Cloud-based framework for inter-comparing submesoscale permitting realistic ocean models

Takaya Uchida[1], Julien Le Sommer[1], Charles Stern[2], Ryan P. Abernathey[2], Chris Holdgraf[3], Aurélie Albert[1], Laurent Brodeau[4,5], Eric P. Chassignet[6], Xiaobiao Xu[6], Jonathan Gula[7,8], Guillaume Roullet[7], Nikolay Koldunov[9], Sergey Danilov[9], Qiang Wang[9], Dimitris Menemenlis[10], Clément Bricaud[11], Brian K. Arbic[12], Jay F. Shriver[13], Fangli Qiao[14], Bin Xiao[14], Arne Biastoch[15,16], René Schubert[7,15], Baylor Fox-Kemper[17], William K. Dewar[1,18], and Alan Wallcraft[6]

[1]Université Grenoble Alpes, CNRS, IRD, Grenoble-INP, Institut des Géosciences de l'Environnement, France
[2]Lamont-Doherty Earth Observatory, Columbia University in the City of New York, USA
[3]2i2c.org, USA
[4]Ocean Next, Grenoble, France
[5]Datlas, Grenoble, France
[6]Center for Ocean-Atmospheric Prediction Studies, Florida State University, USA
[7]Univ. Brest, CNRS, Ifremer, IRD, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, 29280, Plouzané, France
[8]Institut Universitaire de France (IUF), Paris, France
[9]Alfred Wegener Institute (AWI), Helmholtz Center for Polar and Marine Research, Germany
[10]Jet Propulsion Laboratory, National Aeronautics and Space Administration (NASA), USA
[11]Mercator Ocean International, France
[12]Department of Earth and Environmental Sciences, University of Michigan, USA
[13]Oceanography Division, US Naval Research Laboratory, USA
[14]First Institute of Oceanography, and Key Laboratory of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao, China
[15]GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Germany
[16]Kiel University, Kiel, Germany
[17]Department of Earth, Environmental, and Planetary Sciences, Brown University, USA
[18]Department of Earth, Ocean and Atmospheric Science, Florida State University, USA

Correspondence: Takaya Uchida (takaya.uchida@univ-grenoble-alpes.fr)

Abstract. With the increase in computational power, ocean models with kilometer-scale resolution have emerged over the last decade. These models have been used for quantifying the energetic exchanges between spatial scales, informing the design of eddy parametrizations and preparing observing networks. The increase in resolution, however, has drastically increased the size of model outputs, making it difficult to transfer and analyze the data. It remains, nonetheless, of primary importance to assess more systematically the realism of these models. Here, we showcase a cloud-based analysis framework proposed by the Pangeo Project that aims to tackle such distribution and analysis challenges. We analyze the output of eight submesoscale-permitting simulations, all on the cloud, for a crossover region of the upcoming Surface Water and Ocean Topography (SWOT) altimeter mission near the Gulf Stream separation. The cloud-based analysis framework: i) minimizes the cost of duplicating and storing ghost copies of data, and ii) allows for seamless sharing of analysis results amongst collaborators. We describe the framework and provide example analyses (*e.g.*, sea-surface height variability, submesoscale vertical buoyancy fluxes, and

comparison to predictions from the mixed-layer instability parametrization). Basin-to-global scale, submesoscale-permitting models are still at their early stage of development; their cost and carbon footprints are also rather large. It would, therefore, benefit the community to document the different model configurations for future best practices. We also argue that an emphasis on data analysis strategies would be crucial for improving the models themselves.

1 Introduction

Traditionally collaboration amongst multiple ocean modelling institutions and/or the reproducing of scientific results from numerical simulations required the duplication, individual sharing and downloading of data, upon which each party of interest (often an independent group) would analyze the data on their local workstation or cluster. We will refer to this as the 'download' framework (Stern et al., 2022). As realistic ocean simulations with kilometric horizontal resolution have emerged (*e.g.*, Rocha et al., 2016; Schubert et al., 2019; Brodeau et al., 2020; Gula et al., 2021; Ajayi et al., 2021), such a framework has become cumbersome with tera- and peta-bytes of data needed to be transferred and stored as ghost copies. Nevertheless, a real demand exists for collaboration to inter-compare models to examine their fidelity and quantify robust features of submeso- and meso-scale turbulence (the former on the horizontal spatial scales of $O(10\,\mathrm{km})$ and latter on $O(100\,\mathrm{km})$; here on referred to jointly as (sub)mesoscale; Hallberg, 2013; McWilliams, 2016; Lévy et al., 2018; Uchida et al., 2019; Dong et al., 2020). The Ocean Model Intercomparison Project (OMIP), for example, has been successful in diagnosing systematic biases in non-eddying and mesoscale-permitting ocean models used for global climate simulations (Griffies et al., 2016; Chassignet et al., 2020).

Here, we would like to achieve the same goal as OMIP but by inter-comparing submesoscale-permitting ocean models, which have been argued to be sensitive to grid-scale processes and numerical schemes as we increasingly push the model resolution closer to the scales of non-hydrostatic dynamics and isotropic three-dimensional (3D) turbulence (Hamlington et al., 2014; Soufflet et al., 2016; Ducousso et al., 2017; Barham et al., 2018; Bodner and Fox-Kemper, 2020). Considering the enormous computational cost and carbon emission of these submesoscale-permitting models, it would also benefit the ocean and climate modeling community to compile the practices implemented by each modeling group for future runs. In doing so, we analyze eight realistic, submesoscale-permitting ocean simulations, which cover at least the North Atlantic basin, run with the code bases of the Nucleus for European Modelling of the Ocean (NEMO; Madec et al., 2019, https://www.nemo-ocean.eu/), Coastal and Regional Ocean COmmunity model (CROCO; Shchepetkin and McWilliams, 2005, https://www.croco-ocean.org/), Massachusetts Institute of Technology general circulation model (MITgcm; Marshall et al., 1997, https://mitgcm.readthedocs.io/en/latest/), HYbrid Coordinate Ocean Model (HYCOM; Bleck, 2002; Chassignet et al., 2009, https://www.hycom.org/), Finite volumE Sea ice-Ocean Model (FESOM; Danilov et al., 2017, https://fesom2.readthedocs.io/en/latest/index.html), and First Insitute of Oceanography Coupled Ocean Model (FIO-COM, http://fiocom.fio.org.cn/). Considering the amount of data, however, the download framework becomes very inefficient. Therefore, we have implemented the 'data-proximate computing' framework proposed by the Pangeo project where we have stored the model outputs on the cloud and brought the computational resources adjacent to the data on the cloud (Abernathey et al., 2021a; Stern et al., 2022).
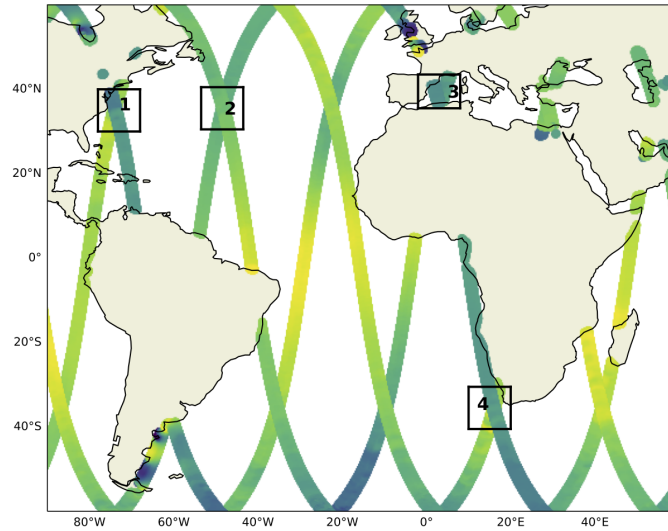
Figure 1. SWOT tracks during its calibration phase and strategic Xover regions in the Atlantic sector. The regions cover the Gulf Stream separation and its extension (Regions 1 and 2), western Mediterranean Sea (Region 3) and Agulhas Rings (Region 4).

Many of these simulations were developed ahead of the Surface Water and Ocean Topography (SWOT) satellite launch (Morrow et al., 2019), now projected to be in November 2022, in order to allow for the instrumental calibration of SWOT
45 (Gomez-Navarro et al., 2018; Metref et al., 2020), and to disentangle the internal wave signals from (sub)mesoscale flows; SWOT is expected to observe the superposed field of the two dynamics (Savage et al., 2017a; Torres et al., 2018; Yu et al., 2021). During its calibration phase, SWOT will pass over the same site every day for six months and have tracks that will cross over with each other. In order to showcase the data-proximate computing framework and its potential for collaborative, open-source and reproducible science, we provide example diagnostics for one of the SWOT Crossover (Xover) regions around
50 the Gulf Stream separation (Region 1 in Figure 1). We leave the detailed diagnostics of (sub)mesoscale flows including other SWOT-Xover regions and the potential impact of modeling numerics on the resolved flow for a subsequent paper.

The paper is organized as follows: We describe the data-proximate computing framework in section 2 and showcase some example analyses using this framework in section 3. Cautionary remarks regarding sustainability into the future for open-source reproducible science are given in section 4 and we conclude in section 5.

55 **2 Data-proximate computing framework**

In order for the data-proximate computing framework to work for collaborative, open-source and reproducible science, it requires two components to work together simultaneously: i) public access to analysis-ready data, and ii) open-source computational resources adjacent to the data.

2.1 Analysis-ready cloud-optimized data

In the field of Earth Science, model outputs are often archived and distributed in binary, HDF5 or NetCDF formats. While we have greatly benefited from these formats, they are not optimized for cloud storage nor for parallelized cloud computing. However, as Earth Scientists, commonly, we do not possess the training in cloud infrastructure nor data engineering required to efficiently convert large scale archival datasets into formats which allow us to leverage the full performance potential of the commercial cloud. Data engineers, on the other hand, do not know the scientific needs of the data. In collaboration with Pangeo Forge (Stern et al., 2022, https://pangeo-forge.readthedocs.io/en/latest/), we have therefore, attempted to fill this niche by streamlining the process of data preparation and submission. To transform their data into analysis-ready cloud optimized (ARCO) formats, data providers (ocean modeling institutions in our case) need only specify the source file location (*e.g.*, as paths on an Ftp, Http or OPeNDAP server) along with output dataset parameters (*e.g.*, particular ARCO format, chunking) in a Python module known as a *recipe*. The recipe module, which is typically a few dozen lines of Python code, relies on a data model defined in the open source `pangeo-forge-recipes` package. Once complete, the recipe is submitted via a Pull Request on Github to the Pangeo Forge `staged-recipes` repository (https://github.com/pangeo-forge/staged-recipes). From here, Pangeo Forge automates the process of converting the data into ARCO format and storing the resulting dataset on the cloud, using its own elastically-scaled cloud compute cluster. The term "analysis-ready" here is used broadly to refer to any dataset that has been preprocessed to facilitate the analysis which will be performed on it (Stern et al., 2022). An example of such recipe for eNATL60 described in section 3 is given in Appendix A. We refer the interested reader to Abernathey et al. (2021a) and Stern et al. (2022) for further details on the technical implementation.

The crowdsourcing approach of Pangeo Forge, to which any data provider can contribute, not only benefits the immediate scientific needs of a single research project, but also the entire scientific community in the form of shared, publicly accessible ARCO datasets which remain available for all to access. This saves each scientist the cost of duplicating and storing ghost copies of the data and allows for reproducible science. The model outputs used for this study are stored on the Open Storage Network (OSN), a cloud storage service provided by the National Science Foundation (NSF) in the U.S. The surface data were saved hourly and interior data in the upper $1000\,\mathrm{m}$ as daily averages (due to cloud storage constraints). To facilitate the access of data from OSN, we have further made them readable via `intake`, a data access and cataloging system which unifies the API to read and load the data (https://intake.readthedocs.io/en/latest/overview.html). Namely, the API to read and load the data is the same for all of the data used in this project, regardless of its distribution format (*e.g.*, binary, HDF5 or NetCDF), because each of the datasets has been converted by Pangeo Forge into the cloud-optimized Zarr format, and subsequently cataloged with `intake`, prior to analysis. This is particularly beneficial for our case where we would like to systematically analyze multiple data collections. The entire process of zarrifying the data, fluxing them to OSN and cataloging scaled well for the four regions shown in Figure 1. Jupyter notebooks for the results shown in section 3, including the Yaml file to access data via `intake`, are given in the Pangeo Data `swot_adac_ogcms` Github repository (https://github.com/roxyboy/swot_adac_ogcms/tree/notebook; a DOI will be added upon acceptance of the manuscript).

2.2 Cloud-based JupyterHub

For data-proximate computational resources, we have implemented a JupyterHub, an open-source platform that provides remote access to interactive sessions in the cloud for many users (Fangohr et al., 2019; Beg et al., 2021), on the Google Cloud Platform (GCP). This infrastructure is run in collaboration with 2i2c.org, a non-profit organization based in the U.S. that manages cloud infrastructure for open source scientific workflows. Authentication for each user/collaborator on the JupyterHub is provided via a white-list of Github usernames, meaning that the hub can be accessed from anywhere and is not tied directly to an institutional account. This has allowed for real-time sharing of Python scripts amongst collaborators and exchanging of feedback on the analytical results we present in section 3. Cloud computing also offers the scaling of resources for improved I/O throughput and optimization of network bandwidth and Central Processing Units (CPUs).

3 Example analyses

The model outputs used for this showcase are from the eNATL60 (Brodeau et al., 2020), GIGATL (Gula et al., 2021), HYCOM50 (Chassignet and Xu, 2017, 2021), FESOM-GS, LLC4320 (Rocha et al., 2016; Stewart et al., 2018), ORCA36 (https://github.com/immerse-project/ORCA36-demonstrator), FIO-COM32 (Xiao et al., 2022), and HYCOM25 (Savage et al., 2017a, b; Arbic et al., 2018) simulations. The detailed configuration of each simulation is given in Appendix B. In order to motivate the reader on the necessity of inter-comparing realistic submesoscale-permitting simulations, we show in Figure 2 the relative vorticity normalized by the local Coriolis parameter on February 1, 00:00 from each model. Despite their similar spatial resolutions, the spatial scales represented vary widely across models. Submesoscale-permitting ocean modeling is in its early stage of development, and each modeling institution is still exploring best practices. Therefore, we did not specify an experimental protocol, as in OMIP, for the model outputs from each institution. Each model uses different atmospheric products and tidal constituents to force the ocean, and the initial conditions and duration of spin up all vary (Appendix B). Nevertheless, we should expect statistical similarity in the oceanic flow at the spatial scales of $O(10\,\mathrm{km})$ if the numerics are robust.

3.1 Surface diagnostics of the temporal mean and variability

In light of the SWOT mission, the primary variable of interest is the ocean dynamic sea level. The AVISO estimate of this quantity is called the Absolute Dynamic Topography (ADT), while the closely related model diagnostic is the Sea Surface Height (SSH) after correcting for the inverse barometer effect if atmospheric pressure variability was simulated. Technically, SSH is defined as the geodetic height of the sea surface above the reference ellipsoid, while ocean dynamic sea level (or ADT) is defined relative to the geoid, but in models typically the geoid and reference ellipsoid coincide so these two definitions are in practice the same (Gregory et al., 2019). Furthermore, in the specific comparisons made here, a regional average of the ocean dynamic sea level estimates is removed first, so that large-scale, slow changes (*e.g.*, ice sheet mass loss contributions) are excluded from the comparison. From an ocean modelling perspective, one of the key features to argue in favor of increasing resolution in the North Atlantic has been the improvement in representing the Gulf Stream (GS) separation and path of the
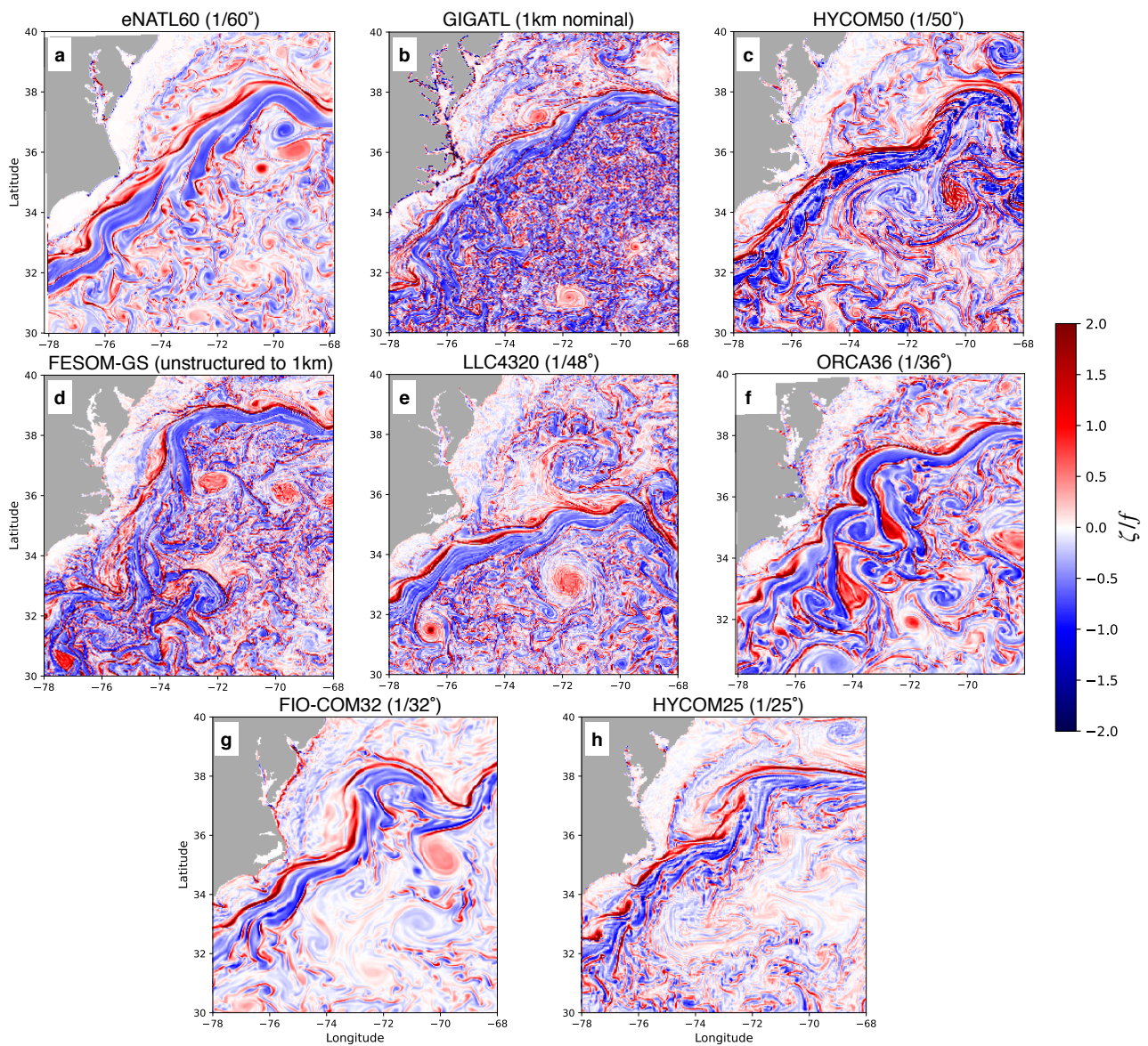
Figure 2. A snapshot of surface relative vorticity normalized by the local Coriolis parameter on February 1, 00:00 from each model in Region 1.

North Atlantic Current (Chassignet and Xu, 2017; Chassignet et al., 2020; Chassignet and Xu, 2021). In assessing the models, it is common to examine the mean state, which we define as the time mean, and variability about the mean. From the perspective of computational cost, the time mean of surface fields is the lightest as the reduction in dimension allows for the download framework where the collaborators can share the averaged data. Variability about the time mean requires access to the temporal dimension, making the computational and data storage cost intermediate. We will further show in section 3.2 an example of 3D diagnostics of the submesoscale flow, which significantly increases the computational cost and burden of transferring data; the 3D diagnostics will highlight the strength of the data-proximate framework where we can consistently apply the same diagnostic methods across different datasets.

In Figure 3, we show the time mean and temporal standard deviation of ocean dynamic sea level from the eight models in the GS separation region. We also show the time-mean ocean dynamic sea level estimated as ADT from the Archiving, Validation and Interpretation of Satellite Oceanographic (AVISO) data for reference. We do not show the standard deviation for AVISO as the spatiotemporal interpolation and smoothing limit its effective resolution to $O(100\,\mathrm{km})$ and $O(10\,\mathrm{day})$ (Chelton et al., 2011; Arbic et al., 2013; Chassignet and Xu, 2017). We provide the modeled standard deviations of ocean dynamic sea level filtered in a manner similar to the smoothing that goes into the AVISO products in Appendix C. The GS in most models tends to separate off of Cape Hatteras on the east coast of the U.S. consistent with AVISO (Figure 3a,c,g,i,k,o,t). In terms of the magnitude of mean SSH, HYCOM50 may be overestimating it relative to AVISO across the path of the separated Gulf Stream. The GS in LLC4320 tends to separate relatively southwards while in FESOM-GS separates northwards relative to AVISO observations respectively (Figure 3e,i). The separation in FESOM-GS may be closer to the observed state in 2014 (Figure 3s) rather than 2012, the actual year of model output. Regarding the standard deviation, while expected, it is interesting that the simulations without tides (FESOM-GS and ORCA36; Figure 3f,l) show significantly lower temporal variability compared to the other models. The low variability in FESOM-GS could also stem from the lack of atmospheric pressure variation in their atmospheric forcing (Table B6). Although HYCOM25 is tidally forced, its standard deviation is relatively low (Figure 3p), which may be due to lower spatial resolution than the region- and basin-scale models used here (Table B2), the computational tradeoff of it being a global simulation. HYCOM25 nevertheless has higher values than FESOM-GS and ORCA36. Considering the difference between simulations tidally forced and not, it is likely that in order to emulate the upcoming SWOT observations, applying tidal forcing is a key aspect in addition to model resolution (Savage et al., 2017a, b; Arbic et al., 2018). The benefit of having tidally forced simulations is that we can develop and test such methods of removing tides.

To complement the temporal standard deviation, in Figure 4, we show the frequency spectra of SSH in the GS separation region. The frequency periodograms were computed every $\sim 10\,\mathrm{km}$ using the `xrft` Python package (Uchida et al., 2021) and then spatially averaged to compute the spectra. The temporal linear trend was removed and a Hann window was applied prior to taking the Fourier transform of SSH as commonly done in studies examining spectra (*e.g.*, Uchida et al., 2017; Savage et al., 2017a; Khatri et al., 2021). At frequencies higher than the Coriolis frequency, LLC4320 shows the highest variability and FESOM-GS the lowest for both winter and summer. FIO-COM32 shows the largest spectral amplitudes at the diurnal and semi-diurnal frequencies, which reflects itself in the large standard deviation (Figure 3n). LLC4320 also shows the largest number of spectral peaks at tidal frequencies, likely due to being forced with full lunisolar tidal potential
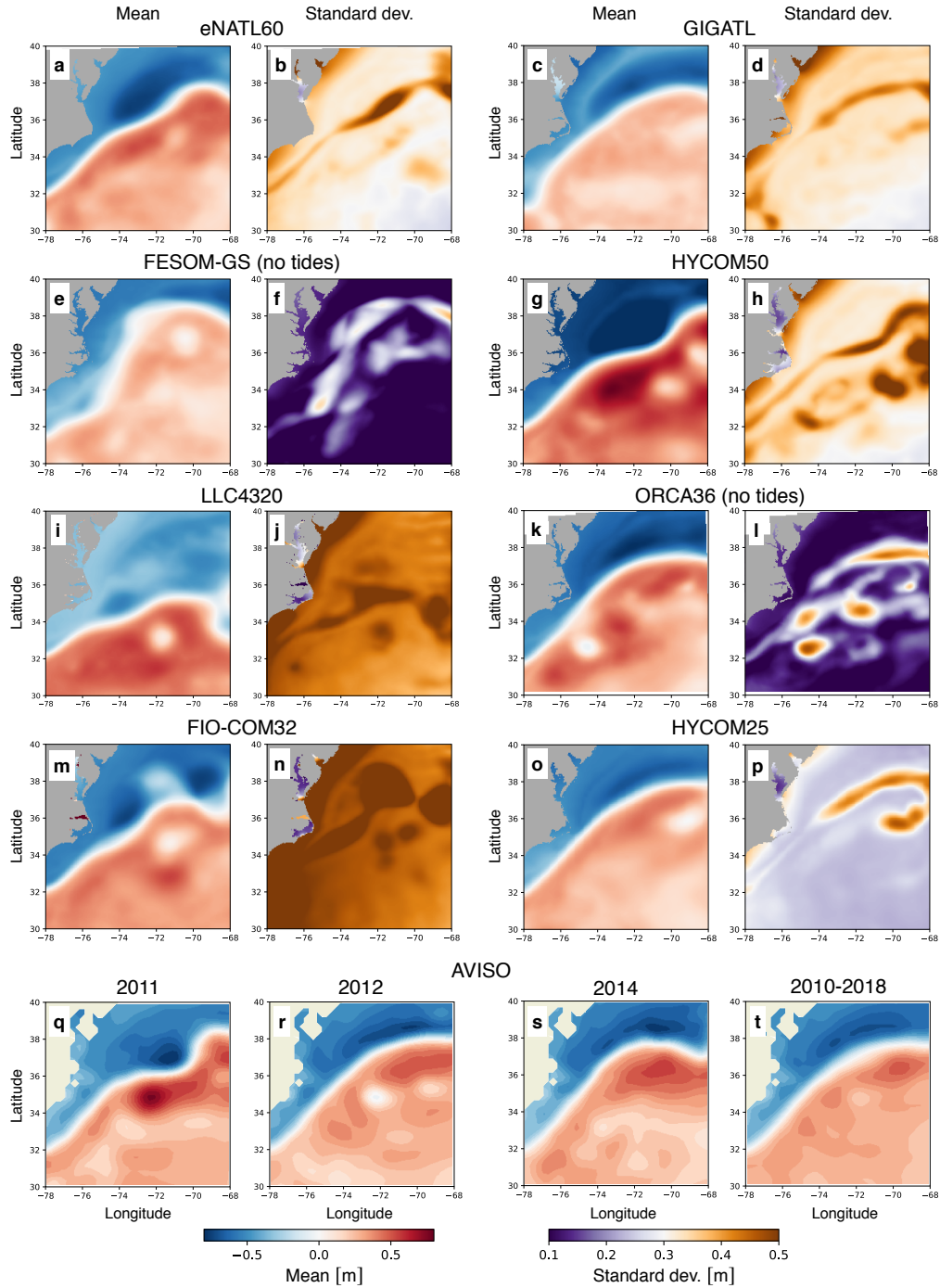
Figure 3. The temporal mean and standard deviation of ocean dynamic sea level in the Gulf Stream separation region (Region 1) during the months of February, March and April using hourly outputs of SSH from the models. The bottom row shows the seasonal mean of ADT fields from AVISO during the months of February, March and April. Daily AVISO data were used to compute the seasonal mean for three individual years (2011, 2012 and 2014) and over 2010-2018. The spatial mean is subtracted from the temporal mean fields from the models and AVISO to ensure that the mean SSH/ADT anomaly fields are comparable (*i.e.*, large-scale contributions have been removed).

as opposed to a discrete number of tidal constituents, as was the case for the other models used here (Table B6). Also note that tidal forcing in the LLC4320 simulation was inadvertently overestimated by a factor of 1.1121. It is not surprising that FESOM-GS lacks spectral peaks at diurnal and semidiurnal frequencies, considering that it is not tidally forced. ORCA36, on the other hand, although not tidally forced displays some activity at diurnal and semidiurnal frequencies possibly due to the inclusion of atmospheric pressure variation in their forcing (Table B6). However, the lower peaks at tidal frequencies in ORCA36 compared to the tidally forced runs reflect themselves in the lower standard deviation as seen in Figure 3l. eNATL60, GIGATL, HYCOM50 and HYCOM25 show similar levels of variability in the diurnal and semidiurnal band. It is interesting to note that at time scales of $O(1\text{-}10\text{ days})$, most runs show higher variability during winter than summer (Figure 4a,c), while the tidally forced runs show higher variability at time scales shorter than $O(1\text{ day})$ during summer (Figure 4b,d). The seasonality at time scales shorter than $O(1\text{ day})$ is reversed for ORCA36, a run with no tidal forcing. It is possible that the increase in forward cascade of energy stimulated by the tides are the culprit for higher SSH variability at time scales shorter than the inertial frequency during summer than winter for the tidally forced runs and vice versa for the non-tidally forced runs (Barkan et al., 2021). The overall higher SSH variability at time scales longer than the inertial frequency during winter than summer, on the other hand, is possibly due to wind-driven inertial waves (Flexas et al., 2019).

3.2 Three-dimensional diagnostics on physical processes

To exemplify 3D diagnostics, we display the submesoscale vertical buoyancy flux from each model using the daily-averaged outputs. Submesoscale vertical buoyancy fluxes in the surface ocean have been of great interest to the ocean and climate modeling community as they modulate the air-sea heat flux, affect mixed-layer depth (MLD), and are a proxy for baroclinic instability taking place within the mixed layer (often referred to as mixed-layer instability (MLI); Boccaletti et al., 2007; Mensa et al., 2013; Johnson et al., 2016; Su et al., 2018; Uchida et al., 2017, 2019; Schubert et al., 2020; Khatri et al., 2021). Ocean models used for climate simulations, however, lack the spatial resolution to resolve MLI due to computational constraints. A recent parametrization proposed by Fox-Kemper et al. (2008) has been operationally implemented by multiple climate modeling groups (Fox-Kemper et al., 2011; Huang et al., 2014; Calvert et al., 2020). While the vertical buoyancy flux predicted by the MLI parametrization has been tested in idealized simulations (Boccaletti et al., 2007; Fox-Kemper and Ferrari, 2008; Brannigan et al., 2017; Callies and Ferrari, 2018), non-eddying and mesoscale-permitting coupled and ocean-only simulations (Fox-Kemper et al., 2011; Calvert et al., 2020), and single-model assessments (*e.g.*, Mensa et al., 2013; Li et al., 2019; Yang et al., 2021; Richards et al., 2021), to our knowledge, a systematic assessment of the MLI parametrization has not been done versus multi-model, submesoscale-permitting, realistic ocean simulations which at least partially resolve the flux in need of parametrization in climate simulations. We take advantage of the unique opportunity provided by our collection of simulations to assess the flux parametrization, *i.e.*, the covariance of the 3D vertical velocity and buoyancy fields versus the modeled mixed layer depth and horizontal buoyancy gradient (3D data were not available for the HYCOM25 simulation).

The MLI parametrization predicts that the submesoscale vertical buoyancy fluxes vertically averaged over the mixed layer ($\overline{(\overline{(\cdot)}^z)}$) can be approximated by the squared horizontal gradient of the mesoscale buoyancy field times the mixed layer depth
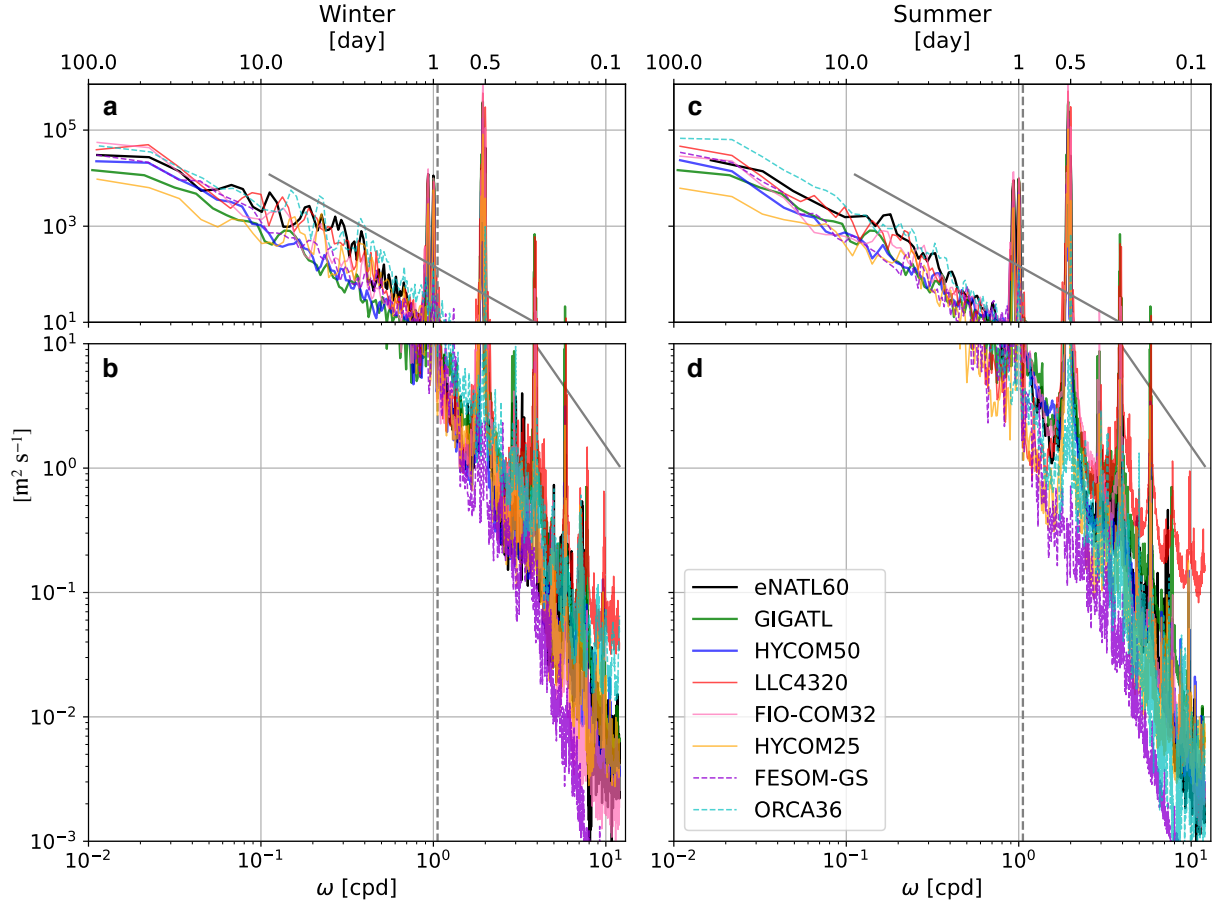
Figure 4. The frequency spectra of hourly SSH for winter (February, March, April; left) and summer (August, September, October; right). The panels are split up at $10\,\mathrm{m}^2\,\mathrm{s}^{-1}$ for visualization purposes. The frequency periodograms were computed every $\sim 10\,\mathrm{km}$ in Region 1 and then spatially averaged. The runs without tidal forcing (FESOM-GS and ORCA36) are shown in dashed lines. The Garrett-Munk spectral slope of ω^{-2} (Garrett and Munk, 1975) is shown as the grey solid line and the domain-averaged Coriolis frequency as the grey dashed line.

squared:

$$\overline{w^s b^s}^z \propto \frac{H_{\mathrm{ML}}^2 \overline{|\nabla_\mathrm{h} b^m|}^{z\,2}}{|f|},$$ (1)

where w, b, f and H_{ML} are the vertical velocity, buoyancy, local Coriolis parameter and MLD. While each model used a different Boussinesq reference density (ρ_0), buoyancy was defined as $b = -g\frac{\sigma_0}{\rho_0}$ where σ_0 is the potential density anomaly with the reference pressure of $0\,\mathrm{dbar}$ and $\rho_0 = 1000\,\mathrm{kg\,m^{-3}}$ for all model outputs. The MLD was defined using the density criterion (de Boyer Montégut et al., 2004), viz. the depth at which σ_0 increased by $0.03\,\mathrm{kg\,m^{-3}}$ from its value at $\sim 10\,\mathrm{m}$ depth. ∇_h is the horizontal gradient and the superscripts s and m indicate the submeso- and meso-scale field respectively. The decomposition between the submeso- and meso-scale were done by applying a Gaussian filter with the standard deviation of $30\,\mathrm{km}$ using the `gcm-filters` Python package (Grooms et al., 2021). Namely, the mesoscale field is defined as the spatially smoothed field with the Gaussian filter and submesoscale as the residual $(\cdot)^s = (\cdot) - (\cdot)^m$. The b^m field includes scales larger than the typical mesoscale but as it is the horizontal gradient of this field we are interested in, $\nabla_\mathrm{h} b^m$ captures the mesoscale fronts. We note that the Gaussian filter, implemented as a diffusive operator, commutes with the spatial derivative (this is an important property as we take the horizontal gradient of b^m; Grooms et al., 2021). While we acknowledge that there may be more sophisticated methods to decompose the flow (Uchida et al., 2019; Jing et al., 2020; Yang et al., 2021), a spatial filter has been commonly applied in examining the submesoscale flow in realistic simulations (*e.g.*, Mensa et al., 2013; Su et al., 2018; Li et al., 2019; Jing et al., 2021). Recently, Cao et al. (2021) argued that in addition to spatial cutoffs, a temporal cutoff improves the decomposition. Upon examining the frequency-wavenumber spectra of relative vorticity and horizontal divergence, however, we found that the daily averaging effectively filtered out the internal gravity waves (not shown). Based on characteristic time scale arguments, it is likely that our daily-averaged submesoscale fields are capturing the component in balance with stratification and Earth's rotation (Boccaletti et al., 2007; McWilliams, 2016), although some of the submesoscale balanced variability and nearly all of the internal gravity wave variability is filtered out by the daily average. Figure 5 shows the decomposition for w and b from eNATL60 on February 1, 2010 at depth $18\,\mathrm{m}$. We see the characteristic feature of the Gulf Stream separation particularly in the buoyancy field (Figure 5d) and submesoscale fronts (Figure 5c,f) superimposed on top of the large scale flow (Figure 5b,e). We will focus on the late winter/early spring months (February, March and April) as the spatial scale of MLI during summer is not well resolved even at kilometric resolution (Dong et al., 2020). We also restrict our diagnostics to the open ocean where the bathymetry is deeper than $100\,\mathrm{m}$ (*e.g.*, Figure 7).

Considering that the Fox-Kemper et al. (2008) MLI parametrization is intended for mesoscale-permitting models (neglecting the dependency on model grid-scale: Δs in Fox-Kemper et al., 2011), we further coarse grained the fields to $\sim 1/12°$ with a box-car operator, which gives:

$$\langle \overline{w^s b^s}^z \rangle \simeq C_e |f|^{-1} \left(\int_{-\langle H_{\mathrm{ML}} \rangle}^{0} \langle |\nabla_\mathrm{h} b^m| \rangle dz \right)^2,$$ (2)

where $\langle \cdot \rangle$ is the coarse-graining operator and C_e a tuning parameter or 'efficiency coefficient' (Fox-Kemper et al., 2011). The Δs scaling to compensate for coarse model resolution was omitted due to all our model outputs partially resolving the
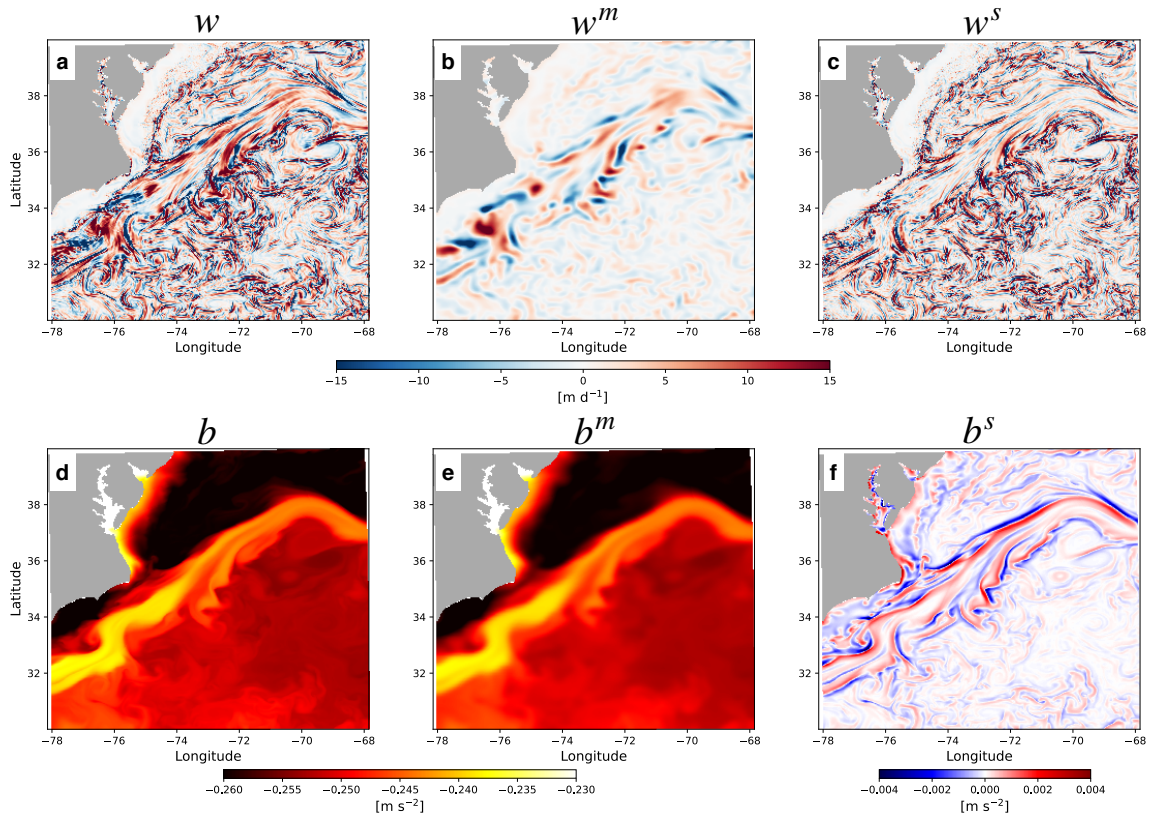
Figure 5. Snapshot from eNATL60 on February 1, 2010 at depth 18 m of the unfiltered daily w and b (left column), filtered fields applying the Gaussian filter (w^m, b^m; middle column), and the residual (w^s, b^s; right column).

submesoscale buoyancy flux. Furthermore, as Δs doesn't vary much among the models, this factor would not contribute much to the overall differences between models, in comparison to the greater variability due to numerics, *etc.*, this manuscript is
225 meant to introduce. We diagnosed C_e by taking the ratio between the right-hand and left-hand side of equation (2) at each grid point and time step (*e.g.*, left columns of Figure E1), and then the horizontal spatial median of it. The diagnosis (2) would differ from the parametrization (1) if there are large vertical variations in the buoyancy gradient, but these are not expected within the frequently-remixed mixed layer. Furthermore, the efficiency coefficient is expected to vary among the multi-model ensemble according to how well-resolved and/or damped the submesoscale instabilities are by model numerics, sub-grid schemes, and
230 daily averaging.

The diagnosed C_e only has a time dependence and fluctuates between the range of $[0.01, 0.07]$ across most models (blue solid curves in Figure 6) in agreement with the value of 0.06 recommended by Fox-Kemper et al. (2008). The time series of the spatial median of $\langle \overline{w^s b^s}^z \rangle$ and its prediction from the MLI parametrization are in sync with each other (black and red solid curves in Figure 6). The order of magnitude of the spatial median of the submesoscale vertical buoyancy flux diagnosed
235 from the models ($O(5 \times 10^{-9} \, \mathrm{m^2 \, s^{-3}})$) also agrees with observational estimates (Mahadevan et al., 2012; Johnson et al., 2016;

Buckingham et al., 2019) with an overall decrease in amplitude towards May except for FIO-COM32, which shows a local maximum around March (black solid curves in Figure 6).

We provide a snapshot of $\langle \overline{w^s b^s}^z \rangle$ and its prediction from the MLI parametrization (*i.e.*, both sides of equation (2)) on February 1 from each model in Figure 7. The joint histograms of the two over the months of February-April are also given in the bottom rows of Figure 7. The joint histograms of the two are concentrated around the one-to-one line indicating spatial correlation. The slight underestimation of magnitude in the MLI parametrization (viz. values falling below the one-to-one line) comes from the fact that while $\langle \overline{w^s b^s}^z \rangle$ can take negative values locally where frontogenesis dominates (*i.e.*, where the isopycnals steepen), the MLI parametrization by construction cannot differentiate between frontogenesis and frontolysis giving only positive values (equation 1). Nonetheless, $\langle \overline{w^s b^s}^z \rangle$ largely takes positive values indicating that processes such as mixed-layer and symmetric instabilities, which yield positive vertical buoyancy fluxes consistent with the extraction of potential energy (Dong et al., 2021), dominate in the surface boundary layer. While we have taken the spatial median to diagnose $C_e(t)$, which yields the best agreement in the time series (Figure 6), one may decide to instead take the spatial mean or mode, which we discuss in Appendix E.

For operational purposes, we would like to have a tuning parameter that is independent of not only space but also time. Therefore, we also display the MLI prediction when C_e is a constant taken to be its time mean. The agreement between $\langle \overline{w^s b^s}^z \rangle$ and the prediction remains surprisingly good (red dashed curves in Figure 6); in other words, the MLI parametrization is relatively insensitive to the temporal variability of $C_e(t)$. Regarding inter-model differences, HYCOM50 and LLC4320 have the smallest buoyancy fluxes predicted by the MLI parametrization (*i.e.*, weaker horizontal gradient magnitude and/or shallower mixed layer depths). The smaller predicted values presents itself as C_e diagnosed from the two taking an order of magnitude larger values than the other models (blue curves in Figure 6c,e); particularly for HYCOM50, using a constant C_e fails to reproduce the magnitude of $\langle \overline{w^s b^s}^z \rangle$ during the early half of February (red dashed curve in Figure 6c). It is possible that the lowest vertical resolution of HYCOM50 amongst the models (Table B2) results in under-representing the MLD despite its fine horizontal resolution particularly south of the Gulf Stream (Figure D1c); the MLI parametrization depends on it quadratically (equation (1)). The MLD from LLC4320 is also relatively shallow (Figure D1e). HYCOM50 and LLC4320 both use the K-profile parametrization (KPP, Table B3; Large et al., 1994) for the boundary-layer closure, which may imply that the KPP parameters warrant further tuning or reformulation for submesoscale-permitting model resolutions (*e.g.*, Bachman et al., 2017; Souza et al., 2020). The shallow MLD may also be due to the differences in the atmospheric products used to force the models (Table B6).

4   Conditions for sustainability

The strength of cloud storage and computing comes from it being decentralized from any specific institution, but this also leaves open the question about who pays for the cost of operating and supporting the cloud infrastructure, as well as paying for the cloud resources. Currently as of writing, the cloud storage provided by OSN is funded by an NSF grant acquired by the Climate Data Science Laboratory at Columbia University, and the JupyterHub on Google Cloud Platform (GCP) by Centre
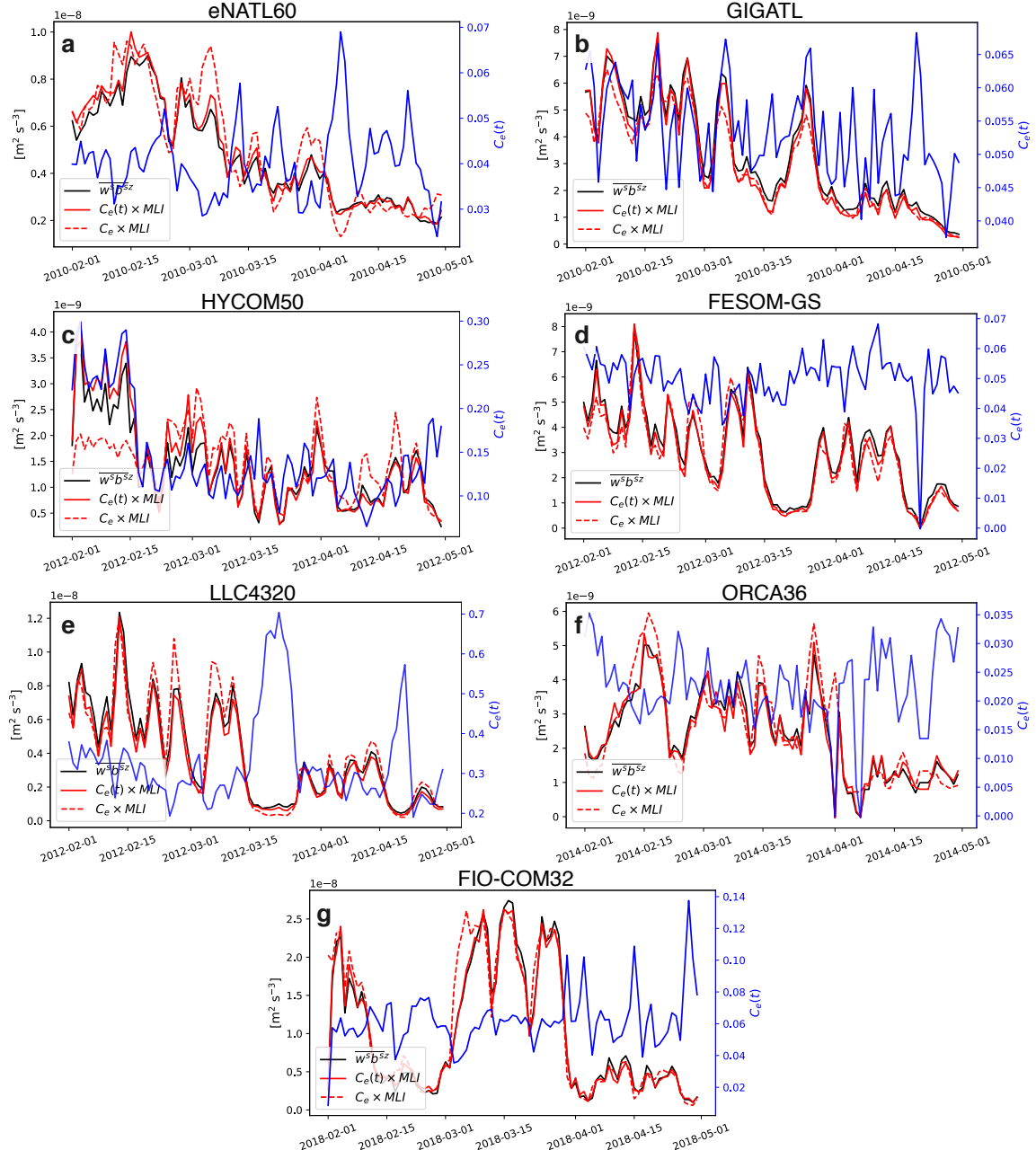
Figure 6. Time series of the spatial median of the submesoscale vertical buoyancy flux averaged over the MLD ($\langle \overline{w^s b^s}^z \rangle$; black solid curve) and its prediction from the MLI parametrization during the months of February to April. Note that the y axes vary depending on the magnitude diagnosed from each simulation in order to highlight its temporal variability. The prediction with temporally varying $C_e(t)$ is shown in red solid curves and with a temporally averaged (constant) C_e in red dashed curves. $C_e(t)$ is plotted against the right y axes in blue. Three-dimensional data were not available for HYCOM25.

14
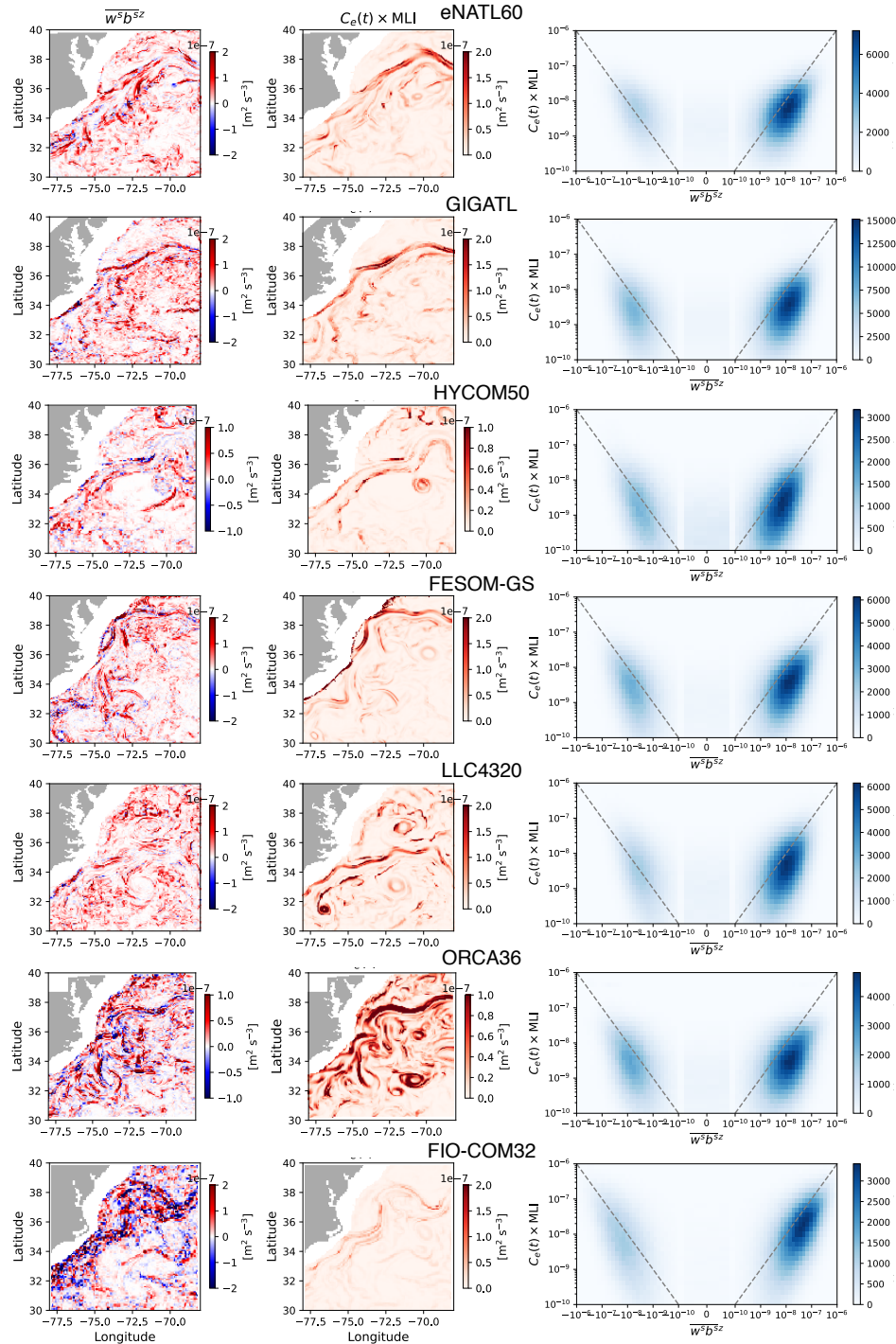
Figure 7. Snapshot of $\left\langle \overline{w^s b^s}^z \right\rangle$ (left column) and $C_e(t) \times$ MLI (middle column) on February 1 for each model. Note that the range of colorbar differs depending on the magnitude diagnosed from each model to highlight their spatial features and comparison between the submesoscale buoyancy flux and its equivalent predicted from the parametrization per simulation. Regions with bathymetry shallower than 100 m are masked out. The right column for each model shows the joint histogram of the two during the months of February to April, and the one-to-one line is shown as the grey dashed line. The histograms were computed using the `xhistogram` Python package (Abernathey et al., 2021c).

National d'Études Spatiales (CNES) funding. The cost of cloud resources for the JupyterHub with parallelized computation adds up to roughly 1000 € per month with the maximum computational resources of 64 cores and 256 gigabytes of memory per user; the resources scale on-demand, while the cost of operating the scalable Kubernetes infrastructure is managed by a vendor (2i2c) for a few thousand dollars a month. Although this may seem expensive compared to the local download framework where the costs of computation on clusters are shouldered upfront upon purchase of the cluster, there are several benefits to a cloud-based approach. First, using cloud infrastructure shifts the burden of hardware maintenance to the cloud provider, and users benefit from regular updates to technology and services that are available, meaning the scientific community can benefit from industry-driven innovations. Second, cloud infrastructure can be managed remotely and may use an inter-operable stack based on standards that are supported by many cloud providers (such as open source tools like Kubernetes and JupyterHub). This makes it easier to port workflows between clouds and get more cost-effective support in operating this infrastructure compared with paying full-time employees that run local hardware for an institution. We would like to note that while we have chosen GCP and OSN for the cloud platform, the core design principles and technology behind Pangeo Forge and the JupyterHub operated by 2i2c are non-proprietary and cloud vendor agnostic (for example, as defined in 2i2c's "Right to Replicate", https://2i2c.org/right-to-replicate). We could re-deploy the entire cloud platform on a different cloud provider with relative ease. This lets the users of this platform benefit from the flexibility and efficiency of the cloud, while minimizing the risk of lock-in and dependence on proprietary technology. As the cloud-based framework spreads within the scientific community, it is also possible that the ocean and climate science community will be able to negotiate better deals with cloud service providers; the framework is apt for Ocean and Climate Model Intercomparison Project (OMIP and CMIP: Griffies et al., 2016; Eyring et al., 2016) studies where tera- and peta-bytes of data need to be shared and analyzed consistently. The systematic storage of ARCO data with open access will also enable reproducible science, a crucial step when evaluating newer simulations against previous runs. While we believe we have showcased the potential for cloud-based computing, the success of the framework will depend on the scientific community to convince its peers and funding organizations to recognize its benefit.

5   Conclusions

In this study, we have implemented a cloud-based framework for collaborative, open-source and reproducible science, and have showcased its potential by analyzing eight submesoscale permitting simulations at a SWOT Crossover (Xover) region around the Gulf Stream separation (Region 1 in Figure 1). We have shown that despite the similar horizontal resolution amongst many models in this study, the spatial scales represented vary widely (Figure 2). This diverse representation likely originates from differences in advective/diffusive schemes, boundary layer parametrizations, atmospheric and tidal forcing, vertical resolution and/or bathymetry, and potentially duration of spin up amongst the simulations used here (Appendix B; cf. Chassignet and Xu, 2021). The need for collaborative work to inter-compare realistic simulations stems from both a scientific interest in the fidelity of submesoscale-permitting ocean models in representing the underlying physics and tracer transport, and an engineering

perspective on the numerics of ocean models. We leave a detailed analysis on the impact of numerics on the resolved dynamics for future work.

We have provided example diagnostics on SSH variability and submesoscale vertical buoyancy fluxes. The temporal standard deviation and spectra of SSH were significantly lower for the simulations without tidal forcing compared to the tidally forced simulations (Figures 3 and 4). This implies that in order to emulate the upcoming SWOT altimetric observations, tidal forcing is a key factor in modeling the surface ocean (Savage et al., 2017a, b; Arbic et al., 2018; Yu et al., 2021; Barkan et al., 2021; Le Guillou et al., 2021). Regarding 3D diagnostics, both the good agreement across multiple models between the tuning parameter C_e in the MLI parametrization and the values recommended by its developers (Figure 6; Fox-Kemper et al., 2008; Fox-Kemper and Ferrari, 2008), and the consistency of the order of magnitude of the flux predicted by the parametrization in the spatially averaged sense with observational estimates (cf. Richards et al., 2021), combine to provide confidence in implementing the MLI parametrization in realistic ocean and climate models. This is in contrast, however, with a recent study by Yang et al. (2021, their Figure 7) where they found (using the Regional Ocean Modeling System, ROMS with KPP) that the time series of $\langle \overline{w^s b^s}^z \rangle$ did not correlate well with the prediction from its parametrization in the Kuroshio extension. While we lack access to their model outputs, we speculate that the differences could be due to the diagnostic methods, domain of interest and/or configuration of their simulation. The contrasting findings all the more highlight the need for collaborative and open data analysis strategies of multi-model ensembles in assessing and improving the simulations themselves. We would like to note that were the modeled domain by Yang et al. (2021) covered Region 1, the cloud-based framework would allow for a straightforward platform to extend the ensemble of simulations (Appendix B) to include their outputs for our inter-comparison and reproducible science.

We end by noting that cloud-based data-proximate computation provides a framework to systematically analyze tera- and peta-bytes of data as we further increase the resolution and complexity of ocean and climate simulations, and as SWOT data becomes available. However, the success of the framework will depend on the ability of scientists to convince funding organizations to recognize its potential. Cloud-based computing differs from the conventional workflow which involves funding local computational resources and storage. While the cloud-based framework does not allow for an individual researcher or group to have prioritized access over the data and analytical tools, we believe that open access to the data will allow for reproducible science and facilitate international collaboration.

Appendix A: Example of `pangeo_forge_recipe` for eNATL60

Here we provide the Pangeo Forge recipe used to flux eNATL60 surface hourly data to OSN for Region 1 during February and April, 2010. The `input_url_pattern` is where the original NetCDF files were hosted on an OPeNDAP server, upon which the files were chunked along the time dimension before being fluxed to the cloud in Zarrified format (Miles et al., 2020). As a contributor to Pangeo Forge, one essentially only needs to specify the `input_url_pattern`. The zarrification and fluxing of the data to the cloud is automated by Pangeo Forge, reducing the infrastructure and cognitive burden on the data provider (Stern et al., 2022).

Listing 1. eNATL60 example

```python
from itertools import product

import pandas as pd
from pangeo_forge_recipes.patterns import pattern_from_file_sequence
from pangeo_forge_recipes.recipes import XarrayZarrRecipe


regions = [1]
season_months = {
    "fma": pd.date_range("2010-02", "2010-05", freq="M")
}


url_base = (
    "https://ige-meom-opendap.univ-grenoble-alpes.fr"
    "/thredds/fileServer/meomopendap/extract/SWOT-Adac"
)


def make_recipe_surface(region, season):
    input_url_pattern = url_base + "/Surface/eNATL60/Region{reg:02d}-surface-hourly_{yymm}.nc"
    months = season_months[season]
    input_urls = [
        input_url_pattern.format(reg=region, yymm=date.strftime("%Y-%m")) for date in months
    ]
    file_pattern = pattern_from_file_sequence(input_urls, "time_counter")

    target_chunks = {"time_counter": 72}
    subset_inputs = {"time_counter": 3}
    recipe = XarrayZarrRecipe(
        file_pattern, target_chunks=target_chunks, subset_inputs=subset_inputs
    )
```

```
            return recipe


370     recipes = {
            f"eNATL60/Region{reg:02d}/surface_hourly/{season}": make_recipe_surface(reg, season)
            for reg, season in product(regions, season_months)
        }
```


Appendix B: Model configurations

375 We provide the model configurations in Tables B1-B6 (blanks indicate the information was not obtainable). The vertical coordinate transformation onto geopotential coordinates for the outputs of GIGATL and HYCOM50, which had terrain-following and isopycnal coordinates as their native grid respectively (Table B2), were done using the `xgcm` Python package (Abernathey et al., 2021b) with linear interpolation. For the sake of storage, only three months of output for summer (Aug., Sep., Oct.) and winter (Feb., Mar., Apr.) respectively are stored on OSN from an arbitrary year per simulation.

Appendix C: Impact of spatiotemporal smoothing on the temporal standard deviation

In this appendix, we examine the effect of spatiotemporal filtering on the modelled SSH standard deviation. In order to mimic a smoothing procedure similar to the AVISO products, we apply a Gaussian spatial filter with the standard deviation of $50\,\mathrm{km}$ using the `gcm-filters` Python package and a 10 day running mean (cf. Chassignet and Xu, 2017). The non-tidally forced runs do not show much difference upon spatiotemporal smoothing from their standard deviation using hourly outputs but, 385 they significantly decrease for the tidally forced runs, particularly LLC4320 and FIO-COM32, with the modelled amplitudes coming closer to the AVISO estimate (Figures 3 and C1). The strong reduction in LLC4320 and FIO-COM32 may be expected as they are the runs with highest SSH variance at frequencies higher than the Coriolis frequency (Figure 4). All simulations agree that there is a local maximum in standard deviation around $37°\mathrm{N}$ where the separated GS situates consistent with AVISO. The SSH variability in GIGATL may be on the lower end considering it is tidally forced (Figure C1b), which could also be due 390 to the lack of pressure variation in the atmospheric forcing (Table B6).

Appendix D: Mixed-layer depth

The MLD averaged between February 1–15 is shown in Figure D1 along with the climatology for the month of February estimated from the Argo floats. We see that the MLD from HYCOM50 and LLC4320 are notably shallower south of the Gulf Stream compared to the other models and Argo estimate.
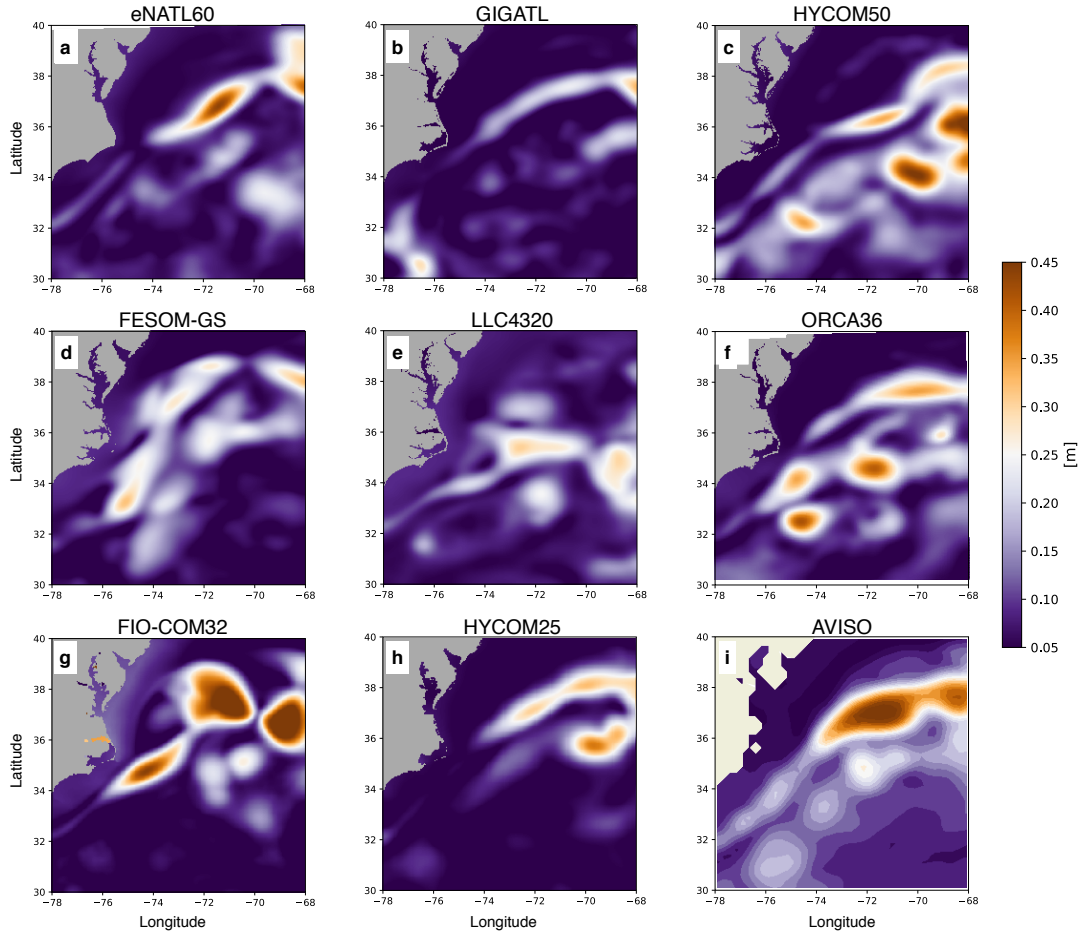
Figure C1. Temporal standard deviation of the spatiotemporally smoothed SSH from the eight models and ADT from AVISO over the years of 2010-2018 during the months of February-April. Note that the colorbar is slightly adjusted from Figure 3 in order to accommodate for lower values.

Appendix E: Efficiency coefficient and the MLI parametrization sensitivity to it

The efficiency coefficient $C_e(t,x,y)$ diagnosed from each simulation is given in the left column of Figure E1 and the joint histogram where $C_e(t)$ is taken as the spatial mean and mode in the right two columns respectively. It is interesting to note that $C_e(t,x,y)$ tends to take small values within fronts (namely, where the magnitude of $\langle \overline{w^s b^s}^z \rangle$ is large), but takes large values, reaching up to $O(1)$, on their periphery (Figures 7 and E1). Comparing the joint histograms in Figures 7 and E1, taking the spatial mean to diagnose $C_e(t)$ tends to overestimate the flux magnitude predicted from the parametrization as the mean is sensitive to extrema than the median; the histograms are concentrated above the one-to-one line (middle column of Figure E1). Diagnosing $C_e(t)$ as the spatial mode seemingly improves the alignment of the histogram with the one-to-one line (right column of Figure E1). However, taking the spatial mode results in $C_e(t)$ reaching values up to two orders of magnitude larger
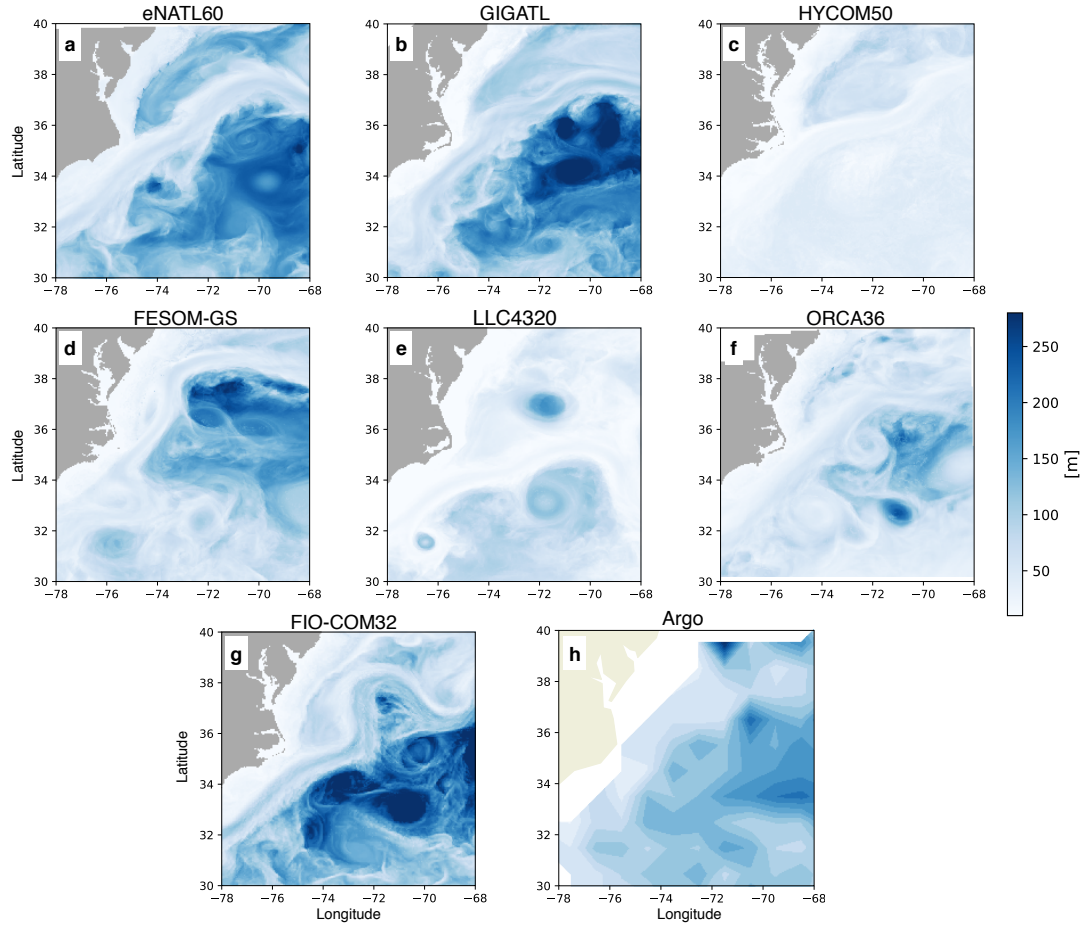
Figure D1. MLD from each model averaged over the duration of February 1–15 when the prediction from the MLI parametrization with a constant C_e in HYCOM50 deviates from the diagnosed submesoscale vertical buoyancy flux. The MLD was defined using the density criteria of de Boyer Montégut et al. (2004). For models with non-geopotential vertical coordinates (*i.e.*, GIGATL and HYCOM50), the MLD was computed using their native coordinates respectively. The climatology for the month of February from the Argo floats is taken from the dataset by Holte et al. (2017). The monthly-mean MLD defined by the density criterion (`mld_dt_mean`) is shown in order to be consistent with our model estimates.

than the values recommended by Fox-Kemper et al. (2008), and the time series predicted from the parametrization results in overestimating the submesoscale buoyancy flux in the spatially averaged sense (Figure E2). The time series predicted from using the spatial mean to estimate $C_e(t)$ further overestimates the buoyancy flux (not shown). We, therefore, recommend the usage of spatial median in estimating $C_e(t)$.

Figure E1. Snapshot of the efficiency coefficient $C_e(t,x,y)$ diagnosed on February 1 from each simulation (left column). The joint histogram of $\langle \overline{w^s b^s}^z \rangle$ and $C_e(t) \times$MLI during the months of February to April (right columns). The middle column shows the histogram when $C_e(t)$ is taken as the spatial mean of $C_e(t,x,y)$ and right column as the spatial mode of $C_e(t,x,y)$. The one-to-one line is shown as the grey dashed line.
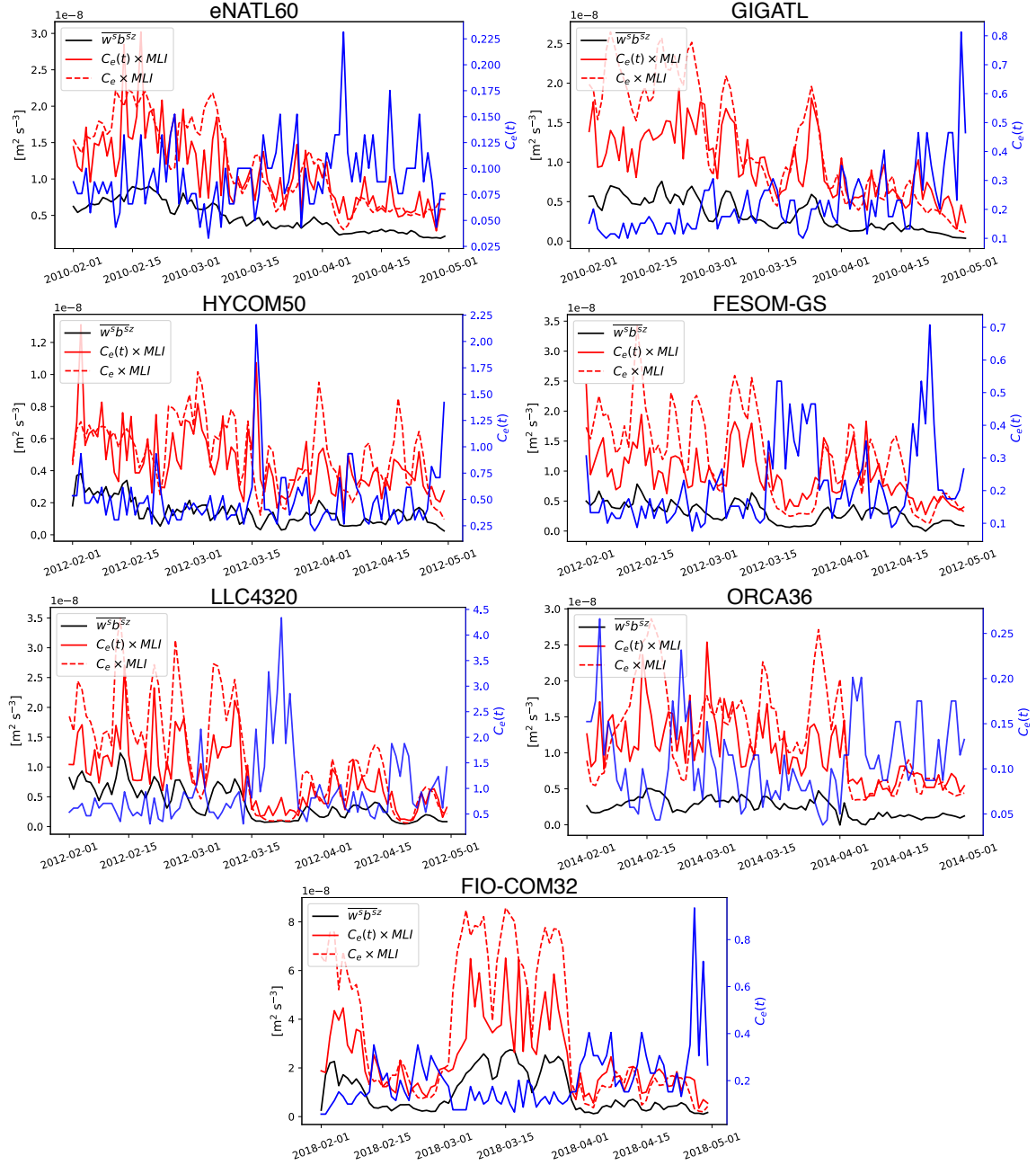
Figure E2. Time series of the spatial median of the submesoscale vertical buoyancy flux averaged over the MLD ($\langle\overline{w^s b^s}^z\rangle$; black solid curve) and its prediction from the MLI parametrization during the months of February to April where $C_e(t)$ is taken as the spatial mode of $C_e(t,x,y)$. Note that the y axes vary depending on the magnitude diagnosed from each simulation in order to highlight its temporal variability. The prediction with temporally varying $C_e(t)$ is shown in red solid curves and with a temporally averaged (constant) C_e in red dashed curves. $C_e(t)$ is plotted against the right y axes in blue.

24

References

Abernathey, R. P.: Petabytes of Ocean Data, Part I: NASA ECCO Data Portal, https://medium.com/pangeo/ petabytes-of-ocean-data-part-1-nasa-ecco-data-portal-81e3c5e077be, 2019.

Abernathey, R. P.: `fastjmd95`: Numba implementation of Jackett & McDougall (1995) ocean equation of state, https://doi.org/10.5281/zenodo.4498376, 2020.

Abernathey, R. P., Augspurger, T., Banihirwe, A., Blackmon-Luca, C. C., Crone, T. J., Gentemann, C. L., Hamman, J. J., Henderson, N., Lepore, C., McCaie, T. A., et al.: Cloud-Native Repositories for Big Scientific Data, Computing in Science & Engineering, 23, 26–35, https://doi.org/10.1109/MCSE.2021.3059437, 2021a.

Abernathey, R. P., Busecke, J., Smith, T., Banihirwe, A., Fernandes, F., Bourbeau, J., Cherian, D., Dussin, R., Swanson-Hysell, N., Constantinou, N., Ponte, A., et al.: `xgcm`: General Circulation Model Postprocessing with xarray, https://doi.org/10.5281/zenodo.3634752, 2021b.

Abernathey, R. P., Dougie, S., Nicholas, T., Bourbeau, J., Joseph, G., Yunyi, Y., Bailey, S., Bell, R., and Spring, A.: `xhistogram`: Fast, flexible, label-aware histograms for numpy and xarray, https://doi.org/10.5281/zenodo.5757149, 2021c.

Abernathey, R. P., Dussin, R., Smith, T., Fenty, I., Bourgault, P., Jones, S., Doddridge, E., Goldsworth, F., Losch, M., Almansi, M., Uchida, T., Cimatoribus, A., Schneider, A. D., Leskis, A., Quintana, A., Ponte, A., Rose, B., Balwada, D., Sérazin, G., Brannigan, L., et al.: `xmitgcm`: Read MITgcm mds binary files into xarray, https://doi.org/10.5281/zenodo.596253, 2021d.

Ajayi, A., Le Sommer, J., Chassignet, E., Molines, J.-M., Xu, X., Albert, A., and Dewar, W. K.: Diagnosing cross-scale kinetic energy exchanges from two submesoscale permitting ocean models, Journal of Advances in Modelling Earth Systems, https://doi.org/10.1029/2019MS001923, 2021.

Arbic, B. K., Garner, S. T., Hallberg, R. W., and Simmons, H. L.: The accuracy of surface elevations in forward global barotropic and baroclinic tide models, Deep Sea Research Part II: Topical Studies in Oceanography, 51, 3069–3101, https://doi.org/10.1016/j.dsr2.2004.09.014, 2004.

Arbic, B. K., Polzin, K. L., Scott, R. B., Richman, J. G., and Shriver, J. F.: On eddy viscosity, energy cascades, and the horizontal resolution of gridded satellite altimeter products, Journal of Physical Oceanography, 43, 283–300, https://doi.org/10.1175/JPO-D-11-0240.1, 2013.

Arbic, B. K., Alford, M., Ansong, J., Buijsman, M., Ciotti, R., Farrar, J., Hallberg, R., Henze, C., Hill, C., Luecke, C., Menemenlis, D., Metzger, E., Müeller, M., Nelson, A., Nelson, B., Ngodock, H., Ponte, R., Richman, J., Savage, A., and Zhao, Z.: A primer on global internal tide and internal gravity wave continuum modeling in HYCOM and MITgcm, in: New Frontiers in Operational Oceanography, edited by Chassignet, E., Pascual, A., Tintoré, J., and Verron, J., pp. 307–392, GODAE Ocean View, https://doi.org/10.17125/gov2018.ch13, 2018.

Bachman, S. D., Fox-Kemper, B., Taylor, J. R., and Thomas, L. N.: Parameterization of Frontal Symmetric Instabilities. I: Theory for Resolved Fronts, Ocean Modelling, 109, 72–95, https://doi.org/10.1016/j.ocemod.2016.12.003, 2017.

Barham, W., Bachman, S., and Grooms, I.: Some effects of horizontal discretization on linear baroclinic and symmetric instabilities, Ocean Modelling, 125, 106–116, https://doi.org/10.1016/j.ocemod.2018.03.004, 2018.

Barkan, R., Srinivasan, K., Yang, L., McWilliams, J. C., Gula, J., and Vic, C.: Oceanic mesoscale eddy depletion catalyzed by internal waves, Geophysical Research Letters, 48, e2021GL094 376, https://doi.org/10.1029/2021GL094376, 2021.

Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiéry, N. M., and Fangohr, H.: Using Jupyter for reproducible scientific workflows, Computing in Science & Engineering, 23, 36–46, https://doi.org/10.1109/MCSE.2021.3052101, 2021.

475   Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates, Ocean modelling, 4, 55–88, https://doi.org/10.1016/S1463-5003(01)00012-9, 2002.

Boccaletti, G., Ferrari, R., and Fox-Kemper, B.: Mixed layer instabilities and restratification, Journal of Physical Oceanography, 37, 2228–2250, https://doi.org/10.1175/JPO3101.1, 2007.

Bodner, A. S. and Fox-Kemper, B.: A breakdown in potential vorticity estimation delineates the submesoscale-to-turbulence boundary in

480   large eddy simulations, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 049, https://doi.org/10.1029/2020MS002049, 2020.

Brannigan, L., Marshall, D. P., Naveira Garabato, A. C., Nurser, A. J. G., and Kaiser, J.: Submesoscale Instabilities in Mesoscale Eddies, Journal of Physical Oceanography, 47, 3061–3085, https://doi.org/10.1175/jpo-d-16-0178.1, 2017.

Brodeau, L., Albert, A., and Le Sommer, J.: NEMO-eNATL60 description and assessment repository,

485   https://doi.org/10.5281/zenodo.4032732, 2020.

Brydon, D., Sun, S., and Bleck, R.: A new approximation of the equation of state for seawater, suitable for numerical ocean models, Journal of Geophysical Research: Oceans, 104, 1537–1540, https://doi.org/10.1029/1998JC900059, 1999.

Buckingham, C. E., Lucas, N., Belcher, S., Rippeth, T., Grant, A., Le Sommer, J., Ajayi, A. O., and Garabato, A. C. N.: The contribution of surface and submesoscale processes, Journal of Advances in Modeling Earth Systems, https://doi.org/10.1029/2019MS001801, 2019.

490   Callies, J. and Ferrari, R.: Note on the Rate of Restratification in the Baroclinic Spindown of Fronts, Journal of Physical Oceanography, 48, 1543–1553, https://doi.org/10.1175/jpo-d-17-0175.1, 2018.

Calvert, D., Nurser, G., Bell, M. J., and Fox-Kemper, B.: The impact of a parameterisation of submesoscale mixed layer eddies on mixed layer depths in the NEMO ocean model, Ocean Modelling, 154, 101 678, https://doi.org/10.1016/j.ocemod.2020.101678, 2020.

Cao, H., Fox-Kemper, B., and Jing, Z.: Submesoscale Eddies in the Upper Ocean of the Kuroshio Extension from High-Resolution Simula-

495   tion: Energy Budget, Journal of Physical Oceanography, 51, 2181–2201, https://doi.org/10.1175/JPO-D-20-0267.1, 2021.

Chassignet, E. P. and Xu, X.: Impact of Horizontal Resolution ($1/12°$ to $1/50°$) on Gulf Stream Separation, Penetration, and Variability, Journal of Physical Oceanography, 47, 1999–2021, https://doi.org/10.1175/JPO-D-17-0031.1, 2017.

Chassignet, E. P. and Xu, X.: On the Importance of High-Resolution in Large-Scale Ocean Models, Advances in Atmospheric Sciences, pp. 1–14, https://doi.org/10.1007/s00376-021-0385-7, 2021.

500   Chassignet, E. P., Hurlburt, H. E., Metzger, E. J., Smedstad, O. M., Cummings, J. A., Halliwell, G. R., Bleck, R., Baraille, R., Wallcraft, A. J., Lozano, C., et al.: US GODAE: Global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM), Oceanography, 22, 64–75, 2009.

Chassignet, E. P., Yeager, S. G., Fox-Kemper, B., Bozec, A., Castruccio, F., Danabasoglu, G., Horvat, C., Kim, W. M., Koldunov, N., Li, Y., et al.: Impact of horizontal resolution on global ocean–sea ice model simulations based on the experimental protocols of the Ocean Model

505   Intercomparison Project phase 2 (OMIP-2), Geoscientific Model Development, 13, 4595–4637, 2020.

Chelton, D. B., Schlax, M. G., and Samelson, R. M.: Global observations of nonlinear mesoscale eddies, Progress in oceanography, 91, 167–216, https://doi.org/10.1016/j.pocean.2011.01.002, 2011.

Danilov, S., Sidorenko, D., Wang, Q., and Jung, T.: The Finite-volumE sea ice–Ocean Model (FESOM2), Geoscientific Model Development, 10, 765–789, https://doi.org/10.5194/gmd-10-765-2017, 2017.

510   de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, Journal of Geophysical Research: Oceans, 109, https://doi.org/10.1029/2004JC002378, 2004.

27

Dong, J., Fox-Kemper, B., Zhang, H., and Dong, C.: The scale of submesoscale baroclinic instability globally, Journal of Physical Oceanography, 50, 2649–2667, https://doi.org/10.1175/JPO-D-20-0043.1, 2020.

Dong, J., Fox-Kemper, B., Zhu, J., and Dong, C.: Application of symmetric instability parameterization in the Coastal and Regional Ocean Community Model (CROCO), Journal of Advances in Modeling Earth Systems, 13, e2020MS002 302, https://doi.org/10.1029/2020MS002302, 2021.

Ducousso, N., Le Sommer, J., Molines, J.-M., and Bell, M.: Impact of the "Symmetric Instability of the Computational Kind" at mesoscale- and submesoscale-permitting resolutions, Ocean Modelling, 120, 18–26, https://doi.org/10.1016/j.ocemod.2017.10.006, 2017.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fangohr, H., Beg, M., Bergemann, M., Bondar, V., Brockhauser, S., Carinan, C., Costa, R., Fortmann, C., Marsa, D. F., Giovanetti, G., et al.: Data exploration and analysis with jupyter notebooks, in: 17th Biennial International Conference on Accelerator and Large Experimental Physics Control Systems, TALK-2020-009, https://doi.org/10.18429/JACoW-ICALEPCS2019-TUCPR02, 2019.

Fernandes, F.: `python-seawater`: Python re-write of the CSIRO seawater toolbox SEAWATER-3.3 for calculating the properties of sea water, https://doi.org/10.5281/zenodo.11395, 2014.

Firing, E., Fernandes, F., Barna, A., and Abernathey, R. P.: `GSW-python`: Python implementation of the Thermodynamic Equation of Seawater 2010 (TEOS-10), https://doi.org/10.5281/zenodo.5214122, 2021.

Flexas, M. M., Thompson, A. F., Torres, H. S., Klein, P., Farrar, J. T., Zhang, H., and Menemenlis, D.: Global estimates of the energy transfer from the wind to the ocean, with emphasis on near-inertial oscillations, Journal of Geophysical Research: Oceans, 124, 5723–5746, https://doi.org/10.1029/2018JC014453, 2019.

Fox-Kemper, B. and Ferrari, R.: Parameterization of mixed layer eddies. Part II: Prognosis and impact, Journal of Physical Oceanography, 38, 1166–1179, https://doi.org/0.1175/2007JPO3788.1, 2008.

Fox-Kemper, B., Ferrari, R., and Hallberg, R.: Parameterization of mixed layer eddies. Part I: Theory and diagnosis, Journal of Physical Oceanography, 38, 1145–1165, https://doi.org/10.1175/2007JPO3792.1, 2008.

Fox-Kemper, B., Danabasoglu, G., Ferrari, R., Griffies, S., Hallberg, R., Holland, M., Maltrud, M., Peacock, S., and Samuels, B.: Parameterization of mixed layer eddies. Part III: Implementation and impact in global ocean climate simulations, Ocean Modelling, 39, 61–78, https://doi.org/10.1016/j.ocemod.2010.09.002, 2011.

Garrett, C. and Munk, W.: Space-time scales of internal waves: A progress report, Journal of Geophysical Research, 80, 291–297, https://doi.org/10.1029/JC080i003p00291, 1975.

Gomez-Navarro, L., Fablet, R., Mason, E., Pascual, A., Mourre, B., Cosme, E., and Le Sommer, J.: SWOT spatial scales in the western Mediterranean sea derived from pseudo-observations and an Ad Hoc filtering, Remote Sensing, 10, 599, https://doi.org/10.3390/rs10040599, 2018.

Gregory, J. M., Griffies, S. M., Hughes, C. W., Lowe, J. A., Church, J. A., Fukimori, I., Gomez, N., Kopp, R. E., Landerer, F., Le Cozannet, G., et al.: Concepts and terminology for sea level: Mean, variability and change, both local and global, Surveys in Geophysics, 40, 1251–1289, https://doi.org/10.1007/s10712-019-09525-z, 2019.

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., Chassignet, E. P., Curchitser, E., Deshayes, J., Drange, H., Fox-Kemper, B., Gleckler, P. J., Gregory, J. M., Haak, H., Hallberg, R. W., Heimbach, P., Hewitt, H. T., Holland, D. M., Ilyina, T., Jungclaus, J. H., Komuro, Y., Krasting, J. P., Large, W. G., Marsland, S. J., Masina, S., McDougall, T. J., Nurser, A. J. G., Orr,

550   J. C., Pirani, A., Qiao, F., Stouffer, R. J., Taylor, K. E., Treguier, A. M., Tsujino, H., Uotila, P., Valdivieso, M., Wang, Q., Winton, M., and Yeager, S. G.: OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project, Geoscientific Model Development, 9, 3231–3296, https://doi.org/10.5194/gmd-9-3231-2016, 2016.

Grooms, I., Loose, N., Abernathey, R., Steinberg, J., Bachman, S. D., Marques, G., Guillaumin, A. P., and Yankovsky, E.: Diffusion-based smoothers for spatial filtering of gridded geophysical data, Journal of Advances in Modeling Earth Systems, p. e2021MS002552,
555   https://doi.org/10.1029/2021MS002552, 2021.

Gula, J., Theetten, S., Cambon, G., and Roullet, G.: Description of the GIGATL simulations, https://doi.org/10.5281/zenodo.4948523, 2021.

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

Hamlington, P. E., Van Roekel, L. P., Fox-Kemper, B., Julien, K., and Chini, G. P.: Langmuir–submesoscale interactions: Descriptive analysis
560   of multiscale frontal spindown simulations, Journal of Physical Oceanography, 44, 2249–2272, 2014.

Holte, J., Talley, L. D., Gilson, J., and Roemmich, D.: An Argo mixed layer climatology and database, Geophysical Research Letters, 44, 5618–5626, http://mixedlayer.ucsd.edu/, 2017.

Huang, C. J., Qiao, F., and Dai, D.: Evaluating CMIP5 simulations of mixed layer depth during summer, Journal of Geophysical Research: Oceans, 119, 2568–2582, https://doi.org/10.1002/2013JC009535, 2014.

565   Jackett, D. R. and McDougall, T. J.: Minimal adjustment of hydrographic profiles to achieve static stability, Journal of Atmospheric and Oceanic Technology, 12, 381–389, https://doi.org/10.1175/1520-0426(1995)012<0381:MAOHPT>2.0.CO;2, 1995.

Jing, Z., Wang, S., Wu, L., Chang, P., Zhang, Q., Sun, B., Ma, X., Qiu, B., Small, J., Jin, F.-F., et al.: Maintenance of mid-latitude oceanic fronts by mesoscale eddies, Science advances, 6, eaba7880, https://doi.org/10.1126/sciadv.aba7880, 2020.

Jing, Z., Fox-Kemper, B., Cao, H., Zheng, R., and Du, Y.: Submesoscale fronts and their dynamical processes associated with symmetric
570   instability in the northwest Pacific subtropical Ocean, Journal of Physical Oceanography, 51, 83–100, https://doi.org/10.1175/JPO-D-20-0076.1, 2021.

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

Khatri, H., Griffies, S. M., Uchida, T., Wang, H., and Menemenlis, D.: Role of mixed-layer instabilities in the seasonal evolution
575   of eddy kinetic energy spectra in a global submesoscale permitting simulation, Geophysical Research Letters, p. e2021GL094777, https://doi.org/10.1029/2021GL094777, 2021.

Large, W. G., McWilliams, J. C., and Doney, S. C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization, Reviews of geophysics, 32, 363–403, https://doi.org/10.1029/94rg01872, 1994.

Le Corre, M., Gula, J., and Tréguier, A.-M.: Barotropic vorticity balance of the North Atlantic subpolar gyre in an eddy-resolving model,
580   Ocean Science, 16, 451–468, https://doi.org/10.5194/os-16-451-2020, 2020.

Le Guillou, F., Lahaye, N., Ubelmann, C., Metref, S., Cosme, E., Ponte, A., Le Sommer, J., Blayo, E., and Vidard, A.: Joint estimation of balanced motions and internal tides from future wide-swath altimetry, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 613, https://doi.org/10.1029/2021MS002613, 2021.

Lévy, M., Franks, P. J., and Smith, K. S.: The role of submesoscale currents in structuring marine ecosystems, Nature communications, 9,
585   1–16, https://doi.org/10.1038/s41467-018-07059-3, 2018.

Li, J., Dong, J., Yang, Q., and Zhang, X.: Spatial-temporal variability of submesoscale currents in the South China Sea, Journal of Oceanology and Limnology, 37, 474–485, https://doi.org/10.1007/s00343-019-8077-1, 2019.

Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., et al.: NEMO ocean engine, https://doi.org/10.5281/zenodo.3878122, 2019.

590 Mahadevan, A., D'asaro, E., Lee, C., and Perry, M. J.: Eddy-driven stratification initiates North Atlantic spring phytoplankton blooms, Science, 337, 54–58, https://doi.org/10.1126/science.1218740, 2012.

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

McWilliams, J. C.: Submesoscale currents in the ocean, Proceedings of the Royal Society A: Mathematical, Physical and Engineering
595 Sciences, 472, 20160 117, https://doi.org/10.1098/rspa.2016.0117, 2016.

Menemenlis, D., Campin, J.-M., Heimbach, P., Hill, C., Lee, T., Nguyen, A., Schodlok, M., and Zhang, H.: ECCO2: High resolution global ocean and sea ice data synthesis, Mercator Ocean Quarterly Newsletter, 31, 13–21, 2008.

Mensa, J. A., Garraffo, Z., Griffa, A., Özgökmen, T. M., Haza, A., and Veneziani, M.: Seasonality of the submesoscale dynamics in the Gulf Stream region, Ocean Dynamics, 63, 923–941, https://doi.org/10.1007/s10236-013-0633-1, 2013.

600 Met Office: Cartopy: a cartographic python library with a Matplotlib interface, Exeter, Devon, https://scitools.org.uk/cartopy, 2010 - 2015.

Metref, S., Cosme, E., Le Guillou, F., Le Sommer, J., Brankart, J.-M., and Verron, J.: Wide-swath altimetric satellite data assimilation with correlated-error reduction, Frontiers in Marine Science, 6, 822, https://doi.org/10.3389/fmars.2019.00822, 2020.

Miles, A., Kirkham, J., Durant, M., Bourbeau, J., Onalan, T., Hamman, J., Patel, Z., Rocklin, M., Dussin, R., et al.: `zarr`: A format for the storage of chunked, compressed, N-dimensional arrays, https://doi.org/10.5281/zenodo.3773450, 2020.

605 Morrow, R., Fu, L.-L., Ardhuin, F., Benkiran, M., Chapron, B., Cosme, E., d'Ovidio, F., Farrar, J. T., Gille, S. T., Lapeyre, G., Traon, P.-Y. L., Pascual, A., Ponte, A., Qiu, B., Rascle, N., Ubelmann, C., Wang, J., and Zaron, E. D.: Global Observations of Fine-Scale Ocean Surface Topography With the Surface Water and Ocean Topography (SWOT) Mission, Frontiers in Marine Science, 6, https://doi.org/10.3389/fmars.2019.00232, 2019.

Richards, K. J., Whitt, D. B., Brett, G., Bryan, F. O., Feloy, K., and Long, M. C.: The impact of climate change on ocean submesoscale
610 activity, Journal of Geophysical Research: Oceans, 126, e2020JC016 750, https://doi.org/10.1029/2020JC016750, 2021.

Rocha, C. B., Chereskin, T. K., Gille, S. T., and Menemenlis, D.: Mesoscale to submesoscale wavenumber spectra in Drake Passage, Journal of Physical Oceanography, 46, 601–620, https://doi.org/10.1175/JPO-D-15-0087.1, 2016.

Savage, A. C., Arbic, B. K., Alford, M. H., Ansong, J. K., Farrar, J. T., Menemenlis, D., O'Rourke, A. K., Richman, J. G., Shriver, J. F., Voet, G., et al.: Spectral decomposition of internal gravity wave sea surface height in global models, Journal of Geophysical Research: Oceans,
615 122, 7803–7821, https://doi.org/10.1002/2017JC013009, 2017a.

Savage, A. C., Arbic, B. K., Richman, J. G., Shriver, J. F., Alford, M. H., Buijsman, M. C., Thomas Farrar, J., Sharma, H., Voet, G., Wallcraft, A. J., et al.: Frequency content of sea surface height variability from internal gravity waves to mesoscale eddies, Journal of Geophysical Research: Oceans, 122, 2519–2538, https://doi.org/10.1002/2016JC012331, 2017b.

Schaffer, J., Timmermann, R., Arndt, J. E., Kristensen, S. S., Mayer, C., Morlighem, M., and Steinhage, D.: A global, high-resolution data
620 set of ice sheet topography, cavity geometry, and ocean bathymetry, Earth System Science Data, 8, 543–557, https://doi.org/10.5194/essd-8-543-2016, 2016.

Schubert, R., Schwarzkopf, F. U., Baschek, B., and Biastoch, A.: Submesoscale Impacts on Mesoscale Agulhas Dynamics, Journal of Advances in Modeling Earth Systems, 2019.

Schubert, R., Gula, J., Greatbatch, R. J., Baschek, B., and Biastoch, A.: The Submesoscale Kinetic Energy Cascade: Mesoscale Absorption of Submesoscale Mixed-Layer Eddies and Frontal Downscale Fluxes, Journal of Physical Oceanography, https://doi.org/10.1175/JPO-D-19-0311.1, 2020.

Shchepetkin, A. and McWilliams, J. C.: The Regional Oceanic Modeling System (ROMS): A split-explicit, free-surface, topography-following- coordinate ocean model, Ocean Modelling, 9, 347–404, https://doi.org/10.1016/j.ocemod.2004.08.002, 2005.

Smith, W. H. and Sandwell, D. T.: Global sea floor topography from satellite altimetry and ship depth soundings, Science, 277, 1956–1962, https://doi.org/doi:10.1126/science.277.5334.1956, 1997.

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

Souza, A. N., Wagner, G., Ramadhan, A., Allen, B., Churavy, V., Schloss, J., Campin, J., Hill, C., Edelman, A., Marshall, J., et al.: Uncertainty Quantification of Ocean Parameterizations: Application to the K-Profile-Parameterization for Penetrative Convection, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 108, https://doi.org/10.1029/2020MS002108, 2020.

Stern, C., Abernathey, R., Hamman, J. J., Wegener, R., Lepore, C., and Harkins, S.: Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production, Frontiers in Climate, p. In rev., https://doi.org/10.31223/X5462G, 2022.

Stewart, A. L., Klocker, A., and Menemenlis, D.: Circum-Antarctic shoreward heat transport derived from an eddy-and tide-resolving simulation, Geophysical Research Letters, 45, 834–845, https://doi.org/10.1002/2017GL075677, 2018.

Su, Z., Wang, J., Klein, P., Thompson, A. F., and Menemenlis, D.: Ocean submesoscales as a key component of the global heat budget, Nature Communications, 9, https://doi.org/10.1038/s41467-018-02983-w, 2018.

Torres, H. S., Klein, P., Menemenlis, D., Qiu, B., Su, Z., Wang, J., Chen, S., and Fu, L.-L.: Partitioning ocean motions into balanced motions and internal gravity waves: A modeling study in anticipation of future space missions, Journal of Geophysical Research: Oceans, 123, 8084–8105, https://doi.org/10.1029/2018JC014438, 2018.

Uchida, T., Abernathey, R. P., and Smith, S. K.: 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, 2017.

Uchida, T., Balwada, D., Abernathey, R. P., McKinley, G. A., Smith, S. K., and Lévy, M.: The contribution of submesoscale over mesoscale eddy iron transport in the open Southern Ocean, Journal of Advances in Modeling Earth Systems, 11, 3934–3958, https://doi.org/10.1029/2019MS001805, 2019.

Uchida, T., Rokem, A., Squire, D., Nicholas, T., Abernathey, R. P., Soler, S., Nouguier, F., Vanderplas, J., Paige, M., Mondal, A., Mayer, A., Halchenko, Y., Wilson, G., Constantinou, N., Ponte, A., Squire, D., Busecke, J., Spring, A., Pak, K., Scott, R., Hoyer, S., Moon, Z., et al.: xrft: Fourier transforms for xarray data, https://doi.org/10.5281/zenodo.1402635, 2021.

Xiao, B., Qiao, F., Shu, Q., Yin, X., Wang, G., and Wang, S.: The development and validation of a global 1/32° surface wave-tide-circulation coupled ocean model: FIO-COM32, Geoscientific Model Development Discussions, pp. 1–25, https://doi.org/10.5194/gmd-2022-52, 2022.

Yang, P., Jing, Z., Sun, B., Wu, L., Qiu, B., Chang, P., and Ramachandran, S.: On the upper-ocean vertical eddy heat transport in the Kuroshio extension. Part I: Variability and dynamics, Journal of Physical Oceanography, 51, 229–246, https://doi.org/10.1175/JPO-D-20-0068.1, 2021.

Yu, X., Ponte, A. L., Lahaye, N., Caspar-Cohen, Z., and Menemenlis, D.: Geostrophy Assessment and Momentum Balance of the Global Oceans in a Tide-and Eddy-Resolving Model, Journal of Geophysical Research: Oceans, 126, e2021JC017 422, https://doi.org/10.1029/2021JC017422, 2021.

Table B1. The model and initial condition used for each simulation, their duration of spin up and year of output stored on OSN. HYCOM50 was spun up from rest and integrated for a total of 20 years. Sensitivity experiments were performed starting from year 15 (Chassignet and Xu, 2017, 2021). LLC4320 used progressive spin-up from a 1/6° state estimate (Menemenlis et al., 2008) followed by 1/12° and 1/24° simulations, as detailed in Table D2 of Rocha et al. (2016).

Simulation	Model (version)	Initial condition	Spin up	Year of output
eNATL60	NEMO (3.6)	GLORYS12	18 months	2010
GIGATL	CROCO	July 2007 from an identical run w/ 3 km resolution	12 months	2010
HYCOM50	HYCOM	GDEM climatology	15 years	year 19
FESOM-GS	FESOM (2.1)	PHC3.0	18 months	2012
LLC4320	MITgcm	ECCO CS510 State Estimate	Progressive	2012
ORCA36	NEMO (4.0)	WOA 2013 (temperature & salinity)	18 months	2014
FIO-COM32	FIO-COM (v2.0_HR32)	June 2016 from FIO-COM 1/10° operational ocean forecast w/ data assimilation	18 months	2018
HYCOM25	HYCOM	WOA 2013		2014

32

Table B2. The horizontal and vertical native coordinate system, spatial resolution and domain coverage for each simulation. The Z^* vertical coordinate is the rescaled geopotential coordinate where the fluctuations of the free surface are taken into account (cf. Griffies et al., 2016). Note that vertical resolution as well as horizontal resolution vary significantly between the models. Outputs from FESOM-GS were interpolated onto a Cartesian grid off-line with a cubic spline.

Simulation	Grid structure	Resolution	Vertical coordinate	Domain (grid points: zonal×meri.)
eNATL60	C-grid	1/60°	Z^* (300 levels)	North Atlantic (8354 × 4729)
GIGATL	C-grid	1 km (nominal)	Terrain following (100 levels)	Atlantic (10500×14000)
HYCOM50	C-grid	1/50°	Hybrid (32 pressure p & isopycnal)	North & Eq. Atlantic (6709×7373)
FESOM-GS	Unstructured	1/2° w/ refinement to 1 km (nominal) in Region 1	Z^* (70 levels)	Global (3000502 vertices)
LLC4320	C-grid	1/48° (nominal)	Z (90 levels)	Global (4320×4320×13 LLC tiles)
ORCA36	C-grid	1/36°	Z^* (75 levels)	Global (12962×9173)
FIO-COM32	B-grid	1/32°	Z^* (57 levels)	Global (11520 × 5504)
HYCOM25	C-grid	1/25°	Hybrid (41 p & isopycnal)	Global (9000 × 7055)

Table B3. The equation of state (EOS), surface boundary layer (SBL) parametrization used, and tidal forcing in each simulation. *Jackett and McDougall (1995, JMD95) in HYCOM is implemented with the approximation by Brydon et al. (1999). The potential densities were computed following each EOS with the reference pressure of 0 dbar (Fernandes, 2014; Abernathey, 2020; Firing et al., 2021). The EOS for FIO-COM32 is available on Github (https://github.com/roxyboy/swot_adac_ogcms/tree/notebook; a DOI will be allocated upon acceptance of the manuscript). Note that FESOM-GS and ORCA36 do not have tidal forcing whilst the others have at least the leading 5 tidal forcings.

Simulation	EOS for density	SBL parametrization	Tidal forcing
eNATL60	TEOS10	TKE	M_2, S_2, N_2, O_1, K_1
GIGATL	JMD95	κ-ϵ closure w/ Canuto A formulation	$M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1$
HYCOM50	JMD95*	KPP	$M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1$
FESOM-GS	EOS80	KPP	N/A
LLC4320	JMD95	KPP	Full lunisolar tidal forcing
ORCA36	EOS80	GLS	N/A
FIO-COM32	preTEOS10	KPP & non-breaking wave induced mixing	$M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1$
HYCOM25	JMD95*	KPP	M_2, S_2, N_2, O_1, K_1

Table B4. The bathymetry configuration of each simulation.

Simulation	Bathymetry
eNATL60	Unsmoothed two-min. Etopo2 file of the National Geophysical Data Center.
GIGATL	SRTM30plus is smoothed using a Gaussian kernel, w/ a width of 4 grid points. Then another step (to avoid pressure gradient errors) is to check that the steepness of the topography does not exceed $r_{max} = \Delta H / H \leq 0.2$ (cf. Le Corre et al., 2020).
HYCOM50	Nearest 5×5 box average of the 15-sec. GEBCO_2019 global dataset to each grid point. Then smoothed once w/ a 1-2-1 9 pt. smoother except within 2 grid points of land.
FESOM-GS	RTopo-2 (Schaffer et al., 2016). Two smoothing cycles by averaging closest grid points.
LLC4320	Unsmoothed Smith and Sandwell (1997) Version 14.1 & IBCAO Version 2.23.
ORCA36	Two paths of Shapiro filter on Etopo08, upon which it is remapped (w/ bi-linear interpolation) onto model grid.
FIO-COM32	A Blackman radial filter (following Arbic et al., 2004) w/ filter radius of about 7 km is used to smooth the GEBCO dataset.
HYCOM25	Nearest 5×5 box average of 30-sec. GEBCO_08 20091120 global dataset to each grid point. Then smoothed once w/ a 1-2-1 9 pt. smoother except within 2 grid points of land.

Table B5. The advection and dissipation scheme used for each simulation. Note that some models have biharmonic viscosities and others do not.

Simulation	Advection scheme (momentum / tracer)	Dissipation scheme (momentum / tracer)
eNATL60	3rd order upwind flux form / 3rd order upwind TVD	Horizontal laplacian / laplacian iso-neutral
GIGATL	3rd order upstream biased flux form / Split and rotated 3rd-order upstream biased	N/A (achieved implicitly via adv. scheme)
HYCOM50	2nd order FCT flux form / 2nd order FCT	Laplacian & biharmonic / laplacian
FESOM-GS	3rd-4th order FCT flux form / 3rd-4th order FCT	Biharmonic (flow aware)
LLC4320	Vector invariant form / 7th order monotonicity preserving	Biharmonic modified Leith / vertical laplacian
ORCA36	3rd order UBS flux form / 4th order FCT	Horizontal laplacian / laplacian iso-neutral
FIO-COM32	2nd order centered flux form / MDPPM	Biharmonic
HYCOM25	2nd order FCT flux form / 2nd order FCT	Laplacian & biharmonic / laplacian

Table B6. The atmospheric and the inclusion of atmospheric pressure variation at the surface.

Simulation	Atmospheric forcing	Atmos. pressure variation (inverse barometer correction)
eNATL60	3-hourly, ERA-interim (DFS5.2) w/ absolute & relative wind stress	Yes (No)
GIGATL	Hourly, CFSR using a bulk formulation w/ relative wind stress	No (No)
HYCOM50	Climatological ERA-40 + 3-hourly wind anomalies from NOGAPS w/ absolute wind stress	No (No)
FESOM-GS	JRA55-do-v1.4.0	No (No)
LLC4320	6-hourly, 0.14° ECMWF analysis starting in 2011	Yes (No)
ORCA36	3-hourly, ECMWF IFS system w/ absolute wind stress, 0.14°	Yes (Yes)
FIO-COM32	3-hourly, NCEP GFS w/ relative wind stress, 0.25°	Yes (Yes)
HYCOM25	3-hourly, NAVGEM w/ relative wind stress, 0.5°	Yes (No)