Authors' replies to RC2

All coauthors thank this reviewer for his/her careful examination and constructive comments. We will follow the reviewer's comments to revise the manuscript and fully address her/his concerns. The following are the point-by-point revision plans, in which we list what we will do in the revision of the manuscript to address each of comments of this reviewer.

General comments:

The manuscript presents a Python-Fortran hybrid programming framework (Hf2pyMDA) designed to enable a more straight-forward integration between AI algorithms (python) and physics-based models (fortran). The framework makes use of f2py to conduct 2-way interaction between Python and Fortran components. The authors test their system using a coupled DA case based on VAE.

Overall, I do believe the topic is timely and relevant, given the growing interest in ML and our vast knowledge of physical modeling and DA. However, I found several issues in clarity, structure, and articulation of the main contributions and so I'm recommending rejection.

1. The manuscript is very difficult to follow and I don't say that lightly. The authors have put in a lot of work into the manuscript yet the presentation significantly limits accessibility and understanding. If this is to be revised, the authors need to put in substantial effort on restructuring and language revision. Here are some of the issues I faced while reading:

a. The text contains many long, complex, and grammatically awkward (sometimes wrong) sentences.

RE: We thank the reviewer for the careful examination. We will perform thorough language examination during the revision and improve the English presentation of the manuscript.

b. Key ideas are introduced rather abruptly without sufficient explanation/context.

RE: We will reorganize the presentation to improve its structure and readability.

c. The logical progression between sections is unclear/misleading with frequent back-and-forth between concepts (e.g., infrastructure, DA method, ML details).

RE: Connecting with 1b, we will pay attention to the logical progression between sections and make silky transitions.

2. I also have an issue with the novelty of the work. At first, I thought this needs a software engineer (rather than a scientist like me) to understand all of the details. But then I saw an integration workflow and later a VAE application with strongly coupled DA. Yet, in all of these components I struggled to find novelty. The use of f2py is well-established. How is this different from existing Python-Fortran coupling approaches? You didn't explain the relationship between the infrastructure and the SCDA application. Imagine I want to adapt this wrapper to an already existing DA library (say PDAF or DART); what are the steps needed? Etc

RE: We understand the reviewer's concern. In the revision, we will clearly present the novelty of the study. When we reorganize the materials, we will organize most of the engineering details in Supplementary Information, and focus on scientific innovation in the main context to address the novelty.

We plan to revise the title to something like "Python-Fortran Hybrid Programming for Deep Incorporation of AI and Physical Models: Examples of AI-LDA applied to climate and weather models,and restructure the presentation accordingly, so that our novel contribution is clearly presented. To that end, we will clarify innovative aspects of integrating large Fortran-based weather/climate modeling and DA systems and Python-based AI algorithms, giving a brief explanation on the specific role of the f2py-based workflow in enabling this integration, while leaving detailed implementation steps to the supplementary information.

3. The presented results show modest improvements (around 4% in some cases), but the significance of these improvements is not discussed in depth. It's unclear whether the improvements are robust across different configurations/datasets. I suggest discussing the statistical significance (probabilistic metrics other than RMSE) and practical impacts of the results. Also, I'd include some text describing any limitations of the approach.

RE: Good advice! We will perform more statistical analyses to demonstrate the robustness of improvement achieved by our new approach and discuss practical impacts of the results in the revision.

4. Figure 6 is kind of hopeless. There is a lot of data and numbers on the figure, making it impossible to read or understand the training procedure. I think this should be simplified or split into a figure and a table.

RE: We agree with the reviewer's comments on figures. We will improve figure presentation in the revision, especially Figures 6 and 11. We will re-plot them to make them more informative and understandable.

5. Line 310: Can you describe more in detail the Cross Attention Step? This seems to be an important detail within the general VAE framework.

RE: We will take care of this concern in the revision. Thanks.

6. Lines 345-347: I am confused on what is the background vs obs loss. The text contradicts the figure. In any case, I was expecting to decrease both loss functions during minimization. Why is the orange curve (not sure if it's obs or background) increasing?

RE: Sorry for the vague statements. We will make a clear description about the behaviors of background loss and observation loss in the revision.

7. Section 4.2.1: Recompiling the model to make subroutine callable in a separate module seems intrusive to me. The beauty about ensemble DA systems is that they treat the physical models as black boxes. But now, it seems with the addition of ML this property is no longer available. In addition, some models cannot be recompiled into a subroutine callable library, so what happens in that case?

RE: In DA applications, the incorporation of ML algorithm into an existing DA system introduces additional increments into the DA results, without changing the numerical model itself. In this study, we incorporate AI-LDA into a high-efficiency multiscale approximate (EnOI-like) filtering DA system, which samples a multiscale ensemble from a single historical time series updated with the DA procedure. We will describe these aspects clearly in the revision. Thanks for pointing out this!

8. What is the classic multiscale DA? Please provide some details or references. There are many approaches to multi-scale DA.

RE: In connection with Comment 7 above, we will clearly describe the multiscale DA scheme we used in this study.