Review of: EnsAI: An emulator for atmospheric chemical ensembles

Reviewer

August 2026

This paper introduces EnsAI, an AI model capable of producing ammonia increment concentrations on a very low computational budget. The ultimate goal of this paper is to use this model to generate ensembles for data assimilation and emission inversions. The manuscript is well written and shows some strong points. I believe this is valuable work in the field of AI-driven atmospheric composition and relevant to the themes of this journal, though, it suffers from a few major concerns detailed below that should be addressed by the author before publication. While the U-Net architecture is well-suited for this spatial task, some specific design choice as well as the loss profile suggest some improvements can be achieved.

1 Major revisions

1.1 Scope of this work.

1.1.1 Generalization to any other chemical context

The manuscript claims to have designed a system for atmospheric chemical ensembles in general, but the actual scope of this work is much narrower: the model produced here is bound to the sole ammonia species in a north American domain. Furthermore, while the author have cleverly taken advantage of the short atmospheric lifespan of ammonia, the design of the model in this paper, i.e. a diagnostic model independent of the background concentration of ammonia (formula (5)) do not generalize easily to a broader scope such as other more long-lived chemical species, or, less even, chemical ensembles in general. The manuscript should better reflect the actual scope of this method.

1.1.2 Generalization to any other spatial or temporal domain

This model has been exclusively trained on one year of data on a northern American domain. This will prevent the model from successfully generalizing to any other spatial domain, and the manuscript should either display proof that the model is indeed able to generalize spatially, or clearly state that EnsAI is not a general emulator but restrained to this domain. Furthermore, while training on a year long of data might capture seasonal variability, it will be insufficient to capture inter-annual variability of meteorological conditions that the model could encounter in its future uses.

1.2 Training and fine tuning choices

The author should specify explicitly which of the network's hyperparameters has been tuned. Specifically, the first layer of the network takes a hard jump from 4 to 128 channels that appear somehow suspicious. The loss curves of figure S1 clearly displays strong signs of overfitting, the validation loss stagnate while the training loss keeps decreasing, it seems the training of the model could have stopped much earlier than 50 epochs. The author should state if any early stopping was used, and if not, precise why.

1.3 Model's physical consistency

I am concerned with the model's physical consistency in two areas:

- Since the model is trained with MSE error without physical constraint, the author should discuss how prone is the model to respect mass preservation or to predict physically impossible negative concentration.
- Since the model is trained on the sole concentrations of ammonia without any information on its chemical precursors on a year of data, the emulator will implicitly assume a constant chemical regime based on that year, and fail to predict accurately ammonia in case of a chemical regime shift (ex: regulations on emissions or wildfire). This might break the physical consistency entirely. The author should discuss these limitations in the manuscript.

1.4 Error comparison in emission inversion

The author present in figure 16 inversion comparison plots $|\Delta e^{GM} - \Delta e^{static}| - |\Delta e^{EnsAI} - \Delta e^{static}|$ which is a excellent choice since most of the increment are dominated by the static baseline, thus these plots show that the model is also sensitive to the dynamics meteorological correlations. The figure in S11 and S12 in the supplements showing the difference between the physical and the AI model bring the rest of the necessary information. While these plots are necessary, they can be hard to read and more representation of these maps showing different metric as normalized mean bias, correlations, and some time dependent display would make the comparison easier to appreciate.

2 Minor revisions

2.1 Ablation studies

It would be particularly suited for this work to perform ablation study of the meteorological parameters. It would show in particular the ability of the model to predict accurately concentrations from meteorological context instead of from spatial statistical interpolation.

2.2 Typos and minor corrections

Section 8.1, Paragraph 1: "...while the correlations are comprised by the univariate...": are **composed of**

"...the univariate emissions correlations ρ_{ee}, the univariate surface concentration ρ_{cc}, and...": surface concentration **correlations**