RC1

Review of "Decreasing trends of ammonia 1 emissions over Europe seen from remote sensing and inverse modelling"

Summary

This paper adds to the growing body of literature on monitoring ammonia from space. The analysis uses the CFPR NH3 product derived from CrIS radiances in an inversion process that uses an LPDM within a Bayesian approach to derive monthly emissions in Europe during the 2013-2020 period. The authors show seasonal variability and calculate trends in emissions over the entire continent (excluding Russia and Turkey) and regionally, and demonstrate that overall there has been a marked decrease in emissions, attributed mainly to due to control strategies adapted by the European Union. This is an important result for policy makers to use in justifying these often unpopular controls. The use of an LPDM to tackle the problem of estimating NH3 emissions from satellite data is, to my knowledge, the first time this approach has been applied to ammonia from space.

The paper is well organized and well written. The quality of the graphics is quite high. It needs only minor revisions to be accepted for publication.

- We appreciate reviewer's comments and his willingness to help improve our manuscript. Below, we have done a big effort to follow his comments and answer his arguments.

Technical issues

Figure 2: the posterior emissions in the upper two panels are in units ng/m2/s while they are in Gg in the two lower panels. This implies integration as well as averaging; please describe how these values were obtained in the text.

- We extend Section 3.1 to describe how the totals (in Gg) are calculated (Please see TrackChanges in L.195-297).

Line 292: the sentence starting with "It should be noted is not correct". If NO_x and SO_2 decreased during the pandemic, more NH3 would remain in the atmosphere, since the there would be less sulphuric and nitric acid for it to react with. NH_3 emissions may well decreased in urban areas (see Cao et al., 2022 (https://pubs.acs.org/action/showCitFormats?doi=10.1021/acs.estlett.1c00730&ref=pdf) but not because SO_2 and NO_X decreased.

- We completely agree with the reviewer here. In fact, we have a manuscript under review saying the same thing. What we mean here is that the method we use to calculate posterior creates a false decrease in the emissions (as described in our manuscript). We rephrased (Track Changes L.334-335).

Line 344: Emission may appear to increase in Scandinavia in winter because emission of OH and O3 concentrations are much lower, so the rate of converting SO2 to sulfate is much slower, less sulfate is made and thus more NH3 stays in the gas form.

- We appreciate for this comment. We have put this explanation when discussing the seasonal variation of the emissions in N. Europe (Track Changes L.370-372)

Line 346: this paragraph is a bit confusing. Does the standard deviation come out of the least squares solution for Equation 4?

- Thank you for this question, we have reformulated this paragraph to clarify the uncertainty calculation. It is exactly as reviewer says, the natural output of the Bayesian solution (in general) is the whole posterior distribution of estimated parameters. Hence, the uncertainty calculated in Section 3.3 is the total standard deviation of the (multivariate) variable following a Gaussian distribution (Track Changes L.223-247).
- We also added the second type of uncertainty calculated from the ensemble of the used prior emissions for the inversion (Track Changes L.424).

Minor edits

Line 130: 10000 retrievals per day per level seems reasonable for this product but what does the number 2920 indicate?

- A typo error was corrected here (Track Changes L.132).

Line 135: ... to 2000 per day per level for 6 vertical levels.

- Corrected (please see Track Changes L.135).

Line 177: ... and one hour temporal resolution.

- Corrected (Track Changes L.180).

Line 185: the more common usage is **difference** rather than **distance**.

- Thank you, we corrected this (Track Changes L.189).

Line 186: between the CrIS vertical profile observations, denoted as v^{sat}, and the simulated retrieved profiles, v^{ret}. The latter are obtained by applying an instrument .....

- Thank you, we reformulated the sentence (see Track Changes L.191).

Line 199: within a circle around each grid cell for computational efficiency. We chose circles with a radius of approximately 445 km, which is shown ....

- Thank you, we reformulated the sentence (see Track Changes L.204).

Line 212: What are the dimensions of the matrices in this equation?

- We have clarified this in Section 2.4. The dimension for one year data batch is 193 (elements in circle around grid cell) times 6 (vertical profiles) times 12 (months), hence 13896 (see Track Changes L.229-231).

Line 218: don't the authors mean into Equation 2 and then into Equation 1?

- It is, indeed, into Eq. 2 and then Eq. 1. We reformulate the sentence to omit misunderstanding (see Track Changes L.254).

Line 226: Does NaN here indicate missing? Please clarify.

- Indeed, we have replaced "NaN" by "missing" (see Track Changes L.286).

Line 236: Does the avgEENV prior vary by year? Or is the inversion done separately each year for computational reasons?

- Both notes are correct. The avgEENV prior vary by year because some priors vary by year. Also, the inversion is done separately for each year which is now commented more clearly in Section 2.4 (see Track Changes L.212).

Figure 1: temporal variability.

- Corrected (see Track Changes L.1018).

Line 331: Rewrite sentence starting with "Especially" as :
The NE emissions dominate the a priori emissions that were used here (avgEENV), because their winter peak in the north is extreme (emissions decline from 35 Gg in winter to 12 Gg in summer). Therefore, due to the strong prior that we use in Northern Europe there is a strong dependence of the posterior seasonality of ammonia on the prior in this region.

- We rewrote the sentence (see Track Changes L.404-406).

Line 358: The current figure 4 should come after the current figures 5 and 6.

- Corrected everywhere in the manuscript (see Track Changes).

Line 411: cannot be resolved at the spatiotemporal resolution of CTM and FLEXPART.

- Corrected (see Track Changes L.556).

Line 447: A uni-directional dry deposition scheme ignores the impacts of ...

- Part of this sentence removed as reviewer suggested (see Track Changes L.597).

Line 475: Rewrite sentence starting at: "Here we examine" as
Here we used satellite observations from CrIS and a novel inverse modelling algorithm to study the spatial variability and seasonality of NH3 emissions over Europe. We then

evaluated the overall impact of such strategies on the emissions of ammonia for the period 2013–2020.

- Sentence modified as suggested (see Track Changes L.628-631).

Line 481: industrial activities

- Corrected (see Track Changes L.634).

Line 479: The highest emissions overall ...

- We were not sure if we understood what is meant to be written here and modified the sentence as “The highest emissions over the 2013–2020…” (see Track Changes L.632).

RC2

In their paper Tichy et al present the emissions of ammonia derived from CrIS satellite observations, and present the trend in these emissions over the period 2013-2020. To my judgement major revisions are needed before the paper can be published, as detailed in the comments below.

- We appreciate reviewer's assistance to improve our manuscript. We have his comments in an effort to optimize our paper.

General comments:

In general the description of the method, inputs, filtering, error modeling are incomplete in the paper, and make it impossible to judge the quality of the results, in particular the reported trends, but also the absolute value of the emissions.

- We have modified the methodology description, adding more information. However, please note that we have intentionally omitted many details to avoid repetition. Our first paper in which we used the same set-up stands as a preprint. Although we do not have a revised or a final version yet, the preprint is cited whenever needed within the manuscript (Tichý et al., 2022). Please see section 2.4.
- As regards to the uncertainty, we now provide more details about how it was calculated, plus that we add another approach that is based on the use of an ensemble of different priors to calculate how sensitive our posterior emissions are with respect to the use of prior (Please see section 3.3 of the Track Change Manuscript).

Tichý, O., Otervik, M. S., Eckhardt, S., Balkanski, Y., Hauglustaine, D. and Evangeliou, N.: NH3 levels over Europe during COVID-19 were modulated by changes in atmospheric chemistry, npj Clim. Atmos. Sci., in review, 1–13, doi:10.21203/rs.3.rs-1930069/v1, 2022.

Trends in ammonia are presented without discussing other trace gases, in particular NOx and SO2, which have a significant trend over the past decade and influence ammonia concentrations. No NOx/SO2 trend results are shown in the paper, and the authors do not provide evidence that the model used (LMDZ-OR-INCA) provide a realistic description of trends and interaction with other chemicals and aerosols.

- We are not sure how to act on this comment. LMDz-OR-INCA model has been used the last 20 years. Its chemistry was implemented by our coauthor and we cite all the details of it in Hauglustaine et al. (2004 and 2012). We do not intend to evaluate the chemical scheme of LMDz-OR-INCA nor repeat technical details here. Both have been done long before and are presented in the references we cite. However, if the Reviewer and Editor insist, we could add a few repetitive sections.
- Also, we evaluate the trends in the emissions and not the trends of atmospheric ammonia in general. Although they are of course linked, NOx/SO2 have been taken by published state-of-the-art inventories (ECLIPSE in the present). As it is explained in our preprint (Tichy et al., 2022) and here too, the chemical loss of NH3 due to chemical reactions with sulfate and nitrate (products of NOx/SO2) is introduced in flexpart with the e-folding lifetime. In other words, we calculate the

lifetime of NH3 in the chemical model first, which then import as a loss parameter in flexpart. All the calculations of the modelled lifetime are presented in our previous work (Evangeliou et al., 2021).

Hauglustaine, D. A., Hourdin, F., Jourdain, L., Filiberti, M.-A., Walters, S., Lamarque, J.-F. and Holland, E. A.: Interactive chemistry in the Laboratoire de Meteorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, J. Geophys. Res., 109(D04314), doi:10.1029/2003JD003957, 2004.
Hauglustaine, D. A., Balkanski, Y. and Schulz, M.: A global model simulation of present and future nitrate aerosols and their direct radiative forcing of climate, Atmos. Chem. Phys., 14(20), 11031–11063, doi:10.5194/acp-14-11031-2014, 2014.
Tichý, O., Otervik, M. S., Eckhardt, S., Balkanski, Y., Hauglustaine, D. and Evangeliou, N.: NH3 levels over Europe during COVID-19 were modulated by changes in atmospheric chemistry, npj Clim. Atmos. Sci., in review, 1–13, doi:10.21203/rs.3.rs-1930069/v1, 2022.
Evangeliou, N., Balkanski, Y., Eckhardt, S., Cozic, A., Van Damme, M., Coheur, P.-F., Clarisse, L., Shephard, M., Cady-Pereira, K. and Hauglustaine, D.: 10–Year Satellite–Constrained Fluxes of Ammonia Improve Performance of Chemistry Transport Models, Atmos. Chem. Phys., 21, 4431–4451, doi:10.5194/acp-21-4431-2021, 2021.

How much does the a-priori emission influence the results? The method description in section 2.4 does not provide the information to judge the influence of the prior compared to the impact of the satellite measurements.

- We extended section 2.4 and completely rewrote section 3.3 to address the question of uncertainty more clearly (Please see the provided manuscript with Track Changes). In the revised version of the manuscript, we study two types of uncertainty, the inversion model uncertainty and the uncertainty arising from prior emissions. The model uncertainty is a natural result from the used Bayesian approach where full Gaussian posterior distribution is available and uncertainty in the form of variance/standard deviation can be easily calculated. The prior emissions uncertainty is calculated using ensemble of the used prior emissions.

The assumptions/modeling of errors of the satellite data (including filtering), in the method (model uncertainty: chemistry, transport) and a-priori emissions are not described. Section 3.3 discusses uncertainties but is very high level and does not provide the details needed to understand the results and related error bars.

- Thank you for pointing out the influence of prior emissions. We rewrote Section 3.3 completely to clearly describe the calculation of uncertainty from the posterior estimates based on property of Gaussian distribution which is the form of our posterior. Moreover, we added the second type of uncertainty calculated from the ensemble of the used prior emissions for the inversion. The abstract and conclusion parts are modified accordingly.

Satellite observations are available once per day, but I assume that the emissions are reported as diurnal mean. What uncertainty does the unknown diurnal cycle introduce?

- We are not sure how to quantify this type of uncertainty and we do not have any indication whether it is significant of not. A diurnal cycle is neither assumed in the Chemistry Transport Model, nor exists in the satellite observations from CrIS,

as pointed by the reviewer. If the Editor or Reviewer think it is necessary to quantify, we would be happy to do it, and await specific instructions.

Detailed comments:

The abstract is long and reads like an introduction, especially the first part. I would propose to shorten it and focus on the actual findings in the paper and new results.

- We agree and have shortened the abstract to 306 words (Please see Track Changes).

The paper has a good introduction with a balanced set of relevant papers.

- We appreciate again for the time the reviewer took to read and try to improve our manuscript.

The paper uses several units for the emissions (per second, per day, per month, per year). This makes it hard to compare the plots. I would suggest to restrict this to one or two choices.

- We have tried to do this; however, there are cases where we cannot have full control of the units. For instance, when comparing our findings with results from the literature, we have to follow the units presented in the literature.

l 40: "Our results are associated with relatively low uncertainties reaching a maximum of 42%" Which result is this? Is it the trend over a region?

- This sentence is now reformulated to address both Gaussian model uncertainty and prior emissions ensemble uncertainty, respectively. Note also that we have updated the Gaussian posterior model uncertainty with higher precision reaching relative uncertainty value.

l 47: "constitute a robust basis for European NH3 estimates". What does "robust basis" refer to. Do the authors claim that the monitoring of pollution levels as set by the regulations can be performed based on satellite observations and inverse modeling (only)?

- We have modified the abstract completely so that it is much shorter now. We explicitly write in L38-40 that “These results indicate that satellite measurements combined with inverse algorithms constitute a robust tool for emission estimates and can infer the evolution of ammonia emissions over large timescales”.

l 47: "de facto". Does this mean that the evolution is based on measurements?

- We have removed this confusing expression (see Track Changes L38-40).

l 87: "Greenhouse Gases Observing Satellite". Please add acronym "GOSAT".

- The acronym has been added (Track Changes L.147).

l 95: "using alternation between CrIS ammonia retrievals performed with the logarithm of concentrations and linearized retrievals." Please explain more clearly: what does the "alternation" between log and linear retrievals mean?

- We completely reformulated this sentence to be more precise in referring to the method by Sitwell et al. (2022) (Track Changes L.155-157).

l 97: "use direct comparisons between the CrIS observations and model retrievals". What are "model retrievals"?

- The sentence has been corrected as to “comparison between CrIS ammonia retrievals and model profiles” (Track Changes L.158).

l 118: "total column random measurement error is estimated in the 10–15% range, with total random errors estimates of ~30%". What is the difference between a "total column random" and a "total random" error? Systematic errors are even more important.

- As Shephard et al. (2020) reported “For the total column amounts, the measurement errors are typically in the 10% to 15% range, whereas the total errors are ~ 30 %”. We now correct for this typo in L.185 (see Track Changes).

Shephard, M. W., Dammers, E., E. Cady-Pereira, K., K. Kharol, S., Thompson, J., Gainariu-Matz, Y., Zhang, J., A. McLinden, C., Kovachik, A., Moran, M., Bittman, S., E. Sioris, C., Griffin, D., J. Alvarado, M., Lonsdale, C., Savic-Jovcic, V. and Zheng, Q.: Ammonia measurements from space with the Cross-track Infrared Sounder: Characteristics and applications, Atmos. Chem. Phys., 20(4), 2277–2302, doi:10.5194/acp-20-2277-2020, 2020.

l 121: "due to the limited vertical resolution". Please mention a typical degrees of freedom of signal for CrIS NH3.

- Missing information were added in L.188 (see Track Changes).

l 130: "Daily CrIS ammonia (version 1.6.3) was interpolated onto a 0.5°×0.5° grid ". How is this done? How are measurements and kernels, defined in log space, averaged? It seems to me that this needs to be done with great care, so more details are required to convince the reader of the correctness of the approach.

- The averaging has been performed under the guidance of the CrIS developers. While we initially thought to use classic oversampling methods, the resolution we wanted to achieve is very coarse, while oversampling is more efficient in urban scales (Zhu et al., 2014). Also, we had tested previously inverse distance weighting interpolation (IDW) in Evangeliou et al. (2021) finding that it creates overestimated gridded NH3 columns.
- Therefore, we decided to use classic linear interpolation and validate our results against ground-based observations from EMEP, which are openly available from https://ebas.nilu.no. Some examples are shown in Fig. R1-R4.

Zhu, L., Jacob, D.J, Mickley, L.J., Marais, E.A., Cohan, D.S., Yoshida, Y., Duncan, B.N., Abad, G.G., and Chance, K.V.: Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns. Environ. Res. Lett. 9, 114004, 2014.
Evangeliou, N., Balkanski, Y., Eckhardt, S., Cozic, A., Van Damme, M., Coheur, P.-F., Clarisse, L., Shephard, M., Cady-Pereira, K. and Hauglustaine, D.: 10–Year Satellite–Constrained Fluxes of Ammonia Improve Performance of Chemistry Transport Models, Atmos. Chem. Phys., 21, 4431–4451, doi:10.5194/acp-21-4431-2021, 2021.

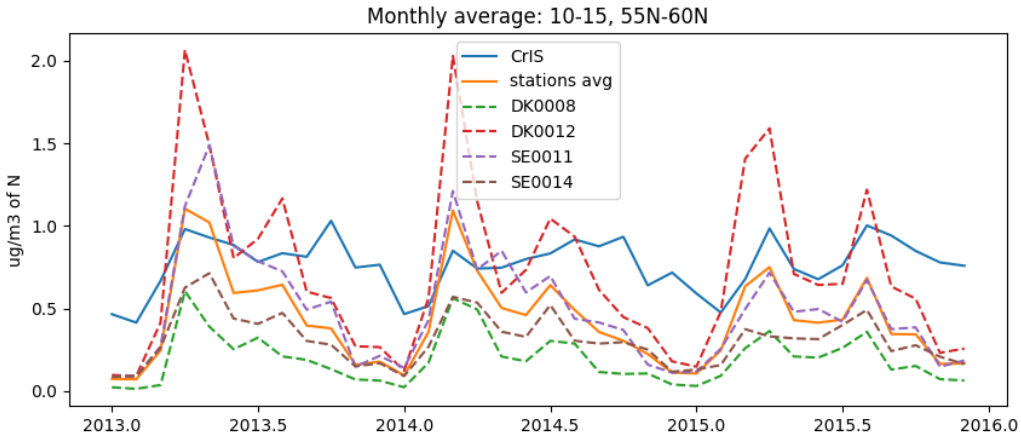

Fig. R 1

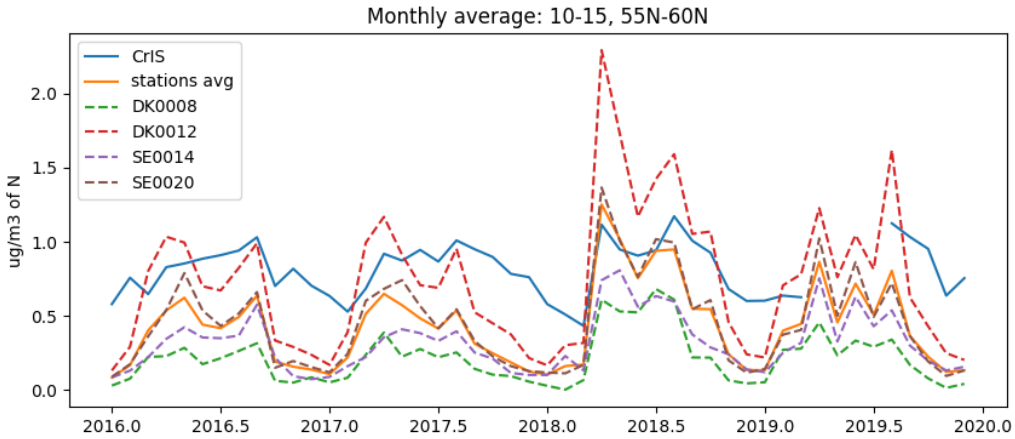

Fig. R 2

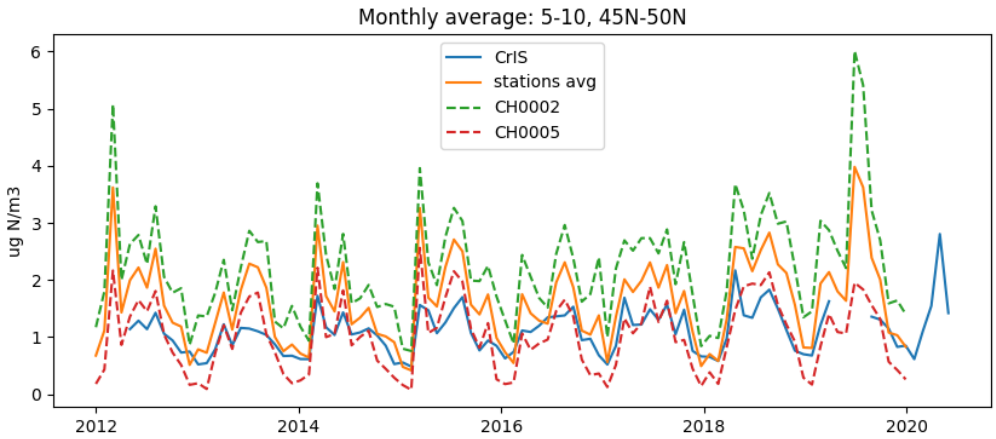

Fig. R 3

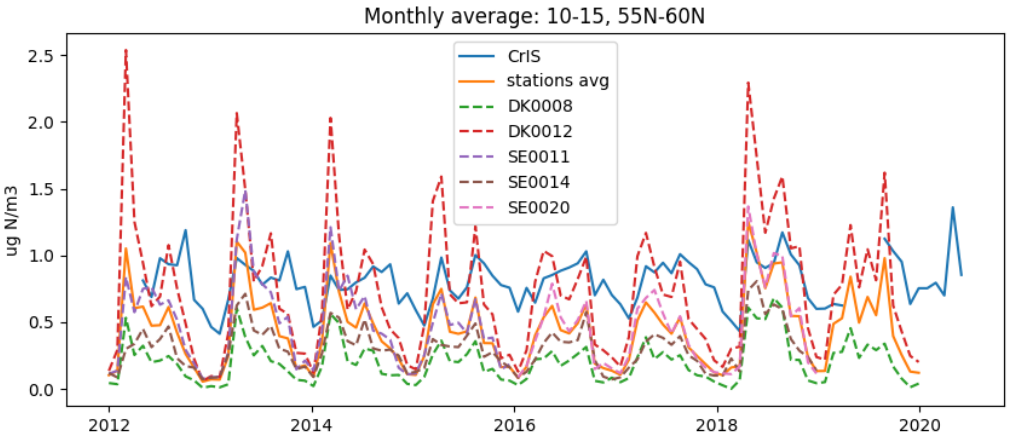

Fig. R 4

l 132-135: I got lost with the number of observations mentioned. Why does 10000 observations consist of 2920 retrievals? How can 10000 observations be reduced to 12000? This sounds like an increase.

- We agree that this did not make sense as it was originally written. We corrected it in L.199-202 (see manuscript with Track Changes).

l 144: For GFED it is clear these are biomass burning emissions. But what is ECLIPSE (which source sectors are included)? Is GFED4 the same as GFEDv4? What sectors does "GEIA" add to the other two emission inventories, and why is GEIA not included in option (i)?

- Here, we are talking about state-of-the-art emission inventories, very well-known among the atmospheric modelling community. All these details are given in the cited literature in section 2.2. We repeat again here:
  - GFED4 is same with GFEDv4 (corrected everywhere in the manuscript) and refers to biomass burning.
  - ECLIPSEv5 and v6 are emissions from anthropogenic sources and include energy/power sector (power plants including combined heat and power, energy production–conversion including district heating plants, fossil fuel distribution), industry (industrial combustion and processes), residential sector (residential combustion sources), transportation (on-road and off-road transport sources) and agriculture (livestock and fertilization).
  - GEIA emission inventory refers to natural emissions from the surface of the ocean.

  About the last point on why GEIA is not included in option (i), please note that throughout this manuscript we calculate land-based emissions and not oceanic ones. The reason is the CrIS retrieval algorithm is still under development above the ocean (see Shephard et al., 2020). Hence, whether GEIA emissions are included in option (i) or not does not play any role in the posterior emissions of NH3 given a 12 h lifetime for NH3.

Shephard, M. W., Dammers, E., E. Cady-Pereira, K., K. Kharol, S., Thompson, J., Gainariu-Matz, Y., Zhang, J., A. McLinden, C., Kovachik, A., Moran, M., Bittman, S., E. Sioris, C., Griffin, D., J. Alvarado, M., Lonsdale, C., Savic-Jovcic, V. and Zheng, Q.: Ammonia measurements from space with the Cross-track Infrared Sounder: Characteristics and

applications, Atmos. Chem. Phys., 20(4), 2277–2302, doi:10.5194/acp-20-2277-2020, 2020.

l 148: Please explain the difference between option (iii) and (iv). What is the reason that they differ so much?

- The differences of the 2 calculated emissions are described in detail in Evangeliou et al. (2021). I explain once again here for clarity. The NE emissions inventory (denoted as iii in the manuscript) was developed using raw data of IASI ammonia, then applying inverse distance weighting (IDW) interpolation onto a 0.5-degree grid and then using modelled lifetimes. The VD emissions inventory (option iv in the manuscript) used 9 years raw data of IASI, gridded onto 0.01 degrees by Van Damme et al. (2018) using classic oversampling methods and then using a constant lifetime of 12 hours everywhere (globally).

Evangeliou, N., Balkanski, Y., Eckhardt, S., Cozic, A., Van Damme, M., Coheur, P.-F., Clarisse, L., Shephard, M., Cady-Pereira, K. and Hauglustaine, D.: 10–Year Satellite–Constrained Fluxes of Ammonia Improve Performance of Chemistry Transport Models, Atmos. Chem. Phys., 21, 4431–4451, doi:10.5194/acp-21-4431-2021, 2021.
Van Damme, M., Clarisse, L., Whitburn, S., Hadji-Lazaro, J., Hurtmans, D., Clerbaux, C. and Coheur, P. F.: Industrial and agricultural ammonia point sources exposed, Nature, 564(7734), 99–103, doi:10.1038/s41586-018-0747-1, 2018.

Section 2.2: What is the reason why the authors created such an elaborate a-priori emission as a average of four estimates? Is the final a-posteriori emission very sensitive to the a-priori?

- The reason why we decided to use an average a priori is because of the two different methodologies used to infer these prior emissions (2 bottom-up and 2 top-down inventories) and because we got the best statistics with respect to independent ground-based observations than with any individual prior dataset.
- Besides, we used this ensemble to answer reviewer's question about how sensitive the posterior emissions are to the use of different priors (see Track Changes L.523-531).

Section 2.2: The a-priori emissions show a very large range of values. So it is unlikely that all of them are realistic. Is there an estimate of typical uncertainties for the four priors individually, or some knowledge of biases? Which one is supposedly the most accurate one?

- Yes, there is a full uncertainty analysis in Klimont et al. (2017) for the GAINS model used in ECLIPSE. They state that "… uncertainties of emission estimates developed with integrated assessment models like GAINS are similar to the estimates for bottom-up inventories discussed above", which are -40% to +100% for different species. For NE and VD there is an uncertainty analysis presented in Evangeliou et al. (2021). In most of Europe, uncertainty is around 45-50%, but it reaches up to 95% in regions characterized by large emissions (Netherlands, Belgium, Denmark).

Klimont, Z., Kupiainen, K., Heyes, C., Purohit, P., Cofala, J., Rafaj, P., Borken-Kleefeld, J. and Schöpp, W.: Global anthropogenic emissions of particulate matter including black carbon, Atmos. Chem. Phys., 17, 8681–8723, doi:10.5194/acp-17- 50 8681-2017, 2017.

Evangeliou, N., Balkanski, Y., Eckhardt, S., Cozic, A., Van Damme, M., Coheur, P.-F., Clarisse, L., Shephard, M., Cady-Pereira, K. and Hauglustaine, D.: 10–Year Satellite–Constrained Fluxes of Ammonia Improve Performance of Chemistry Transport Models, Atmos. Chem. Phys., 21, 4431–4451, doi:10.5194/acp-21-4431-2021, 2021.

Section 2.3: Please comment on the (non-)linearity of the SRMs. How much do they depend on the accurate knowledge of other species, e.g. NOx, SO2. Trends in NH3 will be influenced by trends in concentrations of such species, determining the loss timescale of NH3. Is this accounted for in the study? Without such information it is impossible to judge the quality of the reported trends.

- We present all this discussion in our published preprint.
- However, we now realize that this assessment is crucial for the current methodology and should be mentioned here, as well. We have added an extensive discussion on ammonia's linearities for the present case in (now) section 4.2 (Track Changes L.1097). We appreciate reviewer for pointing this out.

l 193: "This is a useful technique.." Which technique?

- We reformulate this sentence in the revised paper.

l 204: "iterative minimization of distance". Normally a cost function is introduced with terms describing the distance between model and measurement, and model and a-priori emission, or, alternatively, a regularization to avoid the under-determined ill-posed problem. The way it is introduced here it seems there is no penalty for moving away from the a-priori emission? Please explain more clearly what is done, e.g. by introducing the cost function, and specify all terms in detail, including how error covariances are modelled.

- Thank you for pointing out the possible shortcuts in description of the used inversion method, we describe the used inversion algorithm in detail with covariance matrices and their structures defined in the revised paper. The method is a Bayesian counterpart to standard cost function formulation as given, e.g., in (Cao et al., 2020). Here, all unknown parameters are considered as unknown variables to be estimated using the variational Bayes procedure. We significantly extend the definition of the model while we refer to (Tichy et al., 2016) for proper inference of posterior distributions and their shaping parameters using iterative scheme.

l 222; "however, the bounds are large enough to allow for new sources, as well as for attenuation of old sources. For this reason, the choice of prior emission is of great importance in the method." These two sentences seem to contradict each other. If the range is large enough the impact of the a-priori emission should be negligible!? How does the a-priori emission affect the results? This was not clear to me after looking at all the results presented.

- We agree that the presented formulation may be misleading, and we reformulate the paragraph.

l 224: "for some spatiotemporal elements are missing in the dataset." Which percentage of the grid cells is missing on average (on a daily basis)? Please explain the quality filtering for the CrIS data. What quality flag filtering is used here? Does the filtering

remove cloud-covered scenes (is the cloud flag used)? I would say that interpolation is very tricky for large areas without observations.

- From the 2013-2020 period, only 4 months of data are missing due to an instrument failure. This is about 4.2% (4 out of 96 roughly) of the data.
- About the quality flag, as it is specified in the CrIS documentation, it is recommended QF=3 at minimum. QF=4 should be required when using the data in an inverse process, since the averaging kernel will limit the impact of cases with DOF <0.1. When generating maps to look for pattern, QF=5 will return the most intense or persistent features. QF=5 should also be used for comparisons against in situ data, as these observations will have the lowest uncertainty. We used QF=5.
- About the cloud flag, it is recommended in the CrIS documentation that for the highest quality data the pixels with Cloud_Flag=1 can be remove from the analysis. The version 1.6 that we used here contains the same retrievals as previous versions but includes non-detects pixels that are below the detection limit of the sensor (Cloud_Flag=3). This version will provide more representative gridded averaged (level 3) values. We used Cloud_Flag=3 of course since we wanted to calculate gridded averaged values.

Fig 2, panel a: Please use the same color scale as for Figure 1. It is important to see how much this differs from the combined prior avgEENV.

- We agree with the reviewer, and we modify the color scale accordingly.

Fig 2: What does the box plot show? Is it the range of values for the 8 years (8 points)?

- Yes, we specify this in the caption of the figure in the revised version.

Fig 2: Why is there no uncertainty range specified for the posterior emissions in this figure? E.g. in the top-right and bottom-right panels. The uncertainty analysis is presented at the end of the paper.

- Thank you for this remark. The calculated uncertainties (Gaussian) are now plotted together with the average European NH3 posterior_avgEENV in Fig. 2.

l 294: "due to bias created by the decrease of NOx and SO2 " This small sentence is the only mention of NOx/SO2 in the whole paper. The impact of these species on the NH3 emission (concentration) trend should be discussed in much more detail. How are the trends in NOx/SO2 accounted for in this work (e.g. including the impact of COVID-19)? What is the evidence that these trends are described in a realistic way?

- This is a full paper cited as a preprint here (Tichy et al., 2023). It examines a unique case that was encountered over Europe as a result of the recent pandemic, in which SO2 and NOx emissions and levels dropped substantially due to the lockdowns. To account for this unique condition with respect to SO2 and NOx levels, we applied adjustment factors (AFs) to the prior emissions Doumbia et al. (2021) for the January–August 2020 period (0.1°×0.1° resolution on a daily or monthly basis) for the transportation (road, air and ship traffic), power generation, industry and residential sectors. The quantification of AFs is based on activity data collected from different databases and previously published studies. These emission AFs have been applied to the CAMS global inventory, and the changes in emissions of the main pollutants have been assessed for different regions of the world in the first 6 months of 2020. As in any other non-COVID19

year, the Chemistry Transport Model LMDz-OR-INCA with full chemistry of the Nitrogen cycle simulated trends and levels of NH3. Its performance has been assessed by Hauglustaine et al. (2004; 2014).

Tichý, O., Otervik, M. S., Eckhardt, S., Balkanski, Y., Hauglustaine, D. and Evangeliou, N.: NH3 levels over Europe during COVID-19 were modulated by changes in atmospheric chemistry, npj Clim. Atmos. Sci., in review, 1–13, doi:10.21203/rs.3.rs-1930069/v1, 2022.
Doumbia, T. et al. Changes in global air pollutant emissions during the COVID-19 pandemic: A dataset for atmospheric modeling. Earth Syst. Sci. Data 13, 4191–4206 (2021).
Hauglustaine, D. A., Hourdin, F., Jourdain, L., Filiberti, M.-A., Walters, S., Lamarque, J.-F. and Holland, E. A.: Interactive chemistry in the Laboratoire de Meteorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, J. Geophys. Res., 109(D04314), doi:10.1029/2003JD003957, 2004.
Hauglustaine, D. A., Balkanski, Y. and Schulz, M.: A global model simulation of present and future nitrate aerosols and their direct radiative forcing of climate, Atmos. Chem. Phys., 14(20), 11031–11063, doi:10.5194/acp-14-11031-2014, 2014.

l 328: "model likely underperforms" Please also comment on the quality of the CrIS data in wintertime Northern Europe. Are there enough constraints from the satellite observations? How much coverage does CrIS provide after filtering?

- The sentence was requested to change by the first reviewer (see manuscript with Track Changes L.566-568).
- About the question on the coverage after filtering CrIS NH3, we are not really sure “How much coverage” is meant here. How do we define this? In the case of North Europe, we had at minimum 1000 raw measurements per defined grid-cell, but it is difficult to assess whether these measurements are enough or not.

Fig. 3: "decreases in ammonia" How are the numbers on the figure determined. Is it (2020-2013)/2013, or is it derived from a trend line analysis, or something else?

- This information is now added to the manuscript. The relative trends are calculated as (2020-2013) divided by the average emissions from the studied period [2013;2020] for each region.

Fig. S3: (top-left) The order of the trend legends does not match the order of the lines, and a string is missing for central-east. Please update the figure.

- Thank you for noting this typo, we corrected these offsets of texts from lines.
- During this correction, we found that we accidentally plotted avgEENV prior instead of the NE prior at the bottom panel. We corrected this in the current manuscript version and modified the text accordingly.

l 334: "due to the strong prior that we use there" This is one example where it is difficult for the reader to understand the impact of the prior on the posterior results.

- We agree and rephrase the formulation. We also add reference here to extended uncertainty analysis in Section 3.3.

Section 3.3: I propose that uncertainties are discussed before the results, e.g. as part of section 2.

- We would prefer to discuss calculated posterior emissions first and then assess relevant uncertainties, as in most research articles presenting inverse modelling estimates (e.g. more recently published Peng et al., 2022, Vijta et al., 2022, Deng et al., 2022).

Peng, S., Lin, X., Thompson, R.L. et al. Wetland emission and atmospheric sink changes explain methane growth in 2020. Nature 612, 477–482 (2022). https://doi.org/10.1038/s41586-022-05447-w.
Vojta, M., Plach, A., Thompson, R. L., and Stohl, A.: A comprehensive evaluation of the use of Lagrangian particle dispersion models for inverse modeling of greenhouse gas emissions, Geosci. Model Dev., 15, 8295–8323, https://doi.org/10.5194/gmd-15-8295-2022, 2022.
Deng, Z., Ciais, P., Tzompa-Sosa, Z. A., Saunois, M., Qiu, C., Tan, C., Sun, T., Ke, P., Cui, Y., Tanaka, K., Lin, X., Thompson, R. L., Tian, H., Yao, Y., Huang, Y., Lauerwald, R., Jain, A. K., Xu, X., Bastos, A., Sitch, S., Palmer, P. I., Lauvaux, T., d'Aspremont, A., Giron, C., Benoit, A., Poulter, B., Chang, J., Petrescu, A. M. R., Davis, S. J., Liu, Z., Grassi, G., Albergel, C., Tubiello, F. N., Perugini, L., Peters, W., and Chevallier, F.: Comparing national greenhouse gas budgets reported in UNFCCC inventories against atmospheric inversions, Earth Syst. Sci. Data, 14, 1639–1675, https://doi.org/10.5194/essd-14-1639-2022, 2022.

Section 3.3, line 347-356. I could not understand this discussion on how uncertainties are computed. The standard deviation mentioned in line 353 is only valid when the errors in both variables are uncorrelated. Assuming uncorrelated errors in general underestimates the real uncertainty. Equation 5 does not make sense to me, and the variables are not defined (U_elem and u_elem,t).

- We agree that the uncertainty analysis was hard to follow, and we completely reformulated Section 3.3. We calculate two types of uncertainty: (i) Gaussian posterior uncertainty that is an outcome of the used Bayesian inversion method since we have the full posterior distribution and (ii) uncertainty associated with the used ensemble of prior emissions.

The discussion in this section is very high level and does not provide any detail. What needs to be added is a quantitative analysis of the various contributions to the uncertainty: the a-priori emission uncertainty, the uncertainty in the model linking satellite concentrations to emission, and the satellite uncertainties. The discussion should also include the systematic contributions to the error. The posterior uncertainties are very low, especially compared to the range of prior values shown in Fig.1, and this does not give me confidence that the error analysis is conducted in a proper way.

- We agree that an ensemble was missing from the uncertainty analysis to complete the overall picture on how posterior NH3 is affected by use of different priors. We have now added this in the revised form. This has increased the relative uncertainty of the posterior to 101% (see manuscript with Track Changes). As regards to the satellite uncertainty, it is difficult to propagate it in the posterior uncertainty due to the different variables (v^a, A) that are used to calculate v^{ret} (see Eq. 1 in the manuscript).

page 13: How are the comparisons done? Are surface observations available hourly? Are hourly values compared (or daily-mean, or weekly ..)?

- Stations characterized by abundant concentrations measure concentrations hourly, while others weekly or bi-weekly. In any case, all these results are openly available for download from https://ebas.nilu.no.

I had a hard time linking Fig. 6 with Fig S 7. Figure 6 seems to indicate that at many stations there is a very major improvement in the correlation. But the time series in S7 and the statistical quantities presented do not seem to confirm this. S7 indicates only minor improvements in MAE, RMSE, RMSLE. The difficulty to represent individual measured peaks is not really different in the posterior run, so how can the correlations have improved so much, as indicated in Fig. 6?

- Figure S 7 shows only 6 out of the 53 stations that were used in the validation as an example. Of course, posterior NH3 improves in some cases and even doubles in some stations (e.g., DE002 or NO0056 or HU002). However, as seen in Fig. 5 (formerly Fig. 6), only 13 out of the 53 stations where posterior concentrations were compared with the EMEP observations exceeded an R2 value of 0.6, so we do not really understand what is implied here. Of course, top-down algorithms are useful in many cases, but most of the time (depending on the amount of observations and the measurement uncertainties) fail to calculate individual peaks.