ENSAI: AN EMULATOR FOR ATMOSPHERIC CHEMICAL ENSEMBLES

This work presents a valuable advance in AI-based ensemble generation for atmospheric chemistry, but its scope is currently limited to surface ammonia. The title and abstract should be revised to accurately reflect this limitation, or the manuscript should provide evidence for generalization to other chemical species. The discussion of limitations and future directions should be strengthened, particularly regarding the model's applicability beyond ammonia and surface concentrations. The main concerns:

1. Temporal mapping, training period and inversion window

The manuscript should more clearly distinguish among the training period, the temporal resolution of the training samples, and the inversion window. EnsAI is trained using one year of hourly GEM-MACH ensemble outputs. However, Equation (5) appears to represent an instantaneous mapping from the emissions perturbation and corresponding meteorological fields to the surface ammonia concentration perturbation, rather than a one-hour-ahead forecast. Please explicitly clarify the temporal relationship between the inputs and output.

In addition, what is the exact inversion window used for the results presented in this study: one week, several weeks, or one calendar month? The manuscript states that ammonia inversions generally use several weeks of CrIS retrievals and that each inversion estimates a time-mean emissions field over the inversion window. However, Figures S11–S13 present one emissions increment for each month, which appears to imply that monthly inversion windows were used. Please state the start and end dates, or the duration, of each inversion window explicitly.

For example, generating seven days of hourly outputs for 60 ensemble members corresponds to $60 \times 7 \times 24 = 10{,}080$ member-hour inferences. These calculations may of course be performed in batches rather than as 10,080 separate program executions. Please explain whether EnsAI generates the 168 hourly outputs for each member through independent hourly evaluations, a batched multi-time evaluation, or a single trajectory-generating pass. The computational configuration and batching strategy used to obtain the reported seven-second runtime should also be provided.

2. Temporal propagation of emissions perturbations and model errors

The current EnsAI formulation does not include the previous ammonia concentration, previous model state, or an explicit sequence of earlier meteorological conditions as input. It is therefore unclear how the model represents temporal memory associated with transport, chemical transformation, deposition and the accumulated response to a persistent emissions perturbation.

I suggest further investigating this issue using a multi-hour or multi-day sequence model, or by adding lagged meteorological and concentration fields to the EnsAI inputs. Controlled experiments could be performed using either an impulse-like emissions perturbation or a persistent emissions perturbation. The resulting EnsAI and GEM-MACH concentration responses could then be compared after 1, 6, 12, 24 and 72 hours and after one week. Relevant diagnostics could include concentration bias and RMSE, spatial displacement of the plume, ensemble variance, emissions–concentration cross-covariance and correlation length.

This analysis would determine whether the EnsAI errors remain bounded or accumulate with increasing integration time. I would refer to a persistent emissions perturbation over the inversion window, rather than an “emissions error at the initial time”, because the ensemble perturbations in the present study represent time-mean emissions perturbations rather than a perturbation applied only at the initial time.

3. Assessment of suitability for operational inversion

The operational applicability of EnsAI should be evaluated using a continuous or cycling experiment covering more than one inversion window. At present, the analysis mainly demonstrates that EnsAI can reproduce GEM-MACH ensemble covariances and inversion increments for individual periods. This does not fully establish whether the system remains stable and accurate when used repeatedly.

I suggest conducting parallel cycling experiments with EnsAI and GEM-MACH over several consecutive weekly or monthly inversion windows. Both systems should use the same emissions perturbations, meteorological forcing and observations. The posterior emissions obtained in one window could then be used as the prior emissions for the following window. The comparison should include:

- the evolution of ensemble spread and flow-dependent covariances;
- the emissions increments and posterior emissions;
- agreement with assimilated and withheld CrIS retrievals; and
- temporal drift or degradation over repeated cycles.

Because EnsAI currently emulates only the perturbations of one surface chemical field, rather than the complete GEM-MACH atmospheric state, this section may be framed as an assessment of its suitability for **operational ensemble generation and emissions inversion**, rather than as a general operational air-quality forecasting system. The manuscript itself notes that GEM-MACH generates hundreds of chemical fields over 85 vertical levels, whereas EnsAI generates only one surface concentration perturbation field.

4. Direct comparison of the EnsAI- and GEM-MACH-based inversions

Figures S11 and S12 show the GEM-MACH- and EnsAI-based emissions increments separately, making the magnitude, sign and spatial distribution of their differences difficult to assess directly. Please add maps of the signed difference

$$\Delta e^{\mathrm{EnsAI}} - \Delta e^{\mathrm{GM}}$$

for each inversion period, using a consistent colour scale. Maps of the absolute or normalised difference could also be included. These plots would reveal whether EnsAI introduces systematic positive or negative biases and whether the discrepancies are concentrated in particular source regions, seasons or meteorological regimes.

Figure 16 already provides a useful comparison, but it shows the difference between two absolute deviations,

$$|\Delta e^{\mathrm{GM}} - \Delta e^{\mathrm{static}}| - |\Delta e^{\mathrm{GM}} - \Delta e^{\mathrm{EnsAI}}|,$$

rather than the direct signed difference between the EnsAI- and GEM-MACH-based increments. Figure 17 provides an aggregated RMSE, but it does not identify the locations or directions of the discrepancies. Figures S11 and S12 confirm that the two sets of monthly increments are currently displayed only as separate maps.

The direct difference maps could be accompanied by monthly MAE, RMSE, spatial correlation and mean bias, together with regional statistics for major ammonia-emitting areas. January deserves particular attention because the manuscript reports poorer agreement during this period.

Additional minor correction

Near the end of Section 9, the manuscript states that the EnsAI-based increments are “much closer to the increments produced by the static covariances.” Based on the preceding analysis and Figure 17, this presumably should read “much closer to the increments produced using the GEM-MACH ensemble.”