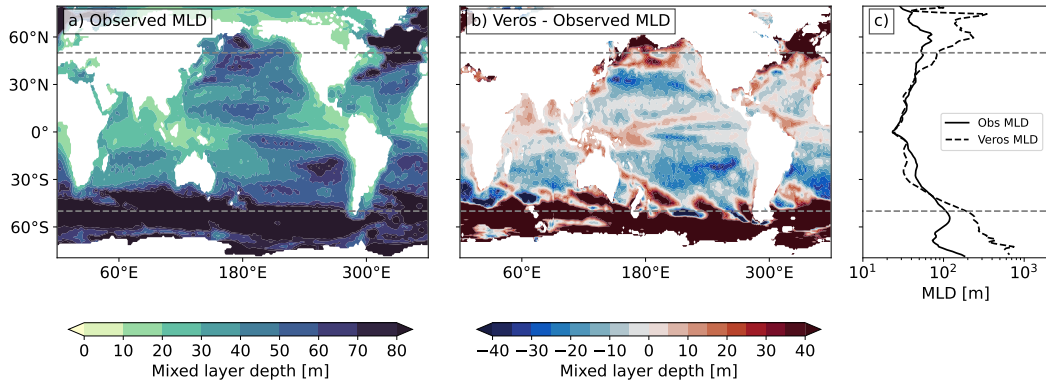
### 267 3.1 Mixed layer depth definition and climatology

268 The depth of the mixed layer does not have a unique definition (de Boyer Montégut  
269 et al., 2004). Density profile observations outnumber diffusivity profiles by orders of mag-  
270 nitude, therefore most MLD definitions use a density threshold criterion of the form:

$$\rho(z) = \rho(z_{\text{ref}}) + \delta\rho, \quad (12)$$

271 where  $\rho$  is the density,  $z$  is the vertical coordinate,  $z_{\text{ref}}$  is the reference depth and  $\delta\rho$  is  
272 a fixed density difference. MLD is thus defined as the maximum depth for which the sea-  
273 water density is  $\delta\rho$  larger than the density at the reference depth.





**Figure 2.** a) The observed mean annual MLD map and b) Default Veros  $1^\circ \times 1^\circ$  MLD bias. In c), the zonal average MLD is compared between observations (solid black line) and Veros (dashed black line). The gray dashed lines indicate the meridional extent of the experiments.

274 We pick the latest dataset from de Boyer Montégut (2022) as the target MLD cli-  
 275 matology. The dataset uses a fixed density threshold of  $\delta\rho = 0.03 \text{ kg m}^{-3}$  and  $z_{\text{ref}} =$   
 276 10 m for computing the MLD. The climatology is based on shipboard observations, drifter  
 277 and buoy data. Almost half of the profiles are from ARGO float measurements. The cli-  
 278 matology and the Default  $1^\circ \times 1^\circ$  Veros MLD bias are shown in Figure 2. MLD in Veros  
 279 is generally too shallow in the tropics, with the exception of the Intertropical Conver-  
 280 gence Zone (ITCZ) region and the northern Indian Ocean.

281 Treguier et al. (2023) find that MLD biases as large as 100 m can arise from us-  
 282 ing model reference depth  $z_{\text{ref}}$  which is different than the one used to compute the MLD  
 283 climatology. We therefore set up the vertical grid in Veros so the  $z_{\text{ref}}$  from the climatol-  
 284 ogy matches the model reference depth closely. In Veros  $1^\circ \times 1^\circ$ ,  $z_{\text{ref}} = 10.7 \text{ m}$ ; in Veros  
 285  $4^\circ \times 4^\circ$ ,  $z_{\text{ref}} = 11.2 \text{ m}$ . MLD is computed at every time step. The depth at which the  
 286 density criterion in Eq. (12) is fulfilled is found by linearly interpolating between the ver-  
 287 tical layers:

$$\text{MLD} = \frac{\rho(\text{MLD}) - \rho(z_b)}{\rho(z_a) - \rho(z_b)} \cdot (z_a - z_b) + z_b, \quad (13)$$

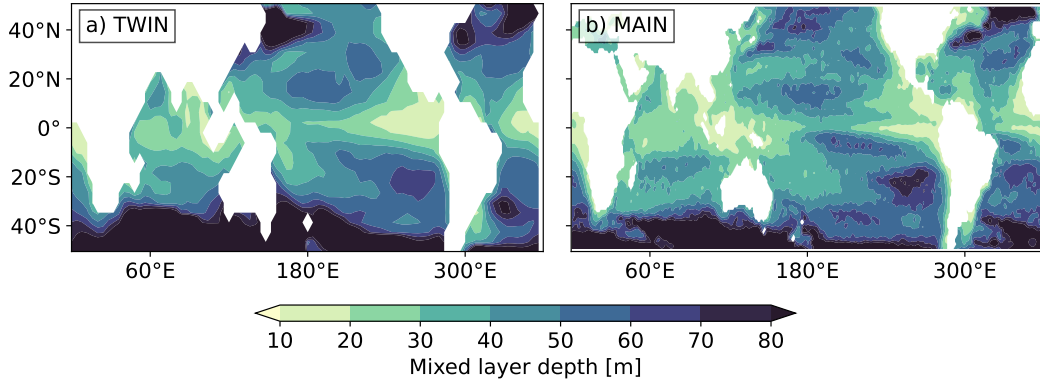
288 where  $z_a$  is the depth layer above MLD, and  $z_b$  is the depth layer below MLD. The den-  
 289 sity  $\rho(\text{MLD}) = \rho(z_{\text{ref}}) + 0.03 \text{ kg m}^{-3}$ .

### 290 3.2 VerOpt setup

291 We present the results of two optimization studies: the TWIN experiment, where  
 292 the target is the Default  $4^\circ \times 4^\circ$  Veros MLD, and the OBS experiment, where the tar-  
 293 get is the MLD climatology. Table 2 outlines the details of the optimization setups.

294 The goal of the TWIN experiment is two-fold: 1) to assess the sensitivity of the  
 295 modeled MLD to the TKE parameters and 2) to determine how well VerOpt can identi-  
 296 fy a set of known parameters. When ocean simulations are compared to climatology,  
 297 a degree of bias is expected even in a perfect model due to observational uncertainty. In  
 298 practice, the model is never perfect, and structural limitations independent of the op-  
 299 timized parameterization are another source of error. The TWIN experiment enables us  
 300 to assess the performance of the optimizer independently of these two factors.

301 The error metric in the TWIN experiment is the natural logarithm of the root mean  
 302 squared relative error of the MLD map simulated with a parameter set  $\mathbf{x}$  ( $\text{MLD}^{\text{sim}}$ ) rel-



**Figure 3.** The target MLD maps for a) the TWIN experiment and b) the OBS experiment.

303 active to the target MLD,  $MLD^{target}$ :

$$RMSRE = \sqrt{\frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M \left( \frac{MLD_{i,j}^{sim} - MLD_{i,j}^{target}}{MLD_{i,j}^{target}} \right)^2}. \quad (14)$$

304 The dimensions of the map are  $N \times M$ , with  $i$  and  $j$  symbolizing the zonal and meridional indices, respectively. The target maps for both experiments are shown in Figure 3.  
 305 MLD is by definition larger than 10 m, so division by zero is not a risk.  
 306

307 The space  $\mathcal{X}_{TWIN} = \{c_k, c_\epsilon, \alpha_{tke}, min_{K_M}\}$  includes the full TKE parameterization.  
 308 The experiment is conducted using the  $4^\circ \times 4^\circ$  Veros setup. The hyper-parameter  $\beta$  of  
 309 the acquisition function (Eq. (5)) is set to 0.8. The choice is rooted in the tendency of  
 310 the optimizer to pick boundary parameter values in the test experiments. The bounds  
 311 for the characteristic length scale in the Matérn kernel (Eq. (3)) are unchanged from VerOpt  
 312 default settings:  $[0.1, 2]$ . The optimization starts with 6 rounds of initialization where  
 313 30  $n_{init}$  points are evaluated, followed by 30 rounds of optimization, with a total of 150  
 314  $n_{Bayes}$  evaluated points.

**Table 2.** The setup of the two optimization studies.

	TWIN	OBS
target MLD map	simulated Veros MLD	MLD climatology
objective function	$\ln(RMSRE)$	RMSRE
horizontal resolution	$4^\circ$	$1^\circ$
latitude bounds	$[52^\circ S, 52^\circ N]$	$[49.5^\circ S, 49.5^\circ N]$
parameter space $\mathcal{X}$	$c_k, c_\epsilon, \alpha_{tke}, min_{K_M}$	$c_k, c_\epsilon$
total number of simulations $n$	180	40
number of iterations	36	20
number of optimization rounds $n_{rounds}$	30	15
number of simulations per iteration $n_{evals}$	5	2
number of initial points $n_{init}$	30	10
number of Bayes points $n_{Bayes}$	150	30
acquisition function parameter $\beta$	0.8	3.0
acquisition function parameter $\gamma$	0.01	0.01
kernel hyper-parameter $\theta$ bounds	$[0.1, 2]$	$[0.1, 2]$

315 In the OBS experiment, the  $1^\circ \times 1^\circ$  Veros setup is used. The goal of the optimiza-  
 316 tion is to minimize the model MLD bias relative to climatology. The parameter space  
 317 is reduced to two dimensions:  $\mathcal{X}_{\text{OBS}} = \{c_k, c_\epsilon\}$  based on the results of the TWIN ex-  
 318 periment further discussed in section 4.1. The reduction of dimensionality is motivated  
 319 by computational limitations, as a single  $1^\circ \times 1^\circ$  30-year simulation takes about 80 times  
 320 longer to complete than a  $4^\circ \times 4^\circ$  run of the same length. A reduced parameter space  
 321 requires fewer evaluated points to explore effectively.

322 The objective function formulation is modified in the OBS experiment to reduce  
 323 its sensitivity to small fluctuations in error. There, we drop the natural logarithm, mak-  
 324 ing  $f_{\text{OBS}}(\mathbf{x})$  equal to the RMSRE from Eq. (14). The total of  $n = 40$  points are eval-  
 325 uated, including 10  $n_{\text{init}}$  points in 5 rounds of initialization and 30  $n_{\text{Bayes}}$  points in 15  
 326 rounds of optimization.

### 327 3.3 Performance and computational resources

328 Veros can run both fully on CPU, fully on GPU (for full float64 calculations), and  
 329 a mixture (for GPUs that only have fast float32 performance, offloading numerically sensi-  
 330 tive float64 solvers to CPU). We conducted the  $4^\circ \times 4^\circ$  TWIN experiment on the Dan-  
 331 ish Center for Climate Computing (DC<sup>3</sup>) CPU cluster *Aegir*, comprised of an Intel Xeon  
 332 E5-2650v4 2.2GHz CPUs, running each simulation with 24 threads. The  $1^\circ \times 1^\circ$  OBS  
 333 experiment was run on our local workstation with mixed calculations on commodity NVIDIA  
 334 RTX 3080 GPUs and AMD Threadripper Ryzen 1950X 16-Core CPUs (*Threadripper*).  
 335 Continued experiments have been conducted on the Lumi-G supercomputer, fully on AMD  
 336 MI250X GPUs (*Lumi*).

337 On a single Lumi MI250X GPU, we can compute about 75  $4^\circ \times 4^\circ$  model years  
 338 per hour, and 1.2  $1^\circ \times 1^\circ$  model years per hour, calculated in full 64-bit floating point  
 339 precision. The average energy consumption for the GPU while running a  $1^\circ \times 1^\circ$  cal-  
 340 culation was 196W, i.e. 163Wh per model year. The float32/float64 mixed GPU/CPU-  
 341 calculations on Threadripper computes at a similar speed, and uses about 300W on the  
 342 GPU and roughly 150W on CPU, i.e.  $\sim 375$ Wh per model year. The fully-CPU calcu-  
 343 lations on Aegir were benchmarked for the same  $1^\circ$  setup by Häfner et al. (2021) for 0.115  
 344 model years using on DC<sup>3</sup> Aegir, completing in 3297s and consuming 320Wh (Tables 1  
 345 and A1 in Häfner et al., 2021), which corresponds to 8 hours and 2782Wh per  $1^\circ \times 1^\circ$   
 346 model year. Thus the same calculation on a single MI250X GPU was not only nearly  
 347 10 times as fast as on the CPUs, but also consume 17 times less energy when running  
 348 fully on Lumi’s MI250X GPUs.

349 Our  $4^\circ \times 4^\circ$  experiments were run on the same system at DC<sup>3</sup>, using a single CPU  
 350 core per run, at 2.8 minutes per model year. Initially we ran the  $1^\circ \times 1^\circ$  experiments  
 351 on the same system, using 16 CPU cores per run, resulting in 11 hours per model year.  
 352 The move to GPUs and reduction to 50 minutes per  $1^\circ \times 1^\circ$  model year made experi-  
 353 mentation much more flexible. As the energy consumption for performing long term cli-  
 354 mate simulations is substantial – especially when performing many runs to optimize pa-  
 355 rameters – the final energy reduction by a factor 17 makes an enormous difference as to  
 356 which kinds of experiments can be justified. Due to the relatively coarse resolution of  
 357 the  $1^\circ \times 1^\circ$ , we ran the experiments on a single GPU per Veros-simulation. However,  
 358 Häfner et al. (2021) showed that Veros scales well up to at least 16 GPUs for  $0.1^\circ \times 0.1^\circ$   
 359 simulations, making high-resolution Veros runs feasible on multi-GPU based systems such  
 360 as Lumi-G.

## 4 Optimization results

### 4.1 TWIN experiment

After 36 optimization steps, the best simulation produces RMSRE  $\approx 1.18\%$ . The parameter values from this simulation,  $\mathbf{x}_{\text{TWIN}}^*$ , are reported in Table 3 together with percentage difference relative to the target values. While  $Ri_c$  is not tuned directly, it turns out to have the strongest influence on the objective function, thus we report it in Table 3 alongside the other parameters.

Two-dimensional scatter plots for each combination of parameters in  $\mathcal{X}_{\text{TWIN}}$  and the histograms of the evaluated points are shown in Figure 4. The histograms of  $c_k$  and  $c_\epsilon$  show well defined peaks, indicating high influence on the objective function. On the contrary, the  $\alpha_{tke}$  histogram has no peaks, and the  $min_{\kappa_M}$  histogram shows two sharp peaks at the boundary, both indicating low influence on the objective function. The lowest error scores are clustered in the upper left corner of the  $c_k$ - $c_\epsilon$  scatter plot.

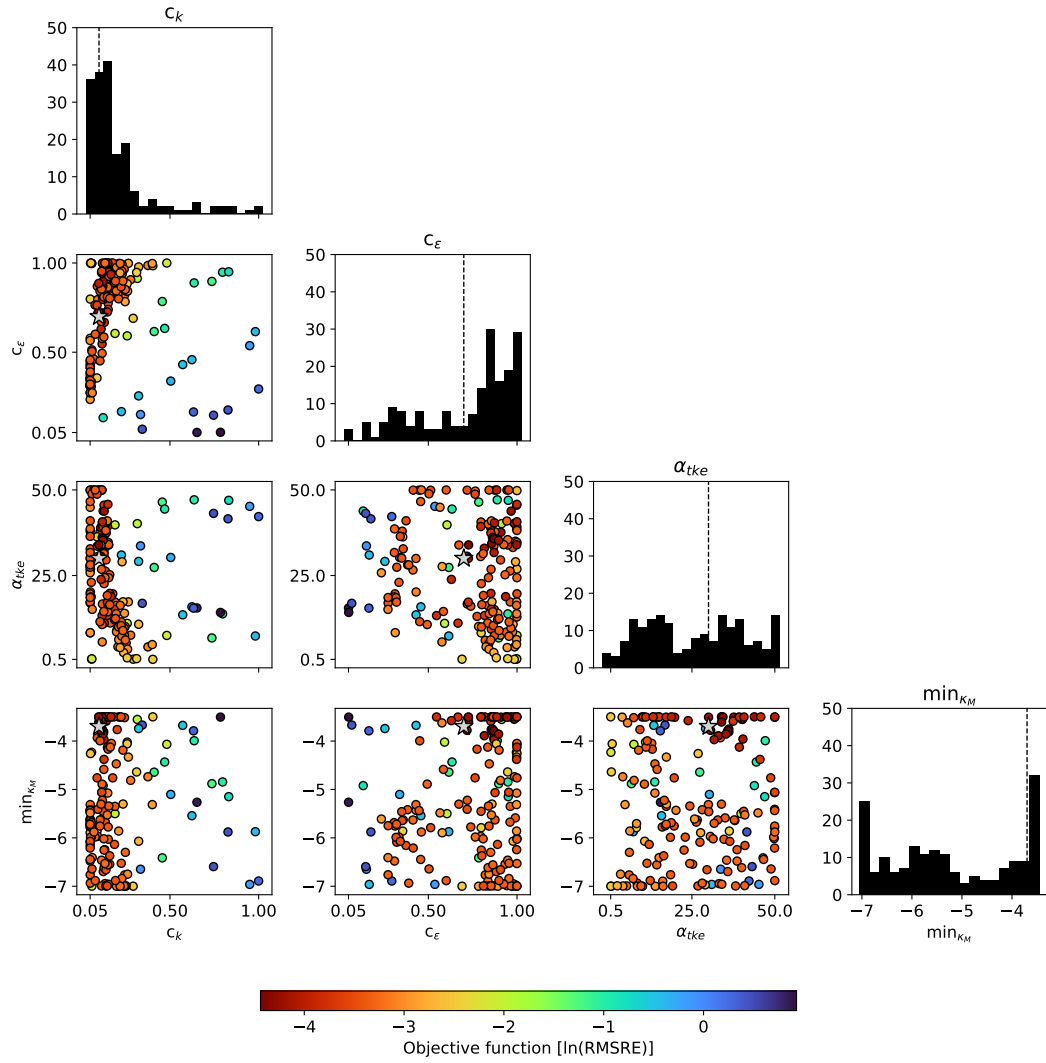
In Figure 5c),  $\ln(\text{RMSRE})$  is plotted as a function of the critical Richardson number  $Ri_c$ . The global MLD is unsurprisingly tightly correlated with  $Ri_c$ , as seen in Fig. 5a). Increasing the  $Ri_c$  has the effect of deepening the global MLD. The deepening is more pronounced at mid-latitudes, where the stratification is weaker. In Fig. 5b), the objective function values are plotted against the number of evaluated points  $n$ . After initialization, suggested points result in RMSRE of less than 13%; 83% of the RMSRE values are below 5%, and 13% of the RMSRE values are below 2%.

The optimization results indicate that large regions of the TKE parameter space simulate MLD maps which are similar. Moreover, since  $Ri_c$  has the highest influence on the MLD, and since it is also a function of  $c_k c_\epsilon^{-1}$ , local RMSRE minima will occur for any parameterization where the  $c_k c_\epsilon^{-1}$  ratio is the same as default. This explains why the center of the  $c_\epsilon$  peak is shifted relative to the target value in Fig. 4. The relatively large parameter errors in Table 3 are also a reflection of this: while  $c_k$  and  $c_\epsilon$  are 21.6% and 24.2% larger than the default values, the ratio - and therefore also  $Ri_c$  - is only 2.1% off from target.

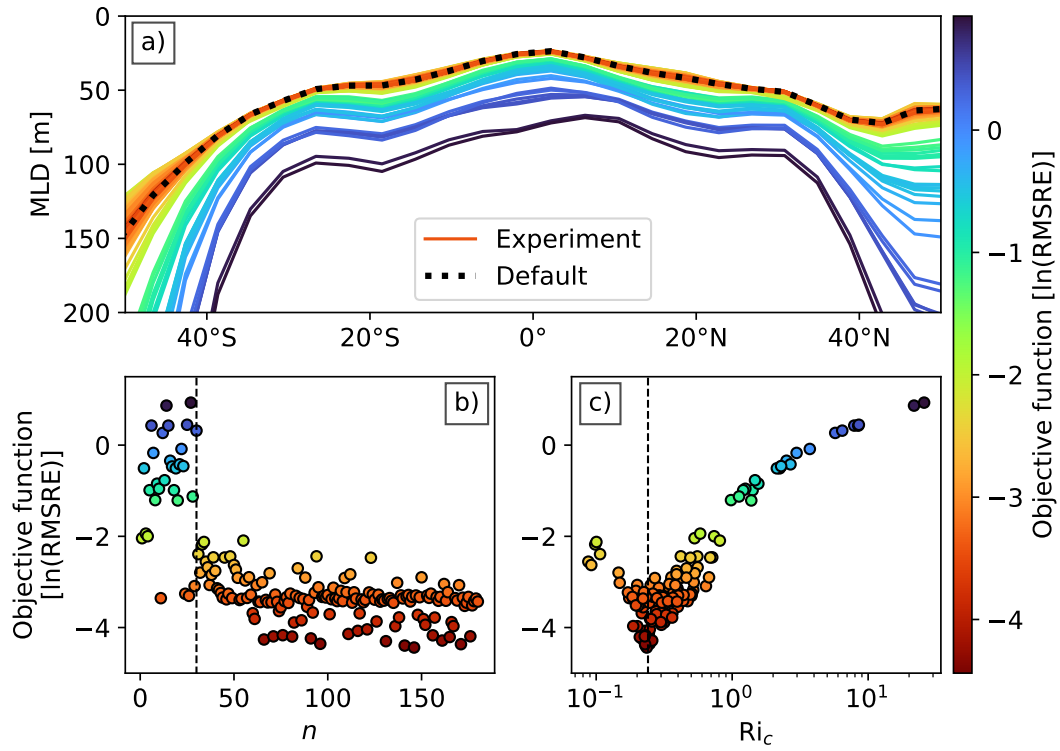
The TWIN experiment demonstrates the robustness of the tool. For the price of only 180 objective function evaluations, the optimizer is able to paint a comprehensive picture of the parameter space, identify the most influential parameters, and find a parameter set which simulates the target MLD to 1.18% accuracy. The ratio  $c_k c_\epsilon^{-1}$  estimated by the optimizer is only 2.1% higher than default. A natural next step which could improve the TWIN estimates for the TKE parameters would be to reduce the parameter space to a region where RMSRE is lowest, further increasing the sensitivity of the optimizer to small objective function fluctuations.

**Table 3.** VerOpt best estimates for the TKE parameterization in the TWIN experiment.

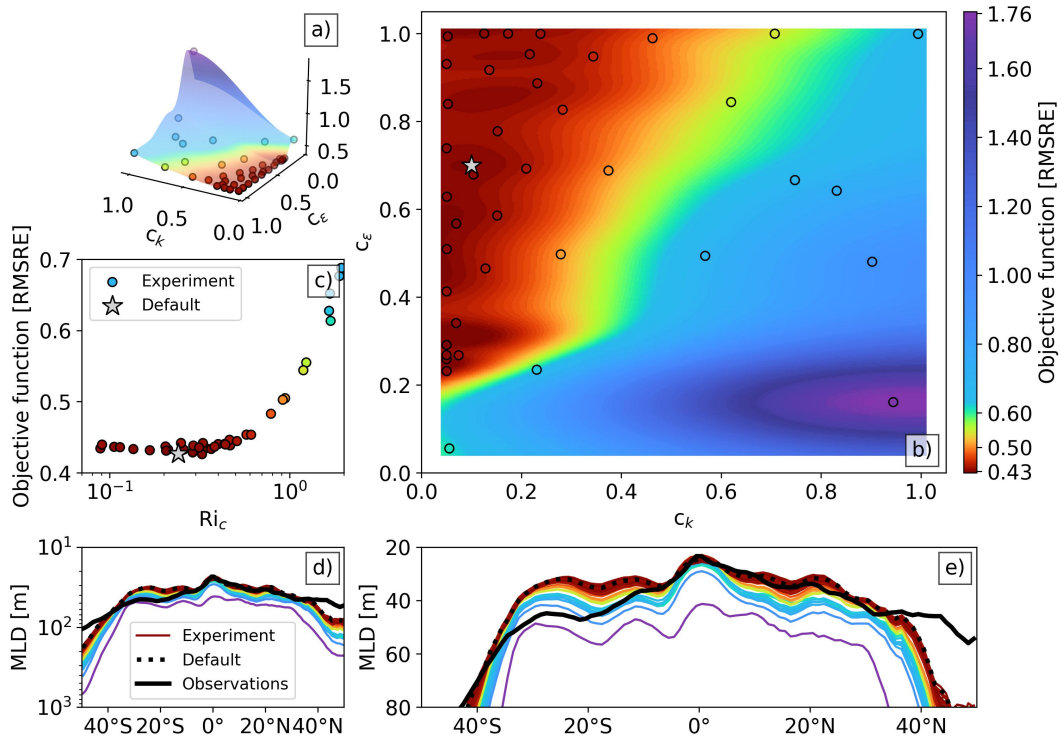
parameter	target value	best TWIN estimate	% difference
$c_k$	0.1	0.1216	21.6
$c_\epsilon$	0.7	0.8693	24.2
$\alpha_{tke}$	30.0	35.7764	19.3
$min_{\kappa_M}$ [ $\text{m}^2 \text{s}^{-1}$ ]	$2 \cdot 10^{-4}$	$1.633 \cdot 10^{-4}$	18.4
$Ri_c$	0.241	0.236	2.1



**Figure 4.** 2D scatter plots and histograms of the TWIN experiment parameters. The target values are indicated on each scatter plot with a gray star and with a dashed line on the histograms.



**Figure 5.** TWIN optimization results: a) zonal MLD averages in the 180 experimental runs and in the Default  $4^\circ \times 4^\circ$  simulation; b) objective function progress after  $n$  points; c)  $\ln(\text{RMSRE})$  as a function of the critical Richardson number  $Ri_c$ . In b), the vertical dashed line indicates the separation between initial and Bayes points. In c), the line indicates the target  $Ri_c \approx 0.24$ .



**Figure 6.** OBS optimization results: a) 3D and b) 2D visualization of the objective function surrogate model mean  $\hat{\mu}$  at the final BO iteration and the 40 evaluated objective function coordinates; c) RMSRE as a function of the critical Richardson number  $Ri_c$ ; d) zonal average MLD in the 40 simulations, Default  $1^\circ \times 1^\circ$  Veros run and observations and e) zonal average MLD as in d) with a reduced y-axis scale. Panel d) is included to show the full extent of MLD bias at higher latitudes; note the logarithmic scale.

397

## 4.2 OBS experiment

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

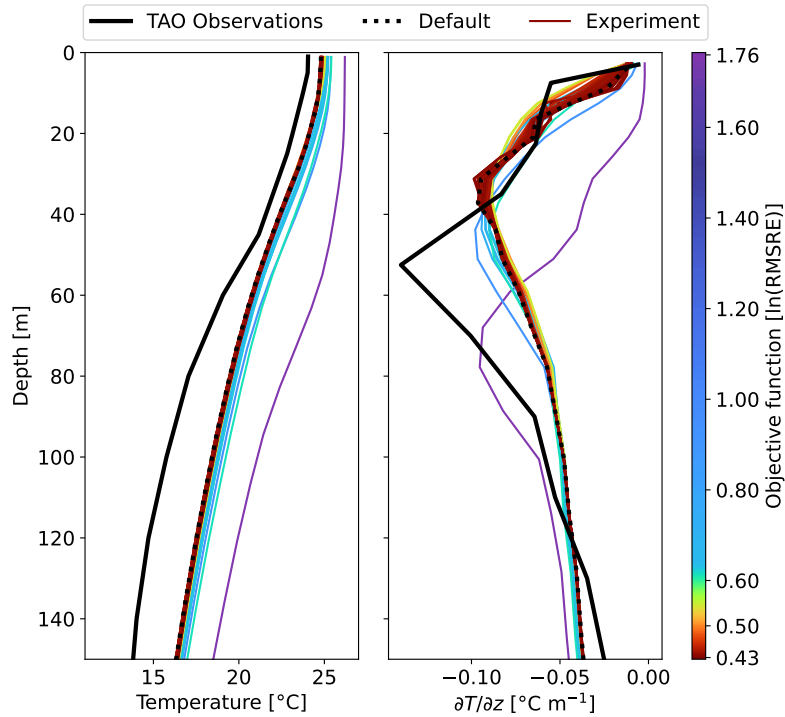
413

414

415

In the OBS experiment, we focus on the two TKE parameters with the most influence over the global MLD in the TWIN experiment:  $c_k$  and  $c_\epsilon$ . Figure 6 shows the overview of the optimization results. The minimum bias in the OBS experiment is 42.62%, which is nearly identical to the Default  $1^\circ \times 1^\circ$  Veros MLD bias of 42.67%. The comparison of the observed and modeled zonal MLD profiles in Fig. 6e) reveals that the TKE scheme parameterization is not likely to be the dominant source of bias for the modeled MLD. The structural limitations of the TKE scheme might be, but in the current form, no set of evaluated parameters brings the global MLD distribution closer to observations. The climatological MLD is asymmetrical about the equator, which is not reflected in the models.

In the run with the lowest error,  $Ri_c = 0.33$ . When the critical Richardson number is less than unity, the MLD bias is relatively insensitive to it. The RMSRE in those runs is  $44.4 \pm 0.2\%$ . For  $Ri_c > 1$ , MLD becomes more sensitive to the parameter values, with MLD bias varying between 54% for  $Ri_c = 1.2$  and 177% for  $Ri_c = 9.9$ . The critical Richardson number shifts the global MLD downwards as it increases, similarly to the  $4^\circ \times 4^\circ$  Veros setup (Fig. 5). The objective function shape in the  $c_k$ - $c_\epsilon$  parameter space in Figure 6a) and b) strongly resembles the  $c_k$ - $c_\epsilon$  scatter plot in Figure 4. The RMSRE is lowest in the upper left quadrant of the parameter space. The similarity be-



**Figure 7.** Mean vertical temperature profile at  $0^\circ\text{N}$ ,  $110^\circ\text{W}$  in TAO mooring observations (thick black line), Default  $1^\circ \times 1^\circ$  Veros simulation (thick dotted line) and across the 40 OBS experiment runs (thin colorful lines).

416 tween the two experiments suggests that including the other two TKE parameters in the  
417 OBS optimization would not improve the RMSRE significantly.

418 The optimization procedure aims to reduce MLD biases, but TKE tuning affects  
419 other output variables as well. An example of the impact of the TKE parameterization  
420 on the vertical temperature profile is shown in Figure 7. The temporal mean of Tropi-  
421 cal Atmosphere Ocean (TAO, McPhaden et al., 2010) mooring temperature observations  
422 at  $0^\circ\text{N}$ ,  $110^\circ\text{W}$  is compared with the Default  $1^\circ \times 1^\circ$  Veros and the 40 OBS simulations.  
423 At this location, a sharp temperature gradient forms at the top of the Equatorial Un-  
424 dercurrent (EUC) as the surface pycnoclines are squeezed by the upwelling. In the ob-  
425 servations, the gradient is sharpest at around 50 m. None of the simulations reproduce  
426 the location of the  $\partial T/\partial z$  peak accurately, but the ones with large MLD bias exhibit closer  
427 resemblance to observations.

## 428 5 Discussion

429 VerOpt is used to identify the set of TKE parameter values that minimize Veros  
430 MLD bias. In the TWIN experiment, the optimization target is MLD simulated using  
431 the Default  $4^\circ \times 4^\circ$  Veros setup. The lowest RMSRE (Eq. (14)) of 1.18% is achieved  
432 with TKE parameters which deviate from the target parameters by about 20-25% (Ta-  
433 ble 3). The ratio  $c_k c_\epsilon^{-1}$ , which sets the critical Richardson number  $Ri_c$  in the model, is  
434 accurately identified by the optimizer, deviating from the target value only by 2.1%.  $Ri_c$   
435 has the highest influence on the modeled MLD (Fig. 5). The target of the OBS exper-  
436 iment with the reduced parameter space  $\mathcal{X}_{\text{OBS}} = \{c_k, c_\epsilon\}$  is MLD climatology (de Boyer Montégut,  
437 2022). The lowest RMSRE in the OBS experiment is 42.62%, only 0.05% lower than the



438 default Veros bias (Fig. 2). In the best simulation,  $Ri_c = 0.33$ . The MLD bias is most  
 439 sensitive to the TKE parameter values when  $Ri_c > 1$ . The zonal averages of MLD in  
 440 the OBS simulations all fail to reproduce the asymmetry of the observed MLD (Fig. 6).  
 441 This means that either the TKE scheme (as it is implemented in Veros) needs extra de-  
 442 grees of freedom to represent mixing in the real ocean, or that the dominant MLD bias  
 443 in Veros originates from another parameterization, a structural limitation, or biases in  
 444 the atmospheric forcing.

445 Contrary to empirical schemes, such as the commonly used KPP (K-profile param-  
 446 eterization, Large et al., 1994), algebraic closures can be embedded into energetically con-  
 447 sistent models (Eden et al., 2014). The TKE scheme consists of only one differential equa-  
 448 tion, as opposed to second order algebraic closures, which contain additional prognos-  
 449 tic equations for length scales or turbulent kinetic energy dissipation  $\epsilon$ . The simplicity  
 450 of the scheme makes it attractive, but it comes at a price of accuracy. Burchard and Bold-  
 451 ing (2001) show, for example, that the parameter  $c_k$  should be a function of stratifica-  
 452 tion and shear. A constant  $c_k$  is therefore likely to be a source of error. However, zonal  
 453 MLD averages in Veros  $1^\circ \times 1^\circ$  are largely similar for  $Ri_c < 1$  (Fig. 6), which shows  
 454 that MLD biases persist for a wide range of  $c_k$  values. With the current TKE implemen-  
 455 tation in Veros, high  $Ri_c > 2$  is required to simulate tropical MLD which resembles ob-  
 456 servations in the Southern hemisphere (Fig. 6e). In the TKE scheme, no mixing can oc-  
 457 cur when the Richardson number is higher than  $Ri_c$ , which has been identified as a struc-  
 458 tural shortcoming of the scheme (e.g Burchard & Bolding, 2001).

459 Another limitation of TKE is the lack of ability to represent non-local mixing pro-  
 460 cesses. IDEMIX (Olbers & Eden, 2013; Nielsen et al., 2018) attempts to alleviate this  
 461 limitation by including an additional source term in the TKE equation (Eq. (7)) to ac-  
 462 count for mixing due to internal wave breaking. Furthermore, sub-mesoscale eddies have  
 463 been shown to restratify the midlatitude oceans (Lévy et al., 2010), but are not currently  
 464 parameterized in Veros.

465 Williamson et al. (2017) propose an alternative method for automated model tun-  
 466 ing which also utilizes GP regression, but employs a different sampling method. They  
 467 search a 24-dimensional parameter space of the NEMO ORCA2 model. Despite differ-  
 468 ences in methodology and the model used, the results of their study are remarkably sim-  
 469 ilar to ours. The TKE parameters  $c_k$  and  $c_\epsilon$  have the highest influence on the temper-  
 470 ature in the upper 300m in ORCA2, alongside the Langmuir cell coefficient and the sur-  
 471 face TKE input coefficient which are not included in Veros. The majority of the param-  
 472 eter space tested in Williamson et al. (2017) results in a warm bias in the upper ocean,  
 473 indicating excessive deepening of the ML. Similarly, we find that increasing the  $c_k c_\epsilon^{-1}$   
 474 ratio in  $1^\circ \times 1^\circ$  Veros leads to global MLD increase, and that MLD is generally the same  
 475 or deeper compared to Default  $1^\circ \times 1^\circ$  Veros in all of the OBS simulations (Fig. 6).

476 Climate models differ not only in tunable parameters, but also fundamental build-  
 477 ing blocks such as grid type, geometry, resolution, components, etc. - all of which affect  
 478 the simulated output. Therefore, parameter tuning is highly model dependent, and an  
 479 optimal choice of parameter values for one model is not guaranteed to be the best choice  
 480 for another. However, the similarities between the results in this work and in Williamson  
 481 et al. (2017) suggest that *some* information about commonly used parameterizations could  
 482 be transferable between different models. MLD deepening as a result of increasing the  
 483  $c_k c_\epsilon^{-1}$  ratio in the TKE scheme is expected to occur in models other than Veros, as a higher  
 484 critical Richardson number necessarily results in enhanced vertical diffusivity.

485 Our work uses a relatively low-dimensional example to demonstrate the utility of  
 486 Bayesian optimization in climate modeling. The goal of this paper is to introduce VerOpt  
 487 in its basic form, but many aspects of the optimization could be enhanced. The BO toolkit  
 488 has grown significantly over the past decades due to its popularity (Wang et al., 2023).

489 The following paragraphs outline the improvements particularly relevant to climate mod-  
490 eling.

491 *Search in higher dimensions.* Our results clearly show that optimizing the TKE  
492 scheme alone is not sufficient to rectify the MLD biases in Veros. A natural next step  
493 would be to extend the search space to other parameterizations, such as IDEMIX men-  
494 tioned above. The bulk formula parameters are also candidates, as they impact the ge-  
495 ographical MLD distribution.

496 High-dimensional optimization problems suffer the curse of dimensionality. The ex-  
497 pensive to evaluate objective functions pose an additional challenge in this regard. If the  
498 number of points necessary to construct an accurate GP model in one dimension is  $n$ ,  
499 then in principle for  $d$  dimensions we would need  $n^d$  points. This is of course in conflict  
500 with the desire to minimize the necessary objective function evaluations. In practice, op-  
501 timizer performance is highly dependent on the shape of the objective function. In the  
502 TWIN experiment, the GP model constructed with only 40 input points is able to ac-  
503 curately locate the region in the parameter space where RMSRE is below 13% (Fig. 5b).  
504 Were the objective function highly nonlinear, this would likely require more Veros sim-  
505 ulations.

506 BO is used in high-dimensional optimization problems across many disciplines, thus  
507 a plethora of methods for coping with the curse of dimensionality has been proposed.  
508 Two most commonly used tactics are variable selection and low-dimensional embeddings.  
509 The former involves reducing the search space based on the relative influence of param-  
510 eters on the objective function (Chen et al., 2012). This strategy has been informally  
511 applied in this work, as we reduced the search space in the OBS experiment based on  
512 the findings from the TWIN optimization. An alternative to the exclusion of inactive  
513 parameters is to optimize a low-dimensional embedding of the high-dimensional param-  
514 eter space, the so-called *active space* (e.g. Nayebi et al., 2019).

515 *Over-sampling at the boundaries.* The *boundary issue* is a common challenge in BO  
516 (Swersky, 2017), where the optimizer over-samples the edges of the parameter space. In  
517 the TWIN experiment, we mitigate this tendency by reducing the value of  $\beta$  in the ac-  
518 quisition function to 0.8, but this comes at a cost of potentially under-exploring the re-  
519 gions in the parameter space where the information about the objective function is lack-  
520 ing. Even with reduced  $\beta$ , the optimizer over-samples the boundaries of  $\min_{K_m}$  (Fig. 4).  
521 The boundary issue is also noticeable in the OBS experiment (Fig. 6b). A successful strat-  
522 egy to combat this tendency has been proposed by Oh et al. (2018). The method, Bayesian  
523 Optimization with Cylindrical Kernels (BOCK), transforms the geometry of the search  
524 space to increase its volume at the center and reduce it at the boundaries. BOCK has  
525 been shown to successfully scale to 500-dimensional optimization problems and is already  
526 implemented in BoTorch (Balandat et al., 2020).

527 *The error metric and MLD uncertainty.* The work of Williamson et al. (2017) un-  
528 derlines the importance of accounting for observational uncertainty in the tuning pro-  
529 cess. In this work, we do not incorporate the information about MLD uncertainty in the  
530 objective function, as it is not included in the dataset we use, which is a significant lim-  
531 itation. However, MLD in the climatology from de Boyer Montégut (2022) is computed  
532 using a constant density threshold of  $0.03 \text{ kg m}^{-3}$ , as opposed to other products which  
533 typically use a variable density threshold to match the  $\pm 0.2^\circ \text{ C}$  temperature criterion  
534 (de Boyer Montégut et al., 2004). This is done to make model comparison with obser-  
535 vations easier, as nearly all OMIP models use constant density threshold for MLD com-  
536 putation (Treguier et al., 2023). The MLD uncertainty could be incorporated into the  
537 objective function by e.g. redefining the error metric. Instead of computing the relative  
538 distance between MLD values in the simulated and target maps (Eq. (14)), the distance  
539 between MLD *distribution* in each grid cell could be used.

540 *Multiple objectives.* Climate model optimization always involves pay-offs between  
 541 the accuracy of different processes at various geographical locations. Tuning to a single  
 542 objective can easily return a parameterization which critically misrepresents another aspect  
 543 of the climate. An example of this effect can be seen in Figure 7, where the observed  
 544  $\partial T/\partial z$  profile is reproduced better in the simulations with high MLD bias. A better ap-  
 545 proach is therefore to tune the model to multiple objectives. In the case of mixing pa-  
 546 rameterizations, important biases aside from MLD are temperature and salinity profiles.  
 547 The multi-objective optimization in VerOpt employs the Expected Hypervolume Improve-  
 548 ment acquisition function (Emmerich et al., 2006).

549 The work presented in this paper only scratches the surface of the potential BO  
 550 has in automated tuning of climate models. Due to its popularity, it has undergone rapid  
 551 development in recent years. This has resulted in the emergence of specialized algorithms,  
 552 which make BO highly adaptable to a wide range of problems. VerOpt makes BO seam-  
 553 lessly applicable to climate modeling. It can work with any model output, but it ben-  
 554 efits significantly from Veros Python/JAX infrastructure which supports GPU compu-  
 555 tation, enabling faster and more energy efficient simulation.

## 556 6 Conclusions

557 We propose Bayesian optimization with VerOpt as a method for automated param-  
 558 eter tuning in OGCMs. Two optimization experiments demonstrate the process by which  
 559 VerOpt identifies a set of TKE parameter values that minimize the simulated MLD er-  
 560 ror relative to a target map. In the idealized TWIN experiment, the best RMSRE is 1.18%.  
 561 In the OBS experiment, where the target is the MLD climatology, the best RMSRE is  
 562 42.62%. The TKE scheme parameters in its current implementation cannot be tuned to  
 563 rectify the MLD biases in Veros. The  $c_k c_\epsilon^{-1}$  ratio, which sets the model critical Richar-  
 564 son number, has the highest influence on MLD. Increasing  $Ri_c$  deepens the mixed layer  
 565 everywhere, and the effect is stronger at locations where the stratification is weak. The  
 566 geographical MLD distribution appears to be set by other model components. In future  
 567 work, higher-dimensional optimization with multiple objectives can potentially be more  
 568 effective in reducing MLD bias in Veros.

## 569 Acronyms

570 **TKE** Turbulent Kinetic Energy  
 571 **MLD** Mixed layer depth  
 572 **GP** Gaussian process  
 573 **BO** Bayesian optimization  
 574 **EKE** Eddy kinetic energy  
 575 **MLL** Marginal log likelihood  
 576 **RMSRE** Root mean square relative error  
 577 **TAO** Tropical Atmosphere Ocean

## 578 Notation

579 **Bayesian optimization**  
 580  $f(\mathbf{x})$  The objective function  
 581  $\mathcal{X}$  The parameter space of the objective function  
 582  $\mathbf{x}$  An input point,  $\mathbf{x} \in \mathcal{X}$   
 583  $\mathbf{x}_*$  A test point,  $\mathbf{x}_* \in \mathcal{X}$   
 584  $\mathbf{X}$  A matrix of input points  
 585  $\mathbf{X}_*$  A matrix of test points

586  $\mathbf{f}(X)$  A vector containing objective function evaluations  
 587  $\mathbf{x}^*$  The global minimum of the objective function  
 588  $\chi^*$  The minimum of evaluated (input) points:  $\chi^* = \arg \min \mathbf{f}(X)$   
 589  $K(X, X')$  The Gram matrix; each element  $k_{i,j} = k(\mathbf{x}_i, \mathbf{x}'_j)$   
 590  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$  Normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$   
 591  $\mathbf{q} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$  The vector  $\mathbf{q}$  is drawn from the normal distribution with mean  $\boldsymbol{\mu}$  and co-  
 592 variance matrix  $\Sigma$   
 593  $\mathcal{GP}(m(\cdot), k(\cdot, \cdot))$  Gaussian process defined by mean  $m(\cdot)$  and kernel  $k(\cdot, \cdot)$   
 594  $g \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$  The function  $g$  is a sample Gaussian process with mean  $m(\cdot)$  and  
 595 kernel  $k(\cdot, \cdot)$   
 596  $\boldsymbol{\theta}$  GP model hyper-parameter vector  
 597  $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\mu}}(\mathbf{x}), \hat{\boldsymbol{\sigma}}(\mathbf{x})$  The hyper-parameter vector, mean and standard deviation of the GP model  
 598 for which the MLL is maximized

599

600 **Oceanography**

601  $\bar{\cdot}$  Mean component in Reynolds decomposition  
 602  $\cdot'$  Turbulent component in Reynolds decomposition  
 603  $x, y, z$  Geographical coordinates: zonal, meridional and vertical direction  
 604  $u, v, w$  Three-dimensional flow: zonal, meridional and vertical velocity  
 605  $\rho$  Sea water density  
 606  $\delta\rho$  Density difference  
 607  $T$  Sea water temperature  
 608  $z_{\text{ref}}$  Reference depth for calculating MLD  
 609  $K_m$  Eddy viscosity  
 610  $\min_{K_m}$  Minimum eddy viscosity  
 611  $K_h$  Eddy diffusivity  
 612  $\min_{K_h}$  Minimum eddy diffusivity  
 613  $e$  Turbulent kinetic energy  
 614  $l_k$  Mixing length scale  
 615  $l_\epsilon$  Dissipation length scale  
 616  $N$  Buoyancy frequency  
 617  $Sh$  Vertical shear  
 618  $c_k$  Mixing coefficient in  $K_m = c_k l_k \bar{e}^{1/2}$   
 619  $c_\epsilon$  Dissipation coefficient in  $\epsilon = c_\epsilon l_\epsilon^{-1} \bar{e}^{3/2}$   
 620  $\alpha_{tke}$  Vertical diffusive flux coefficient in  $K_e = \alpha_{tke} K_m$   
 621  $Ri$  Richardson number  $Ri = N^2 Sh^{-2}$   
 622  $Ri_c$  Critical Richardson number  
 623  $\gamma_{R_f}$  Mixing efficiency coefficient  
 624  $P_{rt}$  Turbulent Prandtl number

625 **7 Open Research**

626 All model simulations in the manuscript have been done using Veros (Häfner et al.,  
 627 2018, 2021), which is available to download on github: <https://github.com/team-ocean/veros>.  
 628 Veros documentation is available at <https://veros.readthedocs.io>. The optimizer  
 629 VerOpt (Stoustrup, 2021) is available on github: <https://github.com/idadx4325/veropt>.  
 630 The mixed layer depth (MLD) climatology (de Boyer Montégut, 2022) can be downloaded  
 631 from <https://www.seanoe.org/data/00806/91774/>. Veros model output presented in  
 632 this paper is available at [https://sid.erda.dk/cgi-sid/ls.py?share\\_id=dGN0t2pSbf](https://sid.erda.dk/cgi-sid/ls.py?share_id=dGN0t2pSbf).

633 **Acknowledgments**

634 The research presented in this manuscript is funded by nextGEMS, a European H2020  
 635 project under the Grant agreement No. 101003470. We would like to thank Nils Brüggemann  
 636 (Max Planck Institute for Meteorology) for multiple discussions about the details of the  
 637 TKE scheme.

638 The authors are grateful to be able to use the Tropical Atmosphere Ocean (TAO)  
 639 mooring array dataset managed by GTMBA Project Office of NOAA/PMEL, as well as  
 640 the mixed layer depth (MLD) climatology from (de Boyer Montégut, 2022) hosted by  
 641 SEA scieNtific Open data Edition (SEANOE).

642 The authors are grateful for computing resources and technical assistance provided  
 643 by the Danish Center for Climate Computing, a facility built with support of the Dan-  
 644 ish e-Infrastructure Corporation, Danish Hydrocarbon Research and Technology Cen-  
 645 tre, VILLUM Foundation, and the Niels Bohr Institute.

646 This project was partially based on computations performed on the LUMI super-  
 647 computer through LUMI project 465000815, awarded by the Danish e-Infrastructure Con-  
 648 sortium (DeIC) and Aarhus University as project DeiC-AU-L5-0022. We Thank DeIC,  
 649 Aarhus University, the EuroHPC Joint Undertaking, and the LUMI consortium for award-  
 650 ing us these resources.

651 **References**

- 652 Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., ... Chin-  
 653 tala, S. (2024, April). PyTorch 2: Faster Machine Learning Through  
 654 Dynamic Python Bytecode Transformation and Graph Compilation. In  
 655 *29th acm international conference on architectural support for program-*  
 656 *ming languages and operating systems, volume 2 (asplos '24)*. ACM.  
 657 Retrieved from <https://pytorch.org/assets/pytorch2-2.pdf> doi:  
 658 10.1145/3620665.3640366
- 659 Antonov, J., Locarnini, R., Boyer, T., Mishonov, A., & Garcia, H. (2006). *World*  
 660 *ocean atlas 2005 volume 2: Salinity* (Vol. 62).
- 661 Balandat, M., Karrer, B., Jiang, D. R., Daulton, S., Letham, B., Wilson, A. G., &  
 662 Bakshy, E. (2020). BoTorch: A Framework for Efficient Monte-Carlo Bayesian  
 663 Optimization. In *Advances in neural information processing systems 33*. Re-  
 664 trieved from <http://arxiv.org/abs/1910.06403>
- 665 Blanke, B., & Delecluse, P. (1993). Variability of the tropical atlantic ocean sim-  
 666 ulated by a general circulation model with two different mixed-layer physics.  
 667 *Journal of Physical Oceanography*, *23*(7), 1363–1388.
- 668 Brüggemann, N., Losch, M., Scholz, P., Pollmann, F., Danilov, S., Gutjahr, O., ...  
 669 Eden, C. (2024). Parameterized internal wave mixing in three ocean general  
 670 circulation models. *Journal of Advances in Modeling Earth Systems*, *16*(6),  
 671 e2023MS003768. Retrieved from [https://agupubs.onlinelibrary.wiley](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003768)  
 672 [.com/doi/abs/10.1029/2023MS003768](https://doi.org/10.1029/2023MS003768) (e2023MS003768 2023MS003768) doi:  
 673 <https://doi.org/10.1029/2023MS003768>
- 674 Burchard, H., & Bolding, K. (2001). Comparative analysis of four second-moment  
 675 turbulence closure models for the oceanic mixed layer. *Journal of Physical*  
 676 *Oceanography*, *31*(8), 1943–1968.
- 677 Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for  
 678 bound constrained optimization. *SIAM Journal on scientific computing*, *16*(5),  
 679 1190–1208.
- 680 Chen, B., Castro, R., & Krause, A. (2012). Joint optimization and variable selection  
 681 of high-dimensional gaussian processes. *arXiv preprint arXiv:1206.6396*.
- 682 de Boyer Montégut, C. (2022). Mixed layer depth climatology computed with a den-  
 683 sity threshold criterion of 0.03 kg/m3 from 10 m depth value [Dataset]. SEA-

- 684 *NOE*, 10, 91774.
- 685 de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004).  
 686 Mixed layer depth over the global ocean: An examination of profile data  
 687 and a profile-based climatology. *Journal of Geophysical Research: Oceans*,  
 688 109(C12).
- 689 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi,  
 690 S., ... Vitart, F. (2011). The era-interim reanalysis: configuration  
 691 and performance of the data assimilation system. *Quarterly Journal*  
 692 *of the Royal Meteorological Society*, 137(656), 553-597. Retrieved from  
 693 <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828> doi:  
 694 <https://doi.org/10.1002/qj.828>
- 695 Eden, C., Czeschel, L., & Olbers, D. (2014). Toward energetically consistent ocean  
 696 models. *Journal of Physical Oceanography*, 44(12), 3160-3184.
- 697 Emmerich, M., Giannakoglou, K., & Naujoks, B. (2006). Single- and multiobjec-  
 698 tive evolutionary optimization assisted by gaussian random field metamod-  
 699 els. *IEEE Transactions on Evolutionary Computation*, 10(4), 421-439. doi:  
 700 10.1109/TEVC.2005.859463
- 701 Gaspar, P., Grégoris, Y., & Lefevre, J.-M. (1990). A simple eddy kinetic en-  
 702 ergy model for simulations of the oceanic vertical mixing: Tests at sta-  
 703 tion papa and long-term upper ocean study site. *Journal of Geophysi-  
 704 cal Research: Oceans*, 95(C9), 16179-16193. Retrieved from [https://](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC095iC09p16179)  
 705 [agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC095iC09p16179](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC095iC09p16179)  
 706 doi: <https://doi.org/10.1029/JC095iC09p16179>
- 707 Gregg, M. C., D'Asaro, E. A., Riley, J. J., & Kunze, E. (2018). Mixing efficiency in  
 708 the ocean. *Annual review of marine science*, 10(1), 443-473.
- 709 Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning,  
 710 C. W., ... others (2016). Omip contribution to cmip6: Experimental and diag-  
 711 nostic protocol for the physical component of the ocean model intercomparison  
 712 project. *Geoscientific Model Development*, 3231.
- 713 Häfner, D., Jacobsen, R., Eden, C., Kristensen, M., Jochum, M., Nuterman, R., &  
 714 Vinter, B. (2018, August 16). Veros v0.1-a fast and versatile ocean simulator  
 715 in pure python [Software]. *Geoscientific Model Development*, 11(8), 3299-3312.  
 716 doi: 10.5194/gmd-11-3299-2018
- 717 Häfner, D., Nuterman, R., & Jochum, M. (2021, December 2). Fast, cheap, and  
 718 turbulent—global ocean modeling with gpu acceleration in python [Soft-  
 719 ware]. *Journal of Advances in Modeling Earth Systems*, 13(12). doi:  
 720 10.1029/2021MS002717
- 721 Huang, C. J., Qiao, F., & Dai, D. (2014). Evaluating cmip5 simulations of mixed  
 722 layer depth during summer. *Journal of Geophysical Research: Oceans*, 119(4),  
 723 2568-2582.
- 724 Jochum, M., & Potemra, J. (2008). Sensitivity of tropical rainfall to banda sea dif-  
 725 fusivity in the community climate system model. *Journal of Climate*, 21(23),  
 726 6445-6454.
- 727 Korn, P., Brüggemann, N., Jungclaus, J. H., Lorenz, S. J., Gutjahr, O., Haak, H.,  
 728 ... Marotzke, J. (2022). Icon-o: The ocean component of the icon earth sys-  
 729 tem model—global simulation characteristics and local telescoping capability.  
 730 *Journal of Advances in Modeling Earth Systems*, 14(10), e2021MS002952.  
 731 Retrieved from [https://agupubs.onlinelibrary.wiley.com/doi/abs/](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002952)  
 732 [10.1029/2021MS002952](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002952) (e2021MS002952 2021MS002952) doi: [https://](https://doi.org/10.1029/2021MS002952)  
 733 [doi.org/10.1029/2021MS002952](https://doi.org/10.1029/2021MS002952)
- 734 Large, W. G., McWilliams, J. C., & Doney, S. C. (1994). Oceanic vertical mixing: A  
 735 review and a model with a nonlocal boundary layer parameterization. *Reviews*  
 736 *of geophysics*, 32(4), 363-403.
- 737 Ledwell, J. R., Watson, A. J., & Law, C. S. (1998). Mixing of a tracer in the py-  
 738 nocline. *Journal of Geophysical Research: Oceans*, 103(C10), 21499-21529.

- Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98JC01738> doi: <https://doi.org/10.1029/98JC01738>
- Lévy, M., Klein, P., Tréguier, A.-M., Iovino, D., Madec, G., Masson, S., & Takahashi, K. (2010). Modifications of gyre circulation by sub-mesoscale physics. *Ocean Modelling*, *34*(1-2), 1–15.
- Liu, Z., Philander, S., & Pacanowski, R. (1994). A gcm study of the tropical-subtropical upper-ocean circulation. *J. Phys. Oceanogr*, *24*, 2606–2623.
- Locarnini, R., Mishonov, A., Antonov, J., Boyer, T., & Garcia, H. (2006). *World ocean atlas 2005, vol. 1: Temperature* (Vol. 68).
- Madec, G., & the NEMO team. (2016). *Nemo reference manual 3.6.stable* (Tech. Rep.). Institut Pierre-Simon Laplace (IPSL).
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., . . . others (2012). Tuning the climate of a global model. *Journal of advances in modeling Earth systems*, *4*(3).
- McPhaden, M. J., Busalacchi, A. J., & Anderson, D. L. (2010). A toga retrospective. *Oceanography*, *23*(3), 86–103.
- Nayebi, A., Munteanu, A., & Poloczek, M. (2019, 09–15 Jun). A framework for Bayesian optimization in embedded subspaces. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th international conference on machine learning* (Vol. 97, pp. 4752–4761). PMLR. Retrieved from <https://proceedings.mlr.press/v97/nayebi19a.html>
- Nielsen, S. B., Jochum, M., Eden, C., & Nuterman, R. (2018). An energetically consistent vertical mixing parameterization in cesm4. *Ocean Modelling*, *127*, 46–54.
- Oh, C., Gavves, E., & Welling, M. (2018). Bock: Bayesian optimization with cylindrical kernels. In *International conference on machine learning* (pp. 3868–3877).
- Olbers, D., & Eden, C. (2013). A global model for the diapycnal diffusivity induced by internal gravity waves. *Journal of Physical Oceanography*, *43*(8), 1759–1779.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148–175.
- Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., . . . Vertenstein, M. (2014). A new synoptic scale resolving global climate simulation using the community earth system model. *Journal of Advances in Modeling Earth Systems*, *6*(4), 1065–1094. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014MS000363> doi: <https://doi.org/10.1002/2014MS000363>
- Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. (2009). Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*.
- Stoustrup, A. (2021). *Automated parameter tuning for the versatile ocean simulator (veros)* (Software, University of Copenhagen). Retrieved from [https://www.gfy.ku.dk/~nuterman/teamocean/docs/Ida.Stoustrup\\_MSc\\_thesis.pdf](https://www.gfy.ku.dk/~nuterman/teamocean/docs/Ida.Stoustrup_MSc_thesis.pdf)
- Swersky, K. (2017). *Improving bayesian optimization for machine learning using expert priors*. University of Toronto (Canada).
- Treguier, A. M., de Boyer Montégut, C., Bozec, A., Chassignet, E. P., Fox-Kemper, B., McC. Hogg, A., . . . others (2023). The mixed-layer depth in the ocean model intercomparison project (omip): impact of resolving mesoscale eddies. *Geoscientific Model Development*, *16*(13), 3849–3872.
- Wang, X., Jin, Y., Schmitt, S., & Olhofer, M. (2023). Recent advances in bayesian

- 794 optimization. *ACM Computing Surveys*, 55(13s), 1–36.
- 795 Waterhouse, A. F., MacKinnon, J. A., Nash, J. D., Alford, M. H., Kunze, E., Sim-  
796 mons, H. L., . . . others (2014). Global patterns of diapycnal mixing from  
797 measurements of the turbulent dissipation rate. *Journal of Physical Oceanogra-*  
798 *phy*, 44(7), 1854–1872.
- 799 Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learn-*  
800 *ing* (Vol. 2) (No. 3). MIT press Cambridge, MA.
- 801 Williamson, D. B., Blaker, A. T., & Sinha, B. (2017). Tuning without over-tuning:  
802 parametric uncertainty quantification for the nemo ocean model. *Geoscientific*  
803 *Model Development*, 10(4), 1789–1816.



## CHAPTER CONTENTS

- 4 Uncharted Depths 75
  - 4.1 MLD for Model Validation 75
  - 4.2 Extended Optimization 76
  - 4.3 Turbulence Closures of Tomorrow 76
    - 4.3.1 ML-based parameterizations 76
    - 4.3.2 Hybrid models and calibration 77
    - 4.3.3 Eddy Rich Models 78
    - 4.3.4 A Final Note 79

## Uncharted Depths

# 4

The two strategies for improving oceanic closure schemes presented in this thesis illustrate how the parameterizations still fail to accurately capture the sub-grid scale processes. This issue is recognized in the community and drives the research into novel methods of attacking the closure problem. A selected collection of these are summarised at the end of this chapter.

Preceding the speculation about what oceanic turbulence closures may look like in the future are two sections that build upon the work in Articles I and II. Section 4.1 is a reflection on practices involved in model validation using MLD fields, and Section 4.2 outlines the extended VerOpt experiments.

### 4.1 MLD FOR MODEL VALIDATION

An important result from Treguier et al. (2023) is the demonstrated MLD sensitivity to its defining criterion. The MLD fields computed using the surface reference layer of a model (1 m or less) and the 10 m depth layer differ by more than 40 m. Since MLD is not commutable, the best practice is to define the model MLD threshold in the same way as in observations and compute MLD at every time step. However, MLD definition is not standardized, thus modelers have to choose which criterion to implement.

As a result, comparing modeled MLD to climatology using the same threshold is sometimes not possible (e.g., Heuzé, 2020). This leads to an unknown fraction of the MLD bias stemming from its definition. When the same criterion is used, it is sometimes not stated whether the modeled MLD was computed based on averaged profiles (Pottapinjara and Joseph, 2022). Article I unfortunately falls within a similar category, where the modeled summer MLD biases in Fig. 2 are computed from averaged monthly temperature profiles.

Even when the best practices in comparing the modeled and observed MLD are followed (e.g., Damerell et al., 2020; Gutjahr et al., 2021), making general statements about MLD biases across studies is complicated by the variety of criteria used. The effort to understand the model response to different vertical mixing parameterizations is undermined by inconsistencies in MLD computation. Community efforts to standardize this variable should be continued.

## 4.2 EXTENDED OPTIMIZATION

The optimization experiments in Article II show how Bayesian optimization can be successfully applied to ocean model tuning. As stated in the manuscript, there are many ways in which the method can be improved. Some of these are addressed in our current continued work with VerOpt.

The parameter space of the extended optimization includes the TKE scheme parameters, the mesoscale eddy closure (Eden, Czeschel, and Olbers, 2014), the drag coefficient in the bulk formula and a scaling parameter for the salinity restoring mask in Veros, altogether forming a modest nine-dimensional parameter space. The optimization runs are performed on the Lumi supercomputer and include 60-year-long integrations of Veros  $1^\circ \times 1^\circ$  setup. Two additional objectives aside from the mixed layer depth are added: the strength of the Antarctic Circumpolar Current (ACC) and the depth of the  $18^\circ\text{C}$  isotherm. The new targets contain observational uncertainty range, addressing the weakness of the previous work where only the climatological MLD mean is considered in the optimization.

The use of Lumi enables significantly more parallel runs per experiment<sup>1</sup>, and one of the current research goals is to establish the number of simulations and optimization steps needed to effectively search the nine-dimensional space. We aim to test the scalability of the method, and to quantify model sensitivity to the extended parameter space.

1. Currently, we are trying 32 parallel Veros runs over 20 optimization steps and one initialization round, amounting to 672 total simulations.

## 4.3 TURBULENCE CLOSURES OF TOMORROW

The two leading groups of ideas for improving the representation of sub-grid scale processes in GCMs can be categorised into machine learning (ML) methods and kilometer-scale modeling (e.g., Balaji et al., 2022; Fox-Kemper et al., 2019). The former involves the combination of data-driven and physics-based components in a single hybrid model, while the latter aims to minimize the error resulting from the truncated energy spectrum by directly simulating mesoscale physics. Both approaches are discussed here, with the focus on vertical mixing.

### 4.3.1 ML-based parameterizations

Zanna and Bolton (2021) identify the four requirements for schemes representing sub-grid processes in climate models:

1. “[A parameterization] accurately captures the physical processes being parameterized,
2. Respects physical principles such as conservation laws,
3. Is numerically stable when implemented into a climate model and
4. Generalizes to new dynamical regimes.”

Data-driven approaches to closing RANS equations may be able to outmatch the physics-based schemes in accurately representing the parameterized processes. Deep networks offer higher predictive skill, but come with new challenges including, but not limited to, points 2-4. Development of ML schemes is focused on addressing these challenges.

The second point concerning consistency with physical principles can be addressed by embedding ML components in physics-based schemes. For example, Sane et al. (2023) expand the ePBL scheme (Reichl and Hallberg, 2018) by replacing the shape function and the velocity scale in the OSBL diffusivity profile with two Artificial Neural Network (ANN) models. The ANNs are trained on SMC output generated with General Ocean Turbulence Model (GOTM, Umlauf and Burchard, 2005). This approach ensures that the new scheme, ePBL\_NN, adheres to the constraints of turbulence theory.

One of the main points of critique for ML methods in climate science is the question of generalizability. Data availability constrain is a bottleneck for all modeling efforts in oceanography. ML parameterizations must therefore learn from model simulations, such as LES or SMCs. The limitation of the NN\_ePBL approach is that the models inherit the SMC limitations. However, SMCs still offer a better physical representation of ocean mixing than simpler closures such as TKE. MOM6 simulations with ePBL\_NN shows considerable reduction of tropical summer MLD biases compared to ePBL, with 10 m deepening at the equator<sup>2</sup>.

The issue of numerical stability and the training procedure of ML parameterizations are addressed in Frezat et al. (2022). Most existing ML schemes implement *a priori* learning, i.e. the models are tuned before being embedded into GCMs. This can lead to unphysical behavior and numerical instabilities (e.g., Zanna and Bolton, 2020). Frezat et al. (2022), show that in a quasi-geostrophic (QG) model, *a posteriori* tuning can improve the performance, stability and generalizability of the ML schemes.

2. Contrary to the tuning in Article II, the improvement *is not* accompanied by increased winter MLD biases.

#### 4.3.2 Hybrid models and calibration

To enable *a posteriori* tuning, automatic differentiation (AD) must be supported in the model. Most contemporary GCMs are not differentiable, which

makes the effective implementation of ML components a challenge. The exceptions include Neural GCM (Kochkov et al., 2024), differentiable QG models (e.g., Frezat et al., 2022; Qu, Bhourri, and Gentine, 2024) and column schemes<sup>3</sup>. The JAX backend (Bradbury et al., 2018) enables automatic differentiation in Veros.

3. See, for example, the tunax model from the MEOM group at Grenoble University.

Shen et al. (2023) envision differentiable modeling as a unifying link between machine learning and physics-based models. Differentiable modeling breaks down the dynamical core of a numerical model into the resolved and sub-grid scale dynamics, both of which involve parameterizations. Using AD and emulation-based tuning, the two sets of parameters can be optimized simultaneously to train the model. The results of Qu, Bhourri, and Gentine (2024) show the success of the method in a QG study.

It should be noted that ML based schemes exacerbate the undesirable GCM feature, which is the need for tuning (Balaji et al., 2022; Mauritsen et al., 2012). While physics-based parameterizations often suffer from structural deficiencies that limit their ability to capture the observed ocean features, ML schemes may in turn be overfitted due to their much higher number of tunable hyper-parameters.

Methods such as history matching Williamson et al. (2015) to quantify parametric uncertainty address the delicate matter of GCM tuning. Rather than parameter optimization, the aim of history-matching is to identify largest possible regions of the parameter space which produce acceptable representation of the observed climate. In “*The art and science of climate model tuning*”, the authors Hourdin et al. (2017) discuss the many pitfalls involved in automated tuning, such as overfitting and compensating biases. They argue for estimating both structural (via model ensembles) and parametric (via parameter perturbations) uncertainty of GCMs.

### 4.3.3 Eddy Rich Models

In contrast to ML methods, kilometer-scale modeling could reduce the amount of OGCM parameters. Mixing induced by shear instability in the ocean will always require parameterizing due to the length scales at which the is dissipated. However, resolving mesoscale and even sub-mesoscale could potentially result in direct simulation of ocean processes which modulate stratification (geostrophic eddies) and induce mixing (e.g., NIWs). This could eliminate the need of adding extra components that simulate the effect of specific physical processes to the “basic” form of OGCM parameterizations.

Article I suggests that 10 km may not be a fine enough resolution to simulate the NIW-induced mixing, but the progressive refining of models such as ICON-Sapphire (Hohenegger et al., 2023) and IFS-FESOM (Rackow et al., 2024) may alleviate this limitation in the future. An alternative path to

resolving the small scales in OGCMs are nested models, where the resolution is increased in critical regions of climate sensitivity (Hewitt et al., 2022).

Following best tuning and modeling practices, model ensembles and parameter uncertainty quantification of future OGCMs are required. This poses a challenge to eddy rich models due to their computational cost and the associated size of model output (e.g., 135TB of output per month for 2.5 km grid spacing in ICON-Sapphire, Hohenegger et al., 2023). Tackling a problem of this magnitude requires community effort. Such is the opinion of (Palmer, 2016), who in his essay “*A personal perspective on modelling the climate system*” calls for international collaborative effort to develop global, fully coupled kilometer-scale models, where enough resources are allocated for long term ensemble studies.

It is safe to say that the prognosis for the future of turbulence closures for ocean models is a significant increase in variety and further adaptation to the needs of the individual modelers. The unifying aspect may be the ambition for objective, transparent and careful tuning procedures.

#### 4.3.4 A Final Note

I would like to bring up this quote from the essay “*Geophysical fluid dynamics: whence, whither and why?*” by Vallis (2016), which has been ringing in my head since the moment I read it a year ago:

*“If we see a hurricane in the tropical mid-Atlantic, we know it will usually move westward, and not because a simulation tells us that.”*

The field of climate modeling is beaming with new creative and exciting methods, which are still in early development stages. In this time, the value of GFD should not be forgotten; it was the deep dive into the history of the TKE scheme which facilitated the understanding and contextualization of the optimization results in Article II. Relying on complex algorithms to do the thinking for us is a waste of resources at best.

## References

- Balaji, V. et al. (2022). “Are general circulation models obsolete?” In: *Proceedings of the National Academy of Sciences* 119.47, e2202075119. DOI: <https://doi.org/10.1073/pnas.2202075119> (Cited on pp. 76, 78).
- Blanke, Bruno and Pascale Delecluse (1993). “Variability of the tropical Atlantic Ocean simulated by a general circulation model with two different mixed-layer physics.” In: *Journal of Physical Oceanography* 23.7, pp. 1363–1388. DOI: [https://doi.org/10.1175/1520-0485\(1993\)023<1363:VOTTAO>2.0.CO;2](https://doi.org/10.1175/1520-0485(1993)023<1363:VOTTAO>2.0.CO;2) (Cited on pp. 12, 48).
- Bourlès, Bernard et al. (2019). “PIRATA: A Sustained Observing System for Tropical Atlantic Climate Research and Forecasting.” In: *Earth and Space Science* 6.4, pp. 577–616. DOI: <https://doi.org/10.1029/2018EA000428> (Cited on p. 6).
- Boyer Montégut, C de (2022). “Mixed layer depth climatology computed with a density threshold criterion of 0.03 kg/m<sup>3</sup> from 10 m depth value.” In: *SEANOE* 10, p. 91774. URL: <https://www.seanoe.org/data/00806/91774/> (Cited on p. 6).
- Boyer Montégut, Clément de et al. (2004). “Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology.” In: *Journal of Geophysical Research: Oceans* 109.C12. DOI: <https://doi.org/10.1029/2004JC002378> (Cited on p. 6).
- Bradbury, James et al. (2018). “JAX: composable transformations of Python+ NumPy programs.” In: URL: <http://github.com/google/jax> (Cited on p. 78).
- Brandt, Peter et al. (2015). “On the role of circulation and mixing in the ventilation of oxygen minimum zones with a focus on the eastern tropical North Atlantic.” In: *Biogeosciences* 12.2, pp. 489–512. DOI: <https://doi.org/10.5194/bg-12-489-2015> (Cited on p. 4).
- Breitburg, Denise et al. (2018). “Declining oxygen in the global ocean and coastal waters.” In: *Science* 359.6371, eaam7240. DOI: <https://doi.org/10.1126/science.aam7240> (Cited on p. 4).
- Brüggemann, Nils et al. (2024). “Parameterized internal wave mixing in three ocean general circulation models.” In: *Journal of Advances in Modeling Earth Systems* 16.6, e2023MS003768. DOI: <https://doi.org/10.1029/2023MS003768> (Cited on p. 16).
- Burchard, Hans and Karsten Bolding (2001). “Comparative analysis of four second-moment turbulence closure models for the oceanic mixed layer.”

- In: *Journal of Physical Oceanography* 31.8, pp. 1943–1968. DOI: [https://doi.org/10.1175/1520-0485\(2001\)031<1943:CAOFSM>2.O.CO;2](https://doi.org/10.1175/1520-0485(2001)031<1943:CAOFSM>2.O.CO;2) (Cited on p. 47).
- Canuto, VM et al. (2002). “Ocean turbulence. Part II: Vertical diffusivities of momentum, heat, salt, mass, and passive scalars.” In: *Journal of Physical Oceanography* 32.1, pp. 240–264. DOI: [https://doi.org/10.1175/1520-0485\(2002\)032<0240:OTPIVD>2.O.CO;2](https://doi.org/10.1175/1520-0485(2002)032<0240:OTPIVD>2.O.CO;2) (Cited on p. 16).
- Cleary, Emmet et al. (2021). “Calibrate, emulate, sample.” In: *Journal of Computational Physics* 424, p. 109716. DOI: <https://doi.org/10.1016/j.jcp.2020.109716> (Cited on p. 42).
- Damerell, Gillian M et al. (2020). “A comparison of five surface mixed layer models with a year of observations in the North Atlantic.” In: *Progress in Oceanography* 187, p. 102316. DOI: <https://doi.org/10.1016/j.pocean.2020.102316> (Cited on pp. 6, 17, 75).
- Dengler, Marcus et al. (2023). *Synoptic upper-ocean mixing events in the tropical Atlantic for evaluating Earth system models*. dataset bundled publication. DOI: 10.1594/PANGAEA.964320. URL: <https://doi.org/10.1594/PANGAEA.964320> (Cited on p. 23).
- Eden, Carsten et al. (2014). “Toward energetically consistent ocean models.” In: *Journal of Physical Oceanography* 44.12, pp. 3160–3184. DOI: <https://doi.org/10.1175/JPO-D-13-0260.1> (Cited on pp. 16, 76).
- Fischer, Tim et al. (2024). “Near-inertial waves mixing and ocean background observations in the eastern tropical North Atlantic (2006–2019) for evaluating Earth system models.” In: PANGAEA. DOI: 10.1594/PANGAEA.964317. URL: <https://doi.org/10.1594/PANGAEA.964317> (Cited on p. 23).
- Foltz, Gregory R. et al. (2018). “An enhanced PIRATA dataset for tropical Atlantic Ocean–atmosphere research.” In: *Journal of Climate* 31.4, pp. 1499–1524. DOI: <https://doi.org/10.1175/JCLI-D-16-0816.1> (Cited on p. 6).
- Fox-Kemper, Baylor et al. (2019). “Challenges and prospects in ocean circulation models.” In: *Frontiers in Marine Science* 6, p. 65. DOI: <https://doi.org/10.3389/fmars.2019.00065> (Cited on pp. 1, 76).
- Frezat, Hugo et al. (2022). “A posteriori learning for quasi-geostrophic turbulence parametrization.” In: *Journal of Advances in Modeling Earth Systems* 14.11, e2022MS003124. DOI: <https://doi.org/10.1029/2022MS003124> (Cited on pp. 77, 78).
- Furuichi, Naoki et al. (2008). “Model-predicted distribution of wind-induced internal wave energy in the world’s oceans.” In: *Journal of Geophysical Research: Oceans* 113.C9. DOI: <https://doi.org/10.1029/2008JC004768> (Cited on p. 21).



- Galbraith, ED and S Eggleston (2017). “A lower limit to atmospheric CO<sub>2</sub> concentrations over the past 800,000 years.” In: *Nature Geoscience* 10.4, pp. 295–298. DOI: <https://doi.org/10.1038/ngeo2914> (Cited on p. 4).
- Garrett, Chris (1996). “Processes in the surface mixed layer of the ocean.” In: *Dynamics of Atmospheres and Oceans* 23.1-4, pp. 19–34. DOI: [https://doi.org/10.1016/0377-0265\(95\)00421-1](https://doi.org/10.1016/0377-0265(95)00421-1) (Cited on p. 3).
- Gaspar, Philippe et al. (1990). “A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station Papa and long-term upper ocean study site.” In: *Journal of Geophysical Research: Oceans* 95.C9, pp. 16179–16193. DOI: <https://doi.org/10.1029/JC095iC09p16179> (Cited on p. 12).
- Gaube, Peter et al. (2019). “Mesoscale Eddies Modulate Mixed Layer Depth Globally.” In: *Geophysical Research Letters* 46.3, pp. 1505–1512. DOI: <https://doi.org/10.1029/2018GL080006> (Cited on p. 17).
- Griffies, Stephen M et al. (2016). “OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project.” In: *Geoscientific Model Development*, p. 3231 (Cited on p. 17).
- Gutjahr, Oliver et al. (2021). “Comparison of ocean vertical mixing schemes in the Max Planck Institute Earth System Model (MPI-ESM1. 2).” In: *Geoscientific Model Development* 14.5, pp. 2317–2349. DOI: <https://doi.org/10.5194/gmd-14-2317-2021> (Cited on pp. 17, 75).
- Häfner, Dion et al. (2018). “Veros vo.1-a fast and versatile ocean simulator in pure Python.” English. In: *Geoscientific Model Development* 11.8, pp. 3299–3312. ISSN: 1991-959X. DOI: [10.5194/gmd-11-3299-2018](https://doi.org/10.5194/gmd-11-3299-2018) (Cited on p. 47).
- Häfner, Dion et al. (2021). “Fast, Cheap, and Turbulent—Global Ocean Modeling With GPU Acceleration in Python.” English. In: *Journal of Advances in Modeling Earth Systems* 13.12. ISSN: 1942-2466. DOI: [10.1029/2021MS002717](https://doi.org/10.1029/2021MS002717) (Cited on p. 47).
- Heuzé, Céline (2020). “Antarctic bottom water and North Atlantic deep water in CMIP6 models.” In: *Ocean Science Discussions* 2020, pp. 1–38. DOI: <https://doi.org/10.5194/os-17-59-2021> (Cited on p. 75).
- Hewitt, Helene et al. (2022). “The small scales of the ocean may hold the key to surprises.” In: *Nature Climate Change* 12.6, pp. 496–499. DOI: <https://doi.org/10.1038/s41558-022-01386-6> (Cited on pp. 20, 79).
- Hohenegger, Cathy et al. (2023). “ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales.” In: *Geoscientific Model Development* 16.2, pp. 779–811. DOI: <https://doi.org/10.5194/gmd-16-779-2023> (Cited on pp. 78, 79).
- Holte, James and Lynne Talley (2009). “A new algorithm for finding mixed layer depths with applications to Argo data and Subantarctic Mode Wa-

- ter formation.” In: *Journal of Atmospheric and Oceanic Technology* 26.9, pp. 1920–1939. DOI: <https://doi.org/10.1175/2009JTECHO543.1> (Cited on p. 6).
- Holte, James et al. (2017). “An Argo mixed layer climatology and database.” In: *Geophysical Research Letters* 44.11, pp. 5618–5626 (Cited on p. 6).
- Hourdin, Frédéric et al. (2017). “The art and science of climate model tuning.” In: *Bulletin of the American Meteorological Society* 98.3, pp. 589–602. DOI: <https://doi.org/10.1175/BAMS-D-15-00135.1> (Cited on p. 78).
- Huang, Chuan Jiang et al. (2014). “Evaluating CMIP5 simulations of mixed layer depth during summer.” In: *Journal of Geophysical Research: Oceans* 119.4, pp. 2568–2582. DOI: <https://doi.org/10.1002/2013JC009535> (Cited on p. 6).
- Hummels, Rebecca et al. (2020). “Surface cooling caused by rare but intense near-inertial wave induced mixing in the tropical Atlantic.” In: *Nature communications* 11.1, p. 3829. DOI: <https://doi.org/10.1038/s41467-020-17601-x> (Cited on p. 22).
- Jochum, Markus et al. (2013). “The impact of oceanic near-inertial waves on climate.” In: *Journal of Climate* 26.9, pp. 2833–2844. DOI: <https://doi.org/10.1175/JCLI-D-12-00181.1> (Cited on pp. 16, 21).
- Kara, A Birol et al. (2000). “An optimal definition for ocean mixed layer depth.” In: *Journal of Geophysical Research: Oceans* 105.C7, pp. 16803–16821. DOI: <https://doi.org/10.1029/2000JC900072> (Cited on p. 6).
- Kochkov, Dmitrii et al. (2024). “Neural general circulation models for weather and climate.” In: *Nature* 632.8027, pp. 1060–1066. DOI: <https://doi.org/10.1038/s41586-024-07744-y> (Cited on p. 78).
- Kolmogorov, Andrei Nikolaevich (1941/1991). “The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers.” In: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 434.1890. Originally published in 1941 in Dokl. Akad. Nauk SSSR (1941) 30(4), pp. 9–13. DOI: <https://doi.org/10.1098/rspa.1991.0075> (Cited on p. 10).
- Kuhlbrodt, Till et al. (2007). “On the driving processes of the Atlantic meridional overturning circulation.” In: *Reviews of Geophysics* 45.2. DOI: <https://doi.org/10.1029/2004RG000166> (Cited on p. 4).
- Kundu, Pijush K et al. (2016). *Fluid mechanics*. 6th ed. Elsevier. ISBN: 978-0-12-405935-1 (Cited on p. 7).
- Large, William G et al. (1994). “Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization.” In: *Reviews of geophysics* 32.4, pp. 363–403. DOI: <https://doi.org/10.1029/94RG01872> (Cited on pp. 12, 16).

- Levitus, S. et al. (2012). “World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010.” In: *Geophysical Research Letters* 39.10. DOI: <https://doi.org/10.1029/2012GL051106> (Cited on p. 4).
- Lévy, Marina et al. (2010). “Modifications of gyre circulation by sub-mesoscale physics.” In: *Ocean Modelling* 34.1-2, pp. 1–15. DOI: <https://doi.org/10.1016/j.ocemod.2010.04.001> (Cited on p. 20).
- Li, Qing et al. (2019). “Comparing Ocean Surface Boundary Vertical Mixing Schemes Including Langmuir Turbulence.” In: *Journal of Advances in Modeling Earth Systems* 11.11, pp. 3545–3592. DOI: <https://doi.org/10.1029/2019MS001810> (Cited on p. 16).
- Marshall, David P and Laure Zanna (2014). “A conceptual model of ocean heat uptake under climate change.” In: *Journal of Climate* 27.22, pp. 8444–8465. DOI: <https://doi.org/10.1175/JCLI-D-13-00344.1> (Cited on p. 4).
- Mauritsen, Thorsten et al. (2012). “Tuning the climate of a global model.” In: *Journal of advances in modeling Earth systems* 4.3. DOI: <https://doi.org/10.1029/2012MS000154> (Cited on p. 78).
- Mellor, George L and Tetsuji Yamada (1982). “Development of a turbulence closure model for geophysical fluid problems.” In: *Reviews of Geophysics* 20.4, pp. 851–875. DOI: <https://doi.org/10.1029/RG020i004p00851> (Cited on pp. 12, 13).
- Meredith, Michael and Alberto Naveira Garabato (2021). *Ocean mixing: drivers, mechanisms and impacts*. Elsevier. ISBN: 978-0-12-821512-8. URL: <https://www.sciencedirect.com/book/9780128215128/ocean-mixing> (Cited on pp. 3, 5, 11).
- Mrozowska, Marta Agnieszka et al. (n.d.). “Fast and efficient: Bayesian optimization with GPU acceleration for ocean models.” Submitted to JGR: Machine Learning and Computing. Preprint available at ESS Open Archive. URL: <https://doi.org/10.22541/essoar.173134338.88249016/v1> (Cited on pp. 15, 47).
- Mrozowska, Marta Agnieszka et al. (2024). “Using NIW Observations to Assess Mixed Layer Parameterizations: A Case Study in the Tropical Atlantic.” In: *Journal of Geophysical Research: Oceans* 129.5. e2024JCo20985. DOI: <https://doi.org/10.1029/2024JCo20985> (Cited on p. 22).
- Nielsen, Søren B et al. (2018). “An energetically consistent vertical mixing parameterization in CCSM4.” In: *Ocean Modelling* 127, pp. 46–54. DOI: <https://doi.org/10.1016/j.ocemod.2018.03.002> (Cited on p. 16).
- Olbers, Dirk and Carsten Eden (2013). “A global model for the diapycnal diffusivity induced by internal gravity waves.” In: *Journal of Physical Oceanography* 43.8, pp. 1759–1779. DOI: <https://doi.org/10.1175/JPO-D-12-0207.1> (Cited on p. 16).

- Oschlies, Andreas (2002). “Improved representation of upper-ocean dynamics and mixed layer depths in a model of the North Atlantic on switching from eddy-permitting to eddy-resolving grid resolution.” In: *Journal of Physical Oceanography* 32.8, pp. 2277–2298. DOI: [https://doi.org/10.1175/1520-0485\(2002\)032<2277:IROUOD>2.O.CO;2](https://doi.org/10.1175/1520-0485(2002)032<2277:IROUOD>2.O.CO;2) (Cited on p. 20).
- Oschlies, Andreas et al. (2018). “Drivers and mechanisms of ocean deoxygenation.” In: *Nature Geoscience* 11.7, pp. 467–473. DOI: <https://doi.org/10.1038/s41561-018-0152-2> (Cited on p. 4).
- Pacanowski, R. C. and S. G. H. Philander (1981). “Parameterization of vertical mixing in numerical models of tropical oceans.” In: *Journal of Physical Oceanography* 11.11, pp. 1443–1451. DOI: [https://doi.org/10.1175/1520-0485\(1981\)011<1443:POVMIN>2.O.CO;2](https://doi.org/10.1175/1520-0485(1981)011<1443:POVMIN>2.O.CO;2) (Cited on p. 16).
- Palmer, T. N. (2016). “A personal perspective on modelling the climate system.” In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 472.2188, p. 20150772. DOI: <https://doi.org/10.1098/rspa.2015.0772> (Cited on pp. 20, 79).
- Peacock, Synte et al. (2006). “A possible sequence of events for the generalized glacial-interglacial cycle.” In: *Global Biogeochemical Cycles* 20.2. DOI: <https://doi.org/10.1029/2005GB002448> (Cited on p. 4).
- Plueddemann, AJ and JT Farrar (2006). “Observations and models of the energy flux from the wind to mixed-layer inertial currents.” In: *Deep Sea Research Part II: Topical Studies in Oceanography* 53.1-2, pp. 5–30. DOI: <https://doi.org/10.1016/j.dsr2.2005.10.017> (Cited on p. 21).
- Pollard, Raymond T and Robert C Millard (1970). “Comparison between observed and simulated wind-generated inertial oscillations.” In: *Deep Sea Research and Oceanographic Abstracts*. Vol. 17. 4. Elsevier, pp. 813–821. DOI: [https://doi.org/10.1016/0011-7471\(70\)90043-4](https://doi.org/10.1016/0011-7471(70)90043-4) (Cited on p. 21).
- Pottapinjara, Vijay and Sudheer Joseph (2022). “Evaluation of mixing schemes in the HYbrid Coordinate Ocean Model (HYCOM) in the tropical Indian Ocean.” In: *Ocean Dynamics* 72.5, pp. 341–359. DOI: <https://doi.org/10.1007/s10236-022-01510-2> (Cited on pp. 6, 17, 75).
- Qu, Yongquan et al. (2024). “Joint parameter and parameterization inference with uncertainty quantification through differentiable programming.” In: *arXiv preprint arXiv:2403.02215*. DOI: <https://doi.org/10.48550/arXiv.2403.02215> (Cited on p. 78).
- Rackow, Thomas et al. (2024). “Multi-year simulations at kilometre scale with the Integrated Forecasting System coupled to FESOM2. 5/NEMOv3. 4.” In: *EGUsphere* 2024, pp. 1–59. DOI: <https://doi.org/10.5194/egusphere-2024-913> (Cited on pp. 20, 78).
- Reichl, Brandon G and Robert Hallberg (2018). “A simplified energetics based planetary boundary layer (ePBL) approach for ocean climate simulations.”

- In: *Ocean Modelling* 132, pp. 112–129. DOI: <https://doi.org/10.1016/j.ocemod.2018.10.004> (Cited on pp. 16, 77).
- Reynolds, Osborne (1895). “IV. On the dynamical theory of incompressible viscous fluids and the determination of the criterion.” In: *Philosophical transactions of the royal society of london.(a.)* 186, pp. 123–164. DOI: <https://doi.org/10.1098/rsta.1895.0004> (Cited on p. 9).
- Richardson, Lewis F (1922). *Weather prediction by numerical process*. Cambridge, University Press. URL: <https://archive.org/details/weatherpredictionrichrich> (Cited on p. 10).
- Rimac, Antonija et al. (2013). “The influence of high-resolution wind stress field on the power input to near-inertial motions in the ocean.” In: *Geophysical Research Letters* 40.18, pp. 4882–4886. DOI: <https://doi.org/10.1002/grl.50929> (Cited on p. 21).
- Rodi, Wolfgang (1987). “Examples of calculation methods for flow and mixing in stratified fluids.” In: *Journal of Geophysical Research: Oceans* 92.C5, pp. 5305–5328. DOI: <https://doi.org/10.1029/JC092iC05p05305> (Cited on p. 13).
- Sane, Aakash et al. (2023). “Parameterizing vertical mixing coefficients in the ocean surface boundary layer using neural networks.” In: *Journal of Advances in Modeling Earth Systems* 15.10, e2023MS003890. DOI: <https://doi.org/10.1029/2023MS003890> (Cited on p. 77).
- Schmidtko, Sunke et al. (2017). “Decline in global oceanic oxygen content during the past five decades.” In: *Nature* 542.7641, pp. 335–339. DOI: <https://doi.org/10.1038/nature21399> (Cited on p. 4).
- Shen, Chaopeng et al. (2023). “Differentiable modelling to unify machine learning and physical models for geosciences.” In: *Nature Reviews Earth & Environment* 4.8, pp. 552–567. DOI: <https://doi.org/10.1038/s43017-023-00450-9> (Cited on p. 78).
- Souza, Andre Nogueira et al. (2020). “Uncertainty quantification of ocean parameterizations: Application to the k-profile-parameterization for penetrative convection.” In: *Journal of Advances in Modeling Earth Systems* 12.12, e2020MS002108. DOI: <https://doi.org/10.1029/2020MS002108> (Cited on p. 42).
- Stevens, Bjorn et al. (2019). “DYAMOND: the DYNAMICS of the Atmospheric general circulation Modeled On Non-hydrostatic Domains.” In: *Progress in Earth and Planetary Science* 6.1, pp. 1–17. DOI: <https://doi.org/10.1186/s40645-019-0304-z> (Cited on p. 20).
- Stoustrup, Aster (2021). “Automated Parameter Tuning for the Versatile Ocean Simulator (VEROS).” MA thesis. University of Copenhagen. URL: [https://www.gfy.ku.dk/~nuterman/teamocean/docs/lda\\_Stoustrup\\_MSc\\_thesis.pdf](https://www.gfy.ku.dk/~nuterman/teamocean/docs/lda_Stoustrup_MSc_thesis.pdf) (Cited on p. 42).

- Thorpe, Stephen A (2007). *An introduction to ocean turbulence*. Vol. 10. Cambridge University Press Cambridge. ISBN: 978-0-521-67680-9 (Cited on pp. 7, 9, 10).
- Treguier, Anne Marie et al. (2023). “The mixed-layer depth in the Ocean Model Intercomparison Project (OMIP): impact of resolving mesoscale eddies.” In: *Geoscientific Model Development* 16.13, pp. 3849–3872. DOI: <https://doi.org/10.5194/gmd-16-3849-2023> (Cited on pp. 6, 17, 20, 75).
- Umlauf, Lars and Hans Burchard (2003). “A generic length-scale equation for geophysical turbulence models.” In: *Journal of Marine Research* 61.2. URL: [https://elischolar.library.yale.edu/journal\\_of\\_marine\\_research/9](https://elischolar.library.yale.edu/journal_of_marine_research/9) (Cited on pp. 13, 15, 47).
- Umlauf, Lars and Hans Burchard (2005). “Second-order turbulence closure models for geophysical boundary layers. A review of recent work.” In: *Continental Shelf Research* 25.7-8, pp. 795–827. DOI: <https://doi.org/10.1016/j.csr.2004.08.004> (Cited on p. 77).
- Umlauf, Lars and Hans Burchard (2020). *Marine Turbulence*. URL: <https://www.io-warnemuende.de/lars-umlau-teaching.html> (Cited on p. 7).
- Vallis, Geoffrey K (2016). “Geophysical fluid dynamics: whence, whither and why?” In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 472.2192, p. 20160140. DOI: <https://doi.org/10.1098/rspa.2016.0140> (Cited on p. 79).
- Van Roekel, Luke et al. (2018). “The KPP boundary layer scheme for the ocean: Revisiting its formulation and benchmarking one-dimensional simulations relative to LES.” In: *Journal of Advances in Modeling Earth Systems* 10.11, pp. 2647–2685. DOI: <https://doi.org/10.1029/2018MS001336> (Cited on p. 16).
- Waterhouse, Amy F et al. (2014). “Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate.” In: *Journal of Physical Oceanography* 44.7, pp. 1854–1872. DOI: <https://doi.org/10.1175/JPO-D-13-0104.1> (Cited on p. 6).
- Wilcox, David C (1988). “Reassessment of the scale-determining equation for advanced turbulence models.” In: *AIAA journal* 26.11, pp. 1299–1310. DOI: <https://doi.org/10.2514/3.10041> (Cited on p. 13).
- Williams, Christopher KI and Carl Edward Rasmussen (2006). *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA. ISBN: 0-262-18253-X. URL: <https://gaussianprocess.org/gpml/> (Cited on p. 42).
- Williamson, Daniel et al. (2015). “Identifying and removing structural biases in climate models with history matching.” In: *Climate dynamics* 45, pp. 1299–1324. DOI: <https://doi.org/10.1007/s00382-014-2378-z> (Cited on pp. 42, 78).
- Williamson, Daniel B et al. (2017). “Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model.” In: *Geoscientific*

- Model Development* 10.4, pp. 1789–1816. DOI: <https://doi.org/10.5194/gmd-10-1789-2017> (Cited on p. 42).
- Wong, Annie PS et al. (2020). “Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats.” In: *Frontiers in Marine Science* 7, p. 700. DOI: <https://doi.org/10.3389/fmars.2020.00700> (Cited on p. 6).
- Zanna, Laure and Thomas Bolton (2020). “Data-driven equation discovery of ocean mesoscale closures.” In: *Geophysical Research Letters* 47.17, e2020GL088376. DOI: <https://doi.org/10.1029/2020GL088376> (Cited on p. 77).
- Zanna, Laure and Thomas Bolton (2021). “Deep learning of unresolved turbulent ocean processes in climate models.” In: *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, pp. 298–306. DOI: <https://doi.org/10.1002/9781119646181.ch20> (Cited on p. 76).
- Zhai, Xiaoming et al. (2005). “Enhanced vertical propagation of storm-induced near-inertial energy in an eddying ocean channel model.” In: *Geophysical research letters* 32.18. DOI: <https://doi.org/10.1029/2005GL023643> (Cited on p. 21).
- Zhai, Xiaoming et al. (2007). “Spreading of near-inertial energy in a 1/12 model of the North Atlantic Ocean.” In: *Geophysical research letters* 34.10. DOI: <https://doi.org/10.1029/2007GL029895> (Cited on p. 21).
- Zhai, Xiaoming et al. (2009). “On the loss of wind-induced near-inertial energy to turbulent mixing in the upper ocean.” In: *Journal of Physical Oceanography* 39.11, pp. 3040–3045. DOI: <https://doi.org/10.1175/2009JPO4259.1> (Cited on p. 21).



This document was typeset using the  $\text{\LaTeX}$   $2_{\epsilon}$  document class `dionsthesis`, based on `uiothesis` developed by Eivind Uggedal. It uses Linux Libertine, developed by the Libertine Open Fonts Project, and Fira Sans, developed by the Mozilla Foundation, as body fonts. `dionsthesis` is available at:

<https://github.com/dionhaefner/dionsthesis/>

The style of `uiothesis` was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". Typographic, structural and graphical decisions in this document follow the ideas presented in Jean-Luc Doumont's book "*Trees, Maps, and Theorems*".