Reply to Referee Comment #1 (RC1):

The authors would like to thank the referee for the invaluable comments and suggestions. The following are the replies for each point of the comment, together with specific revisions that are made. The original comments are in *blue italic* font and listed in paragraphs, with our reply following each paragraph separately. The revisions are also highlighted in the revised manuscript in blue and marked by ***RC1***.

I find the approach proposed in the manuscript to be of interest for the journal. It follows the now discussed road of separation of concerns, whereby the code part dealing with numerical algorithms and the part dealing with the infrastructure (mesh, parallelization) are separated and treated in different ways. However, the manuscript in its present form misses the goal: one expects that the material of the manuscript describing a modeling approach will be sufficient for a reader to learn how to use the approach. I do not see that this goal is reached.

Reply: the authors thank the referee's comment on the relevance of our contribution to the journal. Regarding the comment on the main purpose of the manuscript, we would like to make the following clarifications and corresponding revisions to the manuscript.

First, we would like to clarify that we have multiple purposes for the manuscript: (1) we demonstrate that by refactoring existing model (here NEMO) with JASMIN we can attain extra functionality, while breaking the `silo'-type model development in which modelers make (almost) everything by themselves; (2) the resulting model OMARE is capable to simulate realistic ocean processes with the AMR function working as intended; (3) the introduction of the actual refactorization process, using NEMO and JASMIN as an example. We consider OMARE a worthy try in (re-)constructing the model with a software middleware and satisfactory results are achieved. We agree with the referee that the technical details of how to refactorize the model is an important aspect, but we consider it one of these goals. In response to the referee's suggestion, we provide a technical guide as supplementary material for using JASMIN and compiling/running OMARE (see also the reply to the next paragraph). The text referring to the new technical guide is also added to the revised version of the manuscript.

In order to better explain the overall design idea of OMARE, we intend to add the following figure of the model structure of OMARE and differentiate it with NEMO. The figure, together with the added paragraph to the manuscript is now included in the beginning of Sec. 2 of the revised version of the manuscript.

NEMO, as a typical ocean model, consists of a layered structure. Like many models, it relies on an intermediate layer that contains both a self-developed parallelization software solution, and third-party add-on's (AGRIF, XIOS). As a consequence, certain limitations exist, including limited flexibility in parallelization and adaptivity, especially given that these issues are inherently intertwined. OMARE, on the contrary, relies on JASMIN for managing the parallelization, adaptive refinement, and parallel I/O. Therefore, no extra effort is spent on designing/building the intermediate layer specifically for the model, hence no software `silo' is constructed. More importantly, the model does not suffer from the aforementioned limitations.

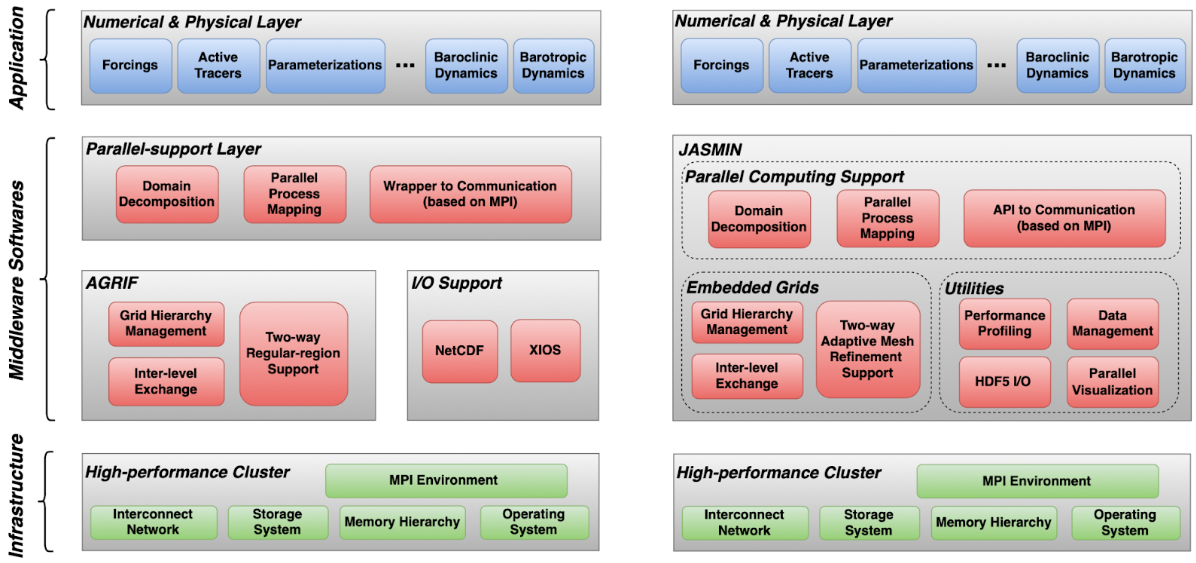

Figure 1. Model structure and abstraction layers of NEMO (left) and OMARE (right).

Of course, there are also limitations associated with OMARE, such as the lack of direct support for accelerator architectures (e.g., GPUs) in JASMIN. Besides, there are potential usability issues due to the current binary release form of JASMIN, which is an issue also raised by other editors. We are aware that these detailed issues should be addressed during future development of OMARE.

The text on lines 165 - 190 shortly describes how the JASMIN is involved, but I doubt a reader can get any understanding of what and why is done. Moreover, it is not at all clear how to use JASMIN in conjunction with the updated code. How the JASMIN environment can be installed, how code is compiled, etc. The description should be essentially extended and be such that those who are willing to follow author's approach can do it.

Reply: we agree with the referee on the importance of the technical aspects (of how JASMIN and the application work together). We do consider the detailed content will be too much to be included in the main part of the paper, considering the overall purpose of the work, as well as the limited relevance of these details only to model developers. Therefore, we supply a concise manual for the installation of JASMIN and the compilation and running of OMARE in the supplementary. The text here is also revised with the addition of the reference to this manual.

The manuscript devotes more than a half of its volume to the description of simple test configuration, going into too much detail, which is hardly optimal. The test case remains the test case, and one can only learn that the approach proposed by the authors is working, yet not without drawbacks related to one-way nesting (the development of errors on fine-coarse boundary). I do not think that this test case is well suited to demonstrate the need of adaptivity. Figure 11 shows clearly that small-scale turbulence occupies 3/4 domain on full mesh, and it occupies only pieces where the resolution is refined in b and c. It is different from the initial phases in Fig. 9 and 10, but 5 or 20 days is a too short time for turbulence to equilibrate, and this transient phase is of no interest (it depends on coarse initial conditions, and does not model any reality). So the conclusion here is that dynamic adaptivity is an interesting, but perhaps not very needed possibility as concerns eddying flows. Static refinement might be doing the work, and one will take a decision where to resolve based on one interest. I foresee, however, one direction, where dynamic refinement still might be of interest -- the simulations of seasonal course of variability. Submesoscale eddies might be suppressed in warm seasons, and a coarser mesh will be sufficient for mesoscale. My recommendation are to make the experimental part more compact. One can hardly learn anything from detailed description of particular eddy features or the comparison of transects (Fig. 13, 14), and there is very little sense in Fig. 9 and 10.

Reply: we would like to thank the referee for the comment on the test case. We would like to make the following replies and revisions to the manuscript.

First, on the choice of Double-Gyre testcase. The Double-Gyre testcase is chosen as a testbed for the major functionalities of OMARE. The case is of intermediate complexity: (1) the model physics is complete, making it capable to produce realistic three-dimensional large-scale ocean processes; (2) it produces an idealized Western Boundary Current system and in particular, the seasonality of submesoscale processes; (3) the testcase omits realistic (or any) bathymetry, therefore avoiding the complex issues such as inconsistent bathymetry between the non-refined and the refined regions. We agree with the referee that the Double-Gyre is a choice among many possible ones. And we chose it mainly because that the aforementioned 3 characteristics. Similar Double-Gyre cases or even more complex idealized cases such as NeverWorld2 (Marques et al., 2022) are also commonly used in the community for model benchmarking or the study of certain processes.

Second, on the demonstration of the need for adaptivity. We agree with the referee that small-scale turbulence is prevalent in the basin for full-basin 0.02-deg experiments, which is actually expected especially during winter. Here we would like to emphasize our perspective on adaptive refinement: where & when we need adaptivity and AMR is an open issue, and the purpose of OMARE is to provide a framework that supports various possible AMR scenarios.

In the Double-Gyre case, the major simulated features include the WBC, the lateral boundaries of the basin, and the associated mesoscale-submesoscale processes. <u>The refinement criteria are designed to capture the kinetically active and submesoscale</u>

processes, and hence based on surface velocity and surface relative vorticity. The two AMR cases have already been shown to capture a majority of the basin's (surface) kinetic energy (Fig. 8), which directly indicates the validity of AMR.

A further proof is provided below in Fig. 2 for the vertical motion induced by submesoscale flow. The enhanced vertical motion and the associated heat (and other tracer) transport is a key characteristics of the submesoscale processes (Taylor & Thompson, 2022). The following figure shows that: (1) the strong vertical motion is strongly concentrated at WBC and some other regions (i.e., ocean fronts, basin's boundary), and (2) the two AMR cases capture these key regions for strong vertical motion. A further investigation of the PDF of vertical speed in Fig. 3 confirms that the strong vertical motion in AMR experiments matches closely with the full-field 0.02-deg experiment.

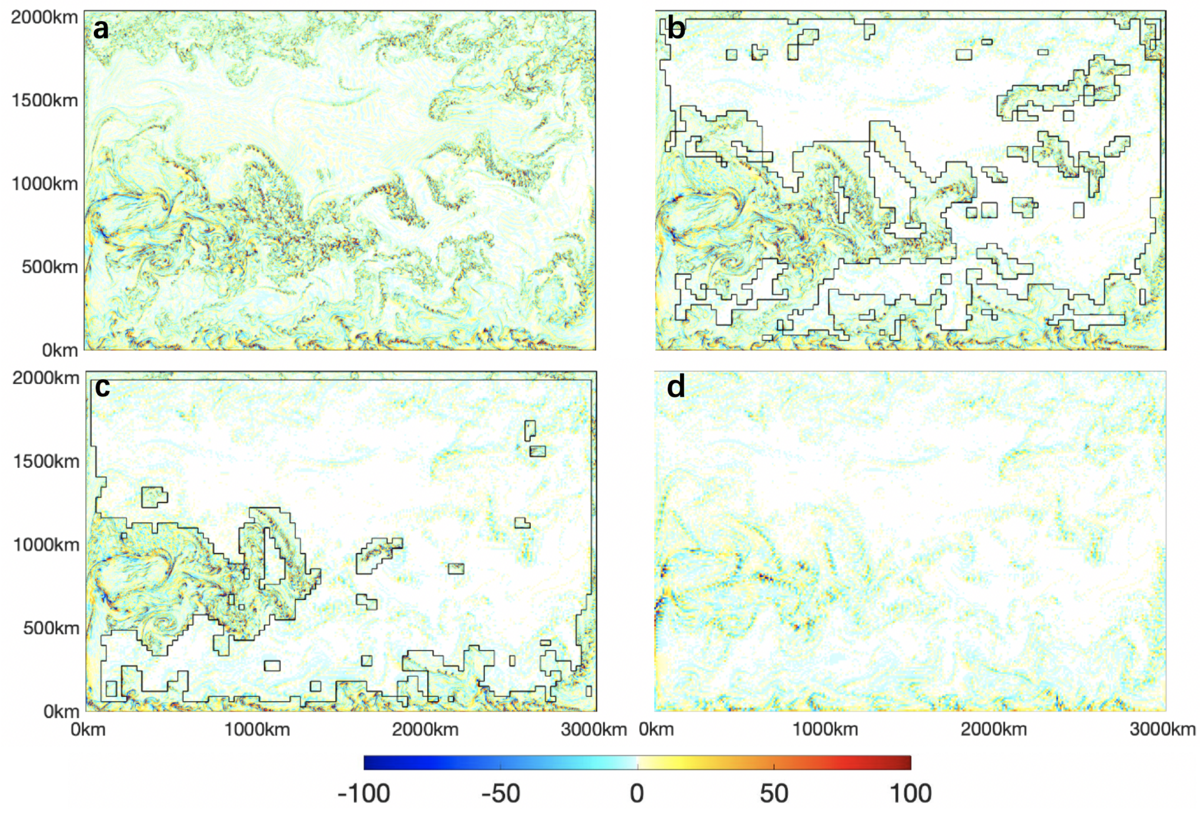

Figure 2. Vertical velocity (w, in m/d) at the 5-th layer after 20 days since Feb.-1st for all experiments. The depth is about 40m. Panel a, b and c shows the result for full-field 0.02˚ experiment (i.e., **S**), that of AMR setting 1 (i.e., **M-S-I**), and that of AMR setting 2 (i.e., **M-S-II**), respectively. In both AMR experiments (panel b and c), the boundary between 0.02˚ and 0.1˚ is marked by black lines. Panel d shows the reference simulation (i.e., **M**) result at 0.1˚ on the same day.

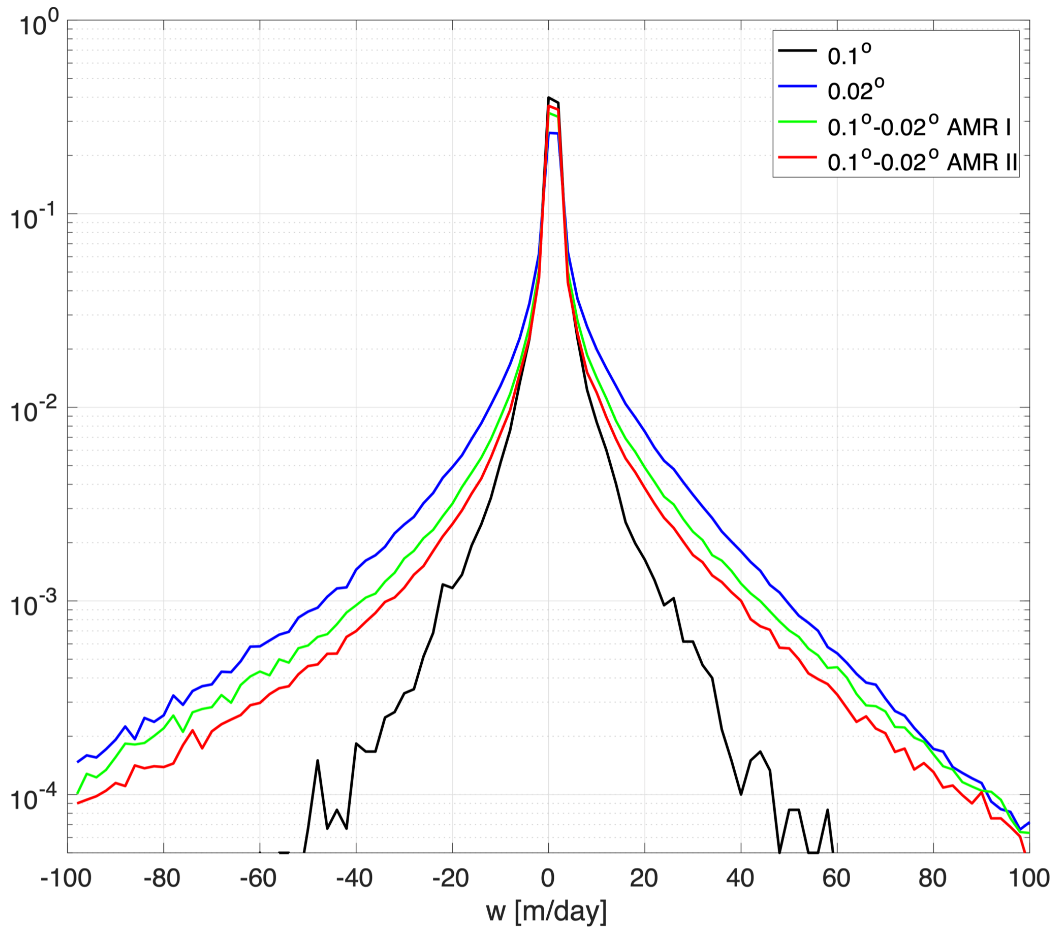

Figure 3. Distribution of vertical velocity (w) of the four experiments in Fig. 2.

Regarding the referee's suggestion on utilizing the seasonal variability to demonstrate adaptivity, we totally agree and consider it a great suggestion to our work. Actually we observe certain seasonality of the overall kinetic energy in our AMR experiments (Fig. 8) and that in the ratio of the refinement region. For 0.1-deg run, the mean KE peaks in summer (Jul-Aug), while the full-field 0.02-deg run definitely shows a different time of the KE peak during the seasonal cycle. Also evident is the seasonally enhanced submesoscale features with 0.02-deg. Therefore we argue that the Double-Gyre could serve as a typical case for demonstrating OMARE's adaptivity capabilities.

To summarize, the criteria of adaptive refinement, in our opinion, should be process dependent as well as study dependent. For example, mesoscale/submesoscale feature identification and tracking is another possible choice for our planned use of the model. <u>In this study, the criteria based on velocity and vorticity, in our opinion, fully served the purpose of demonstrating the capability of AMR and capturing the dynamically changing WBC system</u>. With the suggestions of the referee, we plan to fully explore the design of adaptive refinement for both one-way and two-way (i.e., with feedback) in the future.

Third, on the 5th-, 20-th and 50-th day results of the wintertime AMR run. We agree with the referee that the model status immediately after the refinement to 0.02-deg does not contain valid physics at 0.02-deg, since the spin-up has not fully finished. We want to emphasize that the purpose of showing results on the Feb. 5-th (5 days after refinement) is to: (1) demonstrate that the AMR criteria does capture the instantaneous WBC pattern, and (2) illustrate the spin-up process within the refined regions for all three refined experiments, and the consistency among them. Also it (Fig. 9) compares nicely with that after 20 days' refinement and that after 50 days, showing specific temporal changes in the refined regions. Therefore the comparison across the whole development process further demonstrates that the model captures the dynamically changing WBC with the AMR capability.

According to the referee's comment to make it more concise, we now shorten the result description on Feb. 5th and 20th, but keep the figure (i.e., Fig. 9). We would also like to notice that for the summer case, we only include the 50th-day results.

Please also check the text: there are numerous cases when plural/singular and some terms (e.g. kinematics) are used inappropriately.

Reply: we have gone through the text for corrections of grammatical errors, including the 3 cases of misuse of "kinematics".