Response to reviewer 1:

We thank Reviewer 1 for the insightful comments. We agree that the previous version of the manuscript lacked sufficiently clear physical interpretations of several results, and that some model responses were difficult to interpret. Nevertheless, we believe that the study addresses an important topic and contains valuable analyses and results, which will benefit future modelling. Hopefully, better presented with the revised manuscript. In response to the raised comments, we have substantially revised the manuscript as outlined below:

- Redesigned the ARCTIC025-lowGM experiment, which was the root of our unintuitive results. The capping of the K_{GM} field to a maximum of 75 $m^2 s^{-1}$ gave a very uneven reduction of the K_{GM} field with a very strong reduction in the North Atlantic where K_{GM} is high and a much weaker reduction in the Nordic Seas and Arctic Ocean where K_{GM} is low. In our new ARCTIC025-lowGM we instead scale the K_{GM} field (online) to 25% of its computed value. This gives a much more even reduction spatially and robust results with a response that lies in between the ARCTIC025-noGM and ARCTIC025-highGM experiments for all our metrics.
- Retuned the ARCTIC025-GEOM experiment by reducing the eddy efficiency parameter alpha=0.035 to get a reduced K_{GM} field more in line with observations.
- Additional analysis of the barotopic gyre circulation. By introducing topostrophy and a baroclinic streamfunction we show how the vertically integrated thermal wind contribution changes the along isobath flow with GM parameterization.
- Additional analysis of the horizontally integrated heat budget over the Nordic Seas and Arctic Ocean comparing the different terms of the temperature evolution equation and particularly decomposing the advective contributions into mean and resolved/parameterized eddy temperature fluxes. We are also showing maps of resolved and parameterized eddy temperature flux divergence.

The manuscript describes numerical experiments with a regional model of the Arctic Ocean with two different resolutions and different ways of parameterising unresolved eddies with the GM/Redi scheme.

The introduction is well written and easy to read. The manuscript sets out to address important questions of eddy parameterisation in resolutions where the eddy field is partly resolved (eddy-permitting), impact on notoriously poor model skill in Arctic simulations (with often too warm ocean temperatures), but these questions are never clearly formulated and hence the manuscript turns out to be a long and sometimes confusing description of a comparison between the different model simulations and some observations.

To be fair, the presented results are not straightforward to interpret. For example, sometimes higher resolution reduces meridional overturning, in other places it increases it. Using small GM parameters increases meridional overturning over no GM parameterisation in the coarser setup, but using large GM parameters reduces the overturning. In a different place the model with the lowest overturning is the one without GM parameterisation (but not consistently). These phenomena are described but not explained.

These unintuitive responses were mainly due to the design of the ARCTIC025-lowGM experiment, which now has been changed (see above).

Another example: in the Norwegian and Greenland Seas the “worst” model (low resolution, no GM) gives the closest fit to observations (Fig11), better than the “model truth”, but in the Fram Strait and West-Spitzbergen Current it is among the worst (as expected). In the interior Arctic (Fig12), the high resolution “model truth” has the worst fit to observations.

We now compare simulations over larger regions in the Nordic Seas and Eurasian basin by computing horizontally integrated values for temperature trend terms which yields more robust model responses.

I can only speculate, that the choice of experiments makes it very difficult to extract information specifically about eddy fluxes (no eddy fluxes are explicitly computed, neither for the high resolution model nor the parameterised eddy fluxes which could be diagnosed as well) and eddy activity and GM/Redi, as the “model truth” is eddy resolving only in part of the domain, and the change in resolution also introduces confounding factors such as better resolution of topography (coastlines and bathymetry), that will most likely have a profound effect on the simulation. Different spurious mixing in high and low resolution may also be a problem that is not addressed, also the GM/Redi parameterisation introduced a strong vertical mixing contribution.

Yes, we agree. Unfortunately, it is very difficult to quantify spurious mixing and requires specific diagnostic methods to be implemented in the model (e.g., Banerjee et al., 2023; Burchard and Rennau, 2008; Klingbeil et al., 2014). We have added some text on this.

In our new analysis we now compare how the GM/Redi terms contribute to the horizontally integrated heat budget in the simulations.

As a consequence the results section of the manuscript is very difficult to read (and also less carefully written than the very good introduction) and is very difficult to extract any conclusions. The authors leave the reader with little to no explanation for their sometimes unintuitive results (and they acknowledge the unintuitive results in l494 without further explanation). The final section is not much more than a summary. Many important subjects are touched or implied but not properly addressed (e.g. why does higher resolution lead to an additional warming and salinity bias, Fig12, also observed in other models of the Arctic Ocean), and the modelling reader is left with the conclusion that things are difficult with and without GM/Redi and that we need more work (but that’s something that was also so before reading the manuscript).

We agree and hope that our new experiments and analysis will provide more clarity to the reader.

I recommend that the authors focus on fewer “metrics” and not only describe but also interpret the results (where do eddy fluxes, parameterized or not, but diagnosed from the simulations lead to the differences seen in the diagnostics?), also in the light of benefits and shortcomings of the GM-parameterisation in the Arctic. Is GM an appropriate parameterisation for Arctic eddies? What does one need to do to “get it right”? Maybe diagnose GM coefficients from the high-resolution run. I also recommend additional experiments, e.g. a high-resolution simulation with the coarse resolution topography to exclude additional confounders, a GEOMETRIC experiment with smaller values of alpha to get a better agreement with the observational estimates.

We agree that diagnosing GM coefficient from the high-resolution could in principle provide useful insight. However, since our high-resolution run is not fully eddy-resolving we are concerned that such a diagnosis may not be robust. In addition, we have not identified a suitable diagnostic framework. We would therefore appreciate recommendations on suitable methods.

We also agree that a high-resolution simulation using the ARCTIC025 topography would be an excellent addition to exclude additional confounders. We therefore tested this but could not achieve a stable model configuration using similar settings as our ARCTIC12 and ARCTIC025 configurations. Changing numerical settings such as e.g. momentum advection and viscosity formulation could presumably make the configuration stable, but that would introduce additional confounding factors. In addition, our scope is to use as realistic as possible configurations in our simulation to make it transferable to climate models. We have therefore opted to not include such an experiment. Additionally, using a ARCTIC025 topography in ARCTIC12 experiment would also effectively filter out eddy—topography interactions at scales below the ARCTIC025 roughness scale.

Additional comments (sometimes repetitive of the main points):

page 2

l38: AW reach -> reaches? (Since AW abbreviates “Atlantic Water” and not “waters”)

Changed accordingly.

page 4

l62: attration -> attraction (maybe better attention?)

We changed to ’attention’.

page 5

Caption of Fig2: I would use: $L_d/\min(dx,dy)$ (so min => \mathrm{min} Changed accordingly.

Caption of Fig2:

The abbreviations ARCTIC12, ARCTIC025 only become clear later an maybe should be explained also here in the caption

We have now changed the figure caption to ”The first baroclinic deformation (Rossby) radius L_d [km] computed from the 1/12° configuration ARCTIC12 (left), the ratio of L_d / min(dx, dy) for ARCTIC12 (middle) and for the 1/4° configuration ARCTIC025 (right).”

page 6

l127: the -> a?

The way this description is written implicitly assumes that the reader is familiar with the different options of NEMO (in my opinion, ‘the’ vector-invariant formulation, should be ‘a’ vector-invariant formulation, etc.)

Changed accordingly.

l130: a diffusivity coefficient that is scaled by a defined velocity scale Udiff = 0.0193 and L.

Not clear to me how this works exactly (again a little bit o NEMO insider knowledge is required here). Is the coefficient Udifff*L? Or Udiff*L**2 (which would make more sense to me)? Same for viscosity, where I would argue that one needs to include the timestep as well to guarantee stability for large viscosities.

We have tried to clarify this by adding: "The lateral momentum diffusion uses a bi-laplacian operator applied in the horizontal directions together with a viscosity coefficient that is scaled by a defined velocity scale $U_{visc} = 0.185$ and the horizontal model grid scale L as $A_{visc} = U_{visc} L^3$ giving a domain averaged values of -1.8×10^9 and -4.9×10^{10} $m^4 s^{-1}$ for ARCTIC12 and ARCTIC025, respectively. The lateral tracer diffusion uses a laplacian operator applied in the isopycnal directions (Redi scheme) together with a diffusivity coefficient that is scaled by a defined velocity scale $U_{diff} = 0.0193$ and L as $\kappa_{Redi} = U_{diff} L$ giving a domain averaged values of 46 and 136 $m^2 s^{-1}$ for ARCTIC12 and ARCTIC025, respectively."

and an additional paragraph on the time stepping:

"The model configurations are time stepped with a modified leap frog scheme including a weak Robert-Asselin filter and implicit time stepping of the vertical diffusion. Further, a time splitting approach is used to sub-step the barotropic dynamics over the main model (baroclinic) timestep. The main model timesteps (selected to satisfy the conditional stability constraint for bi-laplacian viscosity $|A_{visc}| < L^4 (64 \Delta T)^{-1}$) are 450 and 1350 seconds for ARCTIC12 and ARCTIC025, respectively."

page 7

l140: The background values for vertical diffusivity and viscosity are tuned down to 5.4·10−6 and 5.4·10−7 m2s−1

Probably not important for this paper (although GM + Redi can have a strong vertical contribution to vertical mixing), but I thought that the TKE scheme of NEMO (Gaspar et al, Blance+Delecluse) uses a be minimum TKE value that accounts for the background mixing (due to internal tides) and no explicit vertical background diffusivities are required. If you use them, then the background mixing is introduced twice (according to my understanding, see also Brueggemann et al, 2024, doi:10.1029/2023MS003768)

In the NEMO4.0.4 implementation of the TKE scheme the background values are used as minimal vertical mixing coefficients to avoid numerical instabilities at weak mixing. To clarify this we have added "These background values are used to cut off the vertical mixing coefficients in the NEMO implementation of the TKE scheme to avoid numerical instabilities associated with weak mixing (Madec and NEMO System Team, 2019)."

l150: a landfast ice parameterization is utilized

Here and in general, I would appreciate references for the various schemes (also Flather, Orlanski, etc.)

We have now added the references to Lemiuex et al (2015), Flather (1994), and Marchesiello et al (2001); and the main model references Madec and NEMO System Team (2019) and Vancopenolle and NEMO Sea Ice Working Group (2023).

l154: ARCTIC12 without GM scheme and this serves as a model truth calculation

Model thruth without full eddy field? How so?

We have added the following to describe what our ”model truth” is in the introduction: ”..configurations, one intermediate resolution (1/4 ◦) configuration that is eddy-permitting and one high resolution (1/12 ◦) configuration that serves as a "model truth", although we note that it is not fully eddy-resolving in parts of the Arctic Ocean (see Fig 2).” In the reminder of the text we instead refer to it as ”references” or ”high resolution” experiment.

l161: higher kappa_GM

Than the “standard GM scheme scaling coefficients”?

l166: while the GEOMETRIC scaling mostly yields values that are too high

As far as I can see the GEOMETRIC values could be tuned down by alpha, why didn’t you use a smaller value of alpha to avoid the overestimation of kappa_GM?

We have retuned the experiment with GEOMETRIC scaling (see answer above).

page 9

l180: Dai and Trenberth dataset -> the Dai and Trenberth dataset ?

Changed accordingly.

page 10

l210: However, to facilitate comparison to observational estimates and other modelling studies we use this common computation of "heat" transport.

I would still call this a temperature transport to avoid confusion.

We have now changed the computation of the heat flux so that it is done on region with a closed volume budget.

page 10

L214 analysis -> analyse/analyze

We have changed to ”analyze”

l214: by looking at

Too colloquial to my taste (you can look at a photo or painting etc). I would use something more specific like inspect, investiage, analyse, etc.

We now use ”.. by *investigating* the kinetic energy partitioning.”

page 11

l237: the contribution of EKE drops off significantly in the high resolution experiments

Here it becomes clear that even the 12th deg simulation does not resolve any eddies (in the interior Arctic) and all that EKE shows is kinetic energy of variability on monthly timescales as the EKE is not much higher for Arctic12 than for the other simulations.

There is indeed eddy activity in the interior Arctic Ocean. In the Nansen Basin, the simulated eddy kinetic energy (EKE) distribution is qualitatively consistent with the very high-resolution study of Müller et al. (2024), although with systematically lower magnitudes. We have added: " We note that the high-resolution simulation exhibits an EKE distribution in the Nansen Basin that is qualitatively consistent with the very high–resolution (1 km) model study of Müller et al. (2024), but with systematically lower EKE levels. This likely reflects the fact that our simulation is not fully eddy-resolving in this region (see Fig. 2).".

page 12

l240: I would rewrite as:

The drop off in eddy activity between the experiments and the dampening effect of the GM scheme on the resolved eddy field can also be clearly detected in snapshots of relative vorticity and potential temperature (Figures A1 and A2).

Thank you, we have changed to your formulation.

l256: start -> starts Changed accordingly.

l253 Nordic Seas.

The results are not quite as expected as the high resolution stream function is weaker than the coarse one without GM. Clearly the eddy activity in this region contributes to the overall hydrographic gradients and hence the mean circulation. Something that would be nice to discuss here (but not done)

l262: For the subpolar gyre, it's even different with Arctic12 being the strongest and ARCTIC025-noGM being relatively weak.

page 14

l268: Section 3.3 Meridional overturning circulation

Results are not conclusive and require explanation

We have removed section 3.3, see previous answer.

l288: At the Iceland–

We have now changed to avoid the line break of hyphenated words.

page 16

l298: In our analysis this contribution is instead included in the Faroe–Scotland gateway making our Scotland–Norway inflow lower.

Not clear why the analysis was adjusted according to the observations before comparison.

page 16

We have rephrased this to " In our analysis we have chosen a closed section between Faroe Islands and Scotland instead of only covering the Faroe Bank Channel because it simplifies transport calculations and comparisons between grids of different resolutions. This leads to an offset with higher (lower) values in our Faroe--Scotland (Scotland--Norway) gateway in or analysis compared to observational estimates."

l302: possibly due to better resolving the channels of the gateways.

It’s important to differentiate between the effects of resolution internal dynamics (eddies) and topography.

Yes agreed, but to some degree they are also linked as increased resolution can impact eddy—topography interactions.

l306: “The northward heat transport that AW …”

Although important for the temperature (bias) of AW (in models) and the entire Arctic, the “heat” flux comparison is difficult and potentially inconclusive, because it does not only show the “heat” flux differences but also differences in mean state, if the volume flux is not balanced. It would be possible to have closed volume budget (for fluxes through all straits of the Arctic) in the model, so I strongly recommend to use that property of the model and have “real” heat fluxes.

We have recomputed the heat fluxes for a closed volume budget across the Nordic Seas. The results are not fundamentally different.

page 18

l324: Although the in- and outflows in the model experiments are higher than estimates reported Tsubouchi et al. (2012).

Main clause is missing

We have now changed it to: ”Compared to observations, all model experiments exhibit a higher net volume transport than the 2.0 Sv estimated by \cite{smedsrud2010}, but a lower transport than the 3.6 Sv inverse-model estimate of \cite{tsubouchi2012}, even though the simulated in- and outflows exceed those reported by \cite{tsubouchi2012}.”.

page 19

l348: leads Changed accordingly.

page 20

l374: is a well-defined Changed accordingly.

l379: The GM scheme impacts Changed accordingly.

page 21

Fig9: no observations? There are multiple observations of temperature alone this section, even current velocity (e.g. Beszczynska-Möller et al., 2012). These direct measurements would be easy to compare to (much easier than, say, inferred “heat” or volume flux)

Yes there are up to 17 moorings in the Fram Strait section but they do not cover the full strait, instruments are deploy during different time periods and at varied depths depending on the period, so it requires expertise in, and extensive of processing of, observational data. We have instead opted to include a recently published dataset of observational estimates based on a combination of existing observations and GLORYS12 reanalysis.

l389: equals 233 meter

Earlier (l221) a different depth (623m) was chosen to identify the AW layer for the MKE and EKE plots in Fig4.

We now use the same depth (233 meter) for temperature and MKE/EKE plots.

page 23

l425: observed values.

Not really "observed" but inferred from observations and an inverse model, so maybe "observation-based estimates"? Changed accordingly.

page 25

show more uniform differences

l454: Also in the entire Eurasian basin is the AW layer thickness over estimated with model experiments showing 980−1370 compared to 630 in PHC3.0.

Grammar (the AWI layer thickness is overestimated); units (meters) are missing

This paragraph has been removed in favour of new analysis.

page 27

l472: The meridional overturning is weaker in the northern North Atlantic using GM, however, no sensitivity is seen in the Nordic Seas overturning

But not in a consistent way!

This sentence have been removed since we removed the meridional overturning metric.

l483: omitting the GM scheme yields too wide AW core

Counterintuitive as GM should flatten isopycnal, here is seems that omitting GM flattens the isopycnal and stretches the AW across the entire strait. Not explained.

Our new results now show a flattening of sloping isopycnals and reduction of thermal wind shear in the both the inflow branch (West Spitzbergen Current) and the outflow branch (East Greenland Current); and the flattening increase with GM strength.

AW code" or yields an AW core that is too wide"

page 28

l503: It should be noted that we did not spend a great deal of effort to tune the new GEOMETRIC scheme.

Maybe this would have been necessary to obtain a more realistic GM-parameter field?

Following our retuning of eddy efficiency parameter α in GEMOETRIC we have now removed this statement.

l505: observational estimates by Kusters et al. (2025) show that the GM diffusivities have a vertical structure with reduction of diffusivities with depth in some regions.

I am not sure if that's the best conclusion from their paper. It rather shows that the assumption of an eddy parameterisation with a uniform GM coefficient may not be appropriate. The fact that the coefficient is not uniform says that the parameterisation is not complete.

We have removed this statement.

page 29

l530: Data availability.

No available data from the simulation results

We will upload the data from the simulations on Zenodo.

page 31

l602: Kusters, N., Balwada, D., and Groeskamp, S.

Incomplete reference, no journal.

We have added the correct reference "Kusters, N., Balwada, D., and Groeskamp, S.: Global Observational Estimates of Mesoscale Eddy-Driven Quasi-Stokes Velocity and Buoyancy Diffusivity, Geophysical Research Letters, 52, e2025GL115 802, https://doi.org/10.1029/2025GL115802, 2025."

Per Pemberton,

On behalf of the other co-authors