Implementation of a data-driven equation-discovery mesoscale parameterization into an ocean model

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

[1]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA
[2]Center for Data Science, New York University, New York, NY, USA
[3]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ 08542, USA

Key Points:

- A data-driven mesoscale eddy parameterization is implemented and evaluated in the GFDL MOM6 ocean model
- Filtering schemes are proposed to improve the numerical and physical properties of the parameterization
- The subgrid parameterization improves the representation of the energy distributions and the climatological mean flow

Corresponding author: P. Perezhogin, pp2681@nyu.edu

Abstract

Mesoscale eddies are poorly represented in climate ocean models, and therefore their effects on the large scale circulation must be parameterized. Classical 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) (ZB20) applied an equation-discovery algorithm to reveal an interpretable expression parameterizing the subgrid mesoscale fluxes 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 stratification and topography. In addition, we propose an approach based on spatial filtering to improve the representation of large-scale energy backscatter and numerical properties of the parameterization. The ZB20 parameterization led to improved climatological mean flow and energy distributions, compared to the current state-of-the-art energy backscatter parameterizations. The ZB20 is scale-aware and can be used with a single value of the non-dimensional scaling coefficient for a range of resolutions. The successful application of the ZB20 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 small ocean currents known as mesoscale eddies. 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 machine-learning parameterization simulating the physics of the mesoscale eddies into the state-of-the-art ocean model. The parameterization is interpretable and captures key physical process 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 flow and energetics. We propose new filtering schemes which improve the physical and numerical properties of the parameterization. Accurate representation of the mesoscale eddies has the potential to resolve long-standing biases present in global ocean models and thus allow for more reliable climate simulations.

1 Introduction

Mesoscale eddies in the ocean 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 kinetic energy (KE) reservoir and are critical for the lateral and vertical transport of tracers (Redi, 1982; Ferrari & Wunsch, 2009; 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 energy budget (Jansen et al., 2019; Bachman, 2019; Loose et al., 2023).

The direct simulation of mesoscale eddies requires resolving the deformation radius with a few grid spacings (Hallberg, 2013) and remains unfeasible in the global climate models (Hewitt et al., 2020; Christensen & Zanna, 2022). Therefore, the systematic effect of the unresolved mesoscale eddies must be *parameterized* to achieve an unbiased climatological state. Classical parameterizations mimic the bulk (that is, mean) effect of the mesoscale eddies on the resolved flow and are commonly based on energetic considerations (D. P. Marshall et al., 2012; Mak et al., 2018). For example, the parameterization of Gent and McWilliams (1990) reduces the potential energy by flattening the isopycnals. Another popular parameterization is the *KE backscatter* (Thuburn et al.,

2014; Jansen & Held, 2014; Grooms et al., 2015; Berloff, 2018; Bachman, 2019; Jansen et al., 2019; Perezhogin, 2020; Juricke et al., 2020; Storto & Andriopoulos, 2021) which increases the KE of the resolved circulation. The backscatter accounts for the unresolved upscale energy cascade and can be diagnosed or parameterized with the upgradient momentum fluxes (that is, fluxes, that sharpen the gradients of the resolved flow).

Recently, machine-learning (ML) methods based on neural networks have been proposed to predict the eddy fluxes directly from data as an alternative to the classical bulk modeling (Krasnopolsky et al., 2010; Rasp et al., 2018; Yuval & O'Gorman, 2020; Guan, Chattopadhyay, et al., 2022; Frezat et al., 2022; Sane et al., 2023; Perezhogin et al., 2023; Yuval & O'Gorman, 2023; Shamekh et al., 2023; Zampieri et al., 2023; Gregory et al., 2023). ML methods capture not only the bulk effect of the subgrid eddies but also the instantaneous fields of the diagnosed eddy forcing (Bolton & Zanna, 2019). These methods are very accurate, but they contain a tremendous number of tunable parameters that obstructs their physical interpretation. The approach proposed by Zanna and Bolton (2020) (hereafter ZB20) provided an alternative to both the traditional bulk approach and the black-box ML modeling by enabling the discovery of a closed-form equation for the eddy parameterization directly from data.

ZB20 parameterized the mesoscale eddy fluxes through a simple interpretable expression that has strong ties with physics-based gradient models (Anstey & Zanna, 2017; Zanna & Bolton, 2020; Jakhar et al., 2023; Khani & Dawson, 2023). 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 ocean model in adiabatic limit. In Section 3, we describe the ZB20 parameterization and propose filtering schemes that reduce the grid-scale numerical instabilities and improve the large-scale KE backscatter. The need for the proposed filtering schemes is clarified in Section 4 where we evaluate the subgrid KE transfer produced by the ZB20 parameterization both offline and online. In Section 4, we further test the ZB20 parameterization in an idealized ocean configuration, Double Gyre, and show a reduction of biases in the mean flow and energetics. In Section 5, we evaluate the ZB20 parameterization for a range of resolutions and show that it does not require retuning the free parameter. This is mainly a consequence of the scale-awareness of the free parameter, which is solely a function of the grid spacing. A more realistic configuration NeverWorld2 (NW2, Marques et al., 2022) mimicking the global ocean is described in Section 6. Testing in the NW2 configuration revealed the need for additional tuning required to improve the numerical stability properties of the parameterization. However, the reduction of biases in energy distributions and mean flow are also more evident compared to the simpler configuration.

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 an idealized setting 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):

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

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

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, 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 relative vorticity; $\nabla = (\partial_x, \partial_y)$ is the horizontal gradient and $\nabla\cdot$ is the horizontal divergence, 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. $\mathbf{S}_k$ is a subgrid momentum parameterization. $\mathbf{V}_k$ is the biharmonic Smagorinsky model (below, we omit index k for brevity):

$$\mathbf{V} = -\nabla \cdot (\nu_4 \nabla(\nabla^2 \mathbf{u})), \tag{3}$$

and $\nu_4 = C_S \Delta^4 \sqrt{\widetilde{D}^2 + D^2}$, where C_S is the non-dimensional Smagorinsky coefficient, Δ is the grid spacing and $\widetilde{D} = \partial_x u - \partial_y v$ is the stretching deformation, $D = \partial_y u + \partial_x v$ is the shearing deformation. We note that the biharmonic operator used in the MOM6 ocean model has a complicated form and refer the reader to S. M. Griffies and Hallberg (2000) for more details.

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 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 using a spatial filtering approach (Large Eddy Simulation, LES, Sagaut, 2006; Fox-Kemper & Menemenlis, 2008; Bachman et al., 2017). We then describe our implementation of ZB20 parameterization of the subgrid mesoscale eddies and baselines.

3.1 Subgrid forcing

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

$$\mathbf{S} = (\overline{\mathbf{u}} \cdot \overline{\nabla})\overline{\mathbf{u}} - \overline{(\mathbf{u} \cdot \nabla)\mathbf{u}}, \tag{4}$$

where $\mathbf{u}$ is the velocity field of the high-resolution model, $\overline{(\cdot)}$ is a spatial filter used in the LES approach and $\overline{\nabla}$ is a discretization of the gradient operator on a coarse grid. This subgrid forcing modifies the governing equations of the coarse ocean model as shown in Eq. (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 Eq. (4). 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*.

The key physical process which is described by the subgrid momentum forcing is the redistribution of the kinetic energy from subgrid to resolved scales which is referred

to as a *KE backscatter* (Frederiksen & Davies, 1997; Berner et al., 2009; Thuburn et al., 2014; Ross et al., 2023). The backscatter energizes the ocean currents on large scales and represents the unresolved on a coarse grid upscale energy cascade (Mana & Zanna, 2014).

3.2 Zanna-Bolton parameterization (ZB20)

Zanna and Bolton (2020) introduced the following *baroclinic parameterization* for subgrid momentum forcing (referred to as ZB20):

$$\mathbf{S} = \begin{pmatrix} S_x \\ S_y \end{pmatrix} = \nabla \cdot \mathbf{T} = \nabla \cdot \begin{pmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{pmatrix} = \begin{pmatrix} \partial_x T_{xx} + \partial_y T_{xy} \\ \partial_x T_{xy} + \partial_y T_{yy} \end{pmatrix}, \tag{5}$$

where the stress tensor $\mathbf{T}$ is given by:

$$\mathbf{T}(\zeta, D, \widetilde{D}) = \underbrace{\kappa_{BC} \begin{bmatrix} -\zeta D & \zeta \widetilde{D} \\ \zeta \widetilde{D} & \zeta D \end{bmatrix}}_{\mathbf{T}_d,\ \text{deviatoric stress}} + \underbrace{\frac{\kappa_{BC}}{2} \left(\zeta^2 + D^2 + \widetilde{D}^2\right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{T}_h,\ \text{hydrostatic stress}}. \tag{6}$$

Zanna and Bolton (2020) estimated the free coefficient κ_{BC} from their simulation, such that $\kappa_{BC} = -8.723 \times 10^8 \mathrm{m}^2$. Here, we follow the approach of Anstey and Zanna (2017) and studies on gradient models (Chen et al., 2003; Meneveau & Katz, 2000) to directly relate the free coefficient to the area of a coarse grid box:

$$\kappa_{BC} = -\gamma \Delta_x \Delta_y \leq 0, \tag{7}$$

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.

We refer the reader to Appendix A for details of the numerical discretization of Eq. (5), while accounting for the curvilinear coordinates and varying layer thickness.

3.2.1 Low-pass filtering of the stress tensor (ZB20-Smooth)

Incorporation of the ZB20 parameterization can result in generation of the numerical noise near the grid scale. It happens because the energetic contribution of the deviatoric stress is zero ($\mathbf{T}_d : (\nabla \mathbf{u}) = 0$, see A3 in Appendix A) while the energetic contribution of the hydrostatic stress ($\mathbf{T}_h \ : \ (\nabla \mathbf{u})$) is not sign-definite (Zanna & Bolton, 2020). Thus, dissipation near the grid scale is not implemented as a hard constraint. Note that the energy contribution of the deviatoric stress is not trivial: it redistributes energy in physical and spectral space (Anstey & Zanna, 2017). The dissipation is often enforced in gradient-based parameterizations by projecting the predicted stress tensor onto the dissipative direction (Bouchet, 2003; Balarac et al., 2013; Vollant et al., 2016). However, these approaches are not suitable for parameterization of mesoscale eddies because their dominant effect is the KE backscatter on large scales. Thus, we suggest to remove 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 (Grooms et al., 2015; Juricke et al., 2019; Bachman, 2019; Perezhogin et al., 2019).

We consider a low-pass convolutional filter defined on 3×3 spatial stencil ("trapezoidal filter" in San (2014)):

$$\mathcal{G} = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \tag{8}$$

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:

$$\mathbf{S} = \nabla \cdot G(\mathbf{T}), \tag{9}$$

where $G = \mathcal{G}^N$ is the low-pass filter which is applied to every component of the stress tensor $\mathbf{T}$ (Eq. (6)), and N is the number of filtering passes of $\mathcal{G}$ (Eq. (8)). We choose $N = 4$ similarly to Juricke et al. (2019) (see Appendix C for sensitivity to the choice of N).

3.2.2 High-pass filtering of the velocity gradients (ZB20-Reynolds)

To further enhance the KE backscatter on large scales, we use an additional filtering scheme.

Perezhogin and Glazunov (2023) show that a particular part of the subgrid stress (Reynolds stress) is responsible for the KE backscatter on large scales. The Reynolds stress is a part of Germano (1986) decomposition of the subgrid stress and represents the effect of the eddy-eddy interactions on the mean flow. Based on their results, we propose a modification of ZB20 parameterization by isolating the backscatter effect through the use of a high-pass filter on the velocity gradients. Therefore, the modified parameterization (hereafter, ZB20-Reynolds) can be expressed as

$$\mathbf{S} = \nabla \cdot G(\mathbf{T}(\zeta', D', \widetilde{D}')), \tag{10}$$

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 Eq. (10) is done as follows. First, the velocity gradients ζ, D, $\widetilde{D}$ are high-pass filtered to obtain ζ', D', $\widetilde{D}'$, respectively. Then we compute the stress tensor $\mathbf{T}(\zeta', D', \widetilde{D}')$ according to Eq. (6) 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 $N = 4$.

3.3 Baseline momentum parameterizations

We consider two backscatter parameterizations as baselines. The first baseline is the KE backscatter of Jansen et al. (2015) (referred to as JH15), already tested in MOM6 (Jansen et al., 2019). It parameterizes the dissipation of the resolved KE with the biharmonic Smagorinsky model. The backscatter of the subgrid KE is parameterized with the Laplacian viscosity having a negative coefficient. The negative viscosity coefficient is informed by a local subgrid KE equation.

The second baseline parameterization is a deep-learning CNN model of Guillaumin and Zanna (2021) (referred to as GZ21). It predicts the subgrid forcing using horizontal velocities and was trained on the global ocean data. It was implemented to the MOM6 ocean model (Zhang et al., 2023) and works together with the biharmonic Smagorinsky model. 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. Zhang et al. (2023) performed such tuning in the Double Gyre configuration at every coarse resolution (in the range $1/4^o - 1/8^o$). We use their setting but double the integration time with a single ensemble member.

4 Experiments in Double Gyre configuration

We test subgrid parameterizations in Double Gyre configuration of the MOM6 ocean model described in detail in Section 3.1 of Zhang et al. (2023). We use their setting but double the integration time (20 years). The model has two fluid layers and is initialized at rest. The circulation is driven by wind and equilibrated by the bottom friction. The

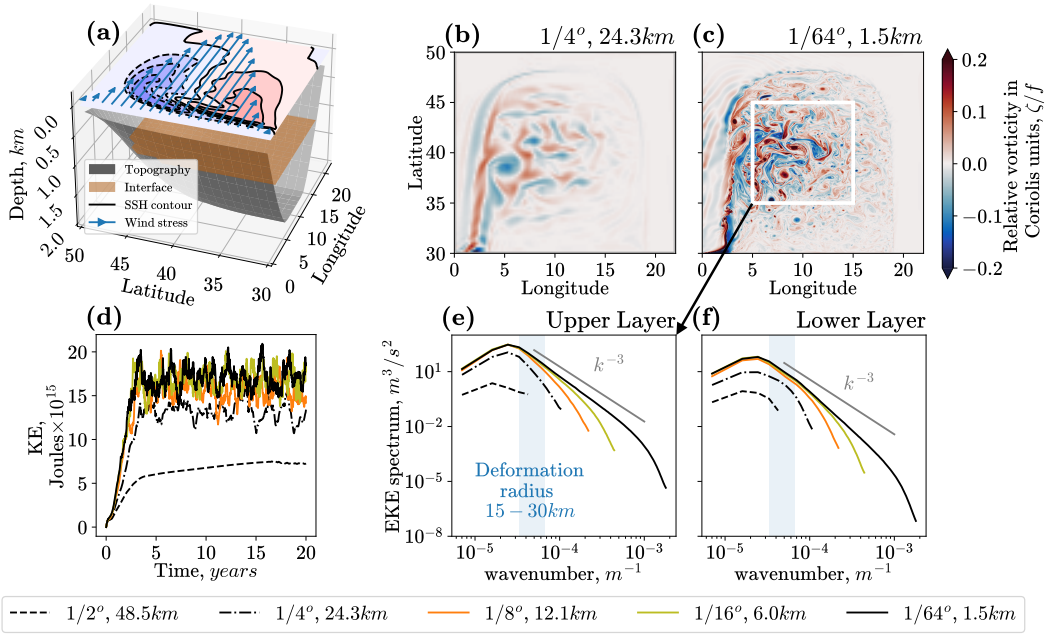

Figure 1. (a) The sketch of the Double Gyre configuration of the MOM6 ocean model; (b) and (c) snapshots of the relative vorticity in the upper layer, (d) total kinetic energy (summed over layers). (e,f) The spatial spectra of the eddy kinetic energy (EKE) in the upper and lower fluid layers. The spectrum is computed in the area shown by the white rectangle. The wavenumbers of the deformation radius ($r_d \approx \frac{1}{f}\sqrt{g'\frac{h_1 h_2}{h_1+h_2}}$) are given by $1/r_d$ and shown by the blue shading.

computational domain is given in a spherical coordinate system with a bowl topography (Figure 1(a)). The Smagorinsky coefficient in all experiments is set to $C_S = 0.06$, similarly to Jansen et al. (2015, 2019); Zhang et al. (2023); see Appendix C for sensitivity to C_S.

The unparameterized simulations ($\mathbf{S} = 0$) for a range of resolutions ($1/64^o$, $1/16^o$, $1/8^o$, $1/4^o$, $1/2^o$) are shown in Figure 1. The high-resolution model ($1/64^o$) has a grid spacing that is 10 times smaller than the Rossby deformation radius, and consequently, it directly simulates the mesoscale eddies (Hallberg, 2013). Coarse ocean models with a grid spacing approximately equal to the Rossby deformation radius ($1/2^o$–$1/8^o$) have a reduced eddy activity at all spatial scales in comparison to the high-resolution model. It follows from the analysis of eddy kinetic energy (EKE) spectrum, see Figure 1(e,f).

	ZB20	ZB20-Smooth	ZB20-Reynolds
γ	0.5	1.0	2.0

Table 1. The default scaling coefficient γ in the different ZB20-based parameterizations.

In this section, we study how subgrid parameterizations improve biases in energetics and mean flow in a coarse ocean model at resolution $1/4^o$. The scaling coefficient γ of the ZB20-based parameterizations which we use in this section is provided in Table 1. The coefficient γ is larger for the filtered than for the unfiltered versions of the ZB20

parameterization because filtering reduces the magnitude of the subgrid forcing prediction.

4.1 Eddy kinetic energy (EKE) spectrum

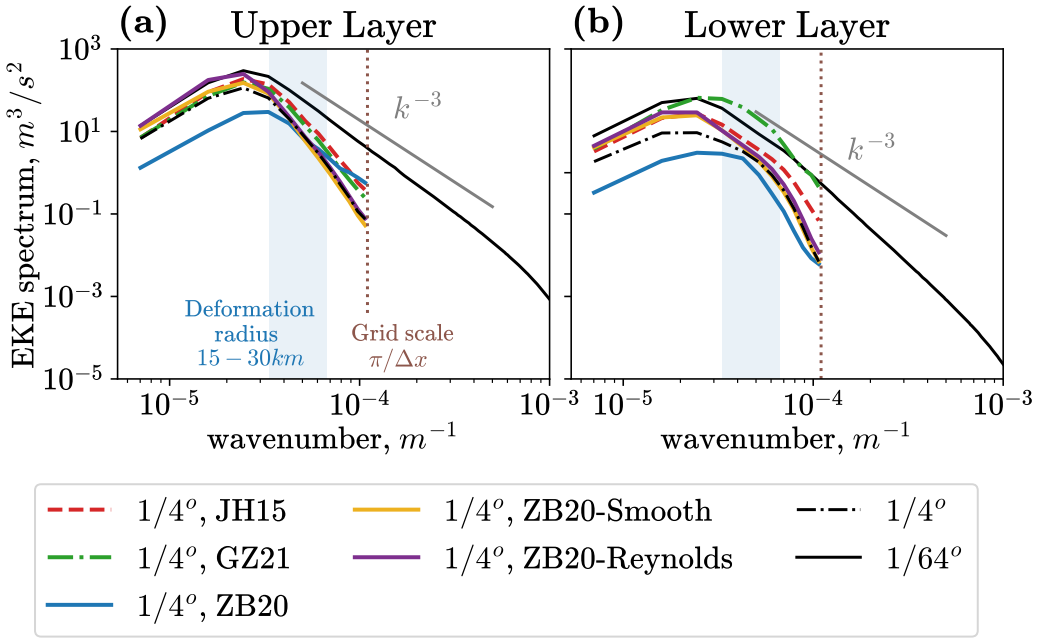

Figure 2. The eddy kinetic energy (EKE) spatial spectrum for the last 10 years of the simulation: (a) upper layer, (b) lower layer. The unparameterized model ($1/4^o$) is compared to parameterized models and high-resolution simulation ($1/64^o$). The spectrum is computed in the region indicated by the white rectangle in Figure 1.

The eddy kinetic energy (EKE) spectrum is one of the metrics that coarse unparameterized ocean models fail to reproduce (Figure 1(e,f)). 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 Double Gyre configuration are computed in the white rectangle shown in Figure 1(c) using a 2D Fourier transform with a Hann window and linear detrending (using xrft package, Uchida et al., 2023).

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, i.e. numerical noise (Figure 2(a)), 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 noise generation. The ZB20-Smooth and ZB20-Reynolds parameterizations improve the EKE spectrum on large scales by increasing it compared to the unparameterized model and closer to the high-resolution model. The ZB20-Reynolds parameterization is more efficient in energizing eddies in the upper layer. The JH15 and GZ21 backscatter parameterizations are more efficient in energizing eddies near the deformation scale. We note that none of the tested subgrid parameterizations reproduce the EKE spectrum on all spatial scales.

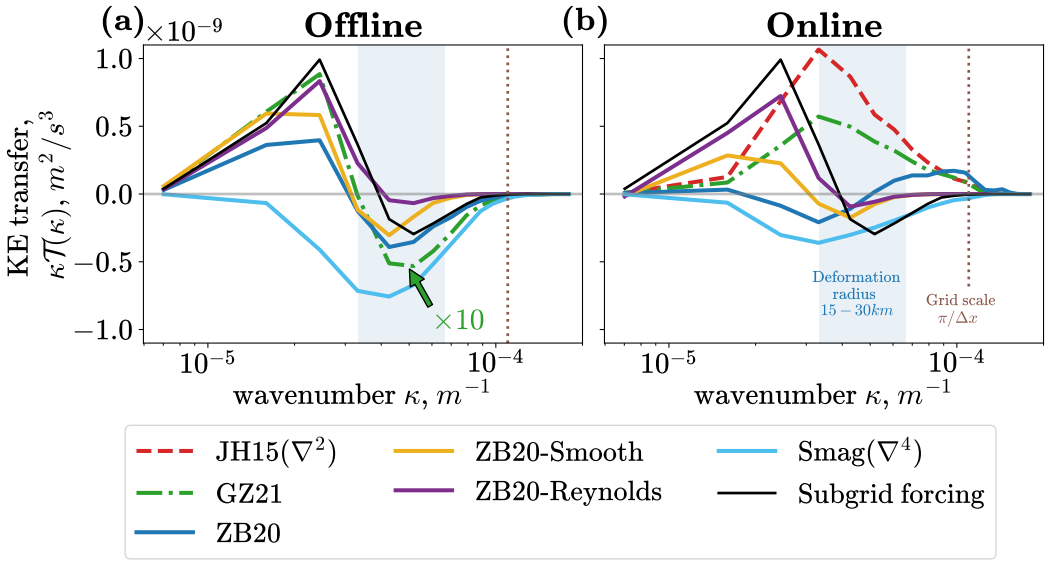

Figure 3. Kinetic energy transfer by subgrid parameterizations: (a) offline performance on fields of the high-resolution model ($1/64^o$) filtered and coarse-grained to $1/4^o$ resolution; (b) subgrid parameterizations are evaluated online at resolution $1/4^o$. The black line (subgrid forcing, Eq. (4)) is the same on both plots. The spectra correspond to the upper fluid layer and are computed in the white rectangle shown in Figure 1; the time-averaging interval is the last 10 years. All subgrid parameterizations are given separately from the biharmonic Smagorinsky model; for example, "JH15(∇^2)" shows only a contribution from the negative Laplacian viscosity. The "Smag(∇^4)" is the biharmonic Smagorinsky model with $C_S = 0.06$; in panel (b) it corresponds to run with $\mathbf{S} = 0$. The prediction given by GZ21 on panel (a) is multiplied by 10 for convenience.

4.2 Subgrid kinetic energy transfer

In this section, we explain the reproducing of the EKE spectrum by different coarse parameterized ocean models via analysis of the subgrid energy transfer. 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 offline analysis evaluates the subgrid parameterization on the filtered and coarse-grained snapshots of the high-resolution model while online analysis evaluates the parameterization once the parameterized ocean model is integrated in time.

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

$$\mathcal{T}(\kappa_x, \kappa_y) = Re\left[\mathcal{F}(u)^* \mathcal{F}(S_x) + \mathcal{F}(v)^* \mathcal{F}(S_y)\right], \tag{11}$$

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 ($\kappa_x^2 + \kappa_y^2 = \kappa^2$) to obtain the isotropic transfer spectrum $\mathcal{T}(\kappa)$. We compute the subgrid forcing according to Eq. (4), where the filtering operator $\overline{(\cdot)}$ is defined as a combination of the Gaussian filter implemented in the package GCM-filters (Grooms et al., 2021; Loose et al., 2022) followed by coarse-graining. 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, Ghosal, 1996; Lund, 1997; Chow & Moin, 2003; Perezhogin & Glazunov, 2023). Here, we use FGR =

2.5, found to optimally reproduce the subgrid KE transfer by the ZB20-based parameterizations in offline analysis.

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 blue line in Figure 3(a). However, in online analysis, the energy transfer is reversed: the ZB20 dissipates energy on large scales and generates energy near the grid scale (Figure 3(b)). This explains the emergence of the grid-scale numerical noise and deterioration of the large scales in the EKE spectrum of the ZB20 model (Figure 2). The filtering techniques (ZB20-Smooth and ZB20-Reynolds) avoid the generation of the numerical noise near the grid scale and preserve the original properties of the ZB20 such that the backscattering region for these filtered parameterizations is located in the large scales (Figure 3(b)). The ZB20-Reynolds parameterization appears more efficient than ZB20-Smooth in reproducing the KE backscatter.

The JH15 parameterization is evaluated only in online simulations; the equation for subgrid KE, evolving in time, complicates the offline analysis of JH15. The backscattering part of JH15 (negative Laplacian viscosity) introduces the positive energy transfer on higher wavenumbers (smaller scales) compared to the filtered ZB20 parameterizations and diagnosed subgrid forcing (Figure 3(b)). The JH15 parameterization is expected to be more accurate when compared to the subgrid forcing diagnosed with more scale-selective filters (Kraichnan, 1976; Perezhogin & Glazunov, 2023). The offline prediction of the GZ21 parameterization is multiplied by 10 for convenience. This factor presumably corrects for the discrepancy in the filter width: the GZ21 was trained with the narrower filter width ($\text{FGR} = \sqrt{12}/2 \approx 1.73 < 2.5$). After applying the correction, the GZ21 parameterization accurately predicts the offline transfer spectrum of the subgrid forcing (Figure 3(a)). However, similarly to the unfiltered ZB20 parameterization, it significantly changes the prediction once evaluated online. The online KE transfer spectrum for the GZ21 is purely positive and lacks the dissipative region; its shape resembles the negative Laplacian viscosity model, see Figure 3(b). The impact of the JH15 and GZ21 parameterizations on the middle scales of the EKE spectrum (Figure 2) can be explained by the more scale-selective backscatter, which may be seen as an advantage; however, it can also lead to numerical instabilities (Grooms et al., 2015; Bachman, 2019; Juricke et al., 2019; Grooms, 2023; Bagaeva et al., 2023).

Note that the small-scale dissipation in online runs is provided by the biharmonic Smagorinsky model and is not captured by ML-based subgrid parameterizations (ZB20, GZ21) contrary to their offline results.

4.3 Kinetic and potential energy

We compute the kinetic energy (KE) and available potential energy (APE) (Appendix B). We split the KE into the kinetic energy of the mean flow averaged over the last 10 years of the simulations (mean kinetic energy, MKE) and residual – the eddy part (eddy kinetic energy, EKE), and similarly for APE with mean potential energy (MPE) and eddy potential energy (EPE).

We compare the coarse models to the high-resolution model and its coarse-grained version (Figure 4). The MKE and APE of the unparameterized model ($1/4^o$) are too large. However, the total kinetic energy (EKE+MKE) is only slightly underestimated and thus, the need for the backscatter parameterization in the Double Gyre configuration is relatively moderate. The ZB20 parameterization improves (reduces) the APE, as expected from the mesoscale parameterization (Figure 4(b)). However, due to the generation of numerical noise, the large-scale eddies are disrupted, and the KE is too low (Figure 4(a)). Note that compared to the physical buoyancy parameterizations (Gent & McWilliams,

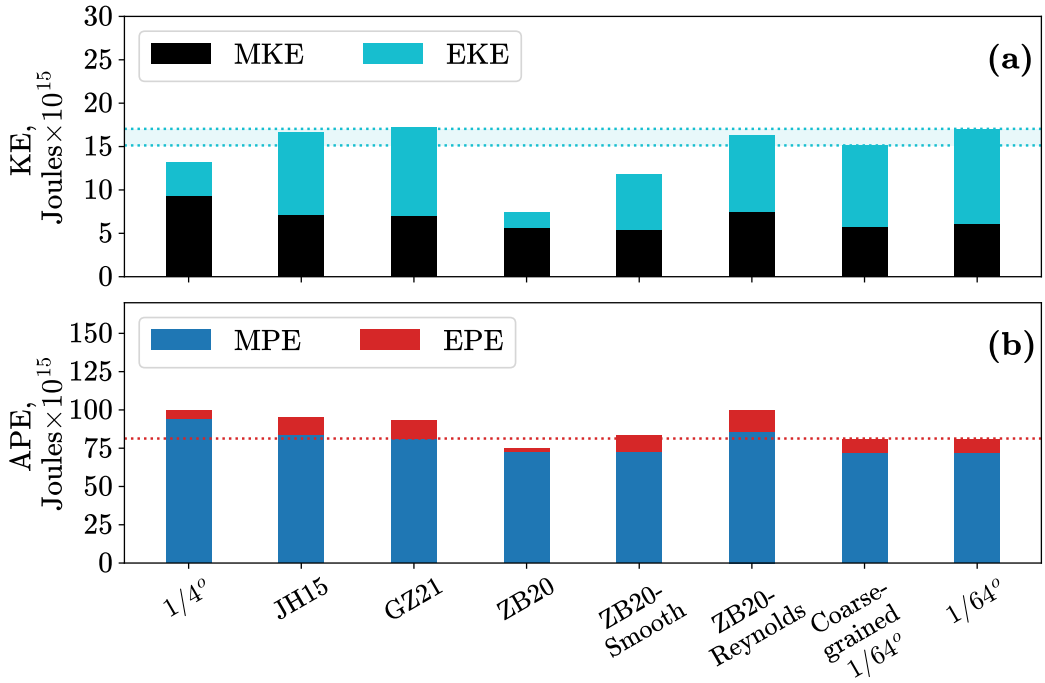

Figure 4. Experiments at resolution $1/4^o$: (a) Kinetic energy (KE) of the mean flow (MKE) and eddies (EKE), (b) Available potential energy (APE) of the mean flow (MPE) and eddies (EPE). In both plots, energy is summed over the fluid layers or interfaces and averaged over the last 10 years. The horizontal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

1990), reduction of APE appears indirectly through the change of the mean flow. 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 4(a)). The backscatter parameterizations (ZB20-Reynolds, JH15 and GZ21) efficiently energize the flow: they increase the kinetic energy and reduce the MKE. However, they have little impact on the APE.

4.4 Mean flow

Figure 5 shows the time-mean sea surface height (SSH). The unparameterized model ($1/4^o$) has a strong persistent recirculation near the boundary marked by the green rectangle, contrary to the prediction of the high-resolution model. In parameterized models (ZB20, ZB20-Smooth, ZB20-Reynolds, JH15, GZ21), we observe an improvement in the representation of the mean flow. The persistent recirculation is less evident, and the region exhibits a meandering jet pattern similar to that observed in the high-resolution model. The ZB20-Smooth model is the most accurate in representing the mean flow, and its RMSE (root mean squared error) is significantly lower than for other parameterized models (see Figure 5 for RMSE SSH values used to describe the mean bias in SSH). Note that the ZB20-Smooth parameterization improved the mean flow without increasing the KE (i.e., without strong backscatter). On the contrary, the most efficient backscatter parameterizations (ZB20-Reynolds, JH15, GZ21) have higher RMSE for SSH. By increasing the scaling coefficient γ, we can further improve the RMSE for the ZB20-Reynolds

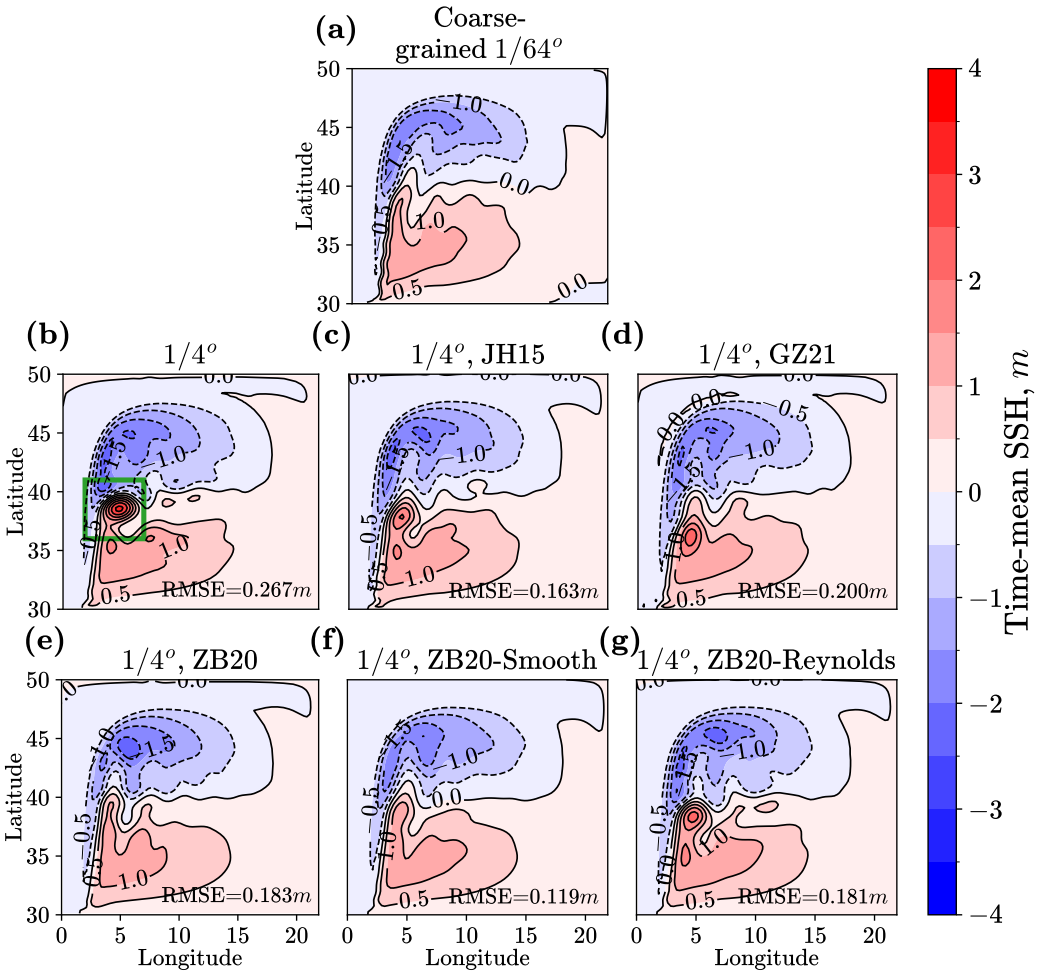

Figure 5. Sea surface height (SSH) averaged over the last 10 years for experiments at $1/4^o$ resolution. For every coarse model, we provide the root mean squared error (RMSE) in the time-mean SSH with respect to the coarse-grained high-resolution model shown in panel (a).

model (Appendix C). However, the value is constrained by numerical stability, 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 properties of the mean flow and potential energy. 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 an eddy-mean flow interaction and has been analyzed in a number of studies (Andrews & Mcintyre, 1976; Hoskins et al., 1983; R. J. Greatbatch, 1998; Wardle & Marshall, 2000; Hughes & Ash, 2001; Kamenkovich et al., 2009; Qiu & Chen, 2010; R. J. Greatbatch et al., 2010; R. Greatbatch et al., 2010; 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 $> 5^o$) and decelerating the jet in the separation region

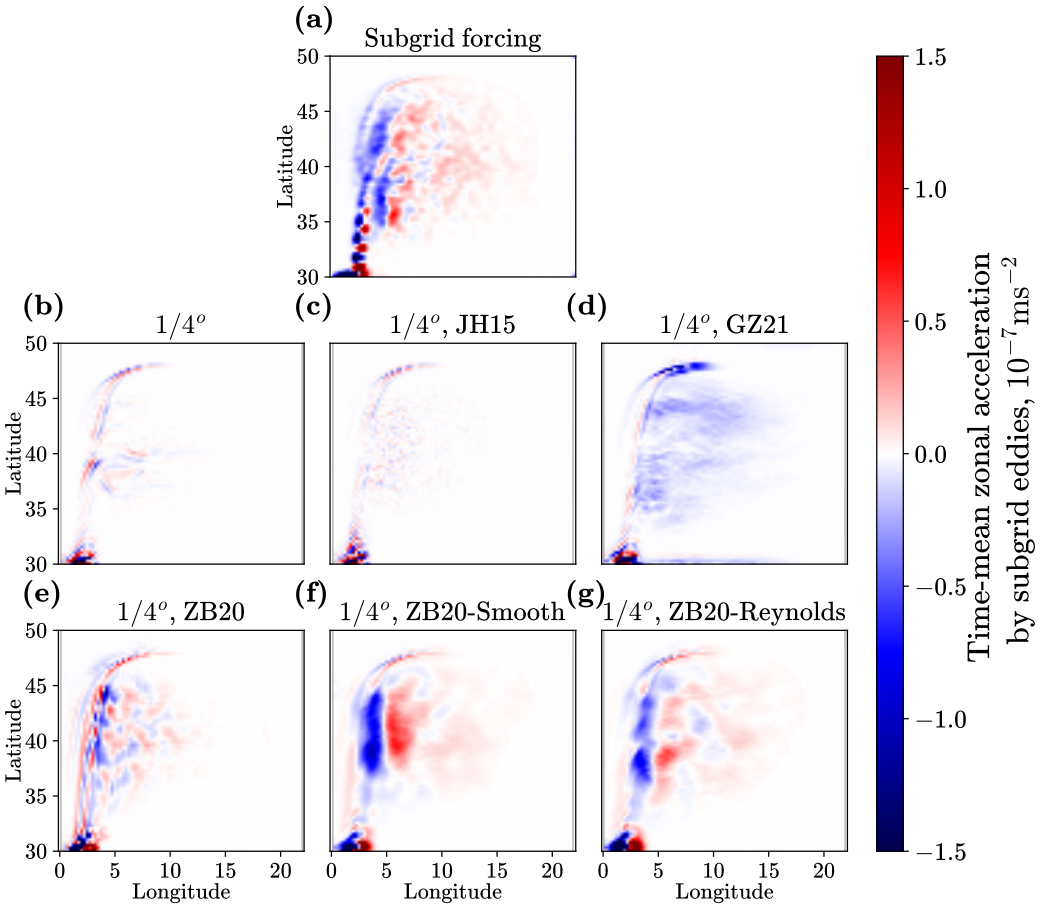

Figure 6. Analysis of the eddy-mean flow interaction following Hughes and Ash (2001): the 10-year mean zonal acceleration in the upper layer produced by the subgrid eddies ($\overline{S}_x^t$). (a) Zonal component of the subgrid forcing (Eq. (4)) diagnosed from the $1/64^o$ model by filtering and coarse-graining; the joint contribution of the subgrid parameterization and biharmonic Smagorinsky model in online simulations is shown in the other panels.

(longitude $< 5^o$), see Figure 6(a). 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 JH15) is too small compared to the diagnosed subgrid forcing (Figure 6(b,c)), 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 6(d)). 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 (Figure 6(f,g)). The zonal acceleration produced by the unfiltered ZB20 parameterization is smaller and less accurate compared to the filtered models (Figure 6(e)). The success of the ZB20-based parameterizations in improving the mean flow and the APE appears to be related to their effect on the time-mean zonal acceleration (Anstey & Zanna, 2017).

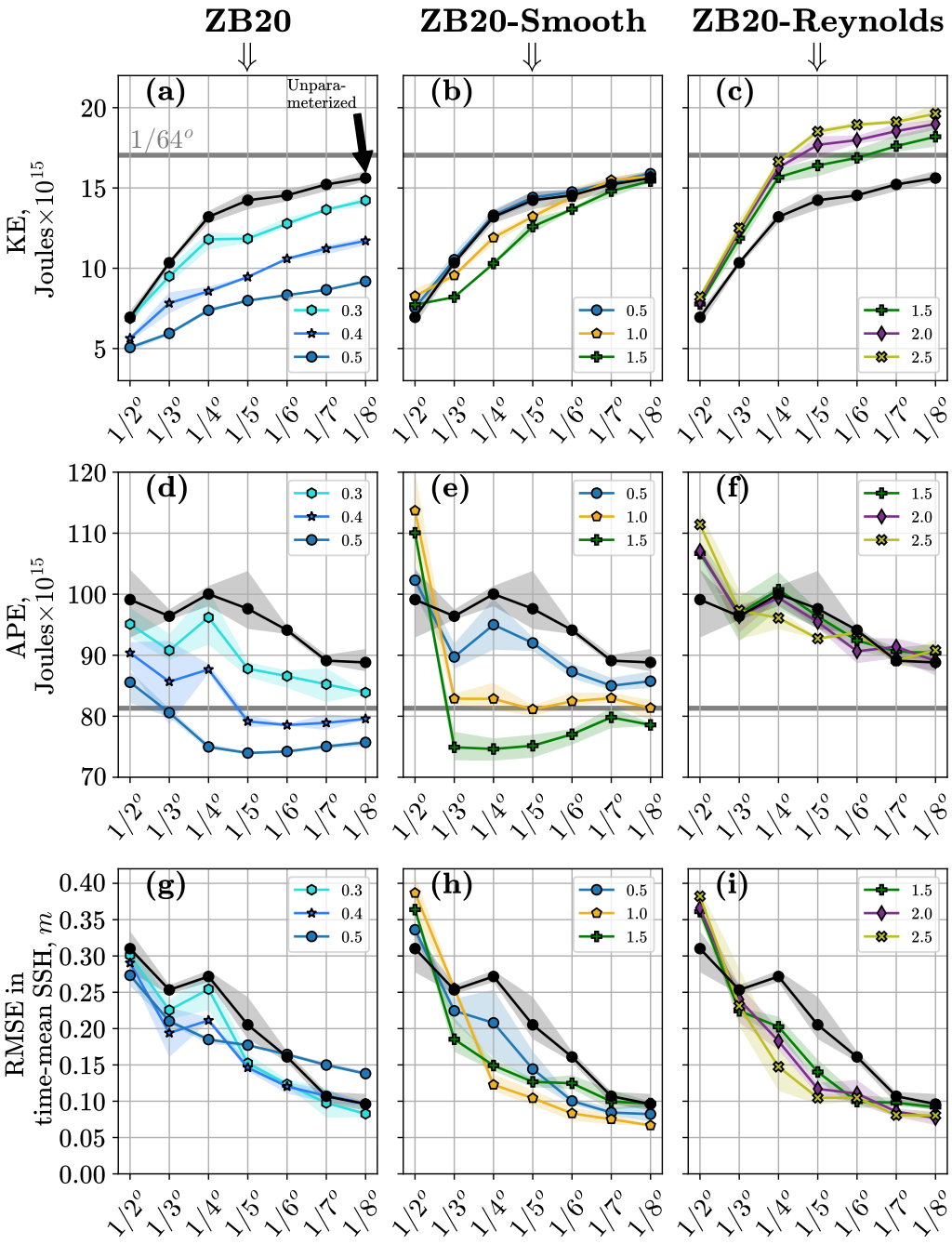

Figure 7. Sensitivity to the scaling coefficient γ (shown in legend). Upper row: the time-mean kinetic energy (KE), middle row: the time-mean available potential energy (APE), lower row: root mean squared error (RMSE) in the time-mean sea surface height (SSH). Left column shows unfiltered ZB20 parameterization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

5 Sensitivity study and scale awareness

The parameterization effect of mesoscale eddies, which are partially resolved, should be reduced as the grid is refined (Haidvogel et al., 2017). This can be achieved by in-

forming the parameterization with the local grid spacing, often referred to as scale-aware parameterizations in the literature (Bachman et al., 2017; Pearson et al., 2017). Note that the free parameter of the ZB20 parameterization, which we use here (Eq. (7)), is scale-aware and controlled by a single non-dimensional coefficient. To quantify the scale awareness, we test the parameterizations in a hierarchy of seven simulations differing in grid resolution.

5.1 Sensitivity to the scaling coefficient

We run experiments with the ZB20-based parameterizations for a range of resolutions from $1/2^o$ to $1/8^o$ 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 high-resolution model $1/64^o$ coarse-grained to the coarse grid. We split the last 15 years of the simulation of the coarse models into three 5-year segments. For every metric, we compute the mean over the segments with confidence intervals provided by the min/max values over the segments.

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

The impact on the kinetic and potential energy for all three subgrid parameterizations is proportional to the scaling coefficient (Figure 7, upper and middle rows). Also, 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, making the developed filtered ZB20 parameterizations scale-aware. An overshoot in some metrics at resolution $1/2^o$ (Figure 7(e)) happens because the mesoscale eddies fall below the grid scale of the coarse model, and thus the main assumption of the LES model, the grid scale corresponds to the inertial range, is violated. In Appendix C we discuss the sensitivity to the filter parameter (N) and to the Smagorinsky coefficient (C_S).

5.2 Comparison to the baseline parameterizations

We compare the proposed subgrid parameterizations (ZB20-Smooth and ZB20-Reynolds) with the default parameter γ to the baseline subgrid parameterizations of JH15 and GZ21. The subgrid parameterizations ZB20-Reynolds, GZ21 and JH15 are equally efficient in energizing the resolved flow, i.e. they parameterize the KE backscatter (Figure 8(a)) but have similar drawbacks. At the lowest resolutions ($1/2^o$−$1/3^o$), the ZB20-Reynolds and JH15 parameterizations underestimate the KE. At the highest resolutions ($1/5^o$−$1/8^o$), the ZB20-Reynolds and JH15 parameterizations slightly overestimate the KE without retuning; note that the GZ21 was tuned at every resolution to reproduce the KE exactly (Zhang et al., 2023). These three backscattering parameterizations have only a slight impact on the APE (Figure 8(b)). The ZB20-Smooth parameterization demonstrates the best representation of the APE for a range of resolutions $1/3^o$−$1/8^o$ (Figure 8(b)) and the best representation of the mean flow for a range of resolutions $1/4^o$−$1/8^o$ (Figure 8(c)).

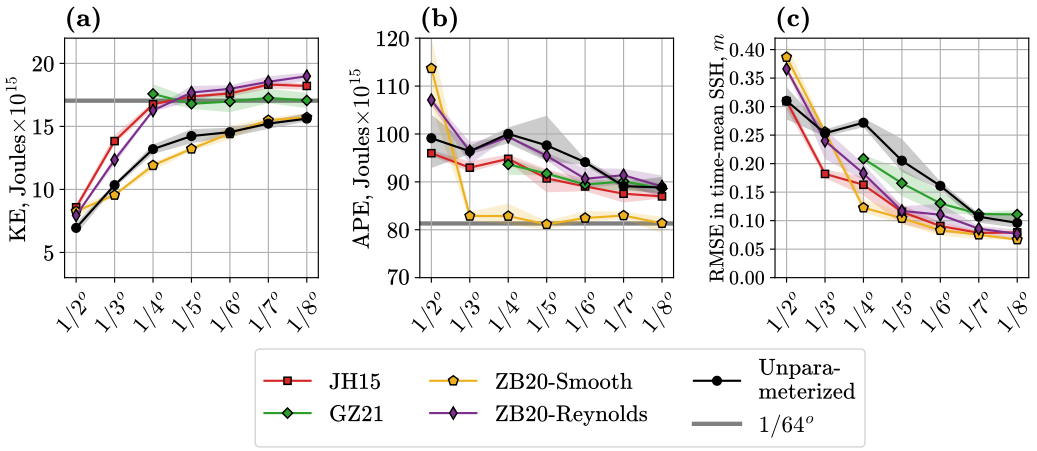

Figure 8. Similar to Figure 7, but comparison of the ZB20 parameterizations to the baselines of Jansen et al. (2015) and Guillaumin and Zanna (2021). The scaling coefficient γ is set to default: 1.0 for ZB20-Smooth and 2.0 for ZB20-Reynolds.

6 Experiments in NeverWorld2 configuration

In this section, we analyze the impact of the different subgrid parameterizations in a considerably more complicated configuration of the MOM6 ocean model – the NeverWorld2 (NW2, Marques et al., 2022). This configuration mimics the global ocean and spans the latitudes from $60^o S$ to $60^o N$, including the equator with topography mimicking an idealized Mid-Atlantic ridge and a Drake Passage. The number of fluid layers is 15. Compared to the Double Gyre configuration, the NW2 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 $1/4^o$ to $1/32^o$ (Marques et al., 2022).

6.1 Numerical issues and tuning

Preliminary experiments with the proposed ZB20 parameterizations demonstrated numerical instabilities accompanied by various runtime errors: ocean velocities are too large, interface height drops 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 ($h_k \approx 10^{-2}$m). To address this numerical instability, we applied two strategies:

- Switching the discretization of the ZB20 parameterization from the energy-conserving (Eq. (A10)) used in the Double Gyre to a non-conserving one (Eq. (A12)). The non-conservative numerical scheme presumably introduces fewer aliasing errors because the multiplication operation follows the interpolation but not vice versa.
- Attenuating the parameterization in regions of geostrophically unbalanced flows. Similarly to Klöwer et al. (2018); Juricke et al. (2019), we introduce the following attenuation function:

$$\left(1 + \frac{\sqrt{D^2 + \widetilde{D}^2 + \zeta^2}}{|f|}\right)^{-1}. \tag{12}$$

The ZB20 stress tensor is multiplied by the attenuation function right before computing its divergence. The inclusion of ζ^2 into Eq. (12) is a proposed modifica-

tion, which was found to improve the numerical stability. The expression $D^2 + \widetilde{D}^2 + \zeta^2$ is proportional to the hydrostatic stress of the ZB20 parameterization, and thus the attenuation allows to bound 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 Yankovsky et al. (2023). The difference between the two approaches is little: we attenuate the parameterization smoothly while they shut off the backscatter abruptly; we use the inverse Coriolis parameter as a time scale while they use the time step (that is, the stability criteria).

We further note that the Smagorinsky coefficient in the coarse model was changed from the value $C_S = 0.2$ used in the original work (Marques et al., 2022) to our default value $C_S = 0.06$. In preliminary experiments with NW2 at resolution $1/4^o$, we noted that the ZB20 parameterization (without filters) could not be tuned to reproduce the KE of the filtered high-resolution model. The increase in the KE is only $\approx 25\%$ for $\gamma = 0.3$ (not shown). The ZB20-Smooth and ZB20-Reynolds parameterizations were tuned to reproduce the KE of the filtered high-resolution model and the corresponding scaling coefficient for both of them is $\gamma = 2.5$.

To sum up, compared to the Double Gyre case, we changed the discretization scheme, introduced the attenuation function, and increased the scaling coefficient. We postpone the development of the parameterization which works in various configurations with a single set of parameters for future studies.

6.2 Results

We run the coarse parameterized and unparameterized models at resolution $1/4^o$ for 30000 days from rest. The high-resolution simulation $1/32^o$ was spun up in multiple stages in Marques et al. (2022). We use snapshots from the last 100 days and time-mean fields over the last 1000 days.

In Figure 9, we compare the coarse models to the high-resolution simulation (black line) and its filtered versions (light blue and pink lines). The filter widths for filtering the $1/32^o$ model are chosen as 1^o and $1/2^o$. That is, we assume that the effective resolution of the coarse models is lower than the nominal resolution given by the coarse grid (Skamarock, 2004; Soufflet et al., 2016). Note that compared to the Double Gyre case, the topography is non-convex, which results in the drop of the interfaces below the topography when the filter is applied near the bottom. We expect that the APE of the filtered $1/32^o$ model should be lower than the APE of the unfiltered solution. However, the APE of the filtered solution was found to be sensitive to the algorithm processing such events.

The unparameterized model ($1/4^o$) has a KE that is approximately 4 times lower than the KE of the $1/32^o$ model and 3 times lower than the KE of the filtered $1/32^o$ model (Figure 9(a)). The APE of the unparameterized model is too high compared to the high-resolution simulation (Figure 9(b)). 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 (Figure 9(a,b)). The analysis of the EKE spectrum at the surface reveals an increase of the energy density on large scales (wavenumbers $\kappa < 5{\cdot}10^{-2}\text{km}^{-1}$) in parameterized models compared to the unparameterized model (Figure 9(c)). The energy density on smaller scales remains mostly unchanged.

According to the analysis of the EKE spectrum shape in the small scales, all the coarse models are closest to the $1/32^o$ model filtered with the filter width $1/2^o$ (Figure 9(c)). However, the KE of the parameterized models matches better the $1/32^o$ model

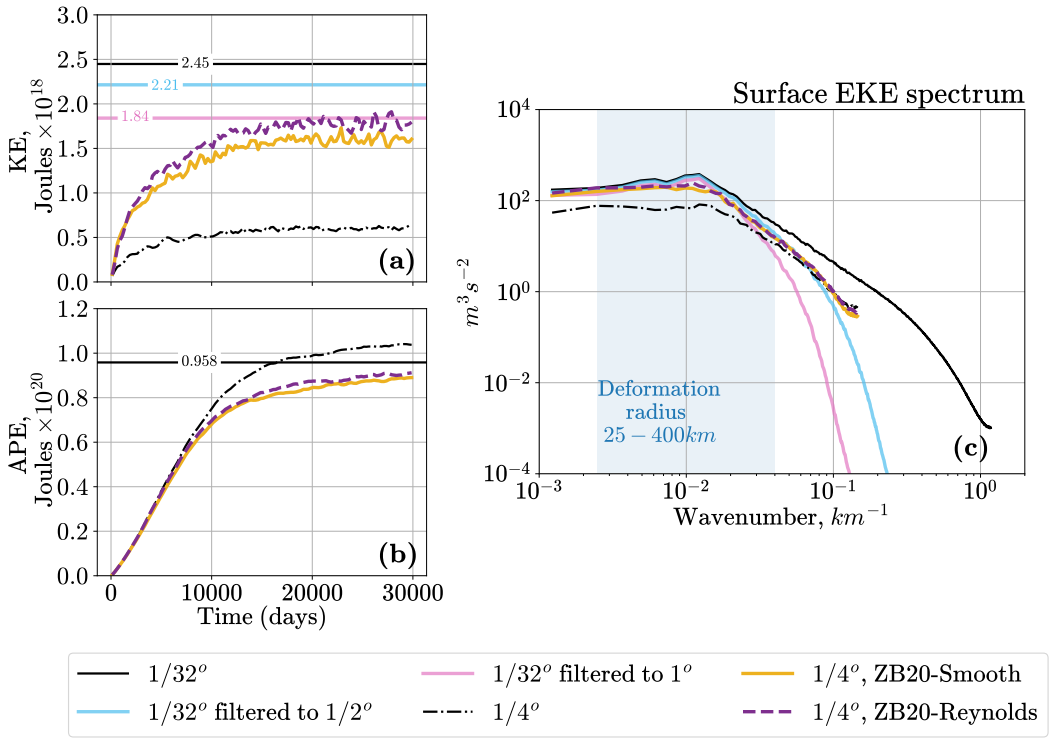

Figure 9. NeverWorld2 configuration. Time series of the (a) kinetic energy (KE) and (b) available potential energy (APE). The time series in $1/4^o$ runs are smoothed in time with a window size of 250 days while in $1/32^o$ run we provide average values over the last 100 days. (c) The EKE zonal spectrum at the surface averaged in time and latitudes. The eddy velocity is defined as a deviation from the 1000-days mean.

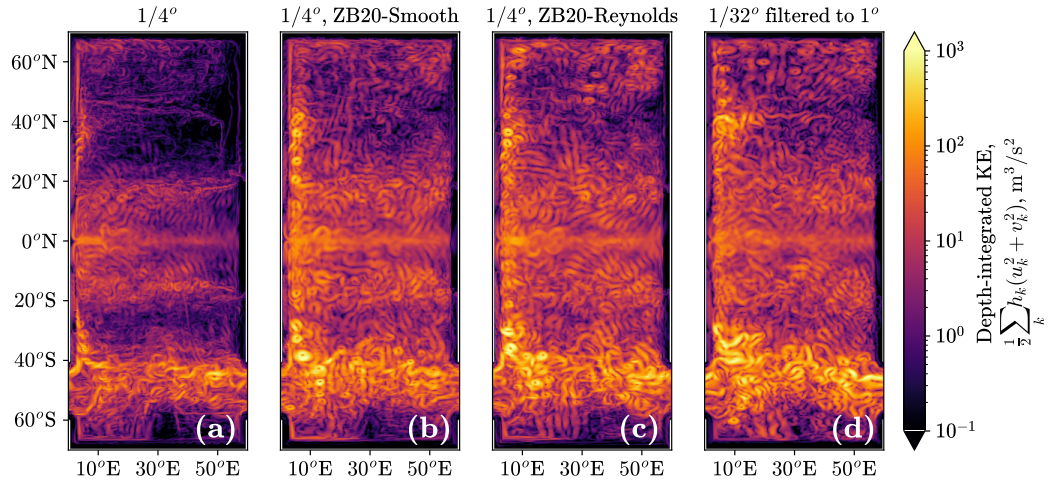

Figure 10. Snapshot of the depth-integrated kinetic energy. The coarse models at resolution $1/4^o$: (a) unparameterized model (biharmonic Smagorinsky), (b) and (c) ZB20 parameterizations. In panel (d) we show the filtered high-resolution model $1/32^o$ with filter width 1^o.

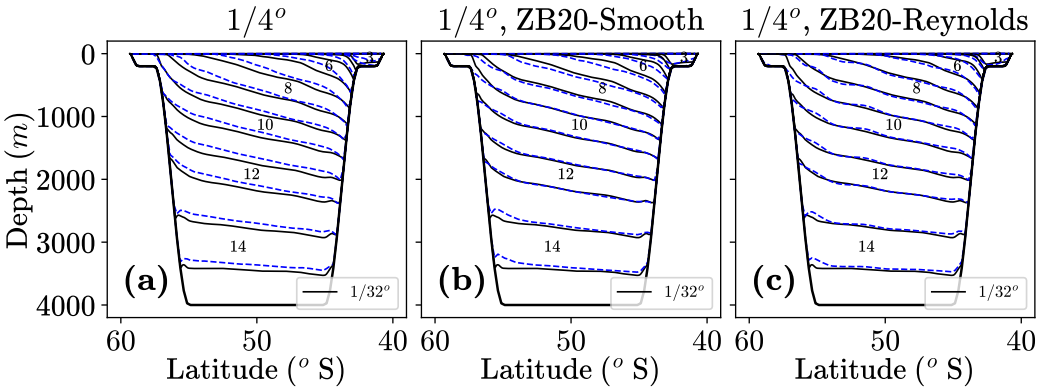

Figure 11. The time-mean interfaces in the Drake Passage (60^oS-40^oS, 0^oE) averaged over 1000 days. The blue dashed lines show the experiments at resolution $1/4^o$: (a) unparameterized model, (b) and (c) ZB20 parameterizations. Black lines show the interfaces of the high-resolution model $1/32^o$. The layer numbers (equal to 14, 12,...) are provided.

filtered with filter width 1^o (Figure 9(a)). A further increase in KE is possible but the APE will decrease even further and presumably will become unphysical. These findings reveal the difficulties in determining the effective resolution of the parameterized simulations.

The snapshots of the depth-integrated KE are shown in Figure 10. Both parameterized models (ZB20-Smooth and ZB20-Reynolds) considerably energize the eddies in all parts of the domain compared to the unparameterized model and in accordance with the high-resolution simulation. The major difference with the high-resolution simulation is the presence of strong coherent eddies near the western boundary in parameterized models (Figure 10(b,c)). Presumably, the attenuation function (Eq. (12)) was needed to bound the growth of these and similar eddies. However, we mention that it is difficult to track how a particular eddy contributes to the development of numerical instabilities.

The largest contribution to the APE is from the isopycnal slopes in the Drake Passage (60^oS-40^oS). In Figure 11, we show the layer interfaces between isopycnal layers in this region. The mesoscale eddies extract the potential energy from the mean flow and act to flatten the isopycnals. The coarse unparameterized model resolves the mesoscale eddies poorly, and consequently, its isopycnals are too steep (Figure 11(a)). Both parameterizations (ZB20-Smooth and ZB20-Reynolds) result in reducing the potential energy and better reproducing of the isopycnals in the Drake Passage (Figure 11(b,c)). The ZB20-Reynolds parameterization is slightly more accurate in the upper layers, while the ZB20-Smooth is slightly more accurate near the northern boundary of the Drake Passage.

7 Conclusions and discussion

In this work, we implemented the data-driven mesoscale eddy parameterization 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 captures the KE backscatter and supplements the upscale energy cascade. Our main findings are as follows:

- The proposed filtering schemes reduce the generation of numerical noise by ZB20 parameterization and enhance the large-scale backscatter.
- The filtered ZB20 parameterizations improve the representation of the energy distributions and the climatological mean flow compared to the current state-of-the-art energy backscatter parameterizations (Jansen et al., 2015; Guillaumin & Zanna, 2021).
- The filtered ZB20 parameterizations are scale-aware. A single non-dimensional parameter $\gamma \approx 1$ can be used for a range of resolutions without retuning.
- The leading order effect of mesoscale eddies depends on the ocean configuration: in the NW2 configuration they significantly energize the flow while in Double Gyre they impose a strong mean subgrid stress.
- The proposed parameterization ZB20-Smooth is able to account for this leading order effect in both ocean configurations and thus demonstrates the most universal properties in presented experiments.

Our methodology paves the way for researchers and ocean modelers to implement and test subgrid ML 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 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. The filtering schemes remain a popular tool in the development of physics-based subgrid models (Juricke et al., 2020; Mak et al., 2023; Grooms, 2023; Bagaeva et al., 2023) and here we demonstrated their utility for ML-based parameterizations. We further learned that testing various discretizations of the subgrid model offers an opportunity to find a numerically stable scheme which is necessary to run online simulations.

While applying the filtering schemes allows us to improve the physical and numerical properties of the ZB20 parameterization, a number of challenges remain. (1) It is difficult to find a single filtering scheme that improves the mean flow and kinetic/potential energies simultaneously. (2) The proposed filtering approach is applied and tuned a posteriori; how to learn it directly from data is an open question. (3) The filtered ZB20 parameterizations do not affect the grid scales, so a supplementary subgrid model that is effective for short waves is needed. In this work, the biharmonic Smagorinsky model plays this role. (4) Without filters, the ZB20 parameterization is computationally cheap ($1-3\%$ of total runtime). However, the filters can increase the computational cost to $5-10\%$ of total runtime.

Possible future improvements to the proposed parameterization include: (1) Including spatial non-locality (P. Wang et al., 2022) during the training process to potentially reduce the need for a posteriori tuning. (2) Using a neural network to improve the prediction of the subgrid stress having the same input features. (3) Informing a subgrid model with local physical parameters to improve generalization to unseen flow regimes, in the spirit of scale-aware parameterizations (Hallberg, 2013; Bachman et al., 2017; Jansen et al., 2019). (4) Imposing physical laws as hard constraints to enhance generalization, as in physics-aware parameterizations (Guan, Subel, et al., 2022; Pawar et al., 2023). For example, in this work we found that constraining the subgrid model to predict the subgrid stress through an interpretable expression streamlines the generalization to ocean configurations with complex coordinates and topography.

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

Appendix A Curvilinear coordinates, varying layer thickness, numerical schemes and boundary conditions

We modify the original parameterization (Eqs. (5), (6)) 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, $\widetilde{D}$, ζ). The computation of these gradients depends on the coordinate system. In generalized curvilinear orthogonal coordinates the expressions for D and $\widetilde{D}$ must be changed to (Appendix A.b in S. M. Griffies and Hallberg (2000) and Section 17.10.2 in S. Griffies (2018)):

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

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

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 Eq. (A3) for simplicity.

A2 Divergence of momentum flux

Following previous work on viscous operators in ocean models (Eqs. (A3) and (A4) in S. M. Griffies and Hallberg (2000) and Section 17.10.3 in S. Griffies (2018)), we modify the divergence of the stress tensor (Eq. (5)):

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

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, i.e. 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:

$$\partial_t \int h\mathbf{u} dxdy = \cdots + \int h \partial_t \mathbf{u} dxdy = \cdots + \int h\mathbf{S} dxdy = \cdots + \tag{A5}$$

$$\cancelto{0}{\int \nabla \cdot (h\mathbf{T}) dxdy}. \tag{A6}$$

A3 Numerical discretization

The rate of change of the KE due to parameterization (Eq. (A4)) up to the boundary fluxes is given by:

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

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

where $\nabla\mathbf{u}$ is the velocity gradient (VG) tensor and (:) is the tensor contraction over two indices (Eyink, 1995). In Cartesian coordinates the components of VG tensor have a simple form $(\nabla\mathbf{u})_{ij} = \partial_j u_i$. The energy contribution from the deviatoric component of $\mathbf{T}$ (Eq. (6)) is zero because it is orthogonal to the VG tensor:

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

However, the Eq. (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 $\widetilde{D}$ are defined in the center of the grid cell. We propose the following energy-conserving discretization of the deviatoric stress:

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

where $\overline{(\cdot)}^{x,y}$ is a linear interpolation from corner to center or vice versa. Terms in the RHS of Eq. (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, i.e. 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 hydrostatic part of $\mathbf{T}$ (Eq. (6)), we use the simplest numerical scheme because no properties are known:

$$\left((\overline{\zeta}^{x,y})^2 + (\overline{D}^{x,y})^2 + \widetilde{D}^2\right)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{A11}$$

The energy-conserving discretization (Eq. (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:

$$\begin{bmatrix} -\overline{\zeta}^{x,y}\overline{D}^{x,y} & \zeta\overline{\widetilde{D}}^{x,y} \\ \zeta\overline{\widetilde{D}}^{x,y} & \overline{\zeta}^{x,y}\overline{D}^{x,y} \end{bmatrix}. \tag{A12}$$

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, it 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:

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

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

$$\text{PE} = \frac{1}{2}\sum_k \int \rho_0 g'_{k+1/2} \eta^2_{k+1/2} dxdy. \tag{B2}$$

The available potential energy (APE) is the PE minus the potential energy of the resting state given by $\eta^{\text{ref}}_{k+1/2} = \max(z^0_{k+1/2}, -H(x,y))$, 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). Mean potential energy (MPE) is the APE of the mean flow $\overline{\eta}^t_{k+1/2}$, and eddy potential energy (EPE) is given by EPE $= \overline{\text{APE}}^t -$ 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 C1(b)). However, the undesirable impact on the KE is also stronger (Figure C1(a)). Considering the error in SSH, the optimal number of filters is $N = 2$ for coarse resolutions ($1/2^o - 1/3^o$) and the optimal number is $N = 4$ for higher resolutions (Figure C1(c)). For the ZB20-Reynolds parameterization, the impact on the KE is higher for higher N (Figure C1(d)). An impact on the SSH metric is also more significant for larger N (Figure C1(f)). 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 the inclusion of 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 run in a numerically stable way for all three parameterizations for a range of scaling coefficients γ. However, the optimal SSH metric is achieved when ZB20-based parameterizations are turned off, i.e. $\gamma = 0$ when $C_S = 0$. This demonstrates that the ZB20-based parameterizations describe only part of the subgrid forcing and they cannot be used without the eddy viscosity. Another important observation – the optimal scaling coefficient γ should be increased when the eddy viscosity coefficient C_S is increased. Finally, the bias in SSH can be reduced approximately twice for the ZB20-Smooth and ZB20-Reynolds parameterizations for all non-zero values of the Smagorinsky coefficient ($C_S = 0.03, 0.06, 0.09$).

Open Research

The NW2 configuration is available via Bhamidipati et al. (2022), for reference simulations in NW2 see Marques, Gustavo (2022). The MOM6 source code (version used to conduct the research) with implemented ZB20 parameterization, configuration files for Double Gyre and NW2, and figure plotting functions is available via Perezhogin (2023b). The selected simulation data is available via Perezhogin (2023a).

Acknowledgments

This research was made possible by Schmidt Futures, a philanthropic initiative founded by Eric and Wendy Schmidt, as part of its Virtual Earth System Research Institute (VESRI).

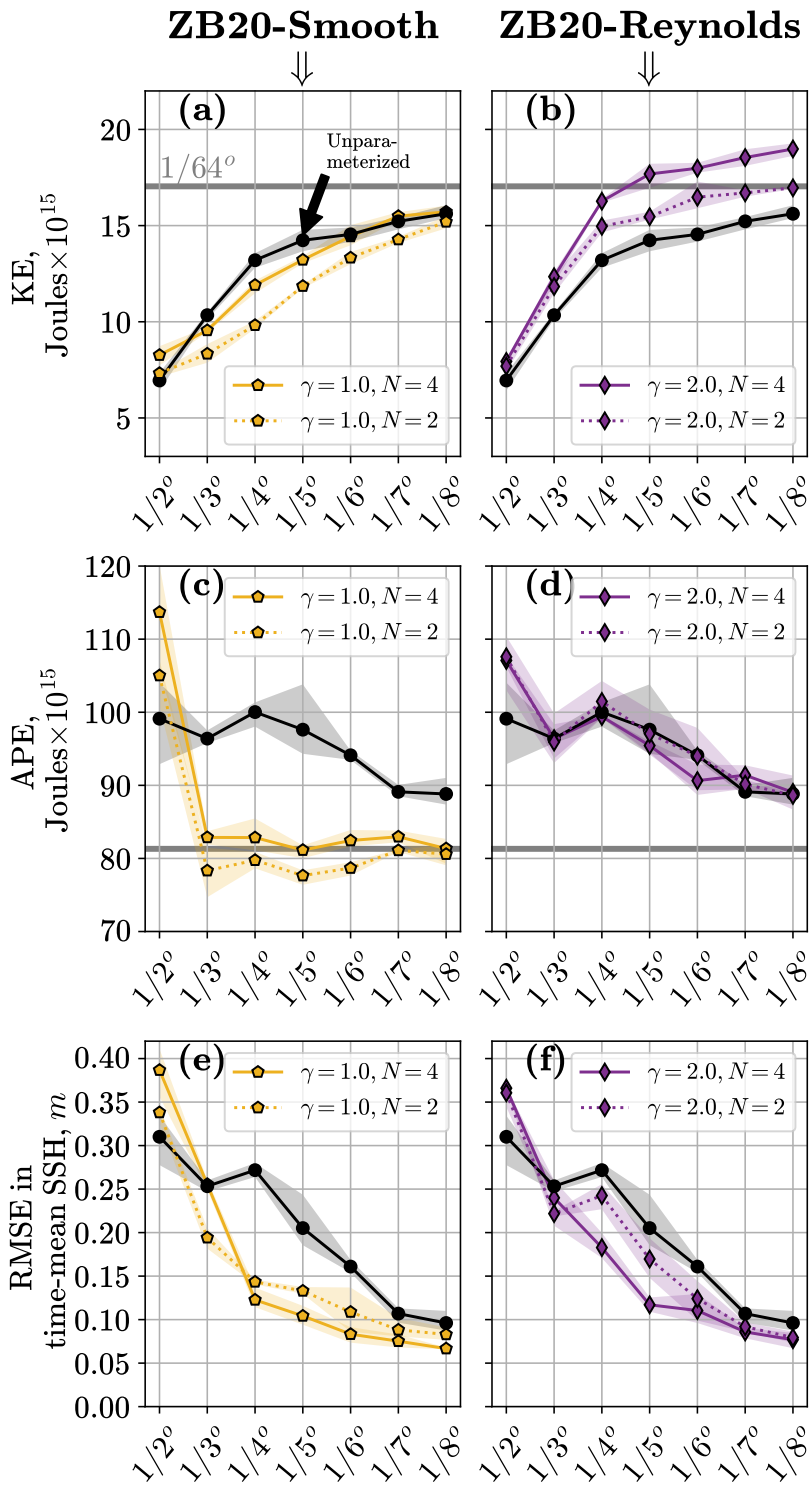

Figure C1. Similar to Figure 7, but sensitivity to the number of the filter passes N in ZB20 parameterizations. Solid lines show the default value of the filter passes ($N = 4$) and dotted lines show a smaller value $N = 2$.

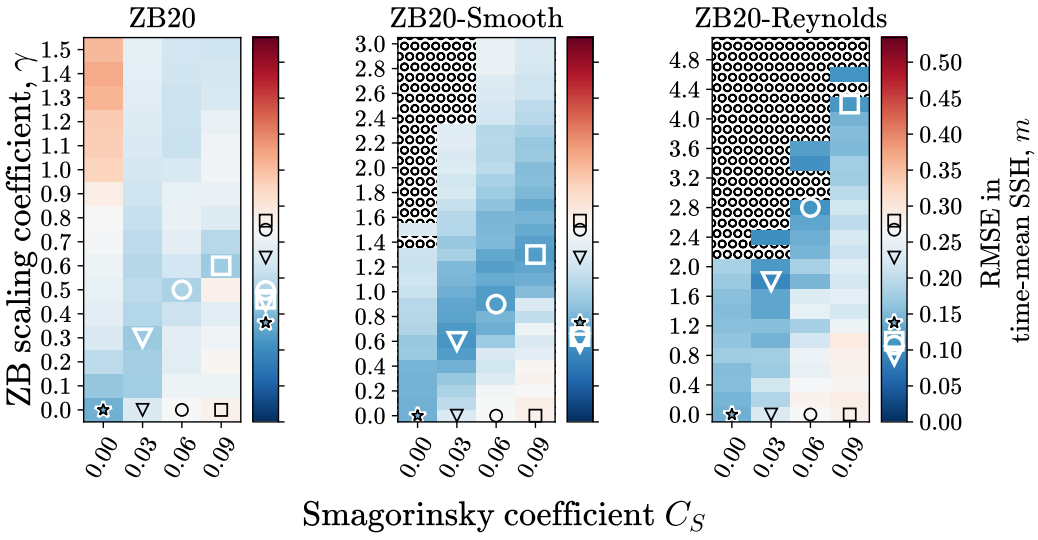

Figure C2. The sensitivity to the Smagorinsky coefficient C_S and scaling coefficient γ in ZB20-based parameterizations. The considered metric is the root mean squared error (RMSE) in the representation of the time-mean sea surface height (SSH). The ocean grid resolution is $1/4^o$. Every small box shows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

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 No. 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; and Elizabeth Yankovsky for help with setting up the backscatter parameterization and NW2 simulations; and Robert Hallberg and Marshall Ward for help with the MOM6 code.

References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., ... others (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.

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 Atmospheric Sciences*, *33*(11), 2031–2048.

Anstey, J. A., & Zanna, L. (2017). A deformation-based parametrization of ocean mesoscale eddy reynolds stresses. *Ocean Modelling*, *112*, 99–111.

Bachman, S. D. (2019). The gm+ e closure: A framework for coupling backscatter with the gent and mcwilliams parameterization. *Ocean Modelling*, *136*, 85–106.

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.

Bagaeva, E., Danilov, S., Oliver, M., & Juricke, S. (2023). Advancing eddy parameterizations: Dynamic energy backscatter and the role of subgrid advection and stochastic forcing. *ESS Open Archive*. doi: 10.22541/essoar.169288735.50187970/v1

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.

Berloff, P. (2018). Dynamically consistent parameterization of mesoscale eddies. part iii: Deterministic approach. *Ocean Modelling*, *127*, 1–15.

Berner, J., Shutts, G., Leutbecher, M., & Palmer, T. (2009). A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ecmwf ensemble prediction system. *Journal of the Atmospheric Sciences*, *66*(3), 603–626.

Bhamidipati, N., Adcroft, A., Marques, G., & Abernathey, R. (2022, August). *ocean-eddy-cpt/NeverWorld2: NeverWorld2-description-paper*. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.6993951` doi: 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. doi: 10.1029/2018MS001472

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

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.

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.

Chow, F. K., & Moin, P. (2003). A further study of numerical errors in large-eddy simulations. *Journal of Computational Physics*, *184*(2), 366–380.

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

Eyink, G. L. (1995). Local energy flux and the refined similarity hypothesis. *Journal of Statistical Physics*, *78*, 335–351.

Ferrari, R., & Wunsch, C. (2009). Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, *41*, 253–282.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., ... others (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.

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.

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.

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.

Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, *20*(1), 150–155.

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.

Ghosal, S. (1996). An analysis of numerical errors in large-eddy simulations of turbulence. *Journal of Computational Physics*, *125*(1), 187–206.

Greatbatch, R., Zhai, X., Claus, M., Czeschel, L., & Rath, W. (2010). Transport driven by eddy momentum fluxes in the gulf stream extension region. *Geophysical Research Letters*, *37*(24).

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.

Greatbatch, R. J., Zhai, X., Kohlmann, J.-D., & Czeschel, L. (2010). Ocean eddy momentum fluxes at the latitudes of the gulf stream and the kuroshio extensions as revealed by satellite data. *Ocean Dynamics*, *60*, 617–628.

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.

Griffies, S. (2018). *Fundamentals of ocean climate models*. Princeton university press.

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.

Grooms, I. (2023). Backscatter in energetically-constrained leith parameterizations. *Ocean Modelling*, *186*, 102265.

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

Grooms, I., Loose, N., Abernathey, R., Steinberg, J., Bachman, S. D., Marques, G., ... Yankovsky, E. (2021). Diffusion-based smoothers for spatial filtering of gridded geophysical data. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002552.

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.

Guan, Y., Subel, A., Chattopadhyay, A., & Hassanzadeh, P. (2022). Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate les. *Physica D: Nonlinear Phenomena*, 133568.

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

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.

Hallberg, R. (2013). Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, *72*, 92–103.

Hewitt, H. T., Roberts, M., Mathiot, P., Biastoch, A., Blockley, E., Chassignet, E. P., ... others (2020). Resolving and parameterising the ocean mesoscale in earth system models. *Current Climate Change Reports*, *6*(4), 137–152.

Hoskins, B. J., James, I. N., & White, G. H. (1983). The shape, propagation and mean-flow interaction of large-scale weather systems. *Journal of Atmospheric Sciences*, *40*(7), 1595–1612.

Hughes, C. W., & Ash, E. R. (2001). Eddy forcing of the mean flow in the southern ocean. *Journal of Geophysical Research: Oceans*, *106*(C2), 2713–2722.

Jakhar, K., Guan, Y., Mojgani, R., Chattopadhyay, A., Hassanzadeh, P., & Zanna, L. (2023). Learning closed-form equations for subgrid-scale closures from high-fidelity data: Promises and challenges. *arXiv preprint arXiv:2306.05014*.

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.

Jansen, M. F., & Held, I. M. (2014). Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling*, *80*, 36–48.

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.

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.

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.

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.

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.

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.

Kraichnan, R. H. (1976). Eddy viscosity in two and three dimensions. *Journal of Atmospheric Sciences*, *33*(8), 1521–1536.

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.

Loose, N., Abernathey, R., Grooms, I., Busecke, J., Guillaumin, A., Yankovsky, E., ... others (2022). Gcm-filters: A python package for diffusion-based spatial filtering of gridded data. *Journal of Open Source Software*, *7*(70).

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.

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). France: Institut Pierre-Simon Laplace (IPSL).

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.

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. *arXiv preprint arXiv:2306.08988*.

Mana, P. P., & Zanna, L. (2014). Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling*, *79*, 1–20.

Marques, G. M., Loose, N., Yankovsky, E., Steinberg, J. M., Chang, C.-Y., Bhamidipati, N., ... others (2022). Neverworld2: An idealized model hierarchy to investigate ocean mesoscale eddies across resolutions. *Geoscientific Model Development*, *15*(17), 6567–6579.

Marques, Gustavo. (2022). *Neverworld2.* UCAR/NCAR - CISL - CDP. Retrieved from `https://www.earthsystemgrid.org/dataset/id/687ed2d9-447c-4f7e-b0f6-0c3214239899.html` doi: 10.26024/F130-EV71

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.

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

Meneveau, C., & Katz, J. (2000). Scale-invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics*, *32*(1), 1–32.

Moser, R. D., Haering, S. W., & Yalla, G. R. (2021). Statistical properties of subgrid-scale turbulence models. *Annual Review of Fluid Mechanics*, *53*, 255–286.

Pawar, S., San, O., Rasheed, A., & Vedula, P. (2023). Frame invariant neural network closures for kraichnan turbulence. *Physica A: Statistical Mechanics and its Applications*, *609*, 128327.

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.

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.

Perezhogin, P. (2023a). *Dataset for paper on implementation of zanna-bolton-2020 subgrid model to mom6 gfdl ocean model.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.10045769` doi: 10.5281/zenodo.10045769

Perezhogin, P. (2023b). *m2lines/implementation-zb20: v1.0.3.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.10045624` doi: 10.5281/zenodo.10045624

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

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. doi: 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.

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.

Redi, M. H. (1982). Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanography*, *12*(10), 1154–1158.

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.

Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction.* Springer Science & Business Media.

Salmon, R. (1980). Baroclinic instability and geostrophic turbulence. *Geophysical & Astrophysical Fluid Dynamics*, *15*(1), 167–211.

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.

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. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003890` doi: 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.

Skamarock, W. C. (2004). Evaluating mesoscale nwp models using kinetic energy spectra. *Monthly weather review*, *132*(12), 3019–3032.

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.

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.

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.

Uchida, T., Abernathey, R., & Smith, S. (2017). Seasonality of eddy kinetic energy in an eddy permitting global climate model. *Ocean Modelling*, *118*, 41–58.

Uchida, T., Rokem, A., Soler, S., Nicholas, T., Abernathey, R., Nouguier, F., ... Moon, Z. (2023, February). *xgcm/xrft: v1.0.1.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7621857` doi: 10.5281/zenodo.7621857

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press.

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.

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.

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.

Wardle, R., & Marshall, J. (2000). Representation of eddies in primitive equation models by a pv flux. *Journal of physical oceanography*, *30*(10), 2481–2503.

Waterman, S., & Jayne, S. R. (2011). Eddy-mean flow interactions in the alongstream development of a western boundary current jet: An idealized model study. *Journal of Physical Oceanography*, *41*(4), 682–707.

Yankovsky, E., Bachman, S., Smith, K. S., & Zanna, L. (2023). Backscatter parameterization of ocean mesoscale eddies informed by vertical structure. *in prep*.

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.

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.

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.

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376.

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. doi: https://doi.org/10.1029/2023MS003697