Review of "MSR v1.0: A High-Resolution Ocean Parameterization Approach via Multiphysics Super-Resolution" by Fuhua Zhu et al. submitted to GMD

June 4, 2026

This paper proposes a method for estimating the high-resolution ocean fields from the coarse data. Novel contributions include various ways to enhance reconstruction of the small-scale features using Dynamic Enhancement Feature (DEF), High-Frequency Enhancement (HFE), Multi-Branch refinement, and physics-informed loss function. This paper represents a general interest, however, it requires considerable rewriting and clarification before publication. That is why I recommend to reject and encourage resubmission of the updated manuscript as a separate paper.

Major comments

1. Scope of the paper. This paper misses a sharp focus of the study. There are way too many topics and details mentioned to truly understand what are the original contributions of the paper.

   (a) I acknowledge that this work was partially inspired by other works on parameterizations, but what authors propose is not a parameterization per se. Thus, "Ocean Parameterization Approach" should be removed from the title and the claim to build a parameterization should be removed from the main text as well. In my belief, authors at best use a well-known parameterization (in this case, Zanna Bolton 2020), in order to come up with a nice loss function which promotes a certain degree of smoothness and consistency among reconstructed high-resolution fields.

   (b) The goal of the high-resolution reconstruction is unclear. It would be nice to relate the suggested reconstruction to any well-known problem in geosciences, such as: ocean state estimation, subgrid parameterization, and so on. As I already said, the connection to subgrid parameterizations is vague. I acknowledge that such connection can be strengthened in future work via deconvolution methods (S Stolz, NA Adams 1999 and Gultekin, C., Subel, A., Zhang, C., Leibovich, M., Perezhogin, P., Adcroft, A., ... & Zanna, L. (2024)). However, at the moment this connection is clearly missing and there is no space in the paper to establish it. Also, I would not recommend following this direction as deconvolution methods are known to have their own issues. The connection to ocean state estimation is also unclear. All quantities authors predict, i.e. velocity gradient tensor, ZB20 stress tensor and its divergence, are derived quantities and do not represent a direct interest in a context of ocean state estimation. Ocean state is roughly an initial condition which can be passed to the ocean model in order to solve the initial value problem. Thus, ocean state is given by ocean velocities, sea surface height, and temperature and salinity. The proposed algorithm can be modified to reconstruct ocean velocities, while all variables predicted right now can be used to ensure that the ocean state represents a certain degree of smoothness and consistency. To sum up, a better justification for the choice of the prediction targets should be given.

(c) I encourage authors to be succinct in presenting their main results and keep only those parts of the architecture that are actually needed.

i. There are many things which can be removed from the paper. For example, MSE, RMSE and coefficient of determination are related to each other through monotonic transformations. I.e., the model which is best in R2, is also best in MSE and RMSE. This means that 66% of all panels in figures 10,11, 12 and 14 can be removed. The remaining parts of the paper should be made more succinct as well.

ii. Authors attempt to use simultaneously as many approaches as possible and suggest way too many new contributions: there are more than 10 little things which make the whole thing work (Haar, Sobel, FDE, Laplacian, Coarse, DEF, Physical loss, Spectral loss, Pixel loss, Multi-branch, PFE, Encode-Decoder). *This is not a generally favorable way of doing science.* That level of model complexity would be justified, if it also gives unprecedented results. However, according to Table 1, the proposed method is only 25% more accurate on average than one of the baselines (EDSR). I expect that simple hyperparameter tuning in EDSR can give a similar improvement in the performance. I would encourage to present only those results which are the most impactful. From Figure 13, Harr and FDE are clearly two the most impactful architectural choices. Further, Figure 14 cannot be objectively used to determine the best architecture as the validation metric (MSE) represents only one component of the multi-objective loss function. Further justification beyond MSE should be given for the choice of the loss function. Authors can consider reconstructing velocity fields and evaluating the interscale energy transfer as a validation metric not included in the loss function. See, for example, https://arxiv.org/pdf/2509.18741.

iii. Many approaches in the paper are repetative. For example, computing derivatives (PFE block) is similar to computing gradients (SGC block) and physical consistency loss. Further, Haar wavelets extract frequency information same as various Fourier transforms used in the paper (block FDE, spectral loss). Also, coarse block serves a similar function as encoder-decoder and main branch. I would enourage to use each general idea only once or at least implement it in a similar way accross the manuscript to reduce redundancy.

2. Significance of the results.

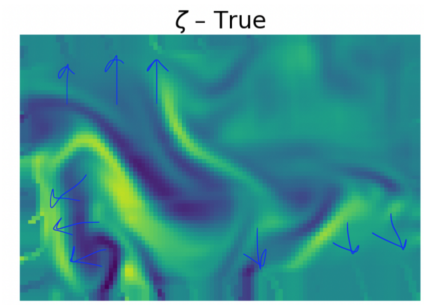
   (a) Authors loosely use word ”generalization” without explaining what kind of generalization is meant. My impression is that in Experiment 2, the model is retrained in a new region. This means, that the same architecture (but not weights and biases) can be used in different regions, which is not a very strong property. In geosciences, generalization commonly implies a different thing – when model is trained on one region and tested on another region. Is it possible to demonstrate that type of generalization, i.e. out of distribution?

   (b) I believe that the proposed architecture was developed as a result of fine tuning for acomplishing a particular task. A little change in the setup or considering a potential application might showcase that some blocks are not that well suited for super-resolution. For example, enhancement of high-frequency content might be sensitivite to the noise in the input data which is often present in real observational data.

Minor comments

1. It should be explained that one of baselines (HighResUNet) learns identity mapping and contains true answers as its inputs. If this is not properly explained, the results of the paper seem to be less significant and it is unclear why this baseline does not appear in all Experiments.

2. I would encourange to check that sharp edges present in Figure 7 have physical meaning as they look strange, see arrows:



3. In Figure 12 important results are behind the legend.