

I'd like to thank the reviewer for their thoughtful comments on the manuscript. Below, the reviewer's comments are displayed in blue and the author responses are in black.

**"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."**

Looking back at the manuscript, I agree that the title and some of the discussion was overly ambitious. I've tried to temper the language used in the title, abstract, and introduction. The title of the paper has been changed to from "EnsAI: An Emulator for Atmospheric Chemical Ensembles" to "EnsAI: An Emulator for Atmospheric Ammonia Ensembles". In the abstract,

"a new AI-based ensemble generation system for atmospheric chemical constituents"

has been changed to

"a new AI-based ensemble generation system for surface ammonia concentrations"

In the introduction,

"The aim of this work is to develop an AI-based model for efficient ensemble generation of tropospheric chemical constituents"

has been changed to

"The aim of this work is to develop an AI-based model for efficient ensemble generation of atmospheric ammonia surface concentrations".

The previous draft of the introduction contained the sentence

"While generating ensembles for ammonia emissions inversions will be used as a test application for the AI-generated ensembles, another goal of this work is to develop an AI system flexible enough to be used in the future for generating ensembles for other chemical constituents relevant for air quality."

has been moved to the 'Conclusions' section (and reworded). Also in the introduction,

"an ensemble of atmospheric concentration perturbations"

has been changed to

"an ensemble of ammonia concentration perturbations"

In the 'Emissions Inversions' section, the sentence

"This study uses the application to ammonia emissions inversions as a testbed for developing the foundations of an AI-based ensemble generation system."

has been changed to

"This study develops an AI-based ensemble generation system to be used in an ammonia emissions inversion system".

“The discussion of limitations and future directions should be strengthened, particularly regarding the model's applicability beyond ammonia and surface concentrations.”

The following paragraphs have been added to the conclusions section:

“ EnsAI was trained on a year's worth of data from the GEM-MACH regional model, which contained data only from North America. Although this exposed EnsAI to regions with a variety of different land-use categories, agricultural types, and meteorological conditions during training, the training data may have not included drastically different meteorological conditions or chemical regimes than those seen in North America during the year used for training. As such, until further tests can be conducted demonstrating its performance in different regions, EnsAI in its present form should be considered to be restricted to its current spatial domain.

While EnsAI was successfully implemented in an emissions inversion system for ammonia, a species with a short atmospheric lifetime, expanding EnsAI to work with longer lived species will require expanding the input capacity to the network. Most significantly, longer lived species will require EnsAI to model a prognostic relationship, which would add fields from previous time steps to the input. Input and output fields for the vertical level above the surface would also be required for long lived species.”

The part about the prognostic relationship mentioned in the second paragraph will be clarified below.

1) “The manuscript should more clearly distinguish among the training period, the temporal resolution of the training samples, and the inversion window ... 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.”

The inversion time window is one calendar month. An unperturbed GEM-MACH model run is run for this month, which produces hourly atmospheric concentrations of ammonia. An ensemble of perturbed GEM-MACH runs were run over this month, producing hourly perturbed atmospheric ammonia fields. These perturbed concentrations fields were used as the training data for EnsAI (which is designed to emulate these perturbed atmospheric ammonia fields), so the temporal resolution of the training data is one hour. 12 months of hourly data was used for training.

The lines in the original manuscript that

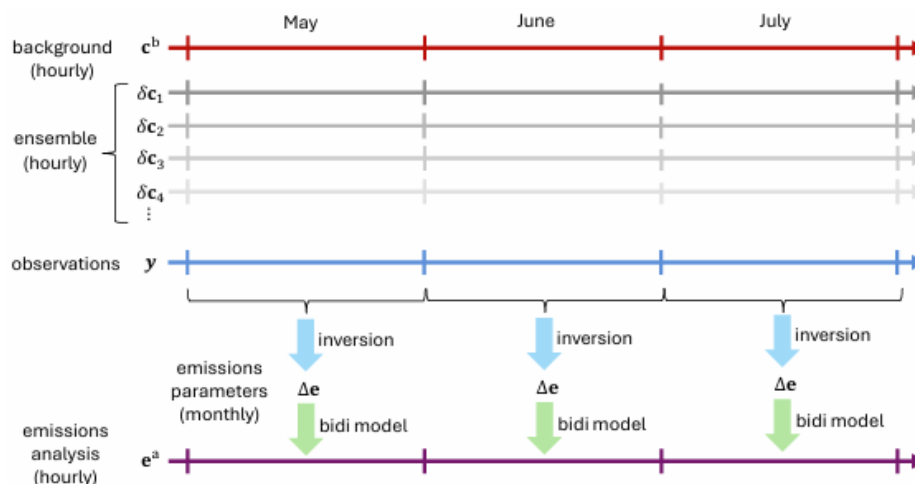
“emissions inversions are generally performed using weeks of retrievals”

were just meant to convey that inversion windows need to be at least a few weeks long, but in principle could be any time period longer than that. But generally these inversions are over one calendar month, so I think this sentence was needlessly confusing and didn't serve a clear purpose. I've removed this sentence in the new draft of the manuscript.

I've changed the manuscript text to contain these lines:

“The hourly fields from an unperturbed GEM-MACH run are used as the a priori (background) model state. The ensemble used in the inversions is comprised of a set of hourly fields produced by perturbed GEM-MACH runs. The emissions inversions use a calendar's month worth of retrievals with the hourly background and ensemble model fields to produce updated emissions model parameters for a particular month. While the product of each inversion is a single field of the emissions parameters for a particular month, the posterior ammonia emissions vary within the inversion time window as the ammonia emissions are dependent on meteorological conditions, most notably the surface temperature.”

Additionally, I've added this new figure to the paper that shows a diagrammatic representation of the inversion process to further clarify the frequency of the data involved:



**Figure 2.** Diagram of the ammonia emissions inversion process. The hourly background concentrations ( $c^b$ ) and ensemble concentration perturbations ( $\delta c$ ) are used with observations ( $y$ ) within a month to produce monthly increments to the emissions parameters  $\Delta e$ . GEM-MACH's bidirectional flux model (represented by the green arrow in the diagram) combines the monthly emissions parameters with meteorological fields to yield hourly updated emissions ( $e^a$ ).

“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.”

The previous version of the manuscript lacked some context which is key to how EnsAI functions. In principle, Eq. (5) should be a prognostic equation, where the atmospheric concentration  $\delta c$  at some particular time is dependent on the concentrations at previous times. However, atmospheric ammonia has a short atmospheric lifetime, so perturbations to the concentrations are short lived, and are primarily driven by emissions. As such, an approximation was made to drop the dependency of  $\delta c$  on previous time steps. Note that the unperturbed ammonia concentrations are

still solved via prognostic equations (as the unperturbed run comes from GEM-MACH) and is it only the concentration perturbations where this approximation is used.

The approximation of not including a dependence of  $\delta c$  on previous time steps was made due to the memory restrictions on the GPU, which greatly limited the number of input fields to the network. As only a limited number of input fields were feasible, only the most important fields for determining the perturbation distribution were used. The results of the manuscript show that EnsAI was able to generate very similar covariances to the original GEM-MACH ensemble despite the fact that  $\delta c$  from previous time steps were not input into the network.

The results on test set show that this is a reasonable approximation for ammonia. However, neglecting  $\delta c$  from previous time steps would likely not work for longer lived species.

I don't think the previous version of the manuscript emphasized this point enough. I added this sentence before Eq. (5)

“Ammonia has a relatively short atmospheric lifetime and individual perturbations to the ammonia concentrations are short-lived. We exploit these properties by neglecting concentration or meteorological fields from previous time steps to further reduce the size of the input. Hence, EnsAI models a diagnostic, rather than a prognostic, relationship.”

and these sentences after Eq. (5)

“As EnsAI describes a diagnostic relationship between variables,  $\delta u$ ,  $\delta v$ ,  $\delta T$ , and  $\delta c$  in the equation above are all valid at the same time (while  $\delta e$  is a monthly mean parameter). We emphasize that the simplification to a diagnostic relationship will only be a reasonable approximation for short-lived species.”

In the conclusions, the paragraph already mentioned above was added as well:

“While EnsAI was successfully implemented in an emissions inversion system for ammonia, a species with a short atmospheric lifetime, expanding EnsAI to work with longer lived species will require expanding the input capacity to the network. Most significantly, longer lived species will require EnsAI to model a prognostic relationship, which would add fields from previous time steps to the input.”

“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.”

As Eq (5) is a diagnostic relationship, each concentration perturbation can be generated independently of one another. In the new version of the manuscript, I've reorganized the paper so that most of the details of architecture and network training are in a Section in the appendix (so that these details are in a single place instead of being scattered around the main body of the paper,

appendix, and supplement). There is a new subsection in this appendix named 'Inference' that has the lines:

“As EnsAI describes a diagnostic relationship, inference can be done in parallel and any number of different time steps could in principle be processed simultaneously. For the 7 second run time to produce a week's worth of simulated data per ensemble member that was quoted in Section 6, inferences were performed in batches of 16.”

2) “The current EnsAI formulation does not include the previous ammonia concentration, previous model state, or an explicit sequence of earlier meteorological conditions as input.”

The additional text added to the manuscript described above clarifies this issue.

“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.”

While EnsAI is not provided the previous concentration perturbation or meteorology, it can still use the input fields it does have (current meteorology and emissions perturbation) to take the processes mentioned above into account.

During training, the response of EnsAI is tuned to map the emissions perturbation to a surface concentration perturbation. Processes such as deposition and chemical transformation affect the atmospheric lifetime of ammonia, which in turn affects the spatial extent of the horizontal correlations. So these types of processing are modeled indirectly through comparing the emissions perturbation to the spatial extent of the surface concentration perturbation. Similarly, information about the transport of ammonia is inferred by examining the spatial pattern of the surface concentration perturbation.

Networks that produce prognostic fields can learn about these processes during training by seeing the temporal evolution of fields. But EnsAI learns these about these processes instead by examining the correlations of the perturbations. For example, there is an intuitive relationship between the spatial correlation length scale of the concentration perturbations and the atmospheric lifetime: In general, the longer the lifetime, the longer the correlation length scale. Not having previous time step does limit this, but this is not a dramatic limitation for ammonia (as shown by EnsAI's ability to reconstruct the error covariance of the ammonia surface concentration). In other words, at least for ammonia, much can be inferred from the covariance structure of the ensemble at a single time step.

“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.”

As EnsAI models a diagnostic relationship that has shown to be successful in reproducing the flow-dependent behaviour of the surface ammonia error covariance, this is addressed in the previously mentioned additions to the manuscript regarding the diagnostic relationship. However, this only applies to short-lived species like ammonia. Extending EnsAI to longer-lived species will necessitate doing as the reviewer suggests here. The previously mentioned addition to the conclusions:

“Most significantly, longer lived species will require EnsAI to model a prognostic relationship, which would add fields from previous time steps to the input.”

addresses this issue for future work.

“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.”

Unlike a cycling assimilation system where forecasts are launched from analyses, the emissions inversion system uses a free model run of GEM-MACH as the background and updates emissions parameters instead of initial conditions. Hopefully the addition of the figure showing the diagram of the inversion process will make this clearer.

The issue of the ‘errors remain bounded’ will be discussed further below in response to comment (3).

As an aside, the comment above did mention ‘Relevant diagnostics could include concentration bias and RMSE’. In the revised version of the manuscript, two new sections have been added: One that examines the posterior surface concentration fields and another that compares the posterior surface concentrations to surface observations from the AMoN and NAPS networks. Although these sections do not deal with the bounding of errors, they do contain these types of diagnostics.

3) “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. 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”

The earlier reviewer comments dealt with some of the language in the previous version of the manuscript being too broad. This may have additionally caused some confusion about setup of the

ensemble, the inversion system, and the differences with a traditional cycling data assimilation system. I hope narrowing the language as mentioned above, in addition to the added description of inversion system, will alleviate this issue.

An operational context for ammonia emissions is very different than for NWP (or for longer lived chemical species). The inversion system always uses a free model run as the background state. The emissions inversion system uses a month-long inversion window to update monthly mean emissions parameters. Ammonia emissions vary significantly depending on the month, as they are driven primarily by agricultural activity. So, if the emissions parameters from a previous inversion were going to be used as the background parameter values of a future inversion, it would be for the same month of a future year, not a subsequent month within the same year. In this case, a single cycle would be a year (or at minimum a single month in a particular year). So to examine if ‘the system remains stable and accurate when used repeatedly’ would require examining inversions of the same month over many different years. To see if the covariance diverges over time would require many years worth of inversions (likely at least a decade), which is well beyond the scope of this study.

If EnsAI is extended for use with other species in a chemical data assimilation system, in which there is a complete assimilation/forecasting cycle, these tests would be advised, but this would be for a future system, not the system being evaluated in this study.

4) “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

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

for each inversion period, using a consistent colour scale. Maps of the absolute or normalized 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,

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

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.”

Plots of  $\Delta e^{\text{EnsAI}} - \Delta e^{\text{GM}}$  and  $\Delta e^{\text{static}} - \Delta e^{\text{GM}}$  have been added to the Supplement (Figs. S15 and S16). Additionally, maps of the bias and standard deviation of the errors for the emissions increments from inversions using EnsAI and the static covariances having been added to the ‘Inversion Results’ section of the manuscript (Fig. 17). This is accompanied by a new paragraph in this section, which reads

“Statistics of these increments taken over the year are shown in Figure 17. The middle and right columns of the figure shows the bias and the standard deviation of errors (STDE) of the emissions increments from the EnsAI and static covariances inversions, respectively, which are defined in Section D of the Appendix. The left column shows the mean and standard deviation of the emissions analyses from the GEM-MACH ensemble inversions for comparison (for notational simplicity, we will not write the <sup>a</sup> superscript for the emission analysis in this section). From this figure, we can see that the inversions using both EnsAI and the static covariance tend to underestimate the increments in many regions, although the extent and magnitude of this underestimation is larger for the static covariances (also see Figures S15 and S16 of the Supplement for plots of individual months). The STDE values for the emissions increment using the static covariances are also noticeably larger than that for inversions using EnsAI in many regions.”

“The direct difference maps could be accompanied by monthly MAE, RMSE, spatial correlation and mean bias”

A new plot of the differences in total emissions within the model domain between the different inversions was added as well, which is now a new panel in the same figure that showed the monthly emissions RMSE (Fig. 18). The difference in total emissions is basically the bias in the emissions but just presented a bit differently. I tried plotting the correlations but the plots were not very informative. The text that accompanies this figure reads:

“ The differences in the total ammonia emissions within the model domain for each month are shown in the top panel of Figure 18. These curves are expressed as percentages by dividing by the total monthly emissions (averaged over 2015) within the domain as determined by the GEM-MACH ensemble inversions. The monthly total emissions from the EnsAI inversions were within 10% of those from the GEM-MACH ensemble inversions for all months. The monthly total emissions differences for the static covariance inversions exceeded a magnitude of 10% for March to May and August to November and had larger differences than the EnsAI inversions for all months except December. The total monthly emissions from both the EnsAI and static covariance inversions were lower than those from the GEM-MACH ensemble inversions for February to November. The total annual emissions from the EnsAI inversions were 2.6% lower than from the GEM-MACH ensemble inversions, but were 6.9% lower for the static covariance inversions.

The root mean square error of the ammonia emissions increment  $RMSE(\Delta e)$  over the domain for each month are shown in the bottom panel of Fig. 18 (expressed as percentages relative to the mean annual emissions given by the GEM-MACH ensemble inversions). The  $RMSE(\Delta e)$  values for EnsAI range from 14% to 35%, and between 29% and 60% for the static covariances. The RMSE for the increments produced using EnsAI have lower values for all months as compared to the static covariances, although only slightly lower for December”

“together with regional statistics for major ammonia-emitting areas”

I’ve added a new section that compares the posterior surface concentrations with surface observations from the AMoN and NAPS networks. This new section includes a plot of changes in RMSE values at each station. Three new plots have been added to the Supplement that shows the

biases of the posterior surface concentrations at each station, and are discussed in this new section. This is not exactly what was suggested by the comment above, but I think it close since with these new plots the reader can examine statistics of the inversion results at different stations (including those near major emissions sources as mentioned above). The new version of the manuscript contains significantly more figures than the previous version (although I've had to place a lot of them in the Supplement as to not overwhelm the main part of the manuscript), so I was hoping the new plots showing the surface observation comparison results could address this comment as well.

[“January deserves particular attention because the manuscript reports poorer agreement during this period.”](#)

While looking into this comment, I found that some of the GEM-MACH ensemble files that were used for the test set contained erroneous values. These files were excluded from the test set in the new version of the manuscript (this problem was with the GEM-MACH files saved on our file system, so was not a problem with EnsAI itself). The poorer agreement was due to the files with erroneous values. The results for January now look in line with the other months of the year.

[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.”](#)

Corrected. Thank you for catching this mistake.