JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS004104

Special Collection:
Machine learning application to Earth system modeling

A Stable Implementation of a Data-Driven Scale-Aware Mesoscale Parameterization

Pavel Perezhogin[1], Cheng Zhang[2], Alistair Adcroft[2], Carlos Fernandez-Granda[1,3], and Laure Zanna[1]

[1]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, [2]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA, [3]Center for Data Science, New York University, New York, NY, USA

Key Points:

- A data-driven mesoscale eddy parameterization is implemented and evaluated in different configurations of the GFDL MOM6 ocean model
- We introduce filtering schemes to reduce the generation of grid-scale noise and enhance the large-scale backscatter
- The subgrid parameterization improves the representation of the energy distributions and the climatological mean state

Correspondence to:
P. Perezhogin,
pp2681@nyu.edu

Citation:
Perezhogin, P., Zhang, C., Adcroft, A., Fernandez-Granda, C., & Zanna, L. (2024). A stable implementation of a data-driven scale-aware mesoscale parameterization. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS004104. https://doi.org/10.1029/2023MS004104

Received 2 NOV 2023
Accepted 1 OCT 2024

Abstract Ocean mesoscale eddies are often poorly represented in climate models, and therefore, their effects on the large scale circulation must be parameterized. Traditional parameterizations, which represent the bulk effect of the unresolved eddies, can be improved with new subgrid models learned directly from data. Zanna and Bolton (2020), https://doi.org/10.1029/2020gl088376 (ZB20) applied an equation-discovery algorithm to reveal an interpretable expression parameterizing the subgrid momentum fluxes by mesoscale eddies through the components of the velocity-gradient tensor. In this work, we implement the ZB20 parameterization into the primitive-equation GFDL MOM6 ocean model and test it in two idealized configurations with significantly different dynamical regimes and topography. The original parameterization was found to generate excessive numerical noise near the grid scale. We propose two filtering approaches to avoid the numerical issues and additionally enhance the strength of large-scale energy backscatter. The filtered ZB20 parameterizations led to improved climatological mean state and energy distributions, compared to the current state-of-the-art energy backscatter parameterizations. The filtered ZB20 parameterizations are scale-aware and, consequently, can be used with a single value of the non-dimensional scaling coefficient for a range of resolutions. The successful application of the filtered ZB20 parameterizations to parameterize mesoscale eddies in two idealized configurations offers a promising opportunity to reduce long-standing biases in global ocean simulations in future studies.

Plain Language Summary This research focuses on improving the accuracy of ocean models by addressing the challenges of representing the mesoscale eddies on coarse grids. These eddies play a crucial role in the Earth's climate system, but traditional climate models struggle to capture their effects. Here, we implemented a new data-driven parameterization simulating the physics of the mesoscale eddies into the state-of-the-art ocean model. The parameterization is interpretable and captures key physical processes related to the mesoscale eddies known as energy backscatter. We tested this parameterization in two idealized ocean scenarios and found that it significantly improves the biases in the representation of the mean state and energetics. We propose new filtering schemes which improve the physical and numerical properties of the parameterization. Accurate representation of the mesoscale eddies by the present scheme has the potential to resolve long-standing biases present in global ocean models and thus allow for more reliable climate simulations.

1. Introduction

Ocean mesoscale eddies emerge on the spatial scale of the Rossby deformation radius (Salmon, 1980; Vallis, 2017), which varies in the global ocean from order 100 km near the equator to 10 km near the poles (Chelton et al., 1998). Mesoscale eddies dominate the ocean kinetic energy (KE) reservoir and are critical for the lateral and vertical transport of tracers (Ferrari & Wunsch, 2009; Redi, 1982; Uchida et al., 2017). The momentum and buoyancy fluxes produced by the mesoscale eddies are crucial in strengthening the mean flow via upgradient fluxes (R. Greatbatch et al., 2010), setting the stratification (Gent & McWilliams, 1990), and closing the global ocean energy budget (Bachman, 2019; Jansen et al., 2019; Loose et al., 2023).

To adequately simulate mesoscale eddies, several grid spacings per deformation radius are required (Hallberg, 2013). Such eddy-resolving simulations remain unfeasible for the foreseeable future in global coupled climate models (Christensen & Zanna, 2022; Hewitt et al., 2020). Therefore, the systematic effect of unresolved mesoscale eddies must be *parameterized* to reduce the biases in the climatology, variability, and response to forcing of the ocean and climate. Traditional parameterizations mimic the bulk (i.e., the mean) effect of the

mesoscale eddies on the resolved flow and are often based on energetic considerations (e.g., Jansen et al., 2015). For example, the parameterization of Gent and McWilliams (1990) and its extensions (D. P. Marshall et al., 2012; Mak et al., 2018, 2022, 2023) reduces the potential energy (PE) by flattening the isopycnals. Another popular parameterization is the *KE backscatter* (Bachman, 2019; Berloff, 2018; Frederiksen, 1999; Frederiksen & Davies, 1997; Grooms et al., 2015; Jansen et al., 2019; Jansen & Held, 2014; Juricke et al., 2020; Perezhogin, 2020; Storto & Andriopoulos, 2021; Thuburn et al., 2014), which returns energy from the subgrid scales to the larger scales. Several of these parameterizations attempt to correct the dissipation associated with eddy viscosity models by returning the KE to the resolved flow. Other parameterizations represent the net inverse energy cascade from subgrid to resolved scales.

Recently, machine-learning methods based on neural networks have been proposed to predict the eddy fluxes directly from data as an alternative to the traditional bulk modeling (Frezat et al., 2022; Gregory et al., 2023; Guan et al., 2022; Krasnopolsky et al., 2010; Perezhogin et al., 2023; Rasp et al., 2018; Sane et al., 2023; Shamekh et al., 2023; Yuval & O'Gorman, 2020; Yuval & O'Gorman, 2023; Zampieri et al., 2023). Neural networks capture the bulk effect of the subgrid eddies and the instantaneous fields of the diagnosed eddy forcing (Bolton & Zanna, 2019). These methods can be accurate but contain many tunable parameters, obstructing their physical interpretation. The approach proposed by Zanna and Bolton (2020) (hereafter ZB20) provided an alternative to both the traditional bulk approach and black-box machine-learning modeling by enabling the discovery of a closed-form equation for the eddy parameterization directly from data.

ZB20 parameterized the mesoscale eddy momentum fluxes through a simple interpretable expression that has strong ties with physics-based gradient models (Anstey & Zanna, 2017; Jakhar et al., 2024; Khani & Dawson, 2023; Zanna & Bolton, 2020). The ZB20 parameterization was trained on the data of the primitive equation ocean model MITgcm (J. Marshall et al., 1997) and accurately predicts the eddy fluxes, including the upgradient fluxes (backscatter), with a skill comparable to neural-network approaches. Once implemented online, the ZB20 parameterization improves the representation of the mean flow and energy distributions (Zanna & Bolton, 2020). However, their online simulations were limited to a simple one-layer shallow water model.

In this work, we implement the ZB20 subgrid parameterization into the primitive-equation GFDL MOM6 ocean model (Adcroft et al., 2019). In Section 2, we describe the ocean model in the adiabatic limit, governed by the stacked shallow water equations (Adcroft & Hallberg, 2006). In Section 3, we describe the ZB20 parameterization and propose modifications via filtering schemes that reduce the resulting grid-scale numerical instabilities and improve the large-scale KE backscatter. In Section 4, we test the ZB20-based parameterizations in an idealized ocean configuration, Double Gyre, and show a reduction of biases in the mean state and energetics. In Section 5, we evaluate the ZB20-based parameterizations for a range of resolutions and show that they do not require retuning of the free parameter. This is mainly a consequence of the scale-awareness (or resolution-awareness) of the free parameter, which is solely a function of the grid spacing. We finally describe tests of the parameterization in a more complex configuration, NeverWorld2 (NW2, Marques et al., 2022), with a cross-equatorial basin and Southern Ocean re-entering channel in Section 6. We find a more pronounced bias reduction in energy distributions and mean state, compared to the simpler Double Gyre configuration. However, our experimentation in the NW2 configuration revealed the need for additional modifications to improve the numerical stability properties of the parameterization, similarly to other backscatter schemes (Yankovsky et al., 2024).

2. Ocean Model

We use the GFDL MOM6 ocean model (Adcroft et al., 2019) in an adiabatic limit with no buoyancy forcing. This allows us to test the direct impact of the new parameterization in idealized settings of a primitive equation model.

The equations of motion are given by the stacked shallow water equations with constant density in each layer (Marques et al., 2022; Zhang et al., 2023).

$$\partial_t \mathbf{u}_k + \frac{f+\zeta_k}{h_k}\hat{\mathbf{z}} \times (h_k \mathbf{u}_k) + \nabla K_k + \nabla M_k = \mathbf{F}_k + \mathbf{V}_k + \mathbf{S}_k, \tag{1}$$

$$\partial_t h_k + \nabla \cdot (\mathbf{u}_k h_k) = 0, \tag{2}$$

where k is the index of the vertical fluid layer, equal to 1 for the surface layer and to $\mathcal{K}$ for the bottom layer; $\mathbf{u}_k = (u_k, v_k)$ is the horizontal velocity, where u_k and v_k are zonal and meridional velocities; h_k is the layer thickness; f is the Coriolis parameter; $\zeta_k = \partial_x v_k - \partial_y u_k$ is the vertical component of the relative vorticity; $\nabla = (\partial_x, \partial_y)$ is the horizontal gradient operator and $\nabla\cdot$ is the horizontal divergence operator, where ∂_x and ∂_y are partial derivatives along zonal and meridional directions; $\hat{\mathbf{z}}$ is the unit vector pointing upward; $\hat{\mathbf{z}} \times \mathbf{u}_k = (-v_k, u_k)$ is the cross product; $K_k = (1/2)\mathbf{u}_k \cdot \mathbf{u}_k$ is the KE per unit mass. The Montgomery potential is given by $M_k = \sum_{l=1}^{k} g'_{l-1/2}\eta_{l-1/2}$, where $\eta_{k+1/2} = -H + \sum_{n=k+1}^{\mathcal{K}} h_n$ is the interface position between layers k and $k+1$ and $H(x,y) \geq 0$ is the ocean depth; $g'_{k+1/2} = g(\rho_{k+1} - \rho_k)/\rho_0$ is the reduced gravity, where ρ_k is the density of the fluid layer, ρ_0 is the reference density and g is the gravitation acceleration. The equations of motion in the horizontal orthogonal curvilinear coordinates are discussed in Adcroft et al. (2019). $\mathbf{F}_k$ represents the wind stress and bottom drag and $\mathbf{S}_k$ is a subgrid momentum parameterization. $\mathbf{V}_k$ is a biharmonic Smagorinsky model (subsequently, we omit index k for brevity), with a viscosity $\nu_4 = C_S \Delta^4 \sqrt{\tilde{D}^2 + D^2}$, where C_S is the non-dimensional Smagorinsky coefficient, Δ is the grid spacing, $\tilde{D} = \partial_x u - \partial_y v$ is the stretching deformation, and $D = \partial_y u + \partial_x v$ is the shearing deformation. We refer the reader to Griffies and Hallberg (2000) for more details regarding the form and implementation of the biharmonic operator used in MOM6.

3. Subgrid Parameterizations

Ocean models at a coarse grid resolution have strong biases in the representation of the mean flow and energetics (Hallberg, 2013; Hewitt et al., 2020). They can be corrected to some degree by parameterizing the effect of the unresolved (subgrid) mesoscale eddies. In this section, we describe how to diagnose the effect of subgrid mesoscale eddies on the resolved flow in the momentum equation using a spatial filtering approach (Large Eddy Simulation, LES, Bachman et al., 2017; Fox-Kemper & Menemenlis, 2008; Sagaut, 2006). We then describe our implementation of the ZB20 parameterization of the subgrid mesoscale eddies and different baselines. The buoyancy fluxes and their parameterizations (Gent & McWilliams, 1990) are absent in all model calculations documented in the paper.

3.1. Subgrid Momentum Forcing

The subgrid mesoscale eddies produce the following subgrid momentum forcing acting on the resolved eddies (Zanna & Bolton, 2020)

$$\mathbf{S} = \left(\overline{\mathbf{u}} \cdot \overline{\nabla}\right)\overline{\mathbf{u}} - \overline{(\mathbf{u} \cdot \nabla)\mathbf{u}}, \tag{3}$$

where $\mathbf{u}$ is the velocity field of the high-resolution model, $\overline{(\cdot)}$ is a spatial filtering and coarse-graining operator. Here $\left(\overline{\mathbf{u}} \cdot \overline{\nabla}\right)\overline{\mathbf{u}}$ is the numerical approximation of the advection operator on a coarse grid with the scheme used in MOM6. Specifically, we employ the Sadourny (1975) energy-conserving scheme formulated in vector-invariant form using the identity $(\mathbf{u} \cdot \nabla)\mathbf{u} = (\zeta/h)\hat{\mathbf{z}} \times (h\mathbf{u}) + \nabla K$. The subgrid forcing (Equation 3) modifies the governing equations of the coarse ocean model as shown in Equation 1. Note that in the LES approach, we should use $\overline{\mathbf{u}}$ whenever referring to the solution of the coarse model, but here, for brevity, we omit this notation everywhere apart from in Equation 3. To enable computations with a coarse ocean model, we should represent the subgrid forcing as a function of the resolved flow, referred to as a *parameterization*.

3.2. Zanna-Bolton Parameterization (ZB20)

In this section, we describe the original ZB20 parameterization and two filtered modifications, referred to as filtered ZB20 parameterizations. All three variants are referred to as ZB20-based parameterizations.

The original ZB20 parameterization for subgrid momentum forcing is given by

$$\mathbf{S} = \begin{pmatrix} S_x \\ S_y \end{pmatrix} = \nabla \cdot \mathbf{T} = \nabla \cdot \begin{pmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{pmatrix} = \begin{pmatrix} \partial_x T_{xx} + \partial_y T_{xy} \\ \partial_x T_{xy} + \partial_y T_{yy} \end{pmatrix}. \tag{4}$$

The parameterization was discovered, using a machine learning algorithm, from data generated from baroclinic ocean simulations Zanna and Bolton (2020). The stress tensor $\mathbf{T}$ is given by

$$\mathbf{T}\left(\zeta, D, \tilde{D}\right) = \kappa_{BC} \underbrace{\begin{bmatrix} -\zeta D & \zeta\tilde{D} \\ \zeta\tilde{D} & \zeta D \end{bmatrix}}_{\mathbf{T}_d,\ \text{deviatoric stress}} + \frac{\kappa_{BC}}{2}\underbrace{\left(\zeta^2 + D^2 + \tilde{D}^2\right)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{T}_I,\ \text{isotropic stress}}. \tag{5}$$

The parameterization is applied independently in every model layer. Here, we follow the approach of Anstey and Zanna (2017) and gradient models studies (Chen et al., 2003; Meneveau & Katz, 2000) to relate the free coefficient κ_{BC} to the area of a coarse grid box:

$$\kappa_{BC} = -\gamma\Delta_x\Delta_y \leq 0, \tag{6}$$

where Δ_x and Δ_y are local grid spacings along x and y directions, respectively, and $\gamma \approx 1$ is a tunable non-dimensional parameter. In 2D incompressible fluids, the ZB20 parameterization (Equation 5) is equivalent to the nonlinear gradient model, which is given by $\mathbf{T} = -l^2(\nabla\mathbf{u})\cdot(\nabla\mathbf{u})^{\dagger}$, where $\nabla\mathbf{u}$ is the velocity gradient tensor, l is the filter width, $\dagger$ is the matrix transpose and $\cdot$ is the matrix multiplication; see Equation 59 in Anstey and Zanna (2017) for details. However, as opposed to the nonlinear gradient model, the ZB20 parameterization excludes the explicit dependence on the horizontal divergence ($\sigma = \partial_x u + \partial_y v$) in stacked shallow water equations (see, for example, Zanna and Bolton (2020) for further discussion on the difference between ZB20 and the nonlinear gradient model). Based on limited numerical experiments in the NW2 configuration, the additional terms in the nonlinear gradient models, which explicitly depend on σ, have a destabilizing effect on the simulations (not shown).

We refer the reader to Appendix A for details of the numerical discretization of Equations 4 and 5, accounting for the curvilinear coordinates and varying layer thickness in MOM6. For interested readers, we note that, as opposed to the original ZB20 model, the implemented ZB20 parameterization may spuriously predict non-zero accelerations for a state of solid body rotation ($\zeta = \text{const}$, $D = \tilde{D} = 0$). These non-zero accelerations are caused by the spatially varying parameterization coefficient κ_{BC}, leading to a non-zero divergence of the isotropic stress tensor ($\nabla\cdot(\kappa_{BC}\mathbf{T}_I) \neq 0$). We anticipate that the non-zero accelerations will be small when the coefficient κ_{BC} varies over large spatial scales, as it is the case for the ZB20-based parameterizations in the current implementation. Other spatially varying coefficients, such as the layer thickness in the stress divergence expression (Equation A4) or the Coriolis parameter in the attenuation function (Equation 11, to be introduced later) can lead to a non-zero divergence of the isotropic stress tensor as well.

3.2.1. Low-Pass Filtering of the Stress Tensor (ZB20-Smooth)

Incorporation of the ZB20 parameterization, which is meant to dissipate energy at small scales, can generate numerical noise near the grid scale. It happens because the dissipation near the grid scale is not a hard constraint of the parameterization. Specifically, the energetic contribution of the deviatoric stress is zero after integration by parts ($\mathbf{T}_d : (\nabla\mathbf{u}) = 0$, see Equation A9 in Appendix A) while the energetic contribution of the isotropic stress ($\mathbf{T}_I : (\nabla\mathbf{u})$) is not sign-definite (Zanna & Bolton, 2020). Dissipation at the grid scale is often enforced in gradient-based parameterizations by projecting the predicted stress tensor onto the dissipative direction (Balarac et al., 2013; Bouchet, 2003; Vollant et al., 2016). However, this approach is unsuitable because it would remove an important effect of the KE backscatter on large scales. Thus, we suggest removing the contribution of the ZB20 parameterization on the grid scale eddies by low-pass filtering the stress tensor. The filtered stress tensor will represent the KE backscatter. Filtering is widely used in mesoscale eddy parameterizations to suppress the numerical noise and increase the spatial scale of the KE backscatter (Bachman, 2019; Grooms et al., 2015; Jurricke et al., 2019; Perezhogin et al., 2019).

We consider a low-pass convolutional filter defined on 3×3 spatial stencil ("trapezoidal filter" in San (2014)), which is applied in every fluid layer independently:

$$\mathcal{G} = \frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \tag{7}$$

The presented filter has the smallest spatial stencil among filters, completely removing the grid harmonics. It will be used as a building block for proposing our filtered ZB20 parameterizations.

The ZB20 parameterization with low-pass filtered stress tensor (hereafter, ZB20-Smooth) is given by:

$$\mathbf{S} = \nabla \cdot G(\mathbf{T}), \tag{8}$$

where $G = \mathcal{G}^N$ is the low-pass filter which is applied to every component of the stress tensor $\mathbf{T}$ (Equation 5), and N is the number of filtering passes of $\mathcal{G}$ (Equation 7). We choose $N = 4$ similarly to Juricke et al. (2019) (see Appendix C for sensitivity to the choice of N). We implement the filtering using the marching-halo algorithm; that is, 4 filter iterations are performed within a single MPI exchange. The computational cost of the original ZB20 parameterization is 2.5% of the ocean model runtime in the NW2 configuration, while the filtered parameterization requires 4% of the runtime. Note that filters more appropriate for tensor elements were presented in Aluie (2019).

3.2.2. High-Pass Filtering of the Velocity Gradients (ZB20-Reynolds)

We consider an additional filtering scheme, which can also enhance the KE backscatter.

Perezhogin and Glazunov (2023) show that the Reynolds stress is responsible for the KE backscatter. The Reynolds stress represents the effect of the eddy-eddy interactions on the mean flow, through the Germano (1986) decomposition of the subgrid stress. To isolate the effect of eddy-eddy interaction from the Reynolds stresses, we propose a modification of the ZB20 parameterization by using a high-pass filter on the velocity gradients. Therefore, the modified parameterization (hereafter, ZB20-Reynolds) can be expressed as

$$\mathbf{S} = \nabla \cdot G\left(\mathbf{T}\left(\zeta', D', \tilde{D}'\right)\right), \tag{9}$$

where $G = \mathcal{G}^N$ is the low-pass filter and $(\cdot)' = I - G$ is the high-pass filter with I being the identity operator. The computation of the stress tensor in Equation 9 is done as follows. First, the velocity gradients $\zeta, D, \tilde{D}$ are high-pass filtered to obtain $\zeta', D', \tilde{D}'$, respectively. Then we compute the stress tensor $\mathbf{T}\left(\zeta', D', \tilde{D}'\right)$ according to Equation 5 using high-pass filtered fields as inputs. Finally, the stress tensor $\mathbf{T}$ is low-pass filtered with the filter G to separate the parameterization tendency from the grid scale. For consistency with the previous section, we choose $N = 4$. This additional filtering results in an increase in the computational cost of the parameterization, which is 6% of the ocean model runtime in the NW2 configuration.

3.3. Baseline Momentum Parameterizations

We consider multiple backscatter parameterizations as baselines.

The first baseline is the KE backscatter of Jansen et al. (2015) (referred to as JHAH15), already tested in MOM6 (Jansen et al., 2019). The JHAH15 parameterization mainly represents the reinjection of KE energy originally dissipated with the biharmonic Smagorinsky model. The backscatter of the subgrid KE is parameterized using a negative Laplacian viscosity model (anti-viscosity). The negative viscosity coefficient is informed by a local equation for vertically averaged subgrid KE. We will also consider an updated version of the JHAH15 parameterization, from Yankovsky et al. (2024) for multi-layer models in Section 6.

The second baseline parameterization is a deep-learning convolution neural network (CNN) model of Guillaumin and Zanna (2021) (referred to as GZ21). It predicts the subgrid forcing using horizontal velocities and was trained on data from a coupled climate simulation. It was implemented in MOM6 (Zhang et al., 2023), together with a

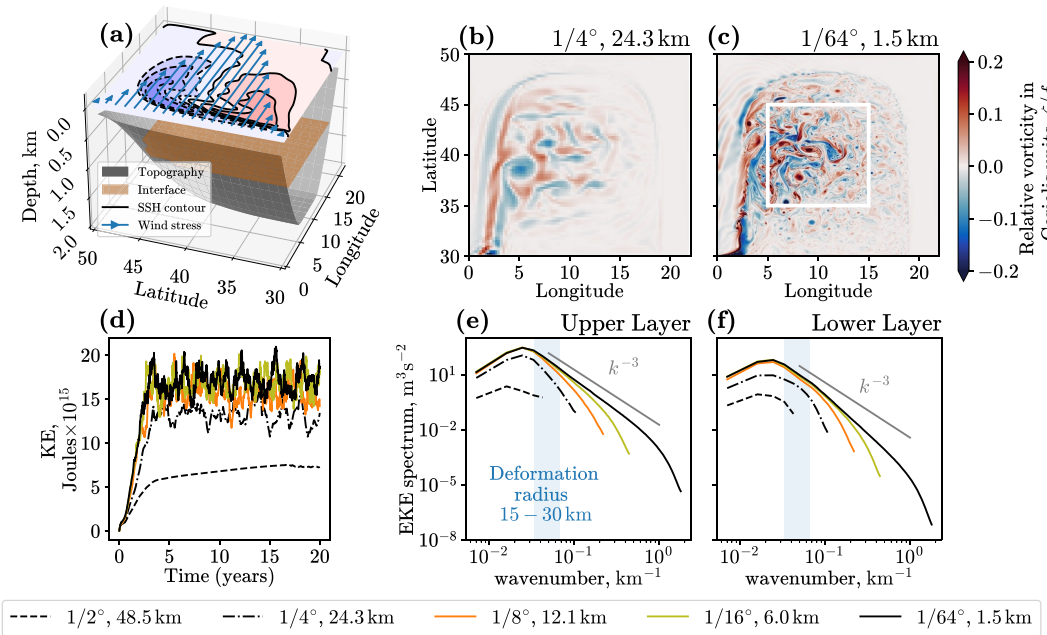

Figure 1. (a) Schematic of the MOM6 Double Gyre configuration; (b) snapshot of relative vorticity in the upper layer in the 1/4° simulation and (c) 1/64° simulation; (d) timeseries of total kinetic energy (KE) (summed over layers) for the different resolutions. The spatial spectra of the eddy kinetic energy as a function of wavenumbers in the (e) upper and (f) lower fluid layers. The spectrum is computed within the area enclosed by the white rectangle. The wavenumbers corresponding to the deformation radius $\left(r_d \approx f^{-1}\sqrt{g' h_1 h_2 (h_1+h_2)^{-1}}\right)$, given by $1/r_d$, are shaded in blue.

biharmonic Smagorinsky model. The GZ21 parameterization energizes the resolved eddies efficiently and consequently can be tuned to increase the KE of the coarse model up to the KE of the high-resolution model.

4. Experiments in Double Gyre Configuration

We first test the subgrid parameterizations in the MOM6 Double Gyre configuration described in Section 3.1 of Zhang et al. (2023). We use their setting but double the integration time (20 years) with a single ensemble member. The model has two fluid layers, initially at rest. Momentum input is via the wind and dissipation by the bottom friction. The computational domain is on a spherical grid with a bowl topography (Figure 1a). The Smagorinsky coefficient in all experiments is set to $C_S = 0.06$, similar to Jansen et al. (2015), Jansen et al. (2019), Zhang et al. (2023), see Appendix C for sensitivity to C_S.

The unparameterized simulations ($\mathbf{S} = 0$) for a range of resolutions (1/64°, 1/16°, 1/8°, 1/4°, 1/2°) are shown in Figure 1. The high-resolution model (1/64°) has a grid spacing (≈ 1.5 km) that is 10 times smaller than the Rossby deformation radius (15 km − 30 km), and consequently, it directly simulates the mesoscale eddies (Hallberg, 2013). Coarse ocean models with a grid spacing in a range from 1/2° (≈ 50 km) to 1/8° (≈ 12 km) barely resolve the Rossby deformation radius. These models have a reduced eddy kinetic energy (EKE) spectrum compared to the high-resolution model (1/64°), with the coarsest models (1/2° and 1/4°) failing to capture the spectrum at all spatial scales, see Figures 1e and 1f.

In this section, we analyze the impact of subgrid parameterizations, in particular, on improving biases in energetics and the mean states in the coarse resolution model with horizontal grid spacing of 1/4°. The values of the scaling coefficient γ of the ZB20-based parameterizations used in this section are reported in Table 1. The coefficient γ is larger for the filtered than for the unfiltered versions of the ZB20 parameterization, since filtering reduces the magnitude of the subgrid forcing. The subgrid parameterizations JHAH15, GZ21 and ZB20-Reynolds are tuned to approximately match the total KE, and the ZB20-Smooth parameterization is tuned to match the available potential energy (APE). Sensitivity to the scaling coefficient and performance at multiple resolutions are discussed in Section 5.

Table 1
Values of the Default Scaling Coefficient γ in the Different ZB20-Based Parameterizations

	ZB20	ZB20-Smooth	ZB20-Reynolds
γ	0.5	1.0	2.0

4.1. Eddy Kinetic Energy (EKE) Spectrum

The EKE spectrum is one of the metrics that coarse unparameterized ocean models fail to reproduce (Figures 1e and 1f). The EKE spectrum is defined as a time-averaged spatial power spectrum of the eddy velocities $\mathbf{u}' = \mathbf{u} - \overline{\mathbf{u}}^t$, where $\overline{(\cdot)}^t$ is a time-average over the last 10 years of the simulations. All spectra in the Double Gyre configuration are computed in the white rectangle shown in Figure 1c using a 2D Fourier transform with a Hann window and linear detrending (using xrft package, Uchida et al., 2023). Throughout the paper, we ignore the contribution of the thickness h in the definition of spectral properties related to KE.

The EKE spectra for runs with different subgrid parameterizations are shown in Figure 2. The ZB20 parameterization without filters has a build-up of energy near the grid scale, that is, numerical noise (Figure 2a), which results in the deterioration of the EKE spectrum at large scales in both fluid layers. The proposed filtering techniques (ZB20-Smooth and ZB20-Reynolds) allow us to attenuate the grid scale noise generation. The ZB20-Smooth and ZB20-Reynolds parameterizations improve the EKE spectrum at large scales by increasing it compared to the unparameterized model, reaching levels closer to the high-resolution model. The ZB20-Reynolds parameterization is more efficient in energizing eddies in the upper layer. The JHAH15 and GZ21 baseline parameterizations are more efficient in energizing eddies near the deformation scale. None of the tested subgrid parameterizations reproduce the EKE spectrum across all spatial scales.

4.2. Subgrid Kinetic Energy Transfer

In this section, we analyze the subgrid energy transfer in coarse parameterized models to explain the shape of the EKE spectrum. The energy transfer is computed both offline and online because the accuracy of the parameterization can change once it is included in the coarse-grid model (Meneveau & Katz, 2000; Ross et al., 2023). Note that the offline analysis evaluates the subgrid parameterization on the filtered and coarse-grained snapshots of the high-resolution model, while the online analysis evaluates the parameterization once the parameterized ocean model is integrated over time.

The KE transfer spectrum of the subgrid forcing or parameterization $\mathbf{S} = (S_x, S_y)$ is given by

$$\mathcal{T}(\kappa_x, \kappa_y) = \mathrm{Re}[\mathcal{F}(u)^* \mathcal{F}(S_x) + \mathcal{F}(v)^* \mathcal{F}(S_y)], \tag{10}$$

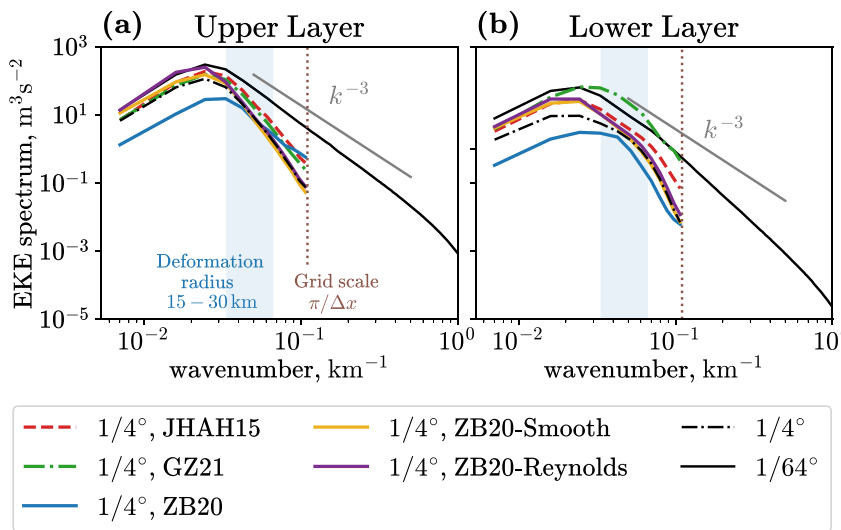

Figure 2. The eddy kinetic energy spatial spectrum as a function of wavenumbers for the last 10 years of the simulation: (a) upper layer, (b) lower layer. The unparameterized model (1/4°, dotted dashed line) is compared to parameterized models (JHAH15 in red, GZ21 in green, ZB20 in blue, ZB20-Smooth in yellow, ZB20-Reynolds in violet) and high-resolution simulation (1/64°, black line). The spectrum is computed in the region indicated by the white rectangle in Figure 1.

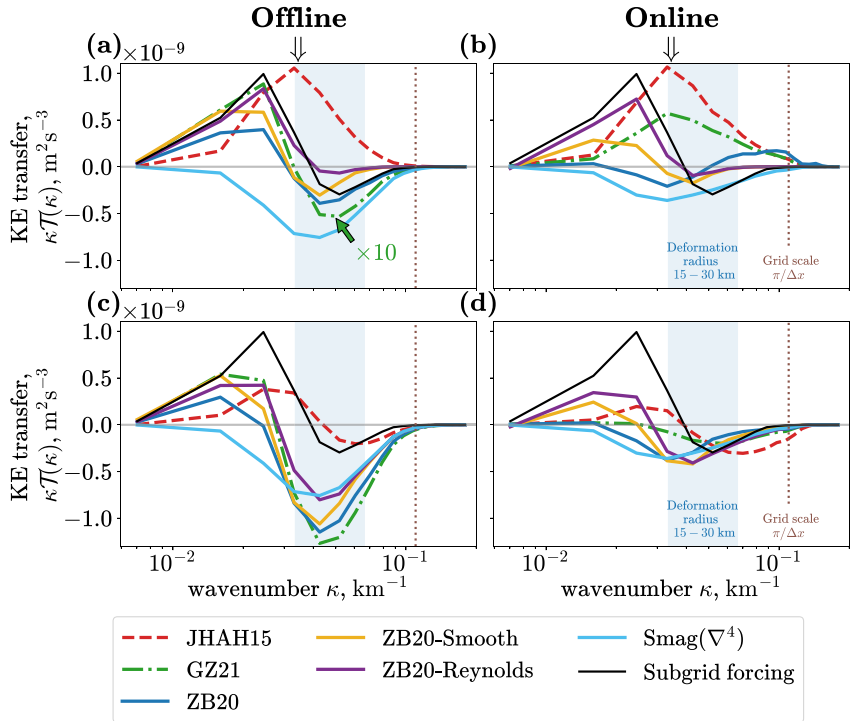

Figure 3. Kinetic energy (KE) transfer in offline (left column) and online (right column) analysis. The target resolution for offline analysis is 1/4° and the online performance of the parameterizations is computed for simulations at 1/4° resolution. (a) Offline contribution from each subgrid parameterization (GZ21 in green, ZB20 in blue, ZB20-Smooth in yellow, ZB20-Reynolds in violet, negative viscosity part of JHAH15 in red) and a biharmonic Smagorinsky model "Smag(∇^4)" in light blue. The black line is the subgrid forcing, Equation 3. (b) The contribution of each parameterization online. (c) and (d): same as panels (a) and (b) but showing the combined contribution of each parameterization with a biharmonic Smagorinsky model offline and online, respectively. The offline analysis is performed on fields of the high-resolution model (1/64°) filtered and coarse-grained to 1/4° resolution. The spectra are computed over the last 10 years of the simulation for the upper fluid layer, in the region enclosed by the white rectangle shown in Figure 1. The KE transfer of GZ21 in panels (a) and (c) is multiplied by 10 for convenience.

where κ_x and κ_y are zonal and meridional wavenumbers, respectively, $\mathcal{F}$ is the 2D Fourier transform, * is the complex conjugate, and Re is the real part. We integrate the two-dimensional transfer spectrum over circles $\left(\kappa_x^2 + \kappa_y^2 = \kappa^2\right)$ to obtain the isotropic transfer spectrum $\mathcal{T}(\kappa)$. We compute the subgrid forcing according to Equation 3, where the filtering operator $\overline{(\cdot)}$ is defined as the Gaussian filter implemented in the GCM-filters package (Grooms et al., 2021; Loose et al., 2022) followed by coarse-graining. The filtering and coarse-graining operations are applied to the 1/64° model output toward a 1/4° resolution. The width of the Gaussian filter is chosen to be proportional to the grid spacing of the coarse resolution model by introducing a filter-to-grid width ratio (FGR, Chow & Moin, 2003; Ghosal, 1996; Lund, 1997; Perezhogin & Glazunov, 2023). The FGR is not a coefficient used for the online implementation of the parameterization; it is only a parameter used to diagnose the offline subgrid transfer. The KE transfer of the subgrid models in the offline analysis depends on the FGR, because, as expected, it depends on grid-scale features. However, the shape of the offline transfer spectrum as a function of wavenumber remains roughly unchanged, while the amplitude varies for different FGR values. For FGR = 2.5, the offline and online KE transfer roughly coincide for the filtered ZB20 and JHAH15 parameterizations (Figure 3). We chose this FGR value in our analysis, which likely corresponds to the filter width that is effectively reproduced in online simulations by the parameterized ocean models.

The subgrid energy transfer contains an interval with small-scale energy and enstrophy dissipation ($\mathcal{T}(\kappa) < 0$) and an interval of large-scale energy backscatter ($\mathcal{T}(\kappa) > 0$), see black line in Figure 3. The unfiltered ZB20 parameterization simulates the positive and negative energy transfer relatively well in the offline analysis; see Figure 3a, blue line. However, in the online simulations, the energy transfer is reversed: the ZB20 parameterization dissipates energy at large scales and generates energy near the grid scale (Figure 3b). This explains the

emergence of the grid-scale numerical noise and deterioration of the large scales in the EKE spectrum for the simulation with the unfiltered ZB20 parameterization (Figure 2). The filtering techniques employed for the ZB20-Smooth and ZB20-Reynolds parameterizations remove the generation of the numerical noise near the grid scale and preserve the original properties of the ZB20 parameterization. Consequently, the backscattering for these filtered parameterizations is located over the large scales (Figure 3b). The ZB20-Reynolds parameterization appears more efficient than ZB20-Smooth in reproducing the KE backscatter.

The dissipation predicted by the filtered ZB20 parameterizations alone is insufficient to parameterize the direct cascade of enstrophy from resolved to subgrid scales because the filtering schemes diminish the effect of these parameterizations at short wavelengths. Thus, all the online experiments are performed with a biharmonic Smagorinsky model, which is effective at short wavelengths. We note that the eddy viscosity model of Leith (1996) may be more suitable for parameterizing the enstrophy dissipation (Bachman et al., 2017). The lower row in Figure 3 shows a net KE transfer between the resolved and subgrid scales, which is given by a combined contribution of ZB20-based parameterizations with a biharmonic Smagorinsky model. The net KE transfer for the filtered ZB20 parameterizations exhibits less accurate correspondence with the diagnosed subgrid KE transfer, displaying an overestimation of dissipation and an underestimation of backscatter, both offline and online (see Figures 3c and 3d). However, the efficacy of the filtered ZB20 parameterizations can be attributed to the scale separation between backscatter and dissipation effects in the net KE transfer, which is crucial for effectively energizing the resolved flow (Bagaeva et al., 2024; Jansen & Held, 2014).

The JHAH15 parameterization returns energy in shorter wavelengths than the filtered ZB20 parameterizations (Figures 3c and 3d). We explain the increased energy density near the deformation scale for the JHAH15 parameterization (Figure 2) by the effect of scale-selective backscatter, which may be seen as an advantage; however, it can also lead to numerical instabilities (Bachman, 2019; Bagaeva et al., 2024; Grooms, 2023; Grooms et al., 2015; Juricke et al., 2019). Note that while the net transfer for the JHAH15 parameterization does not match the diagnosed KE transfer well, the JHAH15 parameterization is expected to be more accurate when compared to the subgrid forcing diagnosed with the cut-off LES filter (Perezhogin & Glazunov, 2023).

For the convenience of plotting, we multiply the offline KE transfer of the GZ21 parameterization by 10 in Figures 3a–3c; this correction is not applied in online runs presented on panels (b,d). Note that the magnitude of offline prediction depends a lot on the FGR, while the shape of the predicted KE transfer is less FGR-sensitive and thus a more crucial metric. The GZ21 parameterization accurately predicts the shape of the KE transfer in offline analysis (Figure 3a). However, like the unfiltered ZB20 parameterization, GZ21 performs differently for online experiments. The online KE transfer spectrum for the GZ21 parameterization is purely positive and lacks a dissipative region. The shape of the predicted KE transfer resembles the negative Laplacian viscosity model (Figure 3b), thus suggesting that GZ21 returns energy in short wavelengths and should be effective in improving the energy density near the deformation scale (Figure 2). A potential drawback of the GZ21 parameterization is the lack of scale separation between dissipation and backscatter models (Figure 3d).

4.3. Kinetic and Potential Energy

We compute the KE and APE (Appendix B) to address how the parameterizations affect the energy partitioning between these reservoirs. We further split each reservoir such that the KE is the sum of the mean kinetic energy (MKE), averaged over the last 10 years of the simulations, and the EKE. We similarly decompose the APE into the mean potential energy (MPE) and eddy potential energy (EPE).

We compare the coarse-resolution models, with and without parameterizations, to the high-resolution model and its coarse-grained output (Figure 4). The APE of the unparametrized model (1/4°) is too large, and its KE is too low compared to the high-resolution simulation. The ZB20 parameterization improves (reduces) the APE, as expected from the mesoscale parameterization (Figure 4b). However, due to the generation of numerical noise, the large-scale eddies are disrupted, and the total KE is too low (Figure 4a). Note that compared to the physical buoyancy parameterizations (Gent & McWilliams, 1990), the reduction of APE by the ZB20 parameterization appears indirectly through the change of the mean state. Directly reducing the APE would require an additional parameterization in the thickness equation (Loose et al., 2023), which we omit here. The ZB20-Smooth parameterization improves (reduces) the APE but with little change to the KE; it also reduces the MKE in agreement with the high-resolution model (Figure 4a). The backscatter parameterizations (ZB20-Reynolds, JHAH15 and GZ21) efficiently energize the flow by increasing the KE. Additionally, they improve (reduce) the

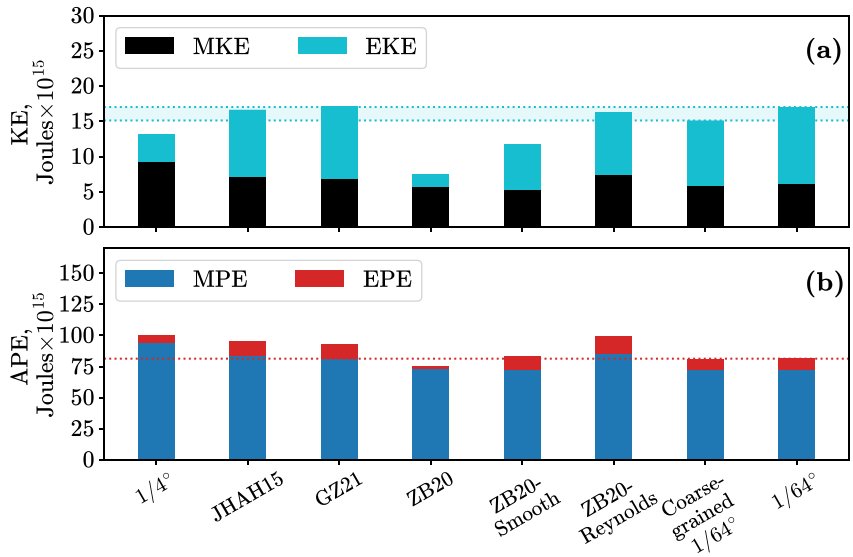

Figure 4. Energy reservoirs in numerical experiments at resolution 1/4°: (a) Kinetic energy (KE) of the mean flow (MKE) and eddies (EKE), (b) Available potential energy (APE) of the mean flow (MPE) and eddies (EPE). In both plots, energy is summed over the fluid layers or interfaces and averaged over the last 10 years. The horizontal dotted lines represent the ground truth value of the total kinetic (a) and total potential (b) energy.

energy of the mean state (MKE and MPE). However, they are less accurate in predicting the APE than the ZB20-Smooth parameterization.

4.4. Mean State

We concentrate on metrics related to the climatological mean state (Figure 5), such as the time-mean sea surface height (SSH). The unparameterized model (1/4°) has a strong persistent recirculation near the western boundary, absent in the high-resolution model. In all parameterized simulations (ZB20, ZB20-Smooth, ZB20-Reynolds, JHAH15, GZ21), we observe an improvement in the representation of the mean state. The persistent recirculation is less evident, and the region exhibits a meandering jet pattern similar to that simulated in the high-resolution model. The ZB20-Smooth model is the most accurate in representing the mean state, and its root mean squared error (RMSE) is significantly lower than for other parameterized models (see Figure 5 for RMSE SSH values used to describe the mean bias in SSH). The ZB20-Smooth parameterization improves the mean state without increasing the KE (i.e., without strong backscatter). However, the most efficient backscatter parameterizations (ZB20-Reynolds, JHAH15, GZ21) have higher RMSE for SSH. By increasing the scaling coefficient γ, we can further improve the RMSE for the ZB20-Reynolds model (Appendix C). However, the value is constrained by numerical stability, and here, we find the optimal coefficient $\gamma = 2.8$ to be on the boundary of the stability region.

4.5. Eddy-Mean Flow Interaction

The two previous sections show that the parameterizations that are best in reproducing the KE backscatter do not demonstrate optimal performance in reproducing the mean SSH and APE. According to Moser et al. (2021), the time-mean subgrid stress is another property of the subgrid eddies that affects the mean flow prediction. In geophysical fluid flows, a similar effect is known as eddy-mean flow interaction and has been analyzed in several studies (Andrews & Mcintyre, 1976; R. J. Greatbatch, 1998, 2010; R. Greatbatch et al., 2010; Hoskins et al., 1983; Hughes & Ash, 2001; Kamenkovich et al., 2009; Qiu & Chen, 2010; Wardle & Marshall, 2000; Waterman & Jayne, 2011). Following Hughes and Ash (2001), we show the 10-year averaged zonal acceleration produced by the subgrid eddies in the upper layer (Figure 6). The subgrid forcing amplifies the resolved jet by accelerating the jet current extension eastward (longitude >5°) and decelerating the jet in the separation region (longitude <5°), see Figure 6a. A similar pattern was shown in Zanna and Bolton (2020); see their Figures S2 and S4. The time-mean contribution of the eddy-viscosity models (biharmonic Smagorinsky and JHAH15) is too small compared to

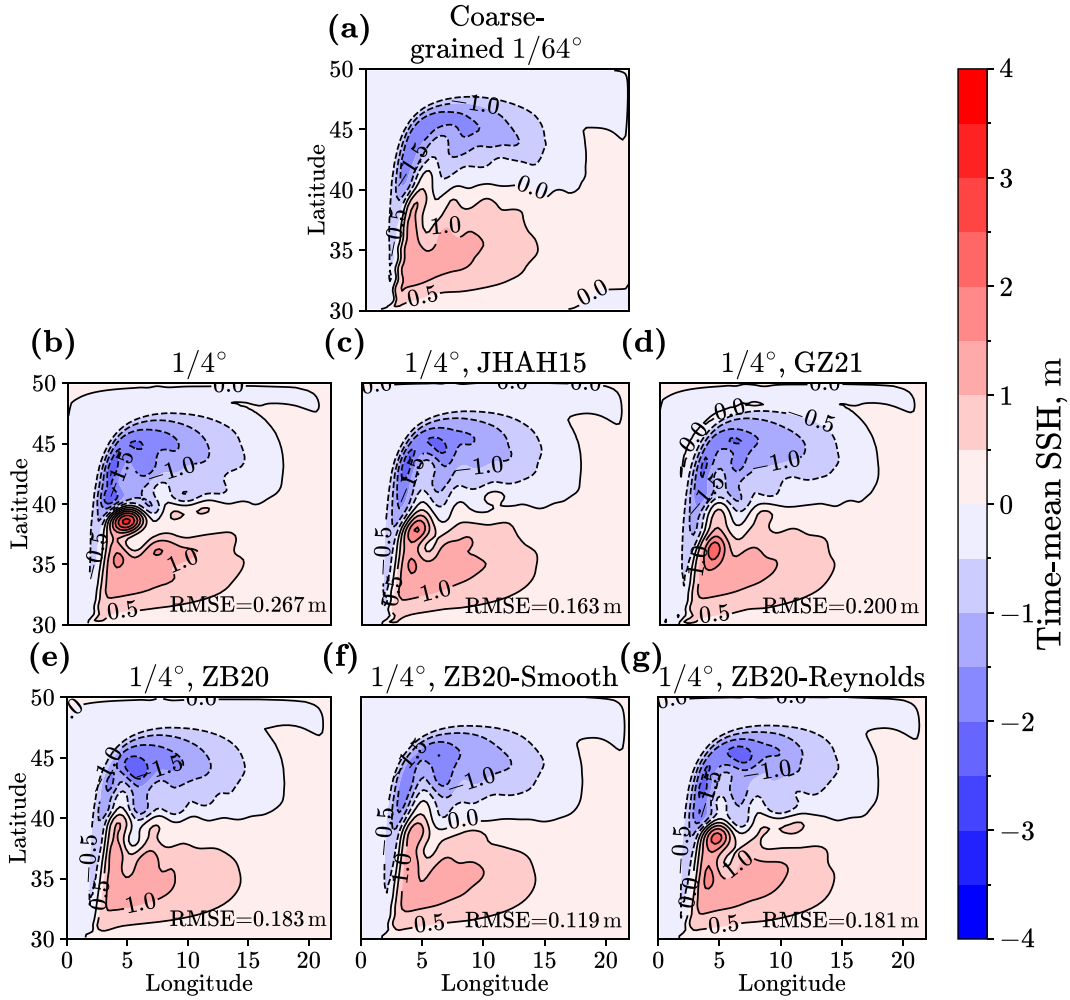

Figure 5. Sea surface height (SSH) averaged over the last 10 years for experiments at 1/4° resolution. For every coarse model, we provide the root mean squared error (RMSE) in the time-mean SSH with respect to the coarse-grained high-resolution model shown in panel (a).

the diagnosed subgrid forcing (Figures 6b and 6c), as expected from Moser et al. (2021). The GZ21 parameterization produces westward accelerations in most of the domain and thus disagrees with the diagnosed subgrid forcing (Figure 6d). The ZB20-Smooth and ZB20-Reynolds parameterizations reproduce the acceleration pattern of the subgrid forcing most accurately, with the ZB20-Smooth having the largest accelerations (Figures 6f and 6g). The zonal acceleration produced by the unfiltered ZB20 parameterization is smaller and less accurate compared to the filtered models (Figure 6e). The success of the ZB20-based parameterizations in improving the mean state and the APE appears to be related to their effect on the time-mean zonal acceleration (Anstey & Zanna, 2017).

5. Sensitivity Study and Scale Awareness

The parameterization effect of mesoscale eddies, which are partially resolved, should diminish as the grid is refined (Haidvogel et al., 2017). We achieve this property by informing the parameterization with the local grid spacing using Equation 6. Parameterizations with such scaling of the free coefficient are often referred to as scale-aware (Bachman et al., 2017; Pearson et al., 2017). To quantify the effect of the scale-aware parameter, we test the parameterizations in the Doube Gyre configuration in a range of seven simulations differing in grid resolution.

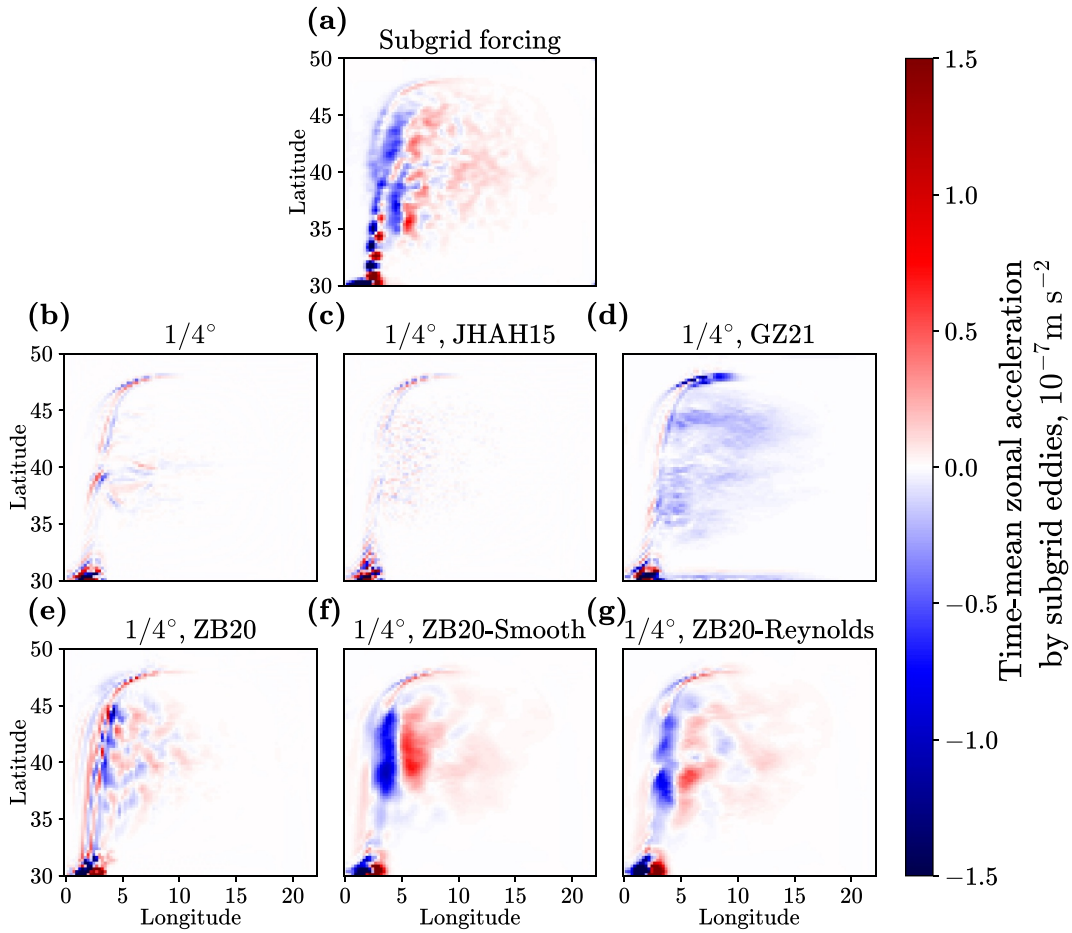

Figure 6. Analysis of the eddy-mean flow interaction following Hughes and Ash (2001): the 10-year mean zonal acceleration in the upper fluid layer produced by the subgrid eddies $\left(\overline{S}_x^t\right)$. (a) The zonal component of the subgrid forcing (Equation 3), diagnosed from the $1/64°$ model by filtering and coarse-graining. Panels (b–g): the zonal acceleration, in online simulations, produced by combining each parameterization with a biharmonic Smagorinsky model.

5.1. Sensitivity to the Scaling Coefficient

We perform experiments with the ZB20-based parameterizations for a range of resolutions from $1/2°$ to $1/8°$ and consider the sensitivity to the scaling coefficient γ. The effect of the subgrid parameterizations is quantified through the following metrics: the time-mean KE and APE, and the RMSE in the time-mean SSH. We compute the RMSE with respect to the 10-year averaged output of the coarse-grained high-resolution model $1/64°$. We split the last 15 years of the simulation of the coarse resolution models into three 5-year segments. We compute the mean over the segments with confidence intervals provided by the min/max values over the different segments for every metric.

The impact of the ZB20-based parameterizations on all metrics is consistent with that described in Section 4 for most resolutions (Figure 7). In particular, the ZB20 and ZB20-Smooth parameterizations spuriously reduce the KE (Figures 7a and 7b) but improve the APE (Figures 7d and 7e). Additionally, the spurious effect on the KE is smaller for the ZB20-Smooth parameterization than for the unfiltered ZB20. The ZB20-Reynolds model leads to an efficient backscatter parameterization: it increases the KE in the simulation (Figure 7c) with little impact on the APE (Figure 7f). All three subgrid parameterizations reduce the bias in SSH, with the simulation using the ZB20-Smooth parameterization having the lowest error and the simulation with the unfiltered ZB20 parameterization having the highest error (Figure 7, lower row).

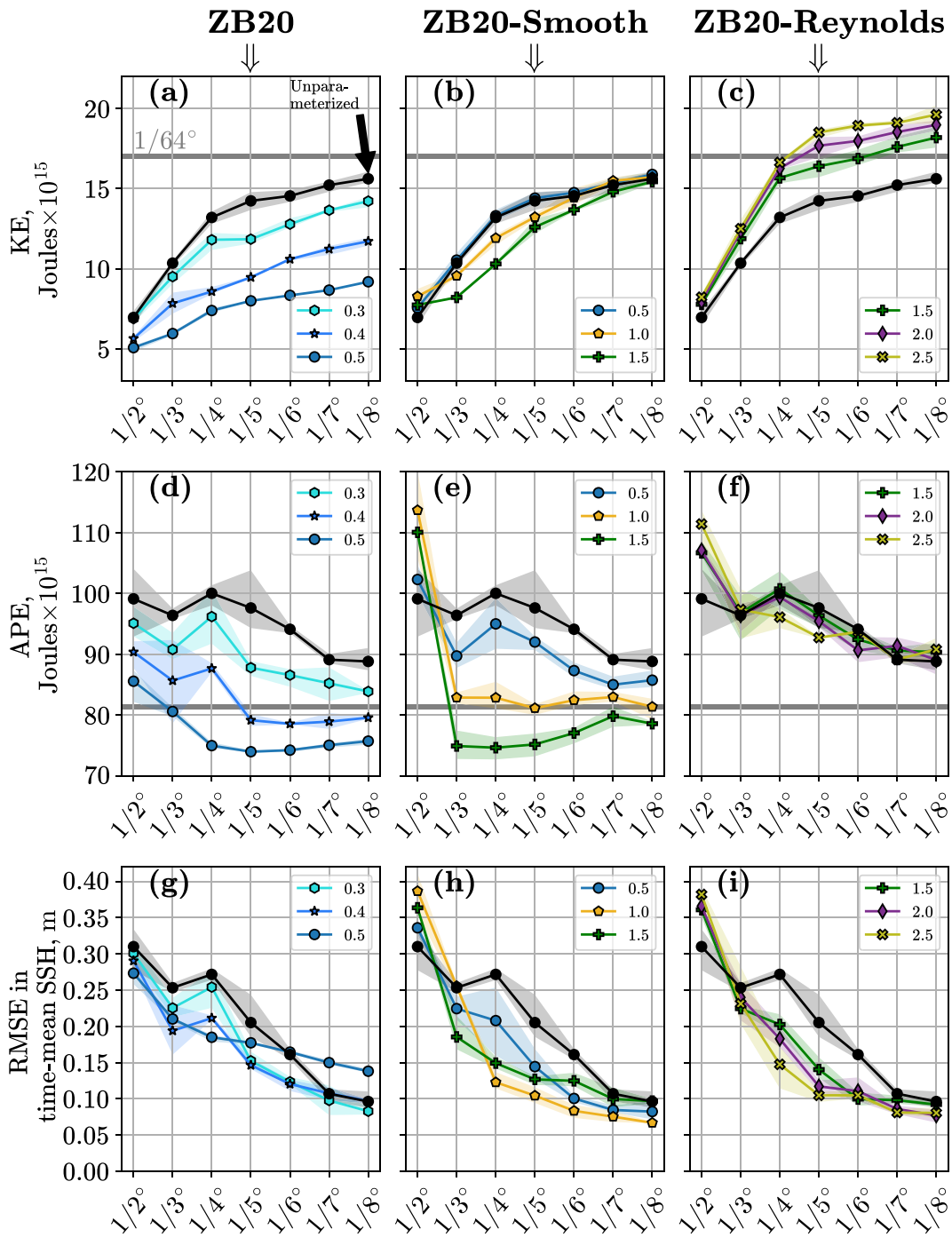

Figure 7. Sensitivity to the scaling coefficient γ (shown in legend). Upper row: the time-mean kinetic energy, middle row: the time-mean available potential energy, lower row: root mean squared error (RMSE) in the time-mean sea surface height. Left column shows unfiltered ZB20 parameterization and two rightmost columns show filtered ZB20 models. The black line shows the unparameterized model (biharmonic Smagorinsky). The gray horizontal line shows the $1/64°$ model. The shading shows confidence intervals (see text).

The impact on the KE and APE for all three subgrid parameterizations is proportional to the scaling coefficient (Figure 7, upper and middle rows). The default scaling coefficients used in the previous section (ZB20-Smooth: $\gamma = 1$, ZB20-Reynolds: $\gamma = 2$) correspond to a compromise in reproducing the presented metrics for a range of resolutions. These non-dimensional coefficients can be kept constant without retuning, thus demonstrating an

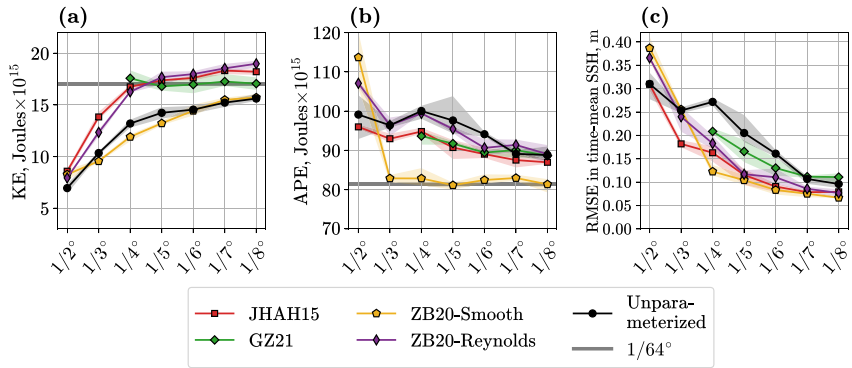

Figure 8. Similar to Figure 7, but comparison of the filtered ZB20 parameterizations to the baselines of Jansen et al. (2015) and Guillaumin and Zanna (2021). The scaling coefficient γ is set to default: 1.0 for ZB20-Smooth and 2.0 for ZB20-Reynolds.

advantage of scale-aware tuning of the free parameter of the filtered ZB20 parameterizations. An overshoot in some metrics at $1/2°$ resolution (Figure 7e) occurs because of the coarse resolution (eddies are not permitted). In Appendix C, we discuss the sensitivity to the number of filtering passes (N) and to the Smagorinsky coefficient (C_S).

5.2. Comparison to the Baseline Parameterizations

We compare the filtered subgrid parameterizations (ZB20-Smooth and ZB20-Reynolds) with the default parameter γ to the baseline subgrid parameterizations of JHAH15 and GZ21. The subgrid parameterizations ZB20-Reynolds, GZ21, and JHAH15 are equally efficient in energizing the resolved flow, that is, they parameterize the KE backscatter (Figure 8a) but have similar drawbacks. At the lowest resolutions ($1/2° - 1/3°$), the ZB20-Reynolds and JHAH15 parameterizations underestimate the KE. At the highest resolutions ($1/5° - 1/8°$), the ZB20-Reynolds and JHAH15 parameterizations slightly overestimate the KE without retuning; note that the GZ21 parameterization was tuned at every resolution to reproduce the KE exactly (Zhang et al., 2023). These three backscattering parameterizations have only a small impact on the APE (Figure 8b). The ZB20-Smooth parameterization demonstrates the best representation of the APE for a range of resolutions $1/3° - 1/8°$ (Figure 8b). The SSH error with ZB20-Smooth is most evidently improved compared to all other parameterizations only at resolution $1/4°$ discussed in Section 4. The ZB20-Smooth leads to the most accurate representation of SSH at resolutions $1/5 - 1/8°$, but its effect is comparable to the baseline simulation with JHAH15 parameterization. The ZB20-Smooth parameterization fails at resolutions coarser than $1/4°$ (Figure 8c).

6. Experiments in the NeverWorld2 (NW2) Configuration

In this section, we analyze the impact of the different subgrid parameterizations in a more complex adiabatic configuration of MOM6—the NeverWorld2 (NW2, Marques et al., 2022) setup. This configuration spans the latitudes from 70°S to 70°N, including the equator, with topography mimicking an idealized Mid-Atlantic ridge and a Drake Passage. The number of fluid layers is 15. This configuration was designed specifically to test mesoscale eddy parameterizations. Compared to the Double Gyre configuration, the NW2 setup has a stronger need for improving the energetics: the KE significantly depends on the resolution and increases by a factor of 4 when resolution is increased from $1/4°$ to $1/32°$ (Marques et al., 2022). Therefore, in this section, we target the $1/4°$ resolution simulation to test the ZB20-based parameterizations and compare them to backscatter baseline parameterizations already implemented in NW2 (Yankovsky et al. (2024), hereafter YBSZ24).

6.1. Numerical Issues and Tuning

Preliminary experiments with the proposed ZB20-based parameterizations demonstrated numerical instabilities accompanied by various runtime errors, including too-large ocean velocities, interface height droping below the bathymetry, and NaN values in the prognostic fields. These issues are partly tied to the structure of isopycnals in NW2; for example, the significant part of the model area (40%) is characterized by isopycnal layers with small thicknesses $\left(h_k \approx 10^{-2}\text{m}\right)$. To alleviate these numerical issues, we implemented two strategies:

- Switch the discretization of the ZB20-based parameterizations from the energy-conserving form (Equation A10) used in the Double Gyre to a non-conserving one (Equation A12). The non-conservative numerical scheme presumably introduces fewer aliasing errors because the multiplication operation follows the interpolation.
- Attenuate the parameterization in regions of geostrophically unbalanced flows. Similar to Klöwer et al. (2018), Juricke et al. (2019), we introduce the following attenuation function:

$$\left(1+\frac{\sqrt{D^2+\tilde{D}^2+\zeta^2}}{|f|}\right)^{-1}. \tag{11}$$

The ZB20 stress tensor is multiplied by the attenuation function before computing its divergence. The inclusion of ζ^2 into Equation 11 is a proposed modification, which was found to improve the numerical stability. The expression $D^2+\tilde{D}^2+\zeta^2$ is proportional to the isotropic stress of the ZB20 parameterization, and thus the attenuation bounds the predicted momentum flux. Note that the attenuation was inspired by a similar technique of shutting off the backscatter in high-strain regions proposed by YBSZ24. The difference between the two approaches is small: we attenuate the parameterization smoothly while they shut off the backscatter abruptly. We use the inverse Coriolis parameter as a threshold time scale in the attenuation function while they use the time step (which is a non-physical parameter but useful because it is embedded in the numerical stability criteria).

We further note that we have modified the Smagorinsky coefficient in the $1/4°$ resolution model, compared to the original NW2 setup. We are using the commonly used (and our default value) of $C_S = 0.06$ instead of the relatively large value of $C_S = 0.2$ in Marques et al. (2022). This change does not affect the reference unparametrized coarse resolution simulation (not shown).

We decide to evaluate the parameterizations (ZB20-based and our baselines) by tuning them to roughly reproduce the total KE of the filtered NW2 high-resolution simulation while minimizing the difference across tuning coefficients. In our preliminary experiments, we noted that the unfiltered ZB20 parameterization could not be tuned to reproduce the KE of the filtered high-resolution model. The increase in the KE compared to the unparameterized simulation was never larger than $\approx 25\%$, with only small improvements in the mean state, thus we do not show results for the unfiltered ZB20 parameterization in this section. The ZB20-Smooth and ZB20-Reynolds parameterizations were tuned by setting the scaling coefficient to $\gamma = 2.5$.

To summarize, compared to the Double Gyre case, we changed the discretization scheme of the ZB20 parameterization, introduced the attenuation function, and increased the scaling coefficient.

6.2. Results

The high-resolution simulation $1/32°$ was spun up in multiple stages (Marques et al., 2022). We use snapshots from the last 100 days and time-mean fields over the last 1,000 days for analysis. We ran the coarse parameterized and unparameterized models at resolution $1/4°$ for 30,000 days, starting from rest. To compare the coarse and high-resolution simulations, we filter the output of the $1/32°$ high-resolution simulations with filter widths $1°$ and $1/2°$. That is, we assume that the effective resolution of the coarse resolution models is lower than the nominal resolution given by the coarse grid (Skamarock, 2004; Soufflet et al., 2016).

The baseline GZ21 parameterization, used in the Double Gyre experiment, struggles to generalize to the NW2 configuration, and thus, we omit the results. Rather than use the JHAH15 parameterization, as for the Double Gyre configuration, we used the most up-to-date version of the JHAH15 parameterization from YBSZ24, which was specifically tuned to represent mesoscale eddies in the NW2 configuration.

The YBSZ24 parameterization attributes a vertical structure to the anti-viscosity coefficient, tuned to match the energetics of the NW2 simulation in their original study. This led to a choice of vertical structure that follows the equivalent barotropic mode raised to the 2nd power. Similarly to the ZB20-based parameterizations, the coarse resolution simulation using the YBSZ24 parameterization was tuned to match the KE of the filtered high-resolution run, using an anti-viscosity scaling coefficient of -0.3. Note, however, that the YBSZ24

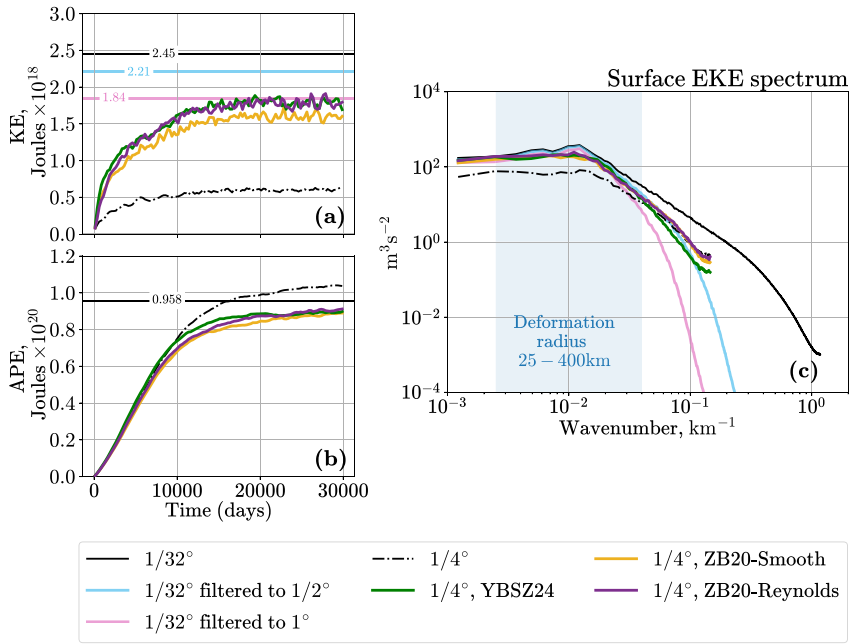

Figure 9. NeverWorld2 configuration. Time series of the (a) kinetic energy and (b) available potential energy. The time series in 1/4° runs are smoothed in time with a window size of 250 days, while in 1/32° run, we provide average values over the last 100 days (c) The EKE zonal spectrum at the surface averaged over 100 days and over latitudes.

parameterization in NW2 requires a much larger Smagorinsky coefficient ($C_S = 0.2$) to remain stable compared to the unparameterized and ZB20-based parameterized models ($C_S = 0.06$).

We compare the coarse models to the high-resolution simulation (Figure 9, black solid line) and its filtered versions (Figure 9, light blue and pink lines). We do not show the APE of the filtered high-resolution model because it is significantly affected by the implementation of the filtering algorithm near the ocean bottom.

The unparameterized model (1/4°) has a KE that is approximately 4 times lower than the KE of the 1/32° model and 3 times lower than the KE of the filtered 1/32° model (Figure 9a). The APE of the unparameterized model is too high compared to the high-resolution simulation (Figure 9b) since the coarse resolution model underestimates the barotropization of the flow (Kjellsson & Zanna, 2017). The ZB20-Smooth and ZB20-Reynolds parameterizations increase the KE approximately 3 times and reduce the APE below the APE of the high-resolution simulation (as we expect for the filtered solution). The ZB20-Smooth parameterization is slightly more efficient in reducing the APE, and the ZB20-Reynolds is slightly more efficient in increasing the KE (Figures 9a and 9b) for the same scaling coefficient γ. The YBSZ24 parameterization is equally efficient in reducing APE and increasing KE. The equilibration of the parametrized run with the YBSZ24 parameterization is faster than the parametrized simulations with the filtered ZB20 closures.

Figure 9c shows the zonal EKE spectrum at the surface, $1/2\left(|\mathcal{F}(u')|^2 + |\mathcal{F}(v')|^2\right)$, averaged in time over 100 days and over all latitudes. The Fourier transform $\mathcal{F}$ is computed in the zonal direction with Hann window and linear detrending. The eddy velocities u' and v' are defined as a deviation from the 1000-day mean velocities at the surface. The filtered ZB20 parameterizations and the YBSZ24 parameterization increase the EKE spectrum density at large scales (Figure 9c). Considering the shape of the EKE spectrum at small scales, the filtered ZB20 models are closest to the 1/32° model filtered with the filter width 1/2° (Figure 9c). However, the total KE better matches the filter scale of 1° (Figure 9a). A further increase in KE is possible but at the expense of the APE presumably becoming unphysical. These findings reveal the difficulties in determining the effective resolution of the parameterized simulations. The simulation with the YBSZ24 parameterization possesses less energy at small scales at the surface, likely due to the high value of the Smagorinsky coefficient, which is necessary for numerical stability (in addition to the local shut-off described in their original study).

The snapshots of the depth-integrated KE are shown in Figure 10. The filtered ZB20 models considerably energize the eddies in all parts of the domain compared to the unparameterized model and in accordance with the

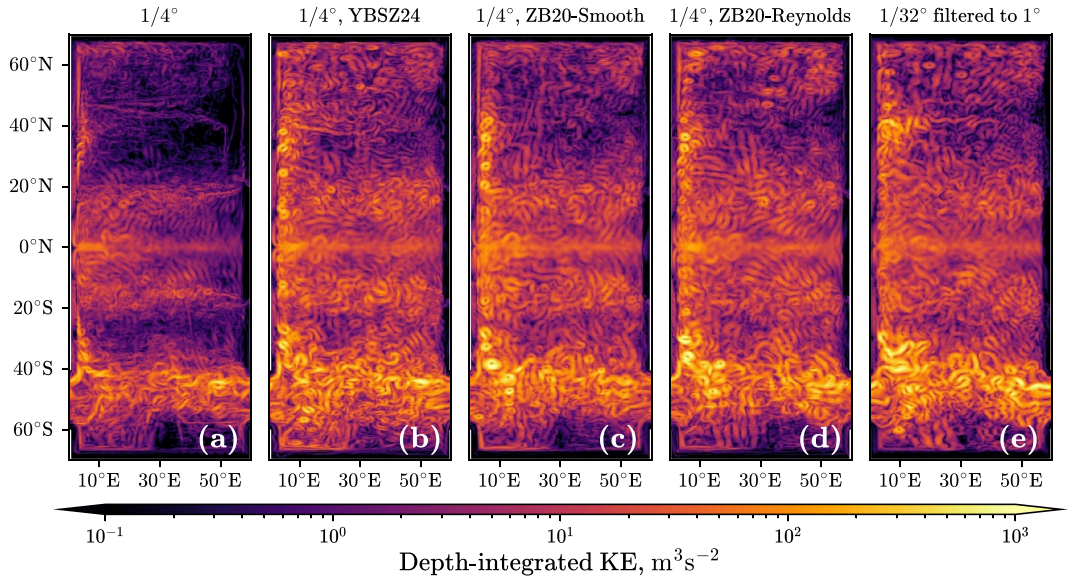

Figure 10. Snapshot of the depth-integrated kinetic energy. The coarse models at resolution 1/4°: (a) unparameterized model (biharmonic Smagorinsky), (b) Yankovsky et al. (2024) (YBSZ24), (c) and (d) filtered ZB20 parameterizations. In panel (e), we show the filtered high-resolution model 1/32° with filter width 1°.

high-resolution simulation. The simulation with the YBSZ24 parameterization is also successful at increasing the KE throughout the domain, with only small differences compared to the ZB20-based parameterized simulations. The major difference between the coarse parameterized models and the high-resolution simulation is the presence of coherent eddies near the western boundary (Figures 10b–10d), including in the YBSZ24 parametrized simulation. Presumably, the attenuation function (Equation 11) was needed to bound the growth of these and similar eddies in the filtered ZB20 models. However, tracking how a particular eddy growth contributes to the development of numerical instabilities is difficult.

The largest portion of the APE of the mean state is described by the vertical structure of the isopycnal interfaces in the Drake Passage (60°S-40°S), shown in Figure 11 at longitude 0°E. The mesoscale eddies extract APE from the mean state and act to flatten the isopycnals. The coarse unparameterized model poorly resolves mesoscale eddies, and consequently, its isopycnals are too steep (Figure 11a). Both filtered ZB20 parameterizations and the YBSZ24 parameterization result in reducing the APE and better reproducing the isopycnal structure in Drake Passage (Figures 11b–11d). In the same figure, we provide the RMSE values with respect to the high-resolution simulation. The biggest improvement is achieved with the ZB20-Reynolds parameterization, which has the lowest RMSE. The improvements are visible in the upper ocean in particular, compared to the ZB20-Smooth and the YBSZ24 parameterizations. However, the bottom two layers are better represented in the YBSZ24 parameterized run than in the ZB20-based simulations. The performance of the filtered ZB20 parameterizations is confirmed according to additional metrics quantifying the mean state. These metrics include the RMSE in the vertical structure of isopycnals in three transects; see Table 2. All parametrized simulations reduce biases, with the ZB20-Reynolds parameterized simulation showing the largest improvement using these metrics. When considering the RMSE in the time-mean SSH (Table 2), all parametrized models decrease the mean bias, with the YBSZ24 parameterization performing best. However, all improvements in the mean state are rather small according to this metric.

7. Conclusions and Discussion

In this work, we implemented the data-driven mesoscale eddy parameterization introduced by Zanna and Bolton (2020) (ZB20) into the GFDL ocean model, MOM6, and tested it in two idealized configurations: the Double Gyre and NeverWorld2 (NW2). The ZB20 parameterization, machine-learned from data, predicts the subgrid momentum fluxes and, in particular, captures the KE backscatter, that is, the inverse energy cascade from the subgrid to resolved scales. Our main findings are as follows.

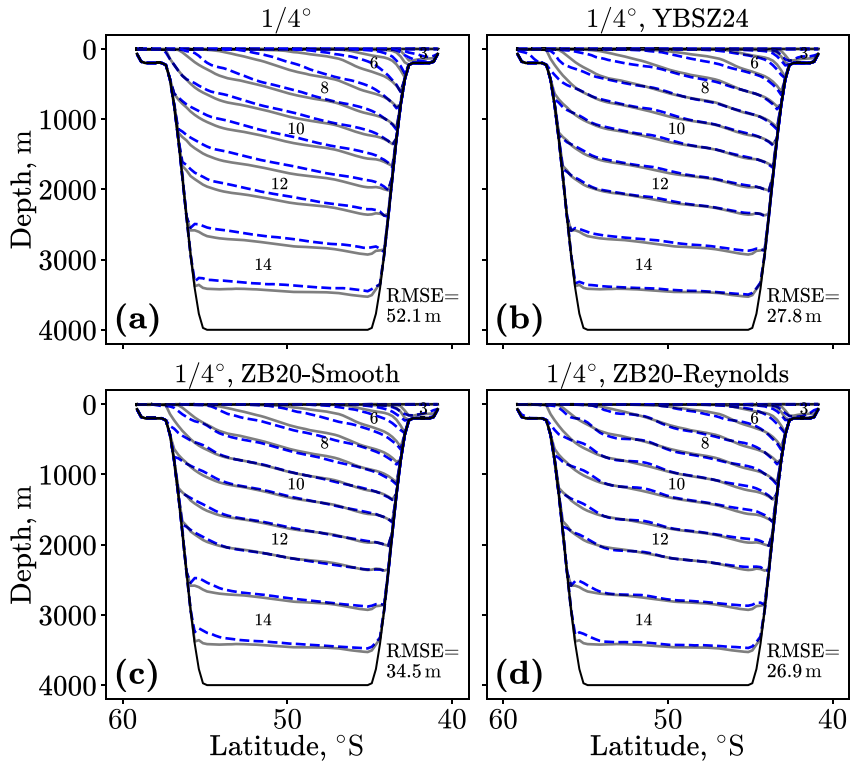

Figure 11. The time-mean interfaces in the meridional transect of Drake Passage (Longitude 0°E) averaged over 1,000 days. The blue dashed lines show the experiments at resolution 1/4°: (a) unparameterized model, (b) Yankovsky et al. (2024) (YBSZ24), (c) and (d) filtered ZB20 parameterizations. Gray lines show the interfaces of the high-resolution model 1/32°. The layer numbers (equal to 14, 12,…) are provided.

- The original parameterization was found to generate numerical noise near the grid scale. We propose two filtering schemes to reduce the generation of numerical noise and isolate the large-scale backscatter effect of the parameterization. The first scheme (ZB20-Smooth) applies a low-pass filter to the ZB20 stress tensor, and the second scheme (ZB20-Reynolds) additionally applies a high-pass filter to the velocity gradients (Perezhogin & Glazunov, 2023).
- The free parameter of the ZB20-based parameterizations is scale-aware. Here, we show that the parameterizations can be used for a range of eddy-permitting resolutions without retuning.
- In the Double Gyre configuration, the ZB20-Reynolds parameterization effectively energizes the resolved flow and performs similarly to the Jansen et al. (2015) (JHAH15) and Guillaumin and Zanna (2021) (GZ21) backscatter parameterizations at eddy-permitting resolutions ($1/4° - 1/8°$). In this configuration, the ZB20-Smooth is less efficient in parameterizing the backscatter but outperforms other parameterizations in the SSH and APE predictions at resolutions $1/4° - 1/8°$. All subgrid parameterizations perform poorly at the coarsest resolutions ($1/2° - 1/3°$), as expected.
- In the NW2 configuration, the filtered ZB20 parameterizations improve the model energetics, namely KE and APE reservoirs, and the energy power spectrum. The ZB20-based parameterizations improve aspects of the climatological mean state (here the vertical structure of the isopycnals and the SSH). The ZB20-based parameterizations perform better or as well as the anti-viscosity parameterization of Yankovsky et al. (2024), which imposes a vertical structure to the anti-viscosity coefficient. The ZB20-based parameterization does not need to impose a vertical structure to its coefficient, the 3D flow dependence is encapsulated in the stress tensor itself.

Table 2
The Root Mean Squared Errors (RMSE) in 1000-Day Averaged Position of Interfaces Over Three Meridional Transects at (0°E, 60°S-40°S) (Drake Passage, Also Shown in Figure 11), (30°E, 70°S-70°N) (Mid-Atlantic Ridge) and (45°E, 70°S-70°N) (Similar to Yankovsky et al. (2024))

	0°E (m)	30°E (m)	45°E (m)	SSH (m)
1/4°	52.1	40.3	34.6	0.101
1/4°, YBSZ24	27.8	20.7	21.3	**0.071**
1/4°, ZB20-Smooth	34.5	20.8	24.9	0.090
1/4°, ZB20-Reynolds	**26.9**	**18.4**	**18.5**	0.080

Note. The rightmost column shows RMSE in 1000-day averaged sea surface height (SSH). The error is computed with respect to 1/32° model. The best values of metrics are shown in bold.

- The direct effect of mesoscale eddy parameterizations in improving the mean state depends on the ocean configuration. In the NW2 configuration, the subgrid parameterization increases the KE of the resolved flow by a factor of 3. This significantly enhances the effect by resolved eddies which act to flatten isopycnals and reduce the APE. In the Double Gyre configuration, on the other hand, the mean state can be improved without energizing the resolved eddies but instead by imposing strong mean subgrid stress (Kjellsson & Zanna, 2017).

Our methodology enables researchers and ocean modelers to implement and test subgrid machine-learning parameterizations in state-of-the-art ocean models. As with all parameterizations in current climate models, the ZB20 parameterization required tuning. The succinct, interpretable form of the subgrid ZB20 model allowed us to study its physical and numerical properties in detail. We leveraged the filtering schemes to extract the parameterization effect on the large-scale flow and to avoid the grid-scale numerical issues. By testing various discretizations of the subgrid model, we were able to find a numerically stable scheme for online simulations. Our filtering schemes can be potentially applied to improve the performance of our baseline parameterizations (JHAH15, BSZ24 and GZ21); though such testing is beyond the scope of this study. We note that filtering schemes have been consistently used in the literature, showing a potential to improve the performance of other mesoscale eddy parameterizations (Bagaeva et al., 2024; Grooms, 2023; Grooms et al., 2015; Juricke et al., 2020; Mak et al., 2023).

While applying the filtering schemes allows us to improve the physical and numerical properties of the ZB20 parameterization, several challenges remain. (a) It is difficult to find a single filtering scheme that simultaneously improves the mean state and kinetic/potential energies. (b) The proposed filtering approach is applied and tuned a posteriori; how to learn it directly from data is an open question. (c) The filtered ZB20 parameterizations do not impact the grid scale flow, therefore a supplementary subgrid dissipative model is needed (in this work, we utilize a biharmonic Smagorinsky model). (d) Without filters, the ZB20 parameterization is computationally cheap (2.5% of total runtime). However, the filtered parameterizations can take $4-6\%$ of total runtime.

Possible future improvements to the proposed parameterization include: (a) Coupling the ZB20-based parameterizations with the subgrid KE equation, as in energetically consistent parameterizations (Jansen et al., 2015; Mak et al., 2018). (b) Including spatial non-locality (P. Wang et al., 2022) or temporal memory (Zanna et al., 2017) during the training process to potentially reduce the need for a posteriori tuning. (c) Using a neural network to improve the prediction of the subgrid stress from the same input features. (d) Informing the subgrid model with local physical parameters to improve generalization to unseen flow regimes (Bachman et al., 2017; Hallberg, 2013; Jansen et al., 2019).

An additional interesting future direction is to apply the developed subgrid parameterizations in eddy-permitting global ocean models to attempt addressing long-standing biases such as the North Atlantic cold bias (Chang et al., 2023; Flato et al., 2014; C. Wang et al., 2014).

Appendix A: Curvilinear Coordinates, Varying Layer Thickness, Numerical Schemes and Boundary Conditions

We modify the original parameterization (Equations 4 and 5) to account for curvilinear coordinates and varying layer thickness. Also, we propose a numerical discretization scheme and boundary conditions.

A1. Computation of the Stress Tensor

The components of the stress tensor $\mathbf{T}$ depend uniquely on the gradients of the velocity field (D, $\tilde{D}$, ζ). The computation of these gradients depends on the coordinate system. In generalized curvilinear orthogonal coordinates the expressions for D and $\tilde{D}$ must be changed to (Appendix A.b in Griffies and Hallberg (2000) and Section 17.10.2 in Griffies (2018)).

$$D = \Delta_y \partial_x (v/\Delta_y) + \Delta_x \partial_y (u/\Delta_x), \tag{A1}$$

$$\tilde{D} = \Delta_y \partial_x (u/\Delta_y) - \Delta_x \partial_y (v/\Delta_x), \tag{A2}$$

where Δ_x and Δ_y are local grid spacings that are proportional to the Lame coefficients.

Following a similar approach, we compute the relative vorticity as follows:

$$\zeta = \Delta_y \partial_x (v/\Delta_y) - \Delta_x \partial_y (u/\Delta_x). \tag{A3}$$

Note that the relative vorticity can be alternatively computed using the contour integral divided by area, that is $\zeta = \Delta_y^{-1} \partial_x (\Delta_y v) - \Delta_x^{-1} \partial_y (\Delta_x u)$, see Section 2.3.2 in Madec and the NEMO team (2008). We found that both approaches give close results and use Equation A3 for simplicity.

A2. Divergence of Momentum Flux

Following previous work on viscous operators in ocean models (Equations A3 and A4 in Griffies and Hallberg (2000) and Section 17.10.3 in Griffies (2018)), we modify the divergence of the stress tensor (Equation 4):

$$\mathbf{S} = \frac{1}{h} \nabla \cdot (h\mathbf{T}) = \frac{1}{h} \begin{pmatrix} \frac{1}{\Delta_y^2} \partial_x \left(\Delta_y^2 h T_{xx} \right) + \frac{1}{\Delta_x^2} \partial_y \left(\Delta_x^2 h T_{xy} \right) \\ \frac{1}{\Delta_y^2} \partial_x \left(\Delta_y^2 h T_{xy} \right) + \frac{1}{\Delta_x^2} \partial_y \left(\Delta_x^2 h T_{yy} \right) \end{pmatrix}, \tag{A4}$$

where we account for curvilinear coordinates with terms including Δ_x and Δ_y and for varying layer thickness with terms including h. The components of the subgrid forcing parameterization $\mathbf{S}$ have the dimensions of acceleration, that is, length^1 time^{-2}. The components of the stress tensor (T_{xx}, T_{xy},…) have the dimensions of length^2 time^{-2}.

Accounting for varying thickness allows to build a parameterization that conserves the integral of momentum up to the boundary fluxes.

$$\partial_t \int h\mathbf{u} \, dx \, dy = \cdots + \int h \partial_t \mathbf{u} \, dx \, dy = \cdots + \int h\mathbf{S} \, dx \, dy = \cdots + \tag{A5}$$

$$\int \nabla \cdot (h\mathbf{T}) \, dx dy. \quad \rightarrow 0 \tag{A6}$$

A3. Numerical Discretization

The rate of change of the KE due to the parameterization (Equation A4) after integration by parts is given by.

$$\partial_t \int \frac{1}{2} h |\mathbf{u}|^2 \, dx \, dy = \cdots + \int (h\mathbf{u}) \mathbf{S} \, dx \, dy = \cdots + \int \mathbf{u} \cdot \nabla \cdot (h\mathbf{T}) \, dx \, dy = \cdots \tag{A7}$$

$$- \int h\mathbf{T} : (\nabla \mathbf{u}) \, dx \, dy, \tag{A8}$$

where $\nabla \mathbf{u}$ is the velocity gradient tensor and (:) is the tensor contraction over two indices (Eyink, 1995). In Cartesian coordinates the components of velocity gradient tensor have a simple form $(\nabla \mathbf{u})_{ij} = \partial_j u_i$. The energy contribution from the deviatoric component of $\mathbf{T}$ (Equation 5) is zero because it is orthogonal to the velocity gradient tensor:

$$2\mathbf{T}_\mathrm{d} : (\nabla \mathbf{u}) = \begin{bmatrix} -\zeta D & \zeta \tilde{D} \\ \zeta \tilde{D} & \zeta D \end{bmatrix} : \underbrace{\begin{bmatrix} \tilde{D} & D \\ D & -\tilde{D} \end{bmatrix}}_{\substack{\text{deviatoric} \\ \text{symmetric} \\ \text{part of } 2\nabla \mathbf{u}}} = -\zeta D \tilde{D} + \zeta \tilde{D} D + \zeta \tilde{D} D - \zeta D \tilde{D} = 0. \tag{A9}$$

However, the Equation A9 does not hold numerically when the Arakawa C grid staggering is used, because D, ζ and T_{xy} are defined in the corner of the grid cell, but T_{xx}, T_{yy} and $\tilde{D}$ are defined in the center of the grid cell. We propose the following energy-conserving discretization of the deviatoric stress $\mathbf{T}_\mathrm{d}$:

$$\underbrace{\begin{bmatrix} -\overline{\zeta D}^{x,y} & \zeta\overline{\tilde{D}}^{x,y} \\ \zeta\overline{\tilde{D}}^{x,y} & \overline{\zeta D}^{x,y} \end{bmatrix}}_{\text{numerical scheme}} : \begin{bmatrix} \tilde{D} & D \\ D & -\tilde{D} \end{bmatrix} = \underbrace{-2\overline{\zeta D}^{x,y}\tilde{D}}_{\text{cell center}} + \underbrace{2\zeta D\overline{\tilde{D}}^{x,y}}_{\text{cell corner}}, \tag{A10}$$

where $\overline{(\cdot)}^{x,y}$ is a linear interpolation from corner to center or vice versa. Terms in the RHS of Equation A10 are defined in different points but cancel after summation over the domain because the interpolation operator is self-adjoint up to the boundary conditions (Section 4.1.2 in Madec and the NEMO team (2008)). Energy-conserving discretization on a non-uniform grid is given by computing the interpolation $\overline{\zeta D}^{x,y}$ in a conservative way, that is, by weighting with the local grid cell area. Note that exact energy conservation property is lost when coefficient κ_{BC} or layer thickness h vary spatially or when the spatial filters are applied.

For the isotropic part of $\mathbf{T}$ (Equation 5), we use the simplest numerical scheme because no properties are known:

$$\mathbf{T}_\mathrm{I} = \left(\left(\overline{\zeta}^{x,y}\right)^2 + \left(\overline{D}^{x,y}\right)^2 + \tilde{D}^2\right)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{A11}$$

The energy-conserving discretization (Equation A10) is used in the Double Gyre configuration. In NW2 runs, we found that the following approximation of the deviatoric component of $\mathbf{T}$ demonstrates better numerical stability properties:

$$\mathbf{T}_\mathrm{d} = \begin{bmatrix} -\overline{\zeta}^{x,y}\overline{D}^{x,y} & \zeta\overline{\tilde{D}}^{x,y} \\ \zeta\overline{\tilde{D}}^{x,y} & \overline{\zeta}^{x,y}\overline{D}^{x,y} \end{bmatrix}. \tag{A12}$$

A4. Boundary Conditions

We apply an analog of the free-slip boundary condition: the momentum flux through the boundary is zero. On the Arakawa C grid, this is achieved by setting $T_{xy} = 0$ on the boundary. Zero boundary conditions are also used for the filtering operations and interpolations on the staggered grid.

Appendix B: Computation of Kinetic and Potential Energy

The KE integrated over the fluid layers and horizontal coordinates in Joules is defined as:

$$\mathrm{KE} = \frac{1}{2}\sum_k \int \rho_0 |\mathbf{u}_k|^2 h_k \, dx\, dy. \tag{B1}$$

The KE of the mean flow $(\overline{\mathbf{u}}^t_k, \overline{h}^t_k)$ is referred to as a MKE. The EKE is defined as $\mathrm{EKE} = \overline{\mathrm{KE}}^t - \mathrm{MKE}$. The PE summed over interfaces in Joules is defined as:

$$\mathrm{PE} = \frac{1}{2}\sum_k \int \rho_0 g'_{k+1/2} \eta^2_{k+1/2} \, dx\, dy. \tag{B2}$$

The APE is the PE minus the potential energy of the resting state given by $\eta^{\mathrm{ref}}_{k+1/2} = \max\left(z^0_{k+1/2}, -H(x,y)\right)$, where $z^0_{k+1/2}$ is the constant nominal position of the interfaces and $H \geq 0$ is the depth (Marques et al., 2022). MPE is the APE of the mean flow $\overline{\eta}^t_{k+1/2}$, and EPE is given by $\mathrm{EPE} = \overline{\mathrm{APE}}^t - \mathrm{MPE}$.

Appendix C: Additional Sensitivity Studies

In Figure C1 we show the sensitivity of the online metrics to the number of filter passes (N) for the ZB20-Smooth and ZB20-Reynolds parameterizations. For the ZB20-Smooth, the effect of reducing APE is slightly stronger for lower N at the same scaling coefficient γ (Figure C1b). However, the undesirable impact on the KE is also stronger (Figure C1a). Considering the error in SSH, the optimal number of filters is $N = 2$ for coarse resolutions ($1/2° - 1/3°$) and the optimal number is $N = 4$ for higher resolutions (Figure C1c). For the ZB20-Reynolds parameterization, the impact on the KE is higher for higher N (Figure C1d). An impact on the SSH metric is also more significant for larger N (Figure C1f). Overall, a slightly more beneficial impact on energetic metrics (KE, APE) is observed in the case of a larger number of filters $N = 4$. However, testing in different configurations is required to determine the optimal parameter N.

In Figure C2, we show the sensitivity to the Smagorinsky coefficient. We consider the bias in SSH prediction because it is sensitive to including any of the parameterizations (ZB20, ZB20-Smooth, ZB20-Reynolds). The unparameterized models with different Smagorinsky coefficients are shown in black markers. White markers show the optimal scaling coefficient γ for a given Smagorinsky constant. Note that we include inviscid simulations ($C_S = 0.00$). The inviscid models can be run stably for all three parameterizations for a range of scaling coefficients γ. However, the optimal SSH metric is achieved when the ZB20-based parameterizations are turned off, that is, $\gamma = 0$ when $C_S = 0$. This demonstrates that the ZB20-based parameterizations describe only part of the subgrid forcing, and cannot be used without an eddy viscosity model. Another important observation—the optimal scaling coefficient γ should be increased when the eddy viscosity coefficient C_S is increased. Finally, the SSH bias can be efficiently reduced by the ZB20-Smooth and ZB20-Reynolds parameterizations for various values of the Smagorinsky coefficient ($C_S = 0.03, 0.06, 0.09$) .

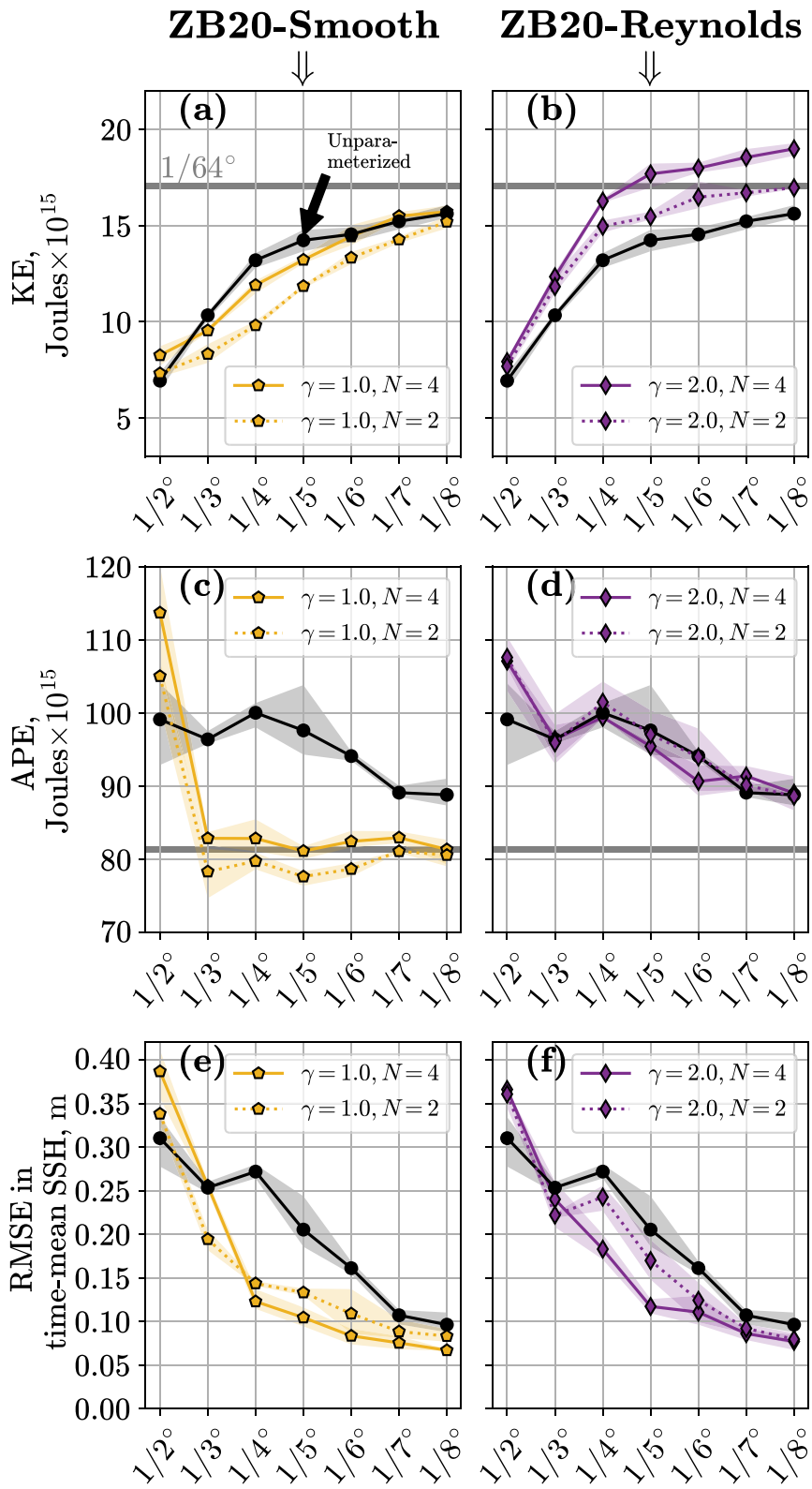

Figure C1. Similar to Figure 7, but showing the sensitivity to the number of the filter passes N for the filtered ZB20 parameterizations. The solid lines show the default value of the filter passes ($N = 4$), and the dotted lines show a smaller value ($N = 2$).

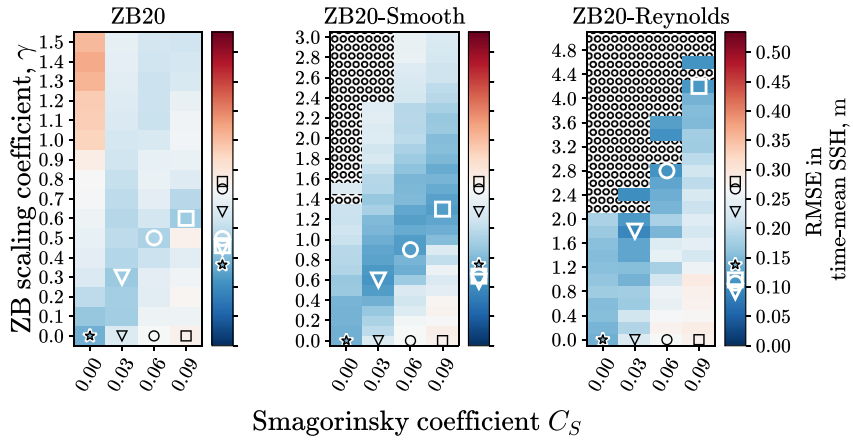

Figure C2. The sensitivity to the Smagorinsky coefficient C_S and scaling coefficient γ in the ZB20-based parameterizations. The considered metric is the root mean squared error (RMSE) in the representation of the time-mean sea surface height. The ocean grid resolution is 1/4°. Every small box shows a single simulation. The hatch area shows unstable simulations.

Data Availability Statement

The original NW2 configuration is available via Bhamidipati et al. (2022) and the data for the NW2 reference simulations, see Marques (2022). The version of the MOM6 source code with the implemented ZB20 parameterization, the configuration files for Double Gyre and NW2, and functions needed for generating the figures in this manuscript are available via Perezhogin (2024b). The simulation data is available via Perezhogin (2024a).

Acknowledgments
This project is supported by Schmidt Sciences, LLC. C.F.G. was partially supported by NSF DMS Grant 2009752. This research was also supported in part through the NYU IT High Performance Computing resources, services, and staff expertise and by the National Science Foundation under Grant NSF PHY-1748958. The authors would like to thank the members of M^{2}LInES, especially Dhruv Balwada and Nora Loose, for their helpful comments and discussions; Elizabeth Yankovsky for help with setting up the backscatter parameterization and NW2 simulations; Robert Hallberg and Marshall Ward for help with the MOM6 code. We also thank the editor, Tapio Schneider, as well as two anonymous reviewers and Julian Mak for their constructive comments that helped to improve the quality and presentation of this paper.

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3167–3211. https://doi.org/10.1029/2019MS001726

Adcroft, A., & Hallberg, R. (2006). On methods for solving the oceanic equations of motion in generalized vertical coordinates. *Ocean Modelling*, *11*(1–2), 224–233. https://doi.org/10.1016/j.ocemod.2004.12.007

Aluie, H. (2019). Convolutions on the sphere: Commutation with differential operators. *GEM-International Journal on Geomathematics*, *10*(1), 9. https://doi.org/10.1007/s13137-019-0123-9

Andrews, D., & Mcintyre, M. E. (1976). Planetary waves in horizontal and vertical shear: The generalized Eliassen-Palm relation and the mean zonal acceleration. *Journal of the Atmospheric Sciences*, *33*(11), 2031–2048. https://doi.org/10.1175/1520-0469(1976)033⟨2031:PWIHAV⟩2.0.CO;2

Anstey, J. A., & Zanna, L. (2017). A deformation-based parametrization of ocean mesoscale eddy Reynolds stresses. *Ocean Modelling*, *112*, 99–111. https://doi.org/10.1016/j.ocemod.2017.02.004

Bachman, S. D. (2019). The GM + E closure: A framework for coupling backscatter with the Gent and McWilliams parameterization. *Ocean Modelling*, *136*, 85–106. https://doi.org/10.1016/j.ocemod.2019.02.006

Bachman, S. D., Fox-Kemper, B., & Pearson, B. (2017). A scale-aware subgrid model for quasi-geostrophic turbulence. *Journal of Geophysical Research: Oceans*, *122*(2), 1529–1554. https://doi.org/10.1002/2016JC012265

Bagaeva, E., Danilov, S., Oliver, M., & Juricke, S. (2024). Advancing eddy parameterizations: Dynamic energy backscatter and the role of subgrid energy advection and stochastic forcing. *Journal of Advances in Modeling Earth Systems*, *16*(4), e2023MS003972. https://doi.org/10.1029/2023MS003972

Balarac, G., Le Sommer, J., Meunier, X., & Vollant, A. (2013). A dynamic regularized gradient model of the subgrid-scale scalar flux for large eddy simulations. *Physics of Fluids*, *25*(7), 075107. https://doi.org/10.1063/1.4813812

Berloff, P. (2018). Dynamically consistent parameterization of mesoscale eddies. Part III: Deterministic approach. *Ocean Modelling*, *127*, 1–15. https://doi.org/10.1016/j.ocemod.2018.04.009

Bhamidipati, N., Adcroft, A., Marques, G., & Abernathey, R. (2022). ocean-eddy-cpt/NeverWorld2: NeverWorld2-description-paper [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.6993951

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*(1), 376–399. https://doi.org/10.1029/2018MS001472

Bouchet, F. (2003). Parameterization of two-dimensional turbulence using an anisotropic maximum entropy production principle. *arXiv preprint cond-mat/0305205*. https://doi.org/10.48550/arXiv.cond-mat/0305205

Chang, C.-Y., Adcroft, A., Zanna, L., Hallberg, R., & Griffies, S. M. (2023). Remote versus local impacts of energy backscatter on the North Atlantic SST biases in a global ocean model. *Geophysical Research Letters*, *50*(21), e2023GL105757. https://doi.org/10.1029/2023GL105757

Chelton, D. B., DeSzoeke, R. A., Schlax, M. G., El Naggar, K., & Siwertz, N. (1998). Geographical variability of the first baroclinic Rossby radius of deformation. *Journal of Physical Oceanography*, *28*(3), 433–460. https://doi.org/10.1175/1520-0485(1998)028⟨0433:GVOTFB⟩2.0.CO;2

Chen, S., Ecke, R. E., Eyink, G. L., Wang, X., & Xiao, Z. (2003). Physical mechanism of the two-dimensional enstrophy cascade. *Physical Review Letters*, *91*(21), 214501. https://doi.org/10.1103/PhysRevLett.91.214501

Chow, F. K., & Moin, P. (2003). A further study of numerical errors in large-eddy simulations. *Journal of Computational Physics*, *184*(2), 366–380. https://doi.org/10.1016/S0021-9991(02)00020-7

Christensen, H., & Zanna, L. (2022). *Parametrization in weather and climate models*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228620.013.826

Eyink, G. L. (1995). Local energy flux and the refined similarity hypothesis. *Journal of Statistical Physics*, *78*(1–2), 335–351. https://doi.org/10.1007/BF02183352

Ferrari, R., & Wunsch, C. (2009). Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, *41*(1), 253–282. https://doi.org/10.1146/annurev.fluid.40.111406.102139

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2014). Evaluation of climate models. In *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change* (pp. 741–866). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.020

Fox-Kemper, B., & Menemenlis, D. (2008). Can large eddy simulation techniques improve mesoscale rich ocean models? *Washington DC American Geophysical Union Geophysical Monograph Series*, *177*, 319–337. https://doi.org/10.1029/177GM19

Frederiksen, J. S. (1999). Subgrid-scale parameterizations of eddy-topographic force, eddy viscosity, and stochastic backscatter for flow over topography. *Journal of the Atmospheric Sciences*, *56*(11), 1481–1494. https://doi.org/10.1175/1520-0469(1999)056⟨1481:SSPOET⟩2.0.CO;2

Frederiksen, J. S., & Davies, A. G. (1997). Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *Journal of the Atmospheric Sciences*, *54*(20), 2475–2492. https://doi.org/10.1175/1520-0469(1997)054⟨2475:EVASBP⟩2.0.CO;2

Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., & Lguensat, R. (2022). A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003124. https://doi.org/10.1029/2022MS003124

Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, *20*(1), 150–155. https://doi.org/10.1175/1520-0485(1990)020⟨0150:IMIOCM⟩2.0.CO;2

Germano, M. (1986). A proposal for a redefinition of the turbulent stresses in the filtered Navier–Stokes equations. *The Physics of Fluids*, *29*(7), 2323–2324. https://doi.org/10.1063/1.865568

Ghosal, S. (1996). An analysis of numerical errors in large-eddy simulations of turbulence. *Journal of Computational Physics*, *125*(1), 187–206. https://doi.org/10.1006/jcph.1996.0088

Greatbatch, R., Zhai, X., Claus, M., Czeschel, L., & Rath, W. (2010a). Transport driven by eddy momentum fluxes in the Gulf Stream Extension region. *Geophysical Research Letters*, *37*(24). https://doi.org/10.1029/2010GL045473

Greatbatch, R. J. (1998). Exploring the relationship between eddy-induced transport velocity, vertical momentum transfer, and the isopycnal flux of potential vorticity. *Journal of Physical Oceanography*, *28*(3), 422–432. https://doi.org/10.1175/1520-0485(1998)028⟨0422:ETRBEI⟩2.0.CO;2

Greatbatch, R. J., Zhai, X., Kohlmann, J.-D., & Czeschel, L. (2010b). Ocean eddy momentum fluxes at the latitudes of the Gulf Stream and the Kuroshio extensions as revealed by satellite data. *Ocean Dynamics*, *60*(3), 617–628. https://doi.org/10.1007/s10236-010-0282-6

Gregory, W., Bushuk, M., Adcroft, A., Zhang, Y., & Zanna, L. (2023). Deep learning of systematic sea ice model errors from data assimilation increments. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003757. https://doi.org/10.1029/2023MS003757

Griffies, S. M. (2018). *Fundamentals of ocean climate models*. Princeton university press. https://doi.org/10.2307/j.ctv301gzg

Griffies, S. M., & Hallberg, R. W. (2000). Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. *Monthly Weather Review*, *128*(8), 2935–2946. https://doi.org/10.1175/1520-0493(2000)128⟨2935:BFWASL⟩2.0.CO;2

Grooms, I. (2023). Backscatter in energetically-constrained Leith parameterizations. *Ocean Modelling*, *186*, 102265. https://doi.org/10.1016/j.ocemod.2023.10226

Grooms, I., Lee, Y., & Majda, A. J. (2015). Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence. *Multiscale Modeling and Simulation*, *13*(3), 1001–1021. https://doi.org/10.1137/140990048

Grooms, I., Loose, N., Abernathey, R., Steinberg, J., Bachman, S. D., Marques, G. M., et al. (2021). Diffusion-based smoothers for spatial filtering of gridded geophysical data. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002552. https://doi.org/10.1029/2021MS002552

Guan, Y., Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2022). Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning. *Journal of Computational Physics*, *458*, 111090. https://doi.org/10.1016/j.jcp.2022.111090

Guillaumin, A. P., & Zanna, L. (2021). Stochastic-deep learning parameterization of ocean momentum forcing. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002534. https://doi.org/10.1029/2021MS002534

Haidvogel, D. B., Curchitser, E. N., Danilov, S., & Fox-Kemper, B. (2017). Numerical modelling in a multiscale ocean. *Journal of Marine Research*, *75*(6), 683–725. https://doi.org/10.1357/002224017823523964

Hallberg, R. (2013). Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, *72*, 92–103. https://doi.org/10.1016/j.ocemod.2013.08.007

Hewitt, H. T., Roberts, M., Mathiot, P., Biastoch, A., Blockley, E., Chassignet, E. P., et al. (2020). Resolving and parameterising the ocean mesoscale in earth system models. *Current Climate Change Reports*, *6*(4), 137–152. https://doi.org/10.1007/s40641-020-00164-w

Hoskins, B. J., James, I. N., & White, G. H. (1983). The shape, propagation and mean-flow interaction of large-scale weather systems. *Journal of the Atmospheric Sciences*, *40*(7), 1595–1612. https://doi.org/10.1175/1520-0469(1983)040<1595:tspamf>2.0.co;2

Hughes, C. W., & Ash, E. R. (2001). Eddy forcing of the mean flow in the Southern Ocean. *Journal of Geophysical Research*, *106*(C2), 2713–2722. https://doi.org/10.1029/2000JC900332

Jakhar, K., Guan, Y., Mojgani, R., Chattopadhyay, A., & Hassanzadeh, P. (2024). Learning closed-form equations for subgrid-scale closures from high-fidelity data: Promises and challenges. *Journal of Advances in Modeling Earth Systems*, *16*(7), e2023MS003874. https://doi.org/10.1029/2023MS003874

Jansen, M. F., Adcroft, A., Khani, S., & Kong, H. (2019). Toward an energetically consistent, resolution aware parameterization of ocean mesoscale eddies. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2844–2860. https://doi.org/10.1029/2019MS001750

Jansen, M. F., & Held, I. M. (2014). Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling*, *80*, 36–48. https://doi.org/10.1016/j.ocemod.2014.06.002

Jansen, M. F., Held, I. M., Adcroft, A., & Hallberg, R. (2015). Energy budget-based backscatter in an eddy permitting primitive equation model. *Ocean Modelling*, *94*, 15–26. https://doi.org/10.1016/j.ocemod.2015.07.015

Juricke, S., Danilov, S., Koldunov, N., Oliver, M., & Sidorenko, D. (2020). Ocean kinetic energy backscatter parametrization on unstructured grids: Impact on global eddy-permitting simulations. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001855. https://doi.org/10.1029/2019MS001855

Juricke, S., Danilov, S., Kutsenko, A., & Oliver, M. (2019). Ocean kinetic energy backscatter parametrizations on unstructured grids: Impact on mesoscale turbulence in a channel. *Ocean Modelling*, *138*, 51–67. https://doi.org/10.1016/j.ocemod.2019.03.009

Kamenkovich, I., Berloff, P., & Pedlosky, J. (2009). Role of eddy forcing in the dynamics of multiple zonal jets in a model of the North Atlantic. *Journal of Physical Oceanography*, *39*(6), 1361–1379. https://doi.org/10.1175/2008JPO4096.1

Khani, S., & Dawson, C. N. (2023). A gradient based subgrid-scale parameterization for ocean mesoscale eddies. *Journal of Advances in Modeling Earth Systems*, *15*(2), e2022MS003356. https://doi.org/10.1029/2022MS003356

Kjellsson, J., & Zanna, L. (2017). The impact of horizontal resolution on energy transfers in global ocean models. *Fluids*, *2*(3), 45. https://doi.org/10.3390/fluids2030045

Klöwer, M., Jansen, M. F., Claus, M., Greatbatch, R. J., & Thomsen, S. (2018). Energy budget-based backscatter in a shallow water model of a double gyre basin. *Ocean Modelling*, *132*, 1–11. https://doi.org/10.1016/j.ocemod.2018.09.006

Krasnopolsky, V., Fox-Rabinovitz, M., Hou, Y., Lord, S., & Belochitski, A. (2010). Accurate and fast neural network emulations of model radiation for the NCEP coupled climate forecast system: Climate simulations and seasonal predictions. *Monthly Weather Review*, *138*(5), 1822–1842. https://doi.org/10.1175/2009MWR3149.1

Leith, C. (1996). Stochastic models of chaotic systems. *Physica D: Nonlinear Phenomena*, *98*(2–4), 481–491. https://doi.org/10.1016/0167-2789(96)00107-8

Loose, N., Abernathey, R., Grooms, I., Busecke, J., Guillaumin, A., Yankovsky, E., et al. (2022). GCM-filters: A Python package for diffusion-based spatial filtering of gridded data. *Journal of Open Source Software*, *7*(70), 3947. https://doi.org/10.21105/joss.03947

Loose, N., Bachman, S., Grooms, I., & Jansen, M. (2023). Diagnosing scale-dependent energy cycles in a high-resolution isopycnal ocean model. *Journal of Physical Oceanography*, *53*(1), 157–176. https://doi.org/10.1175/JPO-D-22-0083.1

Lund, T. (1997). On the use of discrete filters for large eddy simulation. *Annual Research Briefs*, 83–95.

Madec, G., & the NEMO team. (2008). *NEMO ocean engine (Note du Pôle de modélisation No. 27)*. Institut Pierre-Simon Laplace (IPSL). https://doi.org/10.5281/zenodo.3248739

Mak, J., Maddison, J. R., Marshall, D. P., & Munday, D. R. (2018). Implementation of a geometrically informed and energetically constrained mesoscale eddy parameterization in an ocean circulation model. *Journal of Physical Oceanography*, *48*(10), 2363–2382. https://doi.org/10.1175/JPO-D-18-0017.1

Mak, J., Maddison, J. R., Marshall, D. P., Ruan, X., Wang, Y., & Yeow, L. (2023). Scale-awareness in an eddy energy constrained mesoscale eddy parameterization. *Journal of Advances in Modeling Earth Systems*, *15*(12), e2023MS003886. https://doi.org/10.1029/2023MS003886

Mak, J., Marshall, D. P., Madec, G., & Maddison, J. R. (2022). Acute sensitivity of global ocean circulation and heat content to eddy energy dissipation timescale. *Geophysical Research Letters*, *49*(8), e2021GL097259. https://doi.org/10.1029/2021GL097259

Marques, G. M. (2022). NeverWorld2 [Dataset]. *UCAR/NCAR - CISL - CDP*. https://doi.org/10.26024/f130-ev71

Marques, G. M., Loose, N., Yankovsky, E., Steinberg, J. M., Chang, C.-Y., Bhamidipati, N., et al. (2022). NeverWorld2: An idealized model hierarchy to investigate ocean mesoscale eddies across resolutions. *Geoscientific Model Development*, *15*(17), 6567–6579. https://doi.org/10.5194/gmd-15-6567-2022

Marshall, D. P., Maddison, J. R., & Berloff, P. S. (2012). A framework for parameterizing eddy potential vorticity fluxes. *Journal of Physical Oceanography*, *42*(4), 539–557. https://doi.org/10.1175/JPO-D-11-048.1

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

Meneveau, C., & Katz, J. (2000). Scale-invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics*, *32*(1), 1–32. https://doi.org/10.1146/annurev.fluid.32.1.1

Moser, R. D., Haering, S. W., & Yalla, G. R. (2021). Statistical properties of subgrid-scale turbulence models. *Annual Review of Fluid Mechanics*, *53*(1), 255–286. https://doi.org/10.1146/annurev-fluid-060420-023735

Pearson, B., Fox-Kemper, B., Bachman, S., & Bryan, F. (2017). Evaluation of scale-aware subgrid mesoscale eddy models in a global eddy-rich model. *Ocean Modelling*, *115*, 42–58. https://doi.org/10.1016/j.ocemod.2017.05.007

Perezhogin, P. (2020). Testing of kinetic energy backscatter parameterizations in the nemo ocean model. *Russian Journal of Numerical Analysis and Mathematical Modelling*, *35*(2), 69–82. https://doi.org/10.1515/rnam-2020-0006

Perezhogin, P. (2024a). Dataset for paper on implementation of Zanna-Bolton-2020 subgrid model to MOM6 GFDL ocean model [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.12682093

Perezhogin, P. (2024b). m2lines/Implementation-ZB20: v1.0.5 [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.12682335

Perezhogin, P., & Glazunov, A. (2023). Subgrid parameterizations of ocean mesoscale eddies based on Germano decomposition. *Journal of Advances in Modeling Earth Systems*, *15*(10). https://doi.org/10.1029/2023ms003771

Perezhogin, P., Glazunov, A. V., & Gritsun, A. S. (2019). Stochastic and deterministic kinetic energy backscatter parameterizations for simulation of the two-dimensional turbulence. *Russian Journal of Numerical Analysis and Mathematical Modelling*, *34*(4), 197–213. https://doi.org/10.1515/rnam-2019-0017

Perezhogin, P., Zanna, L., & Fernandez-Granda, C. (2023). Generative data-driven approaches for stochastic subgrid parameterizations in an idealized ocean model. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003681. https://doi.org/10.1029/2023MS003681

Qiu, B., & Chen, S. (2010). Eddy-mean flow interaction in the decadally modulating Kuroshio Extension system. *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(13–14), 1098–1110. https://doi.org/10.1016/j.dsr2.2008.11.036

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Redi, M. H. (1982). Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography*, *12*(10), 1154–1158. https://doi.org/10.1175/1520-0485(1982)012⟨1154:OIMBCR⟩2.0.CO;2

Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Benchmarking of machine learning ocean subgrid parameterizations in an idealized model. *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003258. https://doi.org/10.1029/2022MS003258

Sadourny, R. (1975). The dynamics of finite-difference models of the shallow-water equations. *Journal of the Atmospheric Sciences*, *32*(4), 680–689. https://doi.org/10.1175/1520-0469(1975)032⟨0680:TDOFDM⟩2.0.CO;2

Sagaut, P. (2006). *Large eddy simulation for incompressible flows: An introduction*. Springer. https://doi.org/10.1007/b137536

Salmon, R. (1980). Baroclinic instability and geostrophic turbulence. *Geophysical & Astrophysical Fluid Dynamics*, *15*(1), 167–211. https://doi.org/10.1080/03091928008241178

San, O. (2014). A dynamic eddy-viscosity closure model for large eddy simulations of two-dimensional decaying turbulence. *International Journal of Computational Fluid Dynamics*, *28*(6–10), 363–382. https://doi.org/10.1080/10618562.2014.948426

Sane, A., Reichl, B. G., Adcroft, A., & Zanna, L. (2023). Parameterizing vertical mixing coefficients in the ocean surface boundary layer using neural networks. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003890. https://doi.org/10.1029/2023MS003890

Shamekh, S., Lamb, K. D., Huang, Y., & Gentine, P. (2023). Implicit learning of convective organization explains precipitation stochasticity. *Proceedings of the National Academy of Sciences*, *120*(20), e2216158120. https://doi.org/10.1073/pnas.2216158120

Skamarock, W. C. (2004). Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, *132*(12), 3019–3032. https://doi.org/10.1175/MWR2830.1

Soufflet, Y., Marchesiello, P., Lemarié, F., Jouanno, J., Capet, X., Debreu, L., & Benshila, R. (2016). On effective resolution in ocean models. *Ocean Modelling*, *98*, 36–50. https://doi.org/10.1016/j.ocemod.2015.12.004

Storto, A., & Andriopoulos, P. (2021). A new stochastic ocean physics package and its application to hybrid-covariance data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *147*(736), 1691–1725. https://doi.org/10.1002/qj.3990

Thuburn, J., Kent, J., & Wood, N. (2014). Cascades, backscatter and conservation in numerical models of two-dimensional turbulence. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 626–638. https://doi.org/10.1002/qj.2166

Uchida, T., Abernathey, R., & Smith, S. (2017). Seasonality of eddy kinetic energy in an eddy permitting global climate model. *Ocean Modelling*, *118*, 41–58. https://doi.org/10.1016/j.ocemod.2017.08.006

Uchida, T., Rokem, A., Soler, S., Nicholas, T., Abernathey, R., Nouguier, F., et al. (2023). xgcm/xrft: v1.0.1 [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.7621857

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press. https://doi.org/10.1017/9781107588417

Vollant, A., Balarac, G., & Corre, C. (2016). A dynamic regularized gradient model of the subgrid-scale stress tensor for large-eddy simulation. *Physics of Fluids*, *28*(2), 025114. https://doi.org/10.1063/1.4941781

Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. https://doi.org/10.1038/nclimate2118

Wang, P., Yuval, J., & O'Gorman, P. A. (2022). Non-local parameterization of atmospheric subgrid processes with neural networks. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2022MS002984. https://doi.org/10.1029/2022MS002984

Wardle, R., & Marshall, J. (2000). Representation of eddies in primitive equation models by a PV flux. *Journal of Physical Oceanography*, *30*(10), 2481–2503. https://doi.org/10.1175/1520-0485(2000)030⟨2481:ROEIPE⟩2.0.CO;2

Waterman, S., & Jayne, S. R. (2011). Eddy-mean flow interactions in the along-stream development of a western boundary current jet: An idealized model study. *Journal of Physical Oceanography*, *41*(4), 682–707. https://doi.org/10.1175/2010JPO4477.1

Yankovsky, E., Bachman, S., Smith, K. S., & Zanna, L. (2024). Vertical structure and energetic constraints for a backscatter parameterization of ocean mesoscale eddies. *Journal of Advances in Modeling Earth Systems*, *16*(7), e2023MS004093. https://doi.org/10.1029/2023MS004093

Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, *11*(1), 3295. https://doi.org/10.17605/OSF.IO/36YPT

Yuval, J., & O'Gorman, P. A. (2023). Neural-network parameterization of subgrid momentum transport in the atmosphere. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2023MS003606. https://doi.org/10.1029/2023MS003606

Zampieri, L., Arduini, G., Holland, M., Keeley, S. P., Mogensen, K., Shupe, M. D., & Tietsche, S. (2023). A machine learning correction model of the winter clear-sky temperature bias over the Arctic sea ice in atmospheric reanalyses. *Monthly Weather Review*, *151*(6), 1443–1458. https://doi.org/10.1175/MWR-D-22-0130.1

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376. https://doi.org/10.1029/2020GL088376

Zanna, L., Mana, P. P., Anstey, J., David, T., & Bolton, T. (2017). Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Modelling*, *111*, 66–80. https://doi.org/10.1016/j.ocemod.2017.01.004

Zhang, C., Perezhogin, P., Gultekin, C., Adcroft, A., Fernandez-Granda, C., & Zanna, L. (2023). Implementation and evaluation of a machine learned mesoscale eddy parameterization into a numerical ocean circulation model. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003697. https://doi.org/10.1029/2023MS003697