Response to reviewers comments

ACPD manuscript doi: 10.5194/egusphere-2025-5926

Title: 'Improved constraints on ammonia emissions and deposition from co-assimilating NH_3 and NO_2 satellite observations over the Netherlands' by Wizenberg et al.

We would like to thank both reviewers for their helpful comments and suggestions for the manuscript. The reviewer comments are in blue, author responses are in black, and any additions to the text are underlined. The page and line numbers correspond to the version of the manuscript that is available on ACPD.

Responses to reviewer #1

Comment C1.1: I think Section 2.1.1 needs to be revised to make the description of the assimilation system clearer. For instance, on line 98, is only the beta parameters set via the LETKF or 3D gas concentrations are set with the LETKF as well? Please rewrite this sentence to more clearly differentiate between the inversion variables set by the LETKF and all other model variables.

Reply: We thank the reviewer for this comment. We agree that the original wording did not distinguish clearly enough between the variables updated by the LETKF and the prognostic model variables. As part of the revision process, we have made significant changes to Section 2.1.1, including a reorganization of the section for improved clarity and the addition of equations to explicitly define parameters that were not clearly described in the original version.

In our implementation, the LETKF uses an augmented state vector that includes both the trace-gas concentration fields $\mathbf{c}$ and the species-specific emission perturbation factors $\boldsymbol{\beta}$. However, the primary objective of the filter is to estimate the emission perturbation factors $\boldsymbol{\beta}$, which act as multiplicative per-grid-cell scaling factors on the prior emissions. The concentration fields $\mathbf{c}$ are included in the augmented state because they provide the dynamical link between the emission perturbations and the observations. They are not independently prescribed by the LETKF; rather, during the forecast step they evolve through the forward LOTOS-EUROS simulation using emissions modified by the current $\boldsymbol{\beta}$ values. We have revised the text in Section 2.1.1 to make this distinction explicit.

The opening sentence of the paragraph beginning at Page 4, line 98: "The state vector $\mathbf{x}$ of the LETKF includes three-dimensional trace-gas concentrations and two-dimensional perturbation factors (β) for the emissions."

Was moved towards the beginning of Section 2.1.1, and re-written and expanded for clarity to:

"In the present application, the LETKF uses an augmented state vector,

$$X = (c, \beta),$$

where c denotes the three-dimensional trace-gas concentration fields and β denotes one or more two-dimensional emission perturbation fields for the optimized species. The primary objective of the filter is to estimate β, which defines multiplicative per-grid-cell scaling factors applied to the prior emissions."

Comment C1.2: Related to this, when the NO2 retrievals are assimilated and compared against the model output, are all differences attributed to mismatches in NH3 emissions or to other factors as well? If it is the case that the LETKF only adjusts the NH3 emissions, then is it a reasonable assumption to attribute all observed NO2 differences to the NH3 emissions instead of other factors?

Reply: In the main co-assimilation experiment, where IASI and CrIS NH_3 together with TROPOMI NO_2 are assimilated, the LETKF optimizes both the NH_3 and NO_2 emission fields simultaneously. Therefore, differences between the modeled and observed NO_2 are not attributed solely to NH_3 emissions. Instead, the NO_2 observations directly constrain the NO_x (NO_2 + NO) emission field, while the coupled chemistry in LOTOS-EUROS allows these updates to influence the broader reactive nitrogen system, including NH_3, through interactions involving HNO_3 and ammonium nitrate formation. In the additional sensitivity experiments described later in Section 3.5.2, where only NH_3 observations were assimilated, only the NH_3 emission field was optimized. We have clarified this distinction in Section 2.1.1 by adding additional explanatory sentences in the revised manuscript (directly following the additions from C1.1).

Page 4, lines 98: " …while the corresponding concentration fields evolve through the forward LOTOS-EUROS simulation using the perturbed emissions. In the main co-assimilation configuration used in this study, species-specific emission perturbation factors are optimized simultaneously for both NH_3 and NO_2. As a result, assimilated NO_2 observations constrain the NOx emission field directly rather than allowing their signal to be attributed to NH_3 emissions alone."

Comment C1.3: What is the frequency of the analyses produced (i.e. t_k - t_{k-1})? How exactly are different times handled with respect to the temporal correlation in Eq. (1) and what is the assimilation window used? Is the assimilation window for time t_k set as from t_k – delta_t/2 to t_k + delta_t/2, where delta_t is the analyses frequency, and then there is some sort of formula that combines {x_k} for different times using Eq. (1) or does the assimilation window encompass multiple times {t_k} and Eq. (1) is used within a single analysis (or some other method)? Please clarify in the text.

Reply: We agree with the reviewer that the temporal treatment in the original text was not sufficiently clear. In the LETKF configuration used here, the analysis is performed only for the current time step t_k; emissions from previous time steps are not retrospectively adjusted, and the system does not use a multi-time assimilation window in the sense suggested by the reviewer. Instead, the temporal correlation coefficient α_k in Eq. 1 (now Eq. 3 in the revised manuscript) governs the persistence of the emission perturbation factors between successive analysis times t_{k-1} and t_k. In this way, previous emission adjustments influence the current state through temporal persistence, but in the absence of new observations this influence decays exponentially and the system gradually relaxes back toward the a priori emission state. We have expanded the description in the Section 2.1.1 text to clarify this point:

Page 4, lines 96-97: "This formulation ensures that, in the absence of new observations for an extended period, the influence of past updates diminishes and the system progressively returns toward the a priori emission state. <u>The LETKF analysis is applied at hourly analysis times throughout the simulation period. At each analysis time t_k, only the emission state corresponding to that current time step is updated; emissions from previous time steps are not retrospectively adjusted. The temporal correlation coefficient α_k in Equation 1 therefore does not define a multi-time assimilation window, but instead controls the persistence of emission adjustments between successive hourly analysis times.</u>"

Comment C1.4: Lines 102-104 describe the ensemble initialization, but at this point I don't think a definition of beta has been given, so I'm not certain what the stated standard deviation (of one) is the standard deviation of. At the start of the paragraph, it is stated that beta is a "perturbation factor". Please give a precise definition of this. e.g. is this a multiplicative factor or some other type of perturbation. On line 104, it is stated that "The values of the mean and standard deviation for the ensemble can be adjusted if desired." I don't know what this means, please clarify.

Reply: We agree that the original text did not define the emissions perturbation factors β clearly enough before discussing the ensemble initialization. In addition, the initialization values were incorrectly stated in the original manuscript. In the revised manuscript, we now define β explicitly as the species-specific, per-grid-cell multiplicative scaling factors applied to the prior emissions and that they are initialized from a normal distribution with a mean of 1.0 and a standard deviation of 0.5, corresponding to 50% uncertainty around the prior emissions. We also added an equation (Eq. 2 in the revised manuscript) to describe how β is applied to the model emission fields and constrained to remain non-negative. In addition, we removed the sentence stating that the mean and standard deviation "can be adjusted if desired," since this was not sufficiently informative in the manuscript context, and could cause confusion, and was not relevant to the present study because these parameters were not varied here.

Comment C1.5: I'm less familiar with the specifics of the LETKF as compared to other ensemble methods, in which localization is done quite differently, so I was confused by some of the details in Section 2.1.2. I'm aware that for the 'R localization' that is typically used in the LETKF, the optimal localization length scale is often quite a bit smaller than that used in the 'B localization' used in many other ensemble systems (where the Schur product of the ensemble-based B matrix is taken with the localization matrix). However, I wasn't aware that 'R localization' had anything to do with any of the properties of the observations like the instrument's horizontal resolution (for 'B localization' these are unrelated). Could you add some references and maybe some more text explaining this, especially for readers less familiar with ‘R localization’.

Reply: We thank the reviewer for this comment. The reviewer is referring to background-covariance (“B”) localization, in which the ensemble-derived background covariance matrix is tapered directly using a Schur product with a localization matrix. In the LETKF implementation used here, localization is instead applied in observation space following Shin et al. (2016), by selecting and weighting nearby observations for each local analysis. Because of this difference, the interpretation of the localization length is not identical to that in systems using B localization. We have revised the text to make this distinction clearer and to explain why, in the present application, the chosen localization scale was related to the spatial representativeness and footprint size of the satellite retrievals.

Comment C1.6: At the beginning of this section, Shin et al. (2016) is referenced, in which they state that they choose a horizontal localization scale of 2*(10/3)^0.5 * 500 km ~ 1825 km. Although, the Shin et al. (2016) paper is for NWP instead of a relatively short-lived atmospheric gas, I’m surprised that the localization length scales between the two different applications would be so different (compared to the 15 km and 5 km used in for this work).

Reply: We again thank the reviewer for this comment. We cite Shin et al. (2016) for the LETKF localization framework, but not as justification for adopting the same numerical localization length scale. As noted by the reviewer, the ~ 1825 km value used in Shin et al. (2016) was selected for a numerical weather prediction application, where the analyzed variables exhibit broad synoptic-scale correlations. In contrast, the present study concerns short-lived reactive gases, especially NH_3, for which the concentration field is much more locally tied to nearby emissions, chemistry, and deposition. In addition, the assimilated satellite retrievals represent footprint-averaged observations with horizontal resolutions on the order of a few to tens of kilometers. For this reason, substantially smaller localization lengths are appropriate here in order to avoid unrealistically spreading the influence of a single observation over spatial scales not resolved by the measurement. We have clarified this point in Section 2.1.2 in the revised manuscript.

Page 6, lines 138-139: “To ensure computational efficiency and avoid spurious correlations, the LETKF applies spatial and temporal localization in a per grid cell approach following the method described by Shin et al. (2016). <u>In contrast to approaches that apply covariance localization directly to the background error covariance matrix, the LETKF implementation used here applies localization in observation space by selecting and weighting nearby observations for each local analysis. As a result, the localization length in this framework should be interpreted as application-specific and is not expected to match the much larger values used in numerical weather prediction studies such as Shin et al. (2016), where the analyzed variables exhibit broader synoptic-scale spatial correlations.</u>”

Comment C1.7: The ensemble size used in this work (12) is quite small compared to most (non-LETKF) ensemble systems. On line 100, the paper cites Van Der Graaf et al. (2022) as determining this number to be sufficiently large. Van Der Graaf et al. (2022) states “A limited ensemble size of N=12 was found to be sufficient to describe the imposed model uncertainty, which is not too complicated due to short lifetime of NH3 and therefore strong relation between concentrations and nearby emissions.” There doesn’t seem to be a more quantitative analysis of the dependency of the LETKF on ensemble size. If ‘B localization’ was used (in a non-LETKF system), the localization length scale would be much larger than 15 km or 5 km, and so N=12 would be a very small ensemble. N=12 would be sufficient to specify the error covariances within a 15 km x 15 km (or 5 km x 5 km) region, but I don't have an intuitive understanding on why the ensemble size would need to be of very different sizes depending on whether B or R localization was used. Could you also add either some references or reasoning on why this is the case.

Reply: We agree that the work by Van der Graaf et al. (2022) did not present a detailed quantitative sensitivity analysis of ensemble size, and we do not intend to imply that $N = 12$ was established there as a universally optimal value. Rather, $N = 12$ was adopted in the present study because it was used successfully in the same LOTOS-EUROS LETKF configuration by van der Graaf et al. (2022) and because the present LETKF setup differs substantially from large-scale or weakly localized ensemble systems, such as those commonly used in numerical weather prediction. Here, the analysis is performed locally on a per-grid-cell basis using strong spatial localization, so that each local update depends only on a limited subset of nearby observations and a much smaller effective covariance structure. In addition, both NH_3 and NO_2 are short-lived reactive gases, meaning that their concentrations are more locally tied to nearby emissions, chemistry, and deposition than are broad synoptic-scale variables in NWP. Under these conditions, the effective dimensionality of the local analysis problem is substantially reduced, making a modest ensemble size more feasible. We have clarified this reasoning in the revised manuscript and we have adjusted the text accordingly.

The sentence at page 4, lines 100-103: "In this study, $N = 12$ was used as this was found by Van Der Graaf et al. (2022) to be a sufficient trade-off between an appropriate statistical representation of the model uncertainty in the system and overall computational efficiency."

was re-written and expanded to:

"In this study, $N = 12$ was used following van der Graaf et al. (2022), who applied the same LOTOS-EUROS LETKF framework. This relatively modest ensemble size is feasible here because the analysis is strongly localized in space and the short atmospheric lifetimes of NH_3 and NO_2 lead to comparatively compact, local covariance structures."

Comment C1.8: On line 278, it is stated that "cases where the measured and modeled precipitation exceeded a mean absolute deviation of 1-sigma were excluded" from the evaluation statistics. I would have imagined that a 1-sigma cutoff would exclude quite a lot of observations and would have thought setting this at something like >= 2-sigma would be more appropriate. Could you either add a justification of the 1-sigma cutoff or change this to a higher cutoff.

Reply: We agree that a 1σ precipitation-mismatch filter is relatively strict. This threshold was selected after testing looser alternatives during the analysis stage. Relaxing the cutoff to 2σ retained additional cases; for example, at De Zilk the number of observations passing the filter increased from 7200/9408 (76.5%) to 8760/9408 (93.1%), while at other sites the difference was smaller (e.g., Speuld-Garderenseweg: 14016/15864, 88.4%, versus 15072/15864, 95.0%). However, this also degraded the agreement and increased the spread in the wet-deposition comparison, indicating that the additional retained cases were more strongly affected by transient mismatches between measured and modeled precipitation. A further relaxation to 3σ degraded the comparison more substantially. We therefore retained the 1σ filter, as it provided the most robust wet-deposition evaluation in this setup. We have therefore clarified the rationale for this choice in the revised manuscript:

Page 11, line 278: "To minimize their influence on the wet deposition comparisons, a strict filter was applied: cases where the measured and modeled precipitation exceeded a mean absolute deviation of 1 σ were excluded, as were measurements with very low precipitation (<0.1 mm)."

Was re-written to:

"To minimize their influence of transient precipitation mismatches on the wet-deposition comparisons, a relatively strict filter was applied: cases where the mean absolute deviation of the measured and modeled precipitation differed by more than 1σ were excluded, as were measurements with very low precipitation (<0.1 mm). Sensitivity tests with looser thresholds (2 σ and 3 σ) led to poorer agreement and increased spread in the deposition comparison, so the 1 σ filter was adopted for the final wet-deposition evaluation."

Comment C1.9: I don't think Section 3.4 is necessary in the main text as I'm not sure it adds much to the conclusions of the paper. I think it would be best to move this section to the Supplement or an appendix and the authors can add a sentence or two in the main text to referencing the supplement/appendix. Also, I think Eq. (11) might be missing a trace on the right-hand side of the equation.

Reply: We thank the reviewer for this suggestion. We agree that the original presentation of this section could have been clearer. In the revised manuscript, however, we have substantially revised Section 3.4 and corrected the terminology so that the plotted quantity is now consistently described as the averaging kernel sensitivity rather than the DOFS. This terminology and formulation is now also consistent with that of Chen et al. (2023). Following this revision, Eq. (12) (previously Eq. 11 in the reviewed version) is now written in terms of the averaging kernel matrix, and the text explicitly notes that the formal DOFS are given by the trace of this matrix. We have therefore retained the section in the main text, since it provides useful context on the observational constraint and helps interpret the year-to-year differences in the assimilation results, while now doing so in a technically consistent way.

Comment C1.10: There were insufficient details given about the uncertainties of the statistical values presented in Section 3.5. In Fig.7, the slope of the line of best fit has uncertainties associated with it, but the statistics in the lower-right corner (R, mu, sigma) do not. To properly compare these statistics between the base and optimized cases, uncertainties need to be added so the reader can determine whether the differences between these cases is statistically significant or not.

Reply: We agree with the reviewer that small differences in summary statistics should not be overinterpreted. We note that uncertainty estimates for the fitted regression slopes were already included in the regression figures. However, not all quantities reported in the scatter-plot summaries are of the same statistical type. For fitted parameters such as the slope, uncertainty estimates are directly defined and are standard to report. For descriptive metrics such as the mean bias (μ) and spread (σ), uncertainty intervals can in principle also be estimated, but this is less standard and requires additional assumptions about statistical independence and sampling. In the revised manuscript, we therefore addressed this concern primarily by revising the wording throughout Section 3.5 to avoid implying statistical significance from very small differences in quantities such as R, μ, and σ.

Comment C1.11: For instance, on lines 456-457, it states "The correlation in the temporal means also improved slightly from R = 0.84 to R = 0.85, and the slope of the regression also improved from 0.79 to 0.91", but without uncertainties on these numbers the reader cannot tell if these changes are statistically significant or not (as with other subsequent comparisons done later in the paper).

Reply: We agree that the original wording in this sentence was too strong given the very small change in *R*. In the revised manuscript, we have rephrased such statements throughout Section 3.5 to avoid implying statistical significance from marginal differences. For example, changes such as $R = 0.84$ to $R = 0.85$ are now described more cautiously, while greater emphasis is placed on the larger changes in bias and regression slope.

Page 19, line 456-457: "The correlation in the temporal means also improved slightly from *R* = 0.84 to *R* = 0.85, and the slope of the regression also improved from 0.79 to 0.91."

was rephrased to:

"The temporal-mean correlation was very similar in the two cases (*R* = 0.84 for the baseline simulation and *R* = 0.85 for the optimized simulation), whereas the regression slope increased from 0.79 to 0.91."

Comment C1.12: Uncertainties for R, mu, sigma should be added to all scatter plots (Figs. 7, 8, 11-14, 16), Table 2 (also add uncertainties for the slope here), as well to the main text. So line 457 should read '...the slope of the regression also improved from 0.79 +/- 0.03 to 0.91 +/- 0.03' and similarly for whenever R, mu, or sigma values are stated in the main text. Also, in Fig. B2, the difference plot should show the uncertainties (i.e. the standard error) of the mean differences.

Reply: We thank the reviewer for this suggestion. As noted above, uncertainty estimates for the fitted regression slopes were already included in the regression figures, and these have now also been added to Table 2 for compact comparison. We agree that the interpretation of small differences in descriptive statistics should be treated cautiously, and we have therefore revised the text throughout Section 3.5 to avoid overinterpreting marginal changes in R, μ, and σ. We did not add uncertainty estimates for all descriptive statistics in every figure annotation, since this is less standard and requires additional assumptions about statistical independence and sampling. For Appendix Fig. B2, we examined the standard error of the monthly mean differences across sites; however, because each monthly mean is based on a very large number of MAN sites, the resulting standard errors are extremely small and not visually distinguishable at the scale of the figure. We therefore retained the original presentation, in which the shaded regions indicate the standard deviation across sites, as this more clearly conveys the spread in the underlying monthly values while preserving readability.

Comment C1.13: For a lot of the comparisons between the model and surface observations in Section 3.5, scatter plots and their associated statistics were compiled for particular spatial and temporal average separately, for instance in Fig. 7 & 8, Figs. 11 & 12, and Figs. 13 & 14. I think there are more informative ways of computing and presenting these statistics to convey both spatial and temporal information, as well as the statistics overall. It is difficult for the reader to translate the information in the scatter plots of the temporal or spatial mean values into meaningful temporal or spatial information.

Reply: We agree with the reviewer that different visualizations emphasize different aspects of model performance. In the present manuscript, the temporal-mean and spatial-mean scatter plots were shown separately by design in order to isolate temporal agreement at individual sites from agreement in the spatial pattern across the network. We have clarified this motivation by adding an explanatory sentence near the beginning of Section 3.5.1, immediately after the introduction of Figures 7 and 8, so that the intended interpretation of these plots is more explicit.

Page 19, lines 451-453: "A correlation plot of the monthly temporal means is shown in Figure 7 and a corresponding plot of the spatial means (i.e., all sites averaged for a given month) is provided in Figure 8. <u>These two summary views are shown separately to distinguish temporal agreement at individual sites from agreement in the spatial pattern across the Dutch monitoring network. They provide compact statistical summaries, while complementary temporal and spatial context is given by the diurnal-cycle analysis and the mapped concentration fields shown elsewhere in the manuscript.</u> All linear regressions were performed using an ordinary least squares fitting approach."

Comment C1.14: If scatter plots are used, I think they should only display the 'raw' (unaveraged) data. But its often hard for the reader to compare two different scatter plots that have more than ~10 data points, so I would redo all the scatter plots using the unaveraged data and put them in the supplement/appendix (for completeness) and copy the statistical values for R, mu, sigma, and the slope into a Table in the main text (something similar to Table 2, again make sure the uncertainties on each value are stated in the table).

Reply: We thank the reviewer for this suggestion. We agree that raw-data scatterplots can in some cases be informative, but in the present study they would contain a very large number of points and would be difficult to interpret visually, particularly in the main text. We therefore retained the aggregated scatter plots, which provide a more compact summary of temporal and spatial behavior. Rather than replacing these figures, we revised the text to clarify what information each comparison is intended to convey and to avoid overinterpreting small differences in the associated summary statistics.

Comment C1.15: For temporal information, I would plot the data on a time series. For spatial information, I would plot the data on a map, or for LML comparisons since there are only 6 stations, you could also do something like a box and whiskers plot (or something similar) with each station being at a different place on the x-axis. Plotting as a time series or on a map directly conveys the temporal or spatial information.

Reply: We agree that time-series and spatial visualizations are useful for conveying temporal and spatial variability directly. In the present manuscript, however, the purpose of the scatter plots in Section 3.5 is to provide compact statistical summaries of temporal and spatial agreement separately. These therefore complement, rather than replace, the more direct temporal views already shown for the LML diurnal cycles and the spatial maps shown elsewhere in the manuscript. We have clarified this distinction in the revised text.

Comment C1.16: On line 593, it is stated “whereas most MAN sites are located within Natura2000 areas” How many and what percentage are in these conditions? Can you recompute the comparison statistics excluding all of these stations? Or is doing this basically the same at the set of 6 stations colocated with the LML stations? Would it be possible to provide some sort of map showing the land use, maybe like the dominant land use in each grid point, to add support to the statement that LML stations are more representative of the model domain?

Reply: We agree with the reviewer that representativeness is an important issue in the interpretation of the MAN comparisons. Most standard MAN sites are intentionally located in nature areas, often within or near nitrogen-sensitive Natura2000 regions, whereas the six MAN calibration sensors co-located with LML sites provide a more direct comparison with the LML-based evaluation and with more open, model-representative settings. For this reason, we consider the comparison using these six co-located MAN sensors to provide the most practical test of whether the broader MAN bias pattern is primarily a representativeness issue. We have clarified this interpretation explicitly in the revised manuscript by adding a sentence stating that the better agreement for the co-located MAN sensors indicates that the broader positive bias in the full MAN network is dominated primarily by representativeness differences rather than by a uniform model overestimation at all sites. We did not add a separate land-use analysis or recompute the full MAN statistics excluding Natura2000 sites, as this would considerably expand the scope of the present paper.

Page 30, lines 588-590: “In contrast, the full MAN network exhibits positive biases of +1.2 μg m^{-3} and +2.4 μg m^{-3} against the base and optimized simulations, respectively, highlighting the different behavior of the co-located MAN sensors compared with the broader network. <u>The much better agreement obtained for the co-located MAN calibration sensors indicates that the broader positive bias in the full MAN network is dominated primarily by representativeness differences rather than by a uniform model overestimation across all MAN locations.</u>”

Comment C1.17: Line 35: “increased oxidation of NO2 consumes HNO3” NO2 is a precursor to HNO3, so why does this consume HNO3? Please add more details here.

Reply: We agree that the original wording was incorrect. In the revised manuscript, we clarified that oxidation of NO_2 leads to the formation of HNO_3, which can then promote partitioning of NH_3 into ammonium nitrate.

The sentence at page 2, line 35 was changed from:
“In addition, NO_2 indirectly influences the atmospheric lifetime of NH_3 by modulating the availability of nitric acid (HNO_3): increased oxidation of NO_2 consumes HNO_3, leaving less available to form ammonium nitrate from NH_3.

to:

“<u>In addition, NO_2 indirectly influences the atmospheric lifetime of NH_3 by modulating the production of nitric acid (HNO_3): increased oxidation of NO_2 enhances HNO_3 formation, which in turn promotes the partitioning of NH_3 into ammonium nitrate.”</u>

Comment C1.18: Line 92: Is a ‘)’ missing somewhere on this line?

Reply: This error was corrected in the revised manuscript.

Comment C1.19: Lines 99-100: “These are represented” What does ‘these’ refer to here? Later in the sentence it states that the ensemble “capture the uncertainties in both the model and observations”, why does the ensemble capture uncertainties in the observations? This sentence is a bit confusing, please consider rewriting.

Reply: We agree that the original wording was unclear. The sentence was rewritten in the revised manuscript to remove the ambiguous reference to “these” and to clarify that the ensemble represents the uncertainty in the model state, while observation uncertainty is represented separately through the observation error covariance matrix.

Comment C1.20: Line 107: Put the equation defining X^f inline with the text after “the state vector” in line 105. The lines 105 to 112 could be rearranged to make these sentences more clear.

Reply: We agree with the reviewer and in the revised manuscript, we now define the forecast ensemble $\mathbf{X}^f$ inline when it is first introduced, and we reorganized the surrounding text so that the propagation of the forecast members, followed by the definitions of the ensemble mean and forecast error covariance, is presented in a more natural sequence.

Comment C1.21: Line 111: “includes the application of the emissions perturbation factors,” Give more details about this.

Reply: We have clarified this point in the revised manuscript by specifying that the model operator describes the forward LOTOS-EUROS simulation from t_{k-1} to t_k, including the application of the species-specific emission perturbation factors to the prior emissions during the forecast step.

Comment C1.22: Lines 114-116: Would increase clarity if \bar{x} was renamed to \bar{x}^f.

Reply: We agree that this notation is clearer, and $\bar{x}$ was changed to $\bar{x}^f$ in the revised manuscript where appropriate.

Comment C1.23: Line 128: Change “limit” to “approximation”.

Reply: This change was implemented in the revised manuscript.

Comment C1.24: Line 132: Change “representing” to “represents”.

Reply: This change was implemented in the revised manuscript.

Comment C1.25: Line 143: Looking at page 2558 of Shin et al. (2016), they say that they use the fifth order polynomial of Gaspari and Cohn 1999. Is there another part of Shin et al. (2016) where they use this Gaussian function? The Gaspari and Cohn is very similar to a

Gaussian, so there would presumably be little difference between the two but was just confused about the reference here.

Reply: The reviewer is correct that Shin et al. (2016) describe localization using the Gaspari and Cohn taper. In our implementation, the distance weighting is represented using a Gaussian decay function, which serves the same role of smoothly reducing the influence of more distant observations in the local analysis. We revised the text to make clear that Shin et al. (2016) is cited for the LETKF localization framework, while the functional form shown here corresponds to the implementation used in the present study.

Comment C1.26: Section 2.1.3: I think the paper would flow better if the information in this section was moved to Section 2.1.

Reply: We thank the reviewer for this suggestion. We agree that integrating this material into Section 2.1 would be a possible alternative. However, we have retained Section 2.1.3 as a separate subsection in order to keep the model-configuration details distinct from the general LOTOS-EUROS description and the LETKF formulation, and to preserve the overall structure of the methodology section.

Comment C1.27: Section 2.2: I think the total column is being used from all three retrievals (IASI, CrIS, TROPOMI), but this is not clearly stated in the text. References are made to the total column in a few places in Section 2.2, but somewhere it should be clearly stated that the total columns are the observations being used in the LETKF.

Reply: The reviewer is correct that the assimilated observations consist of NH_3 total columns from IASI and CrIS, and NO_2 tropospheric vertical column densities from TROPOMI. We have clarified this explicitly with a sentence added to the start of Section 2.2 in the revised manuscript:

Page 7, line 183 (inserted following the Section 2.2 heading before the start of Section 2.2.1): "In the LETKF configuration used in this study, the assimilated satellite observations consist of NH_3 total columns from IASI and CrIS, and NO_2 tropospheric vertical column densities from TROPOMI."

Comment C1.28: Figures 2, 4, 5: It's difficult to see the differences between rows (a) and (b), and while row (c) show the relative differences, there are several places where the base values are low, so that large percentage differences may correspond to small absolute changes. Can plots of the absolute differences be added as well, or some other way of presenting the results to show where substantial changes occur? The figure might also benefit from changing the units and making the panels larger (they are on the small size and are a bit hard to read).

Reply: We thank the reviewer for this suggestion. We agree that relative-difference plots can visually emphasize regions where the baseline values are small, such that large percentage changes do not necessarily correspond to large absolute changes. In the present figures, however, the relative-difference panels were retained because they provide a compact view of the direction and magnitude of the assimilation-induced changes, while the corresponding base and optimized fields are already shown directly in rows (a) and (b). We therefore did not add an additional row of absolute-difference panels, in order to avoid making already dense multi-panel figures more crowded. Instead, we improved the readability of Figures 2, 4, and 5 by adjusting the figure formatting and the color scaling to increase the visibility of the spatial patterns. We also clarified in the revised text that the relative-difference panels should be interpreted together with the underlying base and optimized fields, and explicitly note where large relative differences correspond to negligible absolute changes.

Page 12, lines 316-317:“ The largest mean increase occurred in 2020 (+14.3%), the smallest in 2021 (+3.0%), with a period-average change of +8.0%.The relative differences in Fig. 2(c) should be interpreted together with the absolute emission fields in Fig. 2(a) and (b), since large percentage changes can still correspond to modest absolute changes where baseline emissions are low.”

Page 16, line 391: “…overestimates deposition processes there. The relative differences in Fig. 5(c) should likewise be interpreted together with the total-column fields in Fig. 5(a) and (b), since the largest percentage changes do not always correspond to the largest absolute concentration changes.”

Comment C1.29: Lines 172-175: This paragraph should be moved to the beginning of Section 3.5, right before Section 3.5.1.

Reply: We believe the cited line numbers from the reviewer may be incorrect, but we interpret this comment as referring to the short transition paragraph on page 15, lines 372-375 introducing the ground-based evaluation. We agree that this material reads more naturally at the start of Section 3.5, and we have adjusted the text accordingly in the revised manuscript.

Comment C1.30: Lines 407-408: Remove the sentence “Independent evaluation against ground-based observations … atmospheric conditions.”

Reply: This sentence was removed in the revised manuscript.

Comment C1.31: Lines 460-461: “This indicates that the spatial pattern ... without assimilation.” Is this sentence referring to Fig. 7 or Fig. 8? Its placement in the paragraph suggests that it is referring to Fig. 8, but would maybe make more sentence inferencing Fig.7? Please reword.

Reply: The sentence was intended to refer specifically to Figure 8, which shows the comparison of the spatial means. We agree that this was not sufficiently explicit from the original wording, and we have rephrased the sentence in the revised manuscript to make this clear.

Page 19, lines 460-461: “This indicates that the spatial pattern of the NH_3 surface concentrations in the Netherlands is being represented more accurately in the LETKF optimized simulation than in the original simulation without assimilation.”

was reworded for clarity to:

“The stronger improvement seen for the spatial means in Figure 8 indicates that the spatial pattern of NH_3 surface concentrations across the Netherlands is represented more accurately in the LETKF-optimized simulation than in the base simulation."

Comment C1.32: 10: It is a bit hard to compare the three different sets (base, assimilation, observations) in these plots. Might be clearer with just 3 curves of the mean, maybe with a shaded standard error region around each curve, instead of the box and whiskers plot.

Reply: We thank the reviewer for this suggestion. We agree that a mean-curve representation with shaded uncertainty bands would be another possible way to present these data. However, we retained the current box-and-whisker style because it conveys not only the mean diurnal cycle, but also the spread and asymmetry of the hourly NH_3 distributions at each site, which would be lost in a mean-only presentation. We have revised the figure caption to make this more explicit and we have made improvements to the figure formatting and presentation for increased visibility and readability. We have also revised the text to make the purpose of the box and whisker plot more explicit.

Page 19-21, lines 470-472: "Since the LML data are provided at an hourly frequency, the impact of assimilation on the diurnal cycles of NH_3 in the model can also be investigated at each site. The mean diurnal cycles from the observations, the base model simulation, and the optimized simulation calculated over the 2018--2022 period are shown in Figure 10. The box-and-whisker representation is used to show not only the central tendency of the diurnal cycle, but also the spread of the hourly concentration distributions at each site."

Comment C1.33: Lines 472-474: "…even though only morning and afternoon satellite overpasses were used" I assume this is the case because the impact from the assimilation remains in the system for at least 6-12 hours. Could you add a sentence about this?

Reply: The reviewer is correct that this is a result of the temporal persistence of the emissions adjustments in the LETKF system which allows the updated state to influence the concentrations at later time-steps beyond the satellite overpass times themselves. We have added an additional sentence to clarify this:

Page 21, lines 472-474: "In most cases, the mean differences in the diurnal cycles showed improvement relative to the observations in the optimized run in comparison to the base simulation even though only morning and afternoon satellite overpasses were used. This is partly because the effect of the assimilation persists between overpass times through the forecast step and temporal persistence of the emission adjustments, allowing the updated state to influence concentrations beyond the observation times themselves."

Comment C1.34: Figure 9: For the legend and the caption, I think 'base' and 'optimized' or a similar wording is better here than 'before' and 'after'. Also, why is the title of the legend 'Sites & Mean', i.e. why 'mean'?

Reply: We agree with the reviewer that the terminology in the Figure 9 caption was inconsistent with "base" and "optimized" that was used throughout the rest of the manuscript. We have adjusted the figure caption and legend. Additionally, the title of the Figure 9 legend was originally "Sites & Mean" due to the presence of one marker which represented the mean across all sites. However, for simplicity, we have now changed the legend title to “LML sites” instead.

Comment C1.35: Line 492: “whereas the Kalman filter adjusts emissions uniformly without differentiating among source types.” But as long as emissions are adjusted on a per grid cell basis, does this matter? For whatever mix of emission source types within a grid cell, there will be a total diurnal pattern that will be a mix of the diurnal patterns of each source. But from the perspective of the assimilation, I would have thought that whatever sort of mix of diurnal profiles wouldn't matter since it is just trying to match the overall diurnal profile of that particular grid cell. Maybe some diurnal profiles are more difficult to capture than others, so maybe the assimilation will be more successful at matching the diurnal pattern from some sources over others, but I would not have thought that this is a feature of the Kalman filter not being able to differentiate between source types.

Reply: We agree that the original wording was too broad. The issue is not that the Kalman filter cannot in principle adjust the total diurnal emission pattern within a grid cell, but rather that in the present setup the assimilation updates a single total emission adjustment factor per species and grid cell, without distinguishing among sector-specific source types that may have different temporal profiles. We have rephrased the sentence accordingly in the revised manuscript.

Page 22, line 492: “… each with distinct diurnal patterns, whereas the Kalman filter adjusts emissions uniformly without differentiating among source types.”

was rephrased to:

“… each with distinct diurnal patterns, whereas the <u>present assimilation setup applies a single total emission adjustment per species and grid cell, without distinguishing among source types that may have different underlying diurnal emission profiles</u>.”

Comment C1.36: Line 512: Remove “individually”.

Reply: This change was implemented in the revised manuscript.

Comment C1.37: Fig 12: Is it a coincidence that the values for mu in Fig. 12 are identical (at least to 1 decimal place) to those in Fig. 11?

Reply: We thank the reviewer for this comment. We checked this carefully and confirm that Figures 11 and 12 are based on different summary statistics. Figure 11 uses site-specific monthly temporal means, whereas Figure 12 uses monthly spatial means obtained by averaging across all sites for each month. The reported mean biases are therefore calculated from different aggregated datasets, but in these cases the resulting values happen to be extremely close numerically. For example, for the first assimilation run the mean bias is 0.1350524 in one case and 0.13505223 in the other, with similarly small differences for the other runs, so the displayed values coincide only because of rounding to one decimal place.

Comment C1.38: Line 454: As with the comments above for Fig. 9, change ‘before’ and ‘after’ to something like ‘base’ and ‘optimized’ or something similar.

Reply: The terminology in the text and relevant figure legends and captions were revised from “before/after” to “base/optimized” for consistency with the rest of the manuscript.

Comment C1.39: 15: What does 'absolute' mean here (second row)? If referring to the absolute value, I would assume that there wouldn't be any negative values in panels (d) to (f). Please clarify what exactly is being plotted in the middle and bottom rows. Is panel (g) supposed to tell use which run (base or opt) is better depending on if the point is blue or red? Please clarify and make a reference to this in the main text.

Reply: We agree that the label “absolute differences” in the second row of Figure 15 could be misinterpreted as referring to absolute values, in which case only non-negative values would be expected. What is actually shown in panels (d) to (f) are signed differences in NH_3 concentration expressed in absolute units of μg m^3, in contrast to the bottom row, which shows the corresponding relative differences in %. We have revised the figure labels and caption to avoid confusion.

Comment C1.40: Lines 571-572: “This suggests that the assimilation enhances large-scale spatial and seasonal variability”. I don't follow where this comes from, please clarify.

Reply: This sentence has been reworded in the revised manuscript for clarity.

Page 29, lines 571-572: “This suggests that
the assimilation enhances large-scale spatial and seasonal variability, but also introduces a systematic offset.”

was rephrased to:

“The increase in spatial correlation suggests that the assimilation better captures the large-scale spatial and seasonal variability seen in the MAN observations, although it also introduces a larger positive offset.”

Comment C1.41: Line 575: “show similar behavior” Does this mean similar to the comparison of MAN sites in Fig. 15 or to the LML observations mentioned in the previous sentence?

Reply: The phrase “show similar behavior” was intended to refer to the broader MAN comparison discussed in the preceding sentence, not to the LML observations. We agree that this was ambiguous and have rephrased the sentence in the revised manuscript to make the reference explicit.

Comment C1.42: Lines 578-590: Starting with the sentence “Additionally, to support … LETKF simulations.” seems a bit out of order. Combine these two paragraphs and move this sentence to the end of the paragraph as a conclusion sentence.

Reply: We agree with the reviewer that the ordering of these two paragraphs could be improved. In the revised manuscript, we combined the introduction of the six co-located MAN calibration sensors with the discussion of their results, so that the comparison and its interpretation are presented in a single, more coherent paragraph. We retained the clarification that these six sensors are not part of the standard MAN dataset, but are used for the monthly calibration of the broader MAN network, since this distinction is important for understanding their role in the comparison.

Comment C1.43: Table 2: Maybe add horizontal lines to separate each base/optimized pair so that it is easier for the reader to compare the numbers between the base and optimized statistics for each case.

Reply: Horizontal lines were added to Table 2 to improve readability and make comparison between each base/optimized pair easier.