

## Authors' replies to RC1

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 our point-by-point revision plans, in which we list what we will do to revise the manuscript to address each of comments of this reviewer.

General comments:

This manuscript uses f2py to couple machine-learning-based latent data assimilation (LDA) with traditional numerical models. The results suggest that the proposed hybrid system can run stably and achieves performance comparable to conventional methods. The topic is timely and potentially valuable, as LDA has shown promising advantages over the model-space DA methods, yet it has rarely been deployed within traditional numerical models. However, the manuscript currently suffers from substantial issues in terms of topic framing, experimental design, algorithmic clarity, and writing quality. Overall, I believe the study has potential, but major revisions are required before it can be considered for publication.

**1. The current scope and framing of the manuscript are not well aligned with its actual contribution.**

(1) The title emphasizes *AI and Physics Modeling and Data Assimilation*, but the manuscript is in practice mainly concerned with coupling ML-based LDA with traditional numerical models.

RE: We agree that the current title the current title does not accurately reflect the context of our study. 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," so that the title better aligns with our actual contribution.

25 (2) The introduction highlights the promise of AI for physical modeling, the main text does not present a concrete example in which AI is used to improve the physical model itself. In addition, the introduction does not clearly explain why ML-based DA is needed in this setting.

RE: We will add more analyses of results to address how AI improves the scientific representation by deeply incorporating the AI algorithm into a traditional DA system. Also, we  
30 will add more statements in the introduction to explain why ML-based DA is needed our case studies.

(3) Section 2 mainly describes Python–Fortran interfacing and f2py-based wrapping in a fairly general way, rather than explaining how AI-based physical modeling and data assimilation are  
35 implemented within the numerical model.

RE: Thank you for pointing this out. We will add the context to describe how AI-based physical modeling and data assimilation are implemented within the numerical model.

I suggest that the manuscript be reframed so that its main focus is clearly defined as the  
40 integration of LDA into traditional numerical modeling systems. In particular, Section 2 should focus more on how LDA is implemented and integrated with the traditional model. LDA is an algorithmic framework, and its current description is too vague.

RE: Thanks a lot for the suggestion and we will revise the MS accordingly.

45 **2. The description and implementation of the LDA algorithm require substantial clarification and revision.**

(1) The description of the algorithm is too limited to support reproducibility. For example, it is unclear how the latent-space background error covariance is specified or estimated.

RE: We will make a clear description of the LDA algorithm in the revision, and discuss how  
50 the latent-space background error covariance is estimated, as well as its impact.

(2) The manuscript should explicitly state that the method used here is Latent-3DVar (L3DVar), since LDA is a broader framework that includes multiple algorithmic variants.

55 RE: Thanks for pointing out this. We will explicitly state that the method used here is Latent-3DVar (L3DVar).

(3) The LDA equations should be presented in the main text (Section 2) rather than embedded in a figure.

60 RE: Thanks. We will do it accordingly.

(4) The notation for the background error covariance is conceptually inconsistent. In Fig. 5,  $B$  is defined as the model-space background error covariance matrix, but the same notation is also used for the latent-space background error covariance in the LDA formulation. To avoid confusion and remain consistent with the current LDA work, the latent-space covariance should  
65 be denoted by  $B_z$ .

RE: Thanks. We will do it accordingly.

(5) A key advantage of applying LDA to high-dimensional atmospheric DA is that the  $B_z$  can become approximately diagonal in latent space. By checking the code, I found that the  
70 manuscript sets this covariance to the identity matrix empirically, which may be problematic. The authors should revise the treatment of  $B_z$  in their experiments and clearly explain how it is estimated in the manuscript. I recommend estimating  $B_z$  with NMC method or from ensembles.

75 RE: Regarding this comment, we have conducted experiments to estimate  $B_z$  using the NMC method, and found that, consistent with the reviewer's advice, it is nearly diagonal. We will provide further discussions of the estimated  $B_z$  in our following replies for 3(2) (CM2-LDA) and 3(3) (WRF-LDA). Thanks for advice on this issue that brings us more thought on the property of  $B_z$ . and we will clarify this issue in our revision accordingly.

(6) The citation of LDA is incomplete. To my knowledge, [1] is the first study showing that  $B_z$  is nearly diagonal in a high-dimensional atmospheric setting, and [2] is the first extending LDA to multivariate atmospheric DA while also showing the same property. In my view, both references should be cited. Moreover, the citation to Fan around Line 245 should refer to Part 2 ([4]) rather than Part 1 ([3]).

RE: Thanks a lot for pointing out this. We will revise the statements accordingly.

(7) In Section 4, the autoencoder is trained using forecast data rather than the reanalysis data more commonly used in LDA studies. This choice should be explained more clearly. I suggest referring to [2] and [3], which both use forecast data to train the autoencoder.

RE: We will clarify this point given the specific motivation of our studies in the revision.

(8) The authors should not refer to their model as a beta-VAE. In the standard usage, a beta-VAE is a generative model with an increased weight on the KL term (i.e., KL weight  $> 1$ ). In the present case, it would be more appropriate to refer to the model simply as a VAE, or more accurately as an AE with slight KL regularization. I consider the latter description to be the most precise.

RE: Thanks a lot! This comment brought us more thought on the concept of beta-VAE, VAE and AE. We will revise relevant statements in the revision.

(9) As discussed in [2], the validity of L3DVar relies on conditions such as the approximately affine behavior of the decoder. The authors should clarify this point in the manuscript.

RE: Thanks. We will clarify this point in the revision.

[1] B. Melinc, Ž. Zaplotnik, 3D-Var data assimilation using a variational autoencoder.

Quarterly Journal of the Royal Meteorological Society 150, 2273–2295 (2024).

105 [2] H. Fan, L. Bai, B. Fei, Y. Xiao, K. Chen, Y. Liu, Y. Qu, F. Ling, P. Gentine, Physically consistent global atmospheric data assimilation with machine learning in latent space. *Sci. Adv.* 12 (2026).

[3] H. Fan, Y. Liu, Z. Huo, Y. Liu, Y. Shi, Y. Li, A Novel Latent Space Data Assimilation Framework with Autoencoder-Observation to Latent Space (AE-O2L) Network. Part I: The Observation-Only Analysis Method. *Monthly Weather Review* 153, 1335–1348 (2025).

110 [4] H. Fan, Y. Liu, Y. Liu, Z. Huo, B. Chen, Y. Qin, A Novel Latent Space Data Assimilation Framework with Autoencoder-Observation to Latent Space (AE-O2L) Network. Part II: Observation and Background Assimilation with Interpretability. *Monthly Weather Review* 153, 1349–1363 (2025).

115 **3. The experimental design is too vague, and in my view the resulting conclusions may not be sufficiently reliable, as the LDA method does not appear to have been used appropriately.**

(1) In both cases, LDA is compared with EnKF, which is ensemble-based. However, I do not see a corresponding ensemble-based design for estimating the background covariance in the LDA implementation. This does not seem to provide a fully fair comparison.

RE: We incorporate AI-LDA into a high-efficiency multiscale approximate (EnOI-like) filtering DA system, which samples multiscale ensemble from a single historical time series updated with the DA procedure. We will make a clear description about the DA system and try to provide scientifically fair analyses for the results in the revision.

125 (2) As noted above, the latent-space background covariance  $B_z$  does not appear to have been properly estimated, which may lead to poor LDA performance.

RE: We initially estimate the latent-space background correlation of CM2-LDA, and find that the correlation coefficient is nearly diagonal. We will use the newly-estimated  $B_z$  to test and re-validate our experiment results.

130

(3) In the WRF experiment, the compression ratio appears to be too small, which may mean that the background error covariance in latent space is not close to diagonal. In that case, using a diagonal  $B_z$  may introduce substantial assimilation error.

135 RE: Thanks for pointing out this. We initially estimate the latent-space background correlation of WRF-LDA, and find that the correlation coefficient is nearly diagonal. Regarding the issue of compression ratio, we will examine how to enhance compression ratio trading off with reconstruction accuracy of high-frequency signals.

(4) The manuscript does not clearly explain how the observation error covariance  $R$  is defined.

140 RE: We will clarify this in the revision. Thanks.

(5) The paper should include a comparison of the computational efficiency of the conventional DA method and LDA.

145 RE: Thanks for pointing out the issue of DA efficiency. In our cases, we donot have the same DA task that is conducted by conventional DA and LDA in a parallel fashion. Instead, we use LDA as an additional DA task (SCDA for climate model, for instance) that is incorporated into the traditional DA system. Although we can't directly compare the computational efficiency of conventional DA and AI-LDA, we still can discuss the efficiency issue in the revision by analyzing the additional cost of the LDA component in terms of percentage of added cost and  
150 benefit.

(6) The manuscript does not provide sufficient statistical evaluation of the autoencoder reconstruction error, but only reports a few individual cases.

155 RE: We will add more analyses and discussions on statistical evaluation of the autoencoder reconstruction error in the revision. Thanks.

(7) The naming of the LDA algorithm should be kept consistent across the different experiments, or at least be made as consistent as possible. The current names for LDA methods are confusing.

160 RE: Thanks a lot! We will take care of this in the revision.

**4. The quality and presentation of the figures require substantial improvement in the revised manuscript.**

(1) Font styles are inconsistent across figures, which gives the figure set an unpolished appearance. A unified font style and size scheme should be used throughout.

RE: We will make them consistent in the revision. Thanks!

(2) Figures 7 and 8 appear to present case studies or example events. If so, the corresponding dates or timestamps should be explicitly labeled.

170 RE: Thanks. We will clarify this.

(3) Several figures are devoted to presenting AE reconstruction errors (e.g., Figures 7, 8, 12, and 13). Some should be moved to the Appendix.

RE: We will do so accordingly. Thanks.

175

(4) Figures 6 and 11 are difficult to read, with text that is too small and a layout that appears overly simplistic. The AE architecture is best presented graphically rather than described only in words. The authors can refer to the presentation style used in existing LDA studies.

RE: We will follow this suggestion and improve their clarity. Thanks a lot!

180

(5) In some figures, the color bars are placed too tightly relative to the main panels, which affects readability. Figure 7 is one example.

RE: Thank you for pointing this out.! We will modify the figures to improve their clarity.

185 (6) The color bars in Figure 9b and 9c appear to be incorrect and should be checked carefully.

RE: Thanks. We will do it.

(7) In Figure 10, the borders around the subpanels should be removed, and the subpanel labels should not be placed directly on the images with a white background box. The authors can refer  
190 to recently published papers in this journal for more standard presentation styles.

RE: Yes, we will re-plot the figure to improve its clarity.

(8) In my view, Figure 2 could be removed and its content briefly summarized in the text.

RE: Yes, we will follow this suggestion.

195

(9) Choose a color bar for showing difference fields (like “bwr”) in Figure 14.

RE: Alright, we will follow.

(10) Overall, the visual quality and aesthetics of the figures should be improved.

RE: We will examine the revision thoroughly to make sure all figures are clear with high  
200 visual quality. Thanks a lot!

## **5. The writing of the manuscript requires substantial improvement.**

(1) Some expressions are uncommon in academic English. For example, at line 319, “From there, we can see that” could be replaced by “It shows that” or, preferably, by stating the  
205 conclusion directly. Similarly, at line 349, “From Figs. 9b-c, we can see that” could be revised to “Figs. 9b-c show that”.

RE: Yes, it will be done accordingly. Thanks.

(2) Some section titles are unclear or not written in standard academic English. For example,  
210 “3.3.3 The VAE’s data-training and latent space minimization” is awkward, and the distinction  
between “3.4 Preliminary results of SCDA” and “3.4.2 The SCDA results” is confusing.

RE: Thanks a lot! We will rephrase them in the revision.

(3) The wording needs to be more precise. For example, at line 248, “ensures” should likely be  
215 replaced by “encourages”, since the KL loss does not ensure the desired latent-space property.

RE: Thank you for this helpful comment. We understand this point and will revise the  
wording accordingly.

(4) Some citations are missing or should be checked more carefully. For example, the first  
220 appearance of VAE around line 241 should be supported by a citation.

RE: Thanks. This is a helpful suggestion, and we will follow it.

(5) There are too many abbreviations, and some of them are not sufficiently clear. I suggest  
reducing or consolidating some of the abbreviations.

225 RE: Alright! We will minimize the usage of abbreviations. Thanks.

(6) In the abstract, some abbreviations are used before being defined. For example, CDA is  
used before being explained, and f2py should also be briefly introduced.

RE: We will address this in the revision. Thanks!

230

The manuscript contains many issues in English expression. I have only listed a few  
representative examples here, and the authors should carefully check the language throughout  
the manuscript.

RE: We will conduct a thorough language check once the revised version has been prepared.

235 Thanks a lot!