Dear Sir/Madam,

The authors would like to thank the editors for comments/suggestions/corrections, helping to improve the present version of the paper. We have carefully revised the manuscript. Parts of the text were rewritten and reorganized. Here, we present a brief context, following point-by-point answers for all questions.

REFEREE-1 COMMENTS:

**** Major ones ****

RC-1: "Assimilation like 4D-Var or EnKF did requires huge computation efforts. However, the 3D-Var calculation complexity is proportion to the size of model or observations, it is usually trivial as illustrated in Table 4 (several seconds). Even handling models with larger size or with super data like remote sensing obsers, the issue could be solved through regional analysis easily. The choice of 3D-Var is faint to support the motivation."

ANSWER-1: Authors fully agree that 4D-Var and EnKF have a higher computational effort than 3D-Var. With 5 observations, the 3D-Var spends 8 secs. However, for assimilating an image 1024 x 1024 pixels, and supposing a linear CPU time (in fact, it isn't true for the 3D-Var, 4D-Var, and EnKF) the assimilation of such image, using the same software/hardware, this takes - $(8/5) \times 1024 \times 1024) \sim 466$ hours ($\sim$ 19 days). However, same assimilation using NN is just the numbers of number of the pseudo-observation points – in other words, the number of model grid points covering the image. For the NN, the CPU-time is exactly linear with the number of grid points. Therefore, by using TensorFlow, this takes $\sim (0.04/5) \times 1024 \times 1024) \sim 2.5$ hours.

RC-2: "In Figure 3 and 4: The author provides very limited samples or snapshots of analysis for testing their trained NN model, without stating the overall performance in the whole testing dataset."

ANSWER-2: Figures 3 and 4 are not intended to represent exhaustive testing. The figures are illustrative showing the analysis results produced with 3D-Var and 2 neural network methods emulating (both of them emulating 3D-Var scheme). Much more examples were executed with synthetic observations from our previous papers – reader can access the papers Cintra and Campos Velho (2012), and Campos Velho et al. (2022) – in the last cited paper, there is a complexity analysis showing a smalller computational complexity by using neural network.

RC-3a: "Page 9, line 206: only 5 airport measurements are assimilated for analysis. Meanwhile, these same data are used for generation of pseudo-observation for validating the analysis?"

ANSWER-3a: Yes. Five airport measurements are used to compute pseudo-observations.

RC-3b: "That is not the corrected way to using the measurements. Crossing validation is required. Please Check Ref: Peter Rayner. Data assimilation using an ensemble of models: a hierarchical approach., 2020, ACP."

ANSWER-3b: Cross-validation is a strategy used during the learning phase for both neural networks (WEKA and TensorFlow). We include a note on cross-validation strategy in the new paper version (Section 3.2). Thank you for your comment.

RC-4: "In Table 3, NN-TensorFlow outperforms the 3D-Var? It is not solid, afterall, 3D-Var analysis is the learning object of NN? Performance should be examined in-depth."

ANSWER-4: Results in Table 3 show better "Mean Error" and "RMSE" for Tensor-Flow than 3D-Var. However, it is not possible to have a final conclusion from the worked example. Statistics with much more examples are in our list of tasks to be carried out in future work.

**** Minor ****

RC-5: "As long as they described the CPU time for assimilation in 3D-Var, NN-TF, NN-Weka in Table 4. It is essential to illustrate the size of the problem, vec x and y in Eq(1), and the solver/environment for 3D-Var and NN. Otherwise, the comparison is unfair".

ANSWER-5: In the new paper version, we added the number of grid points on the directions "x" and "y" (Section 4.3).

RC-6: "How to train the NN is unclear, what is the output actually? the analysis over the whole model domain? Or is it trained grid by grid? How many samples in their 4-year dataset?"

ANSWER-6: The scheme for data assimilation using NN is described in the Section 4.3 - paragraph initiated by "The experiment steps consisted of", in step-(iii): 'observations, 6-hour forecast (background) field and analysis computed for each grid point (for observation and pseudo-observations) are merged in order to obtain a single dataset for each analysis time; ...'.

We agree to write a clearer text to separate the training phase for the NN for the execution phase by applying NN:

- NN Training phase: observations and 6-hour forecast (background) field are used as inputs for the neural networks, with 3D-Var analysis used as a reference, that means, the neural network output must be the analysis for each grid point with observation or pseudo-observation.
- NN Execution phase: inputs are observation (or pseudo-observation) and background, and the output is the analysis.

The analysis is produced for each grid point with observation or pseudo-observation to be assimilated – we introduced a new text in the manuscript (Section 3.2). We use 4 cycles of data assimilation per day for 365 days, performing 1460 samples for dataset of training.