We would like to thank the reviewers for their time, constructive and helpful comments and suggestions. Below, we have replied to each review and revised the draft accordingly. We have listed out the reviewer comments in *blue italic* and the replies in black.

RC1:

This study presents a robust local-scale inversion of agricultural ammonia emissions on the island of Schiermonnikoog in the Netherlands. The authors combine a high-resolution chemistry transport model (LOTOS-EUROS) with a Bayesian inversion framework, incorporating observational constraints from the MAN passive sampling network and synthetic LML-like measurements. The work is timely and relevant, especially considering the national nitrogen crisis and the need for fine-scale emission estimates to support policy and conservation efforts.

The manuscript is well structured and technically sound. It makes contribution by highlighting the limitations of existing monitoring networks and proposing practical strategies for observational enhancement. However, several aspects, including the inversion method, format of the manuscript, interpretation of the inversion results, treatment of uncertainties, and clarity of figures, require major revision before the manuscript can be considered for publication in ACP.

Major Comments、

1. *Manuscript Structure: The current organization resembles a technical report rather than a research paper, especially the introduction section, and please consolidate the abstract into a single continuous paragraph.*

   We have revised the abstract into a single continuous paragraph:

   Quantifying real-world emission reductions is a core goal of atmospheric inversion methods, yet direct validation against known events remains rare, especially for reactive species like ammonia. In this study, we have applied local-scale Bayesian inversions using ground-based measurements and the LOTOS-EUROS air quality model, with high-resolution emission inventories as prior input, not to explore a theoretical scenario, but to evaluate a documented emission reduction. On the island of Schiermonnikoog in the Netherlands, where GVE (grazing livestock units) decreased from 639 to 541, with a particularly notable reduction in dairy cattle, ammonia emissions are expected a 23% reduction between 2019 and 2022. Our inversion captured a similar trend, estimating a 51% decrease, which may be overestimated, largely attributed to uncertainties in the 2019 posterior emissions. The posterior for 2022 shows consistency with the validation and indicates a 27\% reduction compared with the prior emissions of 2019. The associated uncertainty, derived from the posterior error covariance, highlights both the potential of the method and its limitations for policy verification. Moreover, we developed a method to assess the usefulness of individual observations and propose that adding a single high-quality continuous measurement in a strategically chosen location can significantly enhance the inversion performance. This strengthens the observational constraint and enhances the system's ability to resolve temporal variations in emissions.

2. *Inversion Methodology (Lines 185-195): Clarify how Jacobian matrix $K_N = \partial y/\partial x$ is approximated within the iterative Levenberg-Marquardt framework for solving Eq. (3). Given its computational intensity, explicitly state which finite differences methods are used to compute the Jacobian. Additionally, conclusion's claim that the method "assumes a linear relationship between emissions and concentrations" as this contradicts the nonlinearity inherently addressed through iterative K-updating.*

   We have added the definition explicitly at line 182: K_N is the Jacobian matrix at the Nth iteration, which updates accordingly through $K_N=K_{linear}\ \mathrm{diag}(x_{linear\ N})$., following Eq. (4).

   Although the Jacobian matrix K updates every iteration step, the derivative is based on the linear relation between y and x_{linear}, so that the forward model could be approximated to $K_{linear}\ x_{linear}$

3. *The inversion indicates a 51% reduction in ammonia emissions from 2019 to 2022, compared to a 23% reduction based on activity data (GVE and livestock). While the authors briefly acknowledge that the inversion might overestimate the reduction, the manuscript lacks a detailed discussion of possible reasons behind this discrepancy. Recommendation: (1)Test the inversion's sensitivity to prior emission uncertainties (e.g., varying the prior error covariance β). (2) Discuss confounding factors (e.g., unaccounted meteorological influences, changes in farming practices beyond livestock numbers, or biases in the "Other" category).*

   We thank the reviewer for this valuable comment. We conducted the validation and revised the manuscript accordingly (Sect. 3.3, line 356-365):

   The inversion results suggest a 51% reduction in ammonia emissions on Schiermonnikoog between 2019 and 2022. However, this figure may be an overestimate. To address the concern, we performed a leave-one-source-out cross-validation (LOSOCV). This approach is analogous to leave-one-out cross-validation (LOOCV), but instead of omitting one measurement site, we iteratively exclude one state vector element that represents external influences. In each case, we subtract the correlated contribution and then re-conduct the inversion with the remaining elements. The posterior estimate for 2019 exceeded the range of the validation, although the credible interval still encompasses the validated values. This suggests that the apparent overestimation of the emission reduction originates primarily from an overestimation of the 2019 emissions rather than an underestimation in 2022. In contrast, the posterior for 2022 shows consistency with the validation and indicates a 27% reduction compared with the prior emissions of 2019. Thus, the discrepancy between the inversion (51% reduction) and the activity data (23% reduction) can largely be attributed to uncertainties in the 2019 posterior emissions.

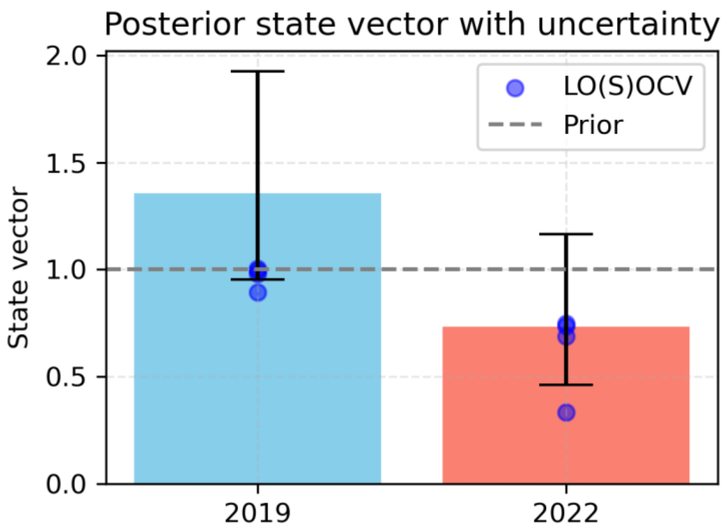


Figure 1 Yearly emission inversion, updated Fig. 8a in the manuscript

For meteorology influences, we validate the meteorology with KNMI measurement sites (see also our response to reviewer 2) and revised the manuscript (Sect. 3.1, line 255-260):

To evaluate the representativeness of the meteorological forcing, model results were compared with observations from nearby KNMI stations on a daily basis (map shown in Fig. S1), including one located close to the island. The agreement was very good, with correlations for wind components, temperature, and pressure consistently above 0.96 and low RMSE values (Table S2, Fig. S2). Precipitation was also well reproduced, with correlations around 0.8. These results indicate that the meteorological fields are reliable and representative for the study area, supporting the robustness of the subsequent analysis.

For the farming practices, the farmers started to use more mineral fertilizer because they wanted to boost the milk production of the remaining animals, since the milk price was good. The other factor was a more complex one: the number of animals decreased, but the surface area of the stable remained the same. Therefore, the excretion of the cows was less but still distributed over the same surface. This led to a lower emission reduction in the report.

4. *Sections 3.1.1 and 3.1.2: The model consistently underestimates peak concentrations near sources (e.g., Meteo Groenglop), attributed to coarse resolution (1.7 km × 2.15 km). However, Schiermonnikoog's emissions are concentrated in a 275-ha polder, likely smaller than the model grid. Recommendation: Conduct a sub-grid sensitivity test (e.g., nesting a higher-resolution domain over the polder) to assess resolution impacts.*

We thank the reviewer for the suggestion.

We have conducted additional sub-grid simulations at a spatial resolution of 500 m * 500 m to test the sensitivity to model resolution. The results, as showinhowever, show only limited improvement: the Pearson correlation and regression slope between observations and model output increase slightly, but the scatter plot indicates that the overall performance remains similar. This is mainly due to the emission inventory, which is available at a coarser resolution than 500 m * 500 m and thus limits the benefit of using finer model grids. We therefore retained the results from the coarser grid simulations (with bi-cubic interpolation for grid extraction) in the manuscript, as they remain representative for the current setup. A description of this additional sensitivity test has been added to the manuscript (line 300-305):

To assess the potential impact of spatial resolution, we conducted additional simulations using a nested domain with 500 m * 500 m resolution over Schiermonnikoog (see the supplement). The results indicate only limited improvement compared to the coarser-resolution setup: the Pearson correlation and regression slope with observations increased slightly, but overall scatter remained similar. This limited gain is mainly due to the coarser resolution of the emission inventory, which constrains the benefit of refining the model grid. Therefore, for consistency, we present the results from the coarser-resolution simulations (with bi-cubic interpolation) in the main analysis.

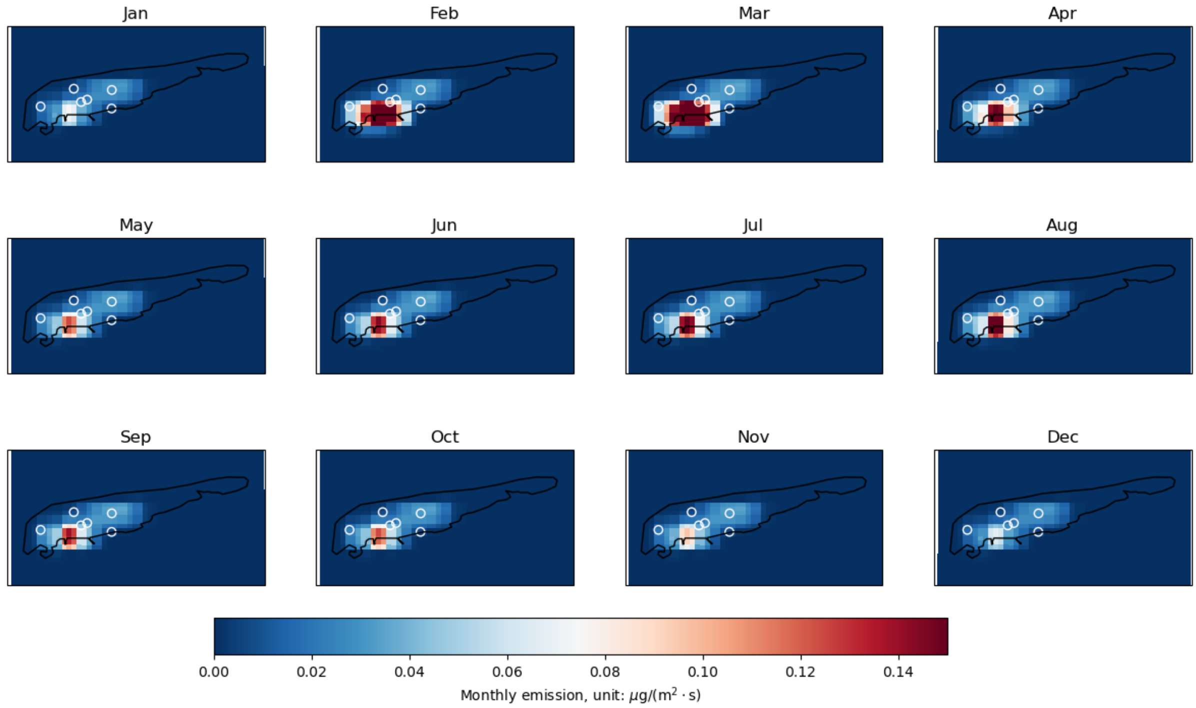


Figure 2 Monthly emission of simulation for 2022 with the prior

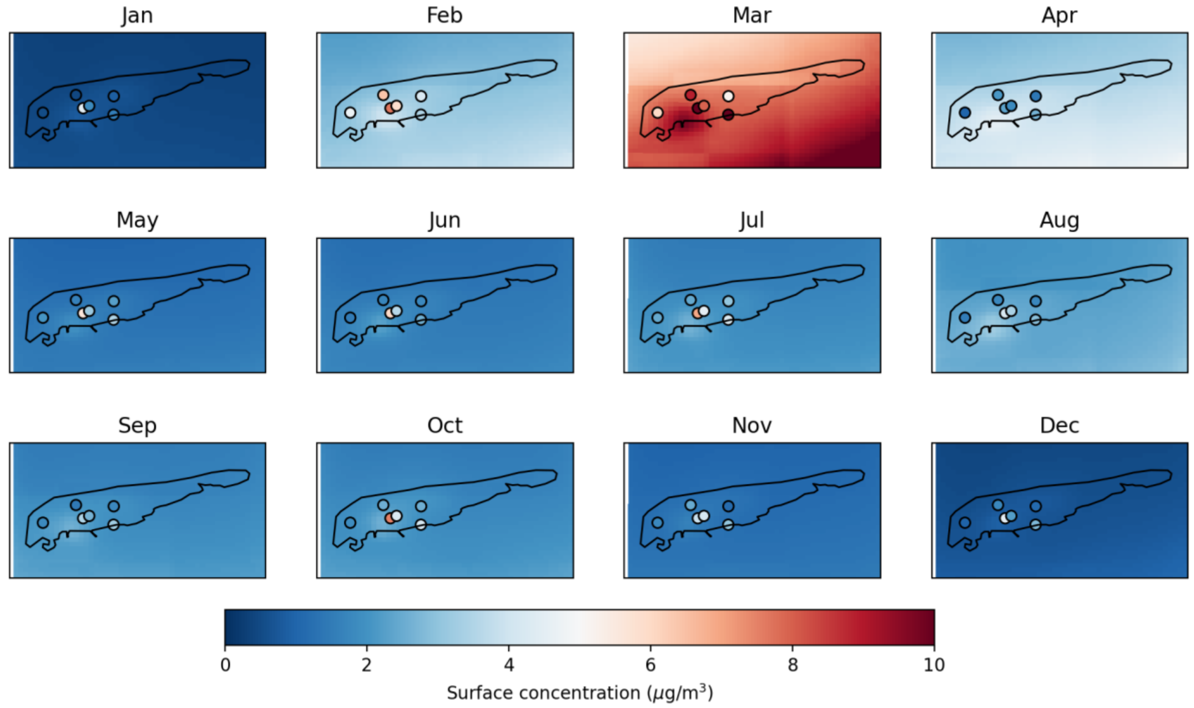


Figure 3 Monthly surface concentration of simulation for 2022 with the prior

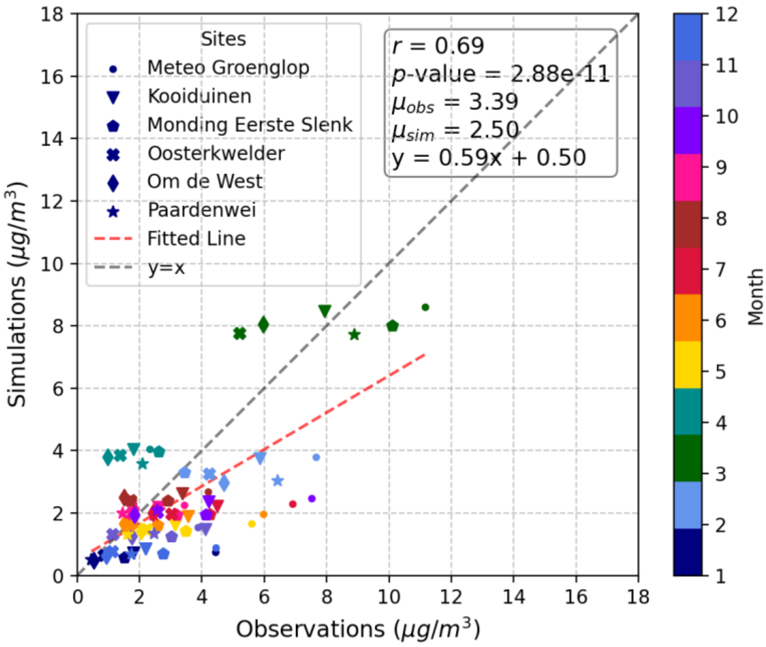


Figure 4 The comparison of model and observation in monthly average of 2022 with the prior emission.

5. *Monthly Inversion Performance (Section 3.3): Monthly inversions fail due to high MAN uncertainties and sparse data. The proposed solution (adding one LML-like site) improves results (Fig. 10) but lacks validation against independent data.*

   We strongly agree with the validation against independent data. But this is also a challenge for our study: the only data available on the island is the monthly ammonia measurement network MAN. We therefore used a theoretical excersise. The validation should therefore be done at a site where such measurements are available, but this is beyond the scope of our manuscript.

Minor Comments

1. *Section 2.4.1: Define how "external influences" (Groningen, Friesland, etc.) were selected for the state vector. Why not include Denmark?*

   We have clarified the selection method for “external influences” in the revised manuscript (line 197). Specifically, any labeled region contributing more than 10% to the total concentration in the labeling step was included explicitly in the state vector.

   While “Denmark” was added in the source apportionment analysis, its contribution remained low and below the threshold. In the inversion setup, Denmark was therefore grouped into the “Other” category, which includes all contributing regions not listed individually.

2. *Section 2.2: Briefly justify the use of CAMS-REG v5.1 (2019) for both 2019 and 2022 prior emissions despite known livestock reductions.*

   Although livestock numbers on the island decreased between 2019 and 2022 (as indicated by the KringloopWijzer report), the official emission inventory for 2022 was not yet available at the start of this study. Since air quality models like LOTOS-EUROS rely on emission inventories as input, and such inventories are typically released with a multi-year delay and often lack annual continuity, we used the latest available dataset, CAMS-REG v5.1 (2019), as a consistent prior for both years. The inversion method is then used to estimate deviations from this prior, effectively capturing emission changes (such as livestock reductions) even when there is no updated inventory reflecting them.

3. *8b: Explain why monthly posterior uncertainties are asymmetric (e.g., wider in spring).*

   We added the elaboration at Sect. 3.3 (line 366-379):

   We also attempted to use the MAN measurements for monthly emission inversion (see Fig 8b). However, the resulting posterior estimates exhibited high emission values during spring and notably low values in autumn. This asymmetry is caused by strong seasonal variability in both ammonia emissions and meteorological conditions. In spring, ammonia emissions increase and become more variable due to fertilizer application. Meteorological factors such as turbulence and boundary layer dynamics also contribute to greater atmospheric variability. Particularly, the posterior value of February exhibited more than four times the prior value, although the result falls within the leave-one-out cross-validation bounds (see Fig. S7). This increase likely reflects the onset of manure application season in February, when agricultural ammonia emissions typically peak due to fertilizer spreading on farmland. While both February and April fall within the spring period, in 2022, February experienced much stronger wind speeds (Fig. S5b), enhancing transport and dispersion. In contrast, April had lower wind speeds, which reduced the spread of ammonia and increased sensitivity to local sources. Additionally, in months like April, August, and September, prevailing north winds placed most observation sites on the leeward side of the source, reducing their sensitivity to local emissions and thus weakening the inversion constraint. In other words, the low results from the inversion may not be due to actually low emissions but rather to the measurements of those months that failed to capture and represent local emissions adequately.

4. *9b: Include the location of the proposed Kooiduinen site on the map.*

   We have updated the propsed site, which is near Waddenhaven Schiermonnikoog (53.472226° N, 6.167259° E) instead of Kooiduinen, and reanalyzed the results (Sect. 3.4, line 417-427). See also our response to reviewer 2.

RC2:

Review of "Local-Scale Inversion of Agricultural Ammonia Emissions: A Case Study on Schiermonnikoog, the Netherlands" *by Li et al.*

Li et al. present an inversion for ammonia emissions in the Netherlands. They use the LOTOS-EUROS model in a flux inversion. They use a network of surface stations measuring ammonia with a focus on improving the representation of agricultural emissions. They include both real and synthetic inversions. It is clear that the authors have done a lot of work. My may comments revolve around the overall framing of the problem and the analysis of the results. Only 6% of the ammonia is local to the island they are studying. Therefore, most of what they measure is transported from distant regions. It's not clear to me that an inversion using 6 sites on the island is really necessary since little of what they measure comes from the island. It seems like we are probably learning more about the upwind sources, but the discussion focuses on Schiermonnikoog.

This ties into my later comments (below) about their recommendations. The authors recommend an additional measurement site on the island. It feels like additional measurements in the upwind region would likely provide more value. In their recommendation for the additional site they seem to disregard their own analysis (Fig 9b). They reject three potential sites suggested by their analysis. It does not seem like the conclusions do not follow from their analysis.

Overall, I think the authors have done a lot of good work. I think the authors should reconsider the conclusions they draw and ensure they are supported by their results. I would recommend major revisions for the manuscript.

We thank the reviewer for this constructive comment. We agree that the observed ammonia on Schiermonnikoog is influenced by upwind sources outside the island, and have revised the framing of the study to reflect this more clearly. Our aim was not only to study local emissions, but also to test the sensitivity of a local-scale inversion to known emission reductions under real conditions.

We have updated the Introduction and Discussion to emphasize the broader objective: assessing whether a reduction in local emissions can be detected despite dominant long-range transport.

Regarding the measurement site recommendation, we have clarified our rationale and addressed the discrepancy with Fig. 9b. While several high-sensitivity sites were identified, our recommendation focused on improving the final inversion on the island for validation purposes. We now explicitly acknowledge the value of upwind measurements and include this in our revised discussion.

Furthermore, we have updated the monthly breakdown for the source contributions (Figs. a and b). Earlier we only included the agricultural contributions, but now we combine the agricultural and other sector emissions in the figure. The original figure could be found in the supplement.

Comments

1) Is this the right tool?

The authors mention in Section 3.1.2 and show in Figure 6 that emissions from Schiemonnikoog only contribute 6% of observed concentrations. The study is motivated by high spatio-temporal variability in ammonia emissions, yet that does not seem to be the case here. This site is dominated by transport. It seems like we would learn more about ammonia emissions from additional measurements in the regions with much larger emissions.

I think the authors can address this through some additional discussion. I.e., justifying why these measurements are still valuable and why this framework is needed to address their questions.

Thank you for your comment. It could be more relevant to do the same excesise in regions with larger emissions. However, the advantage of Schiermonnikoog was its isolation with some high local emissions which make it useful to study.

We have added this in the Discussion as suggested by the reviewer (Sect. 4.2, line 474-477):

The source apportionment analysis revealed that ammonia concentrations on Schiermonnikoog are dominated by long-range transport, which makes emission reductions more difficult to detect due to the relatively high uncertainty in MAN measurements. However, the inversion results provide not only the emission changes but also an improved estimate of the transport contribution. While the model indicated that approximately 50\% of the ammonia originated from Germany, the posterior inversion suggests a lower contribution of about 30\%. Consequently, the local contribution is likely larger than indicated by the original source apportionment.

2) Optimizing the network

The authors set up a nice framework to identify the most useful site on the island to improve the inversion. However they seemingly disregard all of the information from that analysis. It was unclear to me why they go through the effort if they are going to throw out the three most promising sites. Their text is copied below:

"Although Fig. 9(b) indicates that the most informative location would be directly at the island's main emission source, near the Schiermonnikoog-Meteo Groenglop site, this location is not ideal in practice. As discussed in Sect. 3.1.1, this site shows the least agreement between model and observation, likely due to unresolved spatial heterogeneity between reality and model. Moreover, measurement close to the source can lead to a large bias in regional misrepresentation of ammonia concentration (Schulte et al., 2022). The next model-suggested site is Schiermonnikoog-Om de West, at the western edge of the island. However, this site consistently records the lowest ammonia concentrations and is strongly influenced by sea winds, making it less suitable for detecting local agricultural emissions. Another candidate is the Schiermonnikoog-Paardenwei site, located in the north. While the model indicates it as a potentially useful location, this site is surrounded by dense vegetation, which in reality limits its sensitivity to nearby agricultural sources (see Sect. 3.1.1)"

I understand that there are practical considerations that need to be taken into account, but the conclusions don't seem to follow from the analysis.

A minor point regarding Figure 9a and the associated discussion. The observation error will also represent the errors due to the model. I.e., it is more aptly thought of as the "model-data mismatch". I mention this because I wonder what the authors would consider the model error for LOTOS-EUROS.

Improving the instrument precision will eventually be limited by the ability of the model to represent the measurements. Therefore, part of the parameter space in Figure 9a will likely be limited by the model error. This is something that is discussed in Turner et al. (2016; doi:10.5194/acp-16-13465-2016) for a pseudo-data study of urban CO_2 emissions.

We thank the reviewer for this constructive suggestion. We agree that promising sites should be considered within our study regardless of practical constraints. Therefore, we do not restrict ourselves to the existing MAN sites and instead propose a new location near Waddenhaven, Schiermonnikoog (53.472226° N, 6.167259° E). We also updated the manuscript (Sect 3.4, line 417-432):

We then propose adding an LML-like measurement site near Waddenhaven Schiermonnikoog(53.472226° N, 6.167259° E), as shown in Fig. 9(b). The most informative location would be directly at the island's main emission source. However, measurement close to the source can lead to a large bias in regional misrepresentation of ammonia concentration (Schulte et al., 2022). Waddenhaven offers a compromise: it is near the source but lies outside Banckspolder, a reclaimed polder valued for farming. In addition, all existing MAN sites are situated to the north of the source, which limits their effectiveness under northerly winds. By contrast, Waddenhaven lies within the footprint of the local emission and is typically downwind of the main source, making it the most suitable candidate for improving inversion performance.

Figure 10 illustrates the effect of incorporating a single high-quality, LML-like observation on monthly ammonia emission inversion for 2022. The synthetic observational error for the LML-like site is set at 3.5 %, based on values reported by by Dammers(2017) and Blank (2001). Compared to the current MAN-only network, the addition of one strategically placed LML-like measurement significantly reduces uncertainty in most months. Posterior estimates not only align closely with target values but also exhibit a narrower percentile spread across most months, indicating improved stability. Overall, this result demonstrates that even a single, well-placed, high-precision observation can substantially improve inversion performance, enhancing the system's ability to track temporal variability and increase DFS. With high-frequency, low-error measurements, it becomes feasible to detect near-real-time emission changes.

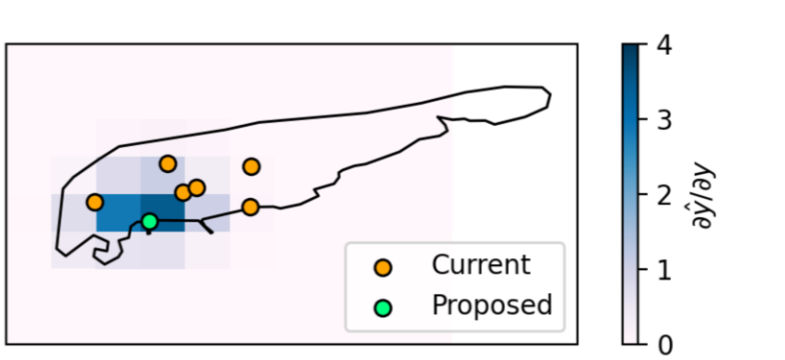


Figure 5 Where to put the site? Updated also in the draft, Fig. 9b

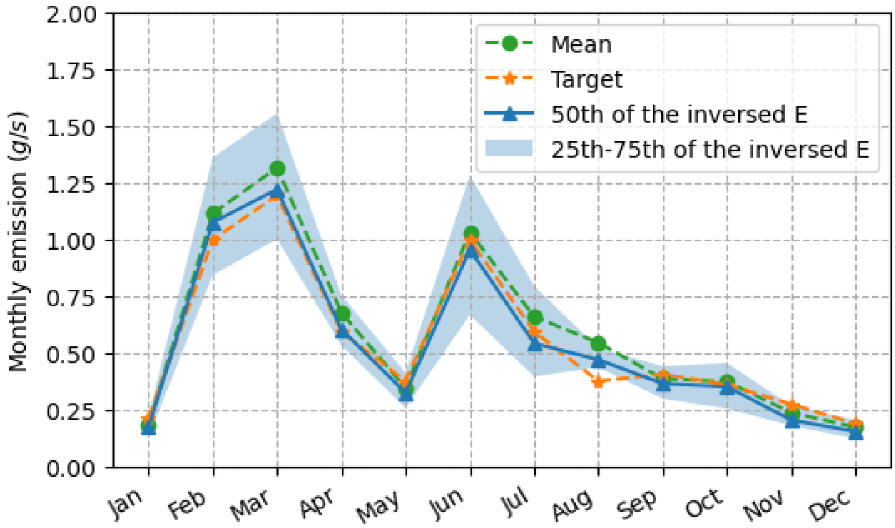


Figure 6 The posterior emission after adding one LML-like measurement, updated also in the draft Fig. 10b

We also acknowledge the reviewer’s point that a denser monitoring network may be limited by the ceiling of the model itself. We have added a discussion of this in the manuscript (Sect. 3.4, line 401-405):

Note that for denser monitoring networks (e.g., the bottom-right corner of Fig. 9a), the achievable improvement may eventually be limited by the ceiling of the model itself. In this case, part of the parameter space could be biased by model error, as also discussed in Turner et. Al (2016). Nevertheless, for the current monitoring network, as well as for moderate improvements in measurement precision, the analysis still provides valuable insights.

3) Description of the inversion

There are some key details in the inversion that seem to be lacking. For example, it was not clear from the methods section what the temporal resolution of the inversion was. Are the authors solving for 5 parameters (single scaling factors for their regions)? From Section 3.2, it seems like the authors have two setups: an annual inversion and a monthly inversion. I am not entirely sure though.

The inversion is conducted for both yearly and monthly ammonia emissions. We have added the resolution at line 151. And indeed we are optimizing 5 parameters, 1 for emission of the island and another 4 to fix the boundary conditions. We added this to the manuscript (line 200).

As an aside, it would be helpful to indicate where some of their different regions are in the Netherlands. I would suggest adding that to Figure 1 or Figure 2. I am not familiar with the geography of the Netherlands and do not know the spatial extent of Groningen, Friesland, etc. This would be very helpful for understanding the actual setup of their inversion and interpreting the results.

We have added Fig. 2(b) to show the Dutch provinces and the region for inversion:

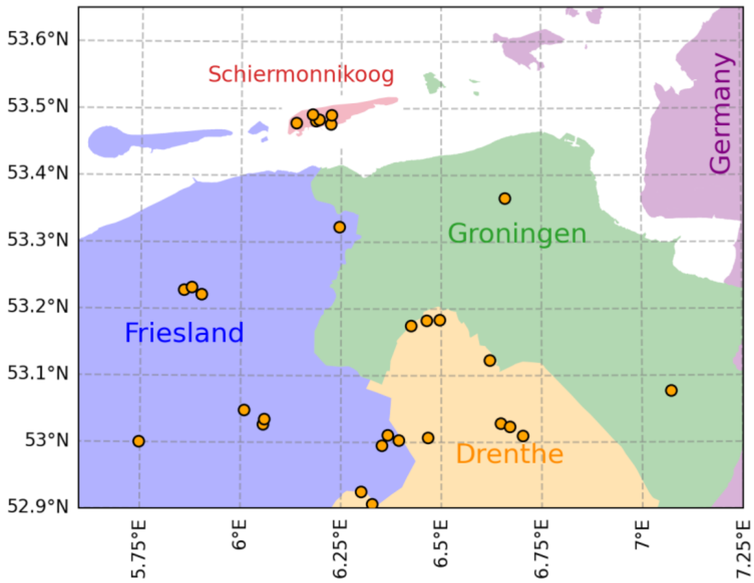


Figure 7 Dutch provinces

Boundaries: *It was not clear to me what the domain of the inversion was. What domain is being simulated and where are the boundaries. How are the boundary conditions specified?*

The domain of inversion is shown in Fig 2(b). The nested domains employed for LOTOS-EUROS simulation are Domains 1, 2, and 3 displayed in Fig. 2(a).

We define the external fluences based on its relative contribution. And the external fluences then are used for fixing the boundary conditions of the inversion (line 197-205).

Jacobian: *How was the the Jacobian constructed? Do you recompute **K** at each iteration in the LM method? Is **K** being constructed from perturbations? If so, is this a one-sided, two-sided, how big is the perturbation, etc.*

We have added the definition at line 182:

K_{linear} is the Jacobian in linear space, which is constructed with an one-sided perturbation of 40%. Notably, for monthly emission inversions, the Jacobian is defined as a block diagonal matrix so that each month is considered independent.

Prior error covariance matrix: *the description of* $\mathbf{S_a}$ *seems to be missing. How large are the uncertainties in* $\mathbf{S_a}$*? Are there off-diagonal error covariances in the monthly inversion? The authors mention that the monthly inversion performs poorly, but this strikes me as odd because a monthly inversion with temporal error correlations should give something similar.*

We defined the prior error covariance matrix S_a at Sect. 2.4.1. The matrix is constructed diagonally with terms defined as $(\ln \beta)^2$, between lines 207 and 211.

There are no off-diagonal terms in the prior error covariance. Also we constructed the Jacobian matrix as block diagonal so that each month is considered independent, reflecting ammonia's short lifetime and high meteorological sensitivity. Unlike long-lived greenhouse gases, ammonia emissions show limited month-to-month correlation due to rapid atmospheric processing and highly variable seasonal source patterns, which is the reason that the uncertainty of monthly ammonia emissions is set to a factor of 4 and yearly to 2.

Observational error covariance matrix: *I appreciate that the authors have tried to develop an* $\mathbf{S_o}$ *matrix that they think is a better representation of their network. However, it seems that they neglect off-diagonal terms in* $\mathbf{S_o}$*. This may be fine for the monthly inversion, but I think it could be problematic for the annual inversion. The authors show seasonal biases in the simulation of ammonia. This seasonal bias will manifest itself as a correlated error in the annual inversion because the model has errors in the seasonality (*$\mathbf{S_o}$ *is the model-data mismatch, and there is an error in the model as mentioned in the* *previous comment). I think a justification for neglecting off-diagonal errors is needed or a test using off-diagonal errors.*

We thank the reviewer for pointing this out. The observational error covariance matrix S_o in our study was constructed from two components: residual errors and documented measurement errors.

The residual errors were used to construct the main part of S_o. In our framework these residual errors represent the random error, which we attribute primarily to uncertainties in the emissions, given that the ECMWF meteorology and LOTOS-EUROS model have been extensively validated. The extended MAN error was applied due to its documented high uncertainties.

According to Lolkema et al. (2015) and Noordijk et al. (2020), the measurement uncertainty in the MAN network mainly consists of (i) random errors of the passive sampler, (ii) calibration method errors, and (iii) calibration standard errors. The first two are random and therefore uncorrelated across sites, while the third is a systematic component (~8%). Compared to the magnitude of the random and residual errors, this systematic part is small, and thus its contribution to off-diagonal terms is limited.

To further test potential spatial correlations, we examined pairs of stations located within the same grid cell (showing below). The comparison shows that the documented measurement errors are larger than the representation error between nearby stations. This supports the assumption that correlations between stations (off-diagonal terms) are of secondary importance. Moreover, we applied bicubic interpolation when sampling model values at the station locations to reduce representativity errors.

Based on this analysis we consider the neglect of off-diagonal terms in S_o to be a reasonable approximation for our inversion framework. We agree that for long-term (annual) inversions, correlated errors from seasonal biases are a potential issue, and we will clarify this limitation in the manuscript (Sect. 4.2, line 483-487):

In this study we neglected off-diagonal terms in the observational error covariance matrix, effectively assuming errors are uncorrelated between sites. This assumption is supported by the dominance of random measurement errors and by the relatively small representation errors observed between nearby stations. Nevertheless, neglecting correlations means that potential systematic errors, such as seasonal biases in the model, are not explicitly accounted for. While this simplification is unlikely to strongly affect monthly inversions, correlated errors could become more relevant for annual-scale inversions.

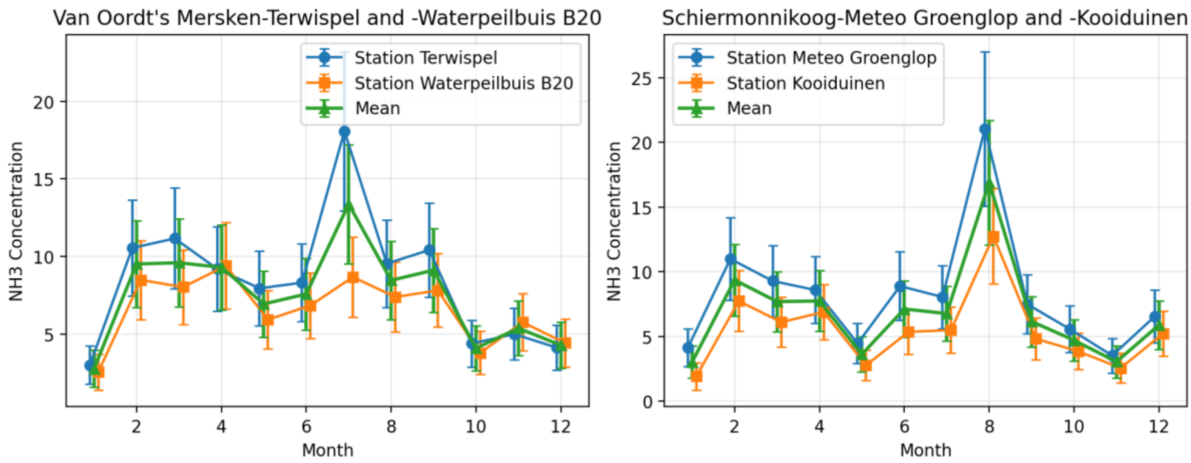


Figure 8 Stations within the same grid cell, 2019

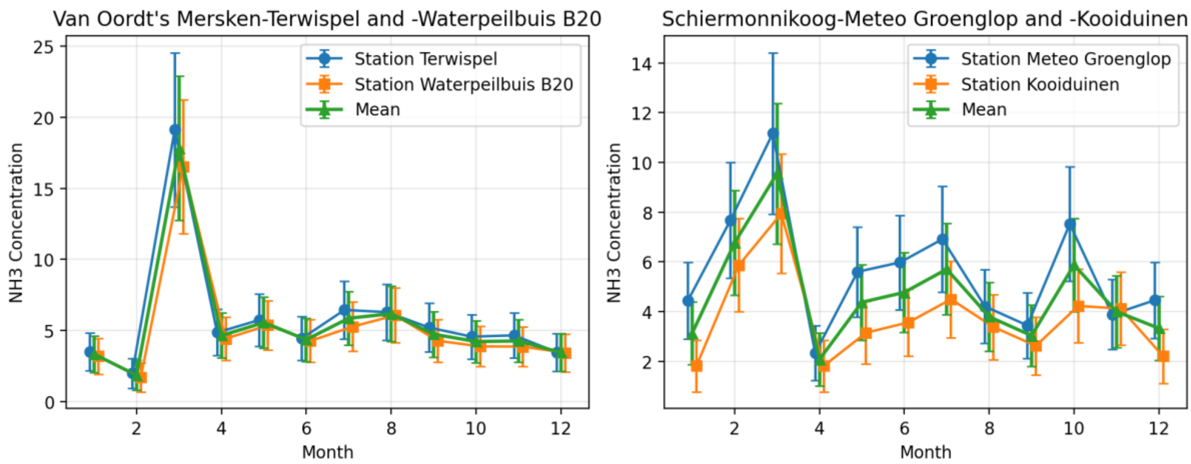


Figure 9 Stations within the same grid cell, 2022

Meteorology: *How good is the meteorology for this region? Getting the PBL height correct is likely important. Is this well-represented at the site? Models often times have difficulty getting coastal areas, is this an issue for your island? Given the large contribution of distant sources, getting the transport correct seems critical. Assessment of the meteorology seems important for this application.*

We agree with the reviewer that PBL height is important and usually an issue in coastal regions. However, we currently don't have local data on the boundary layer height as observations are lacking, and secondly we're limited to the datasets that are available. Intercomparison with Harmonie or WRF would be interesting but outside of the scope of this paper as its not standard meteorology thats coupled to LOTOS-EUROS.

To further evaluate the meteorological representation in our model, we analyzed the boundary layer height (BLH) patterns for 2022. The seasonal variability showing below, which reflects meteorologically consistent patterns rather than model artifacts. The elevated February BLH corresponds to multiple cyclonic systems that transited the region during this period, enhancing vertical mixing through mechanical turbulence and convective processes. Conversely, the lower summer values (June and August) are consistent with our island's position between Arctic cyclonic systems and European continental anticyclones, where subsiding motion suppresses boundary layer development. The March minimum aligns with high-pressure dominance typical of early spring conditions. This seasonal BLH variability reflects the maritime influence and limited continental thermal contrasts characteristic of island locations, demonstrating that our model appropriately captures the local meteorological regime.

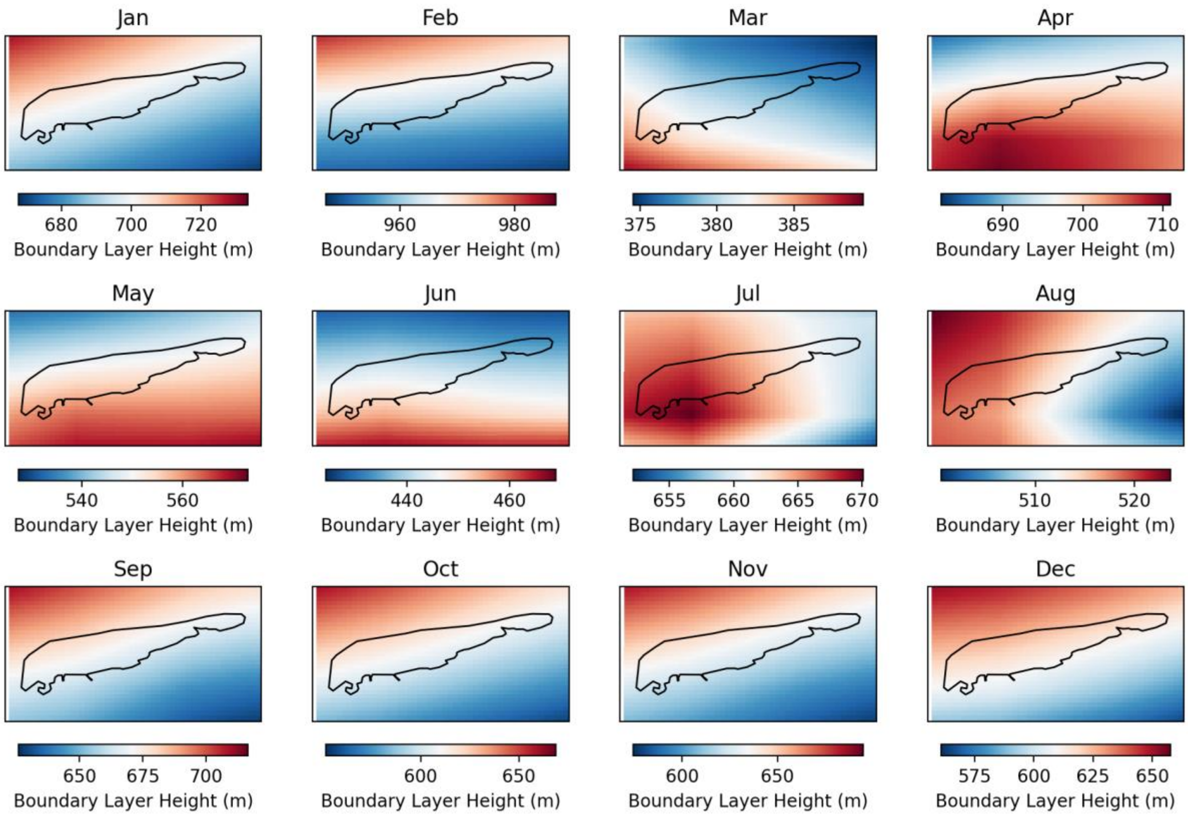


Figure 10 Contour maps of boudary layer height (BLH) in 2022

While direct BLH observations at our site are not available for validation, the model's skill in representing surface meteorological conditions provides confidence in the overall meteorological framework. We validate the meteorology with KNMI measurement sites available within the inversion domain: Leeuwarden, Lauwersoog, Eelde and Nieuw Beerta, of which map is shown below and in the supplement:

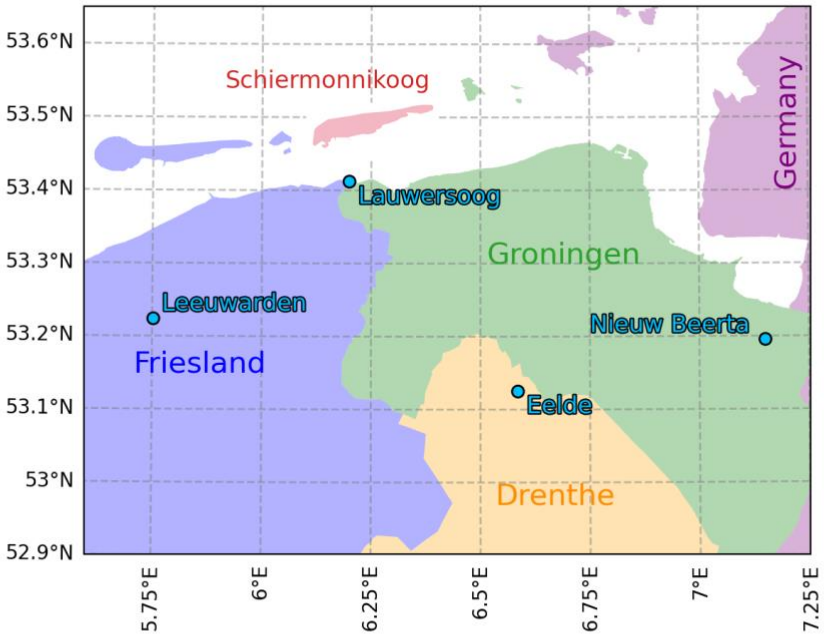


Figure 11 Meteorological measurement sites available for the validation (cyan circles)

The comparision is showing below, which is also added to the supplement. Note that the abbrevations correspond to the site names: Leeuwarden, Lauwersoog, Eelde, and Nieuw Beerta. And the measurement of pressure is not available at Lauwersoog and Nieuw Beerta.

The agreement was very good, with correlations for wind components, temperature, and pressure consistently above 0.96 and low RMSE values. Precipitation was also well reproduced, with correlations around 0.8. These results indicate that the meteorological fields are reliable and representative for the study area, supporting the robustness of the subsequent analysis.

Table 1 The statistics between the modeled and the measured meteorological parameters: horizontal components of 10m wind (U and V), precipitation, temperature, and pressure, including Pearson's correlation coefficient (r), root-mean-square error (RMSE)

			LW	LS	EE	NB
2019	U(m/s)	r	0.99	0.98	0.99	0.99
		RMSE	0.53	0.83	0.47	0.58
	V(m/s)	r	0.98	0.99	0.97	0.97
		RMSE	0.54	0.79	0.63	0.76
	T(K)	r	1.00	0.99	1.00	1.00
		RMSE	0.59	0.86	0.62	0.62

	R(mm)	r	0.75	0.83	0.81	0.82
		RMSE	3.34	2.18	2.19	2.06
	P(hPa)	r	1.00		1.00	
		RMSE	1.43		1.81	
2022	U(m/s)	r	0.99	0.99	0.99	0.99
		RMSE	0.66	0.86	0.47	0.64
	V(m/s)	r	0.98	0.98	0.96	0.97
		RMSE	0.58	0.79	0.69	0.71
	T(K)	r	0.99	0.99	0.99	1.00
		RMSE	0.62	0.85	0.63	0.52
	R(mm)	r	0.80	0.87	0.84	0.77
		RMSE	2.89	2.36	2.30	2.64
	P(hPa)	r	1.00		1.00	
		RMSE	1.32		1.64	

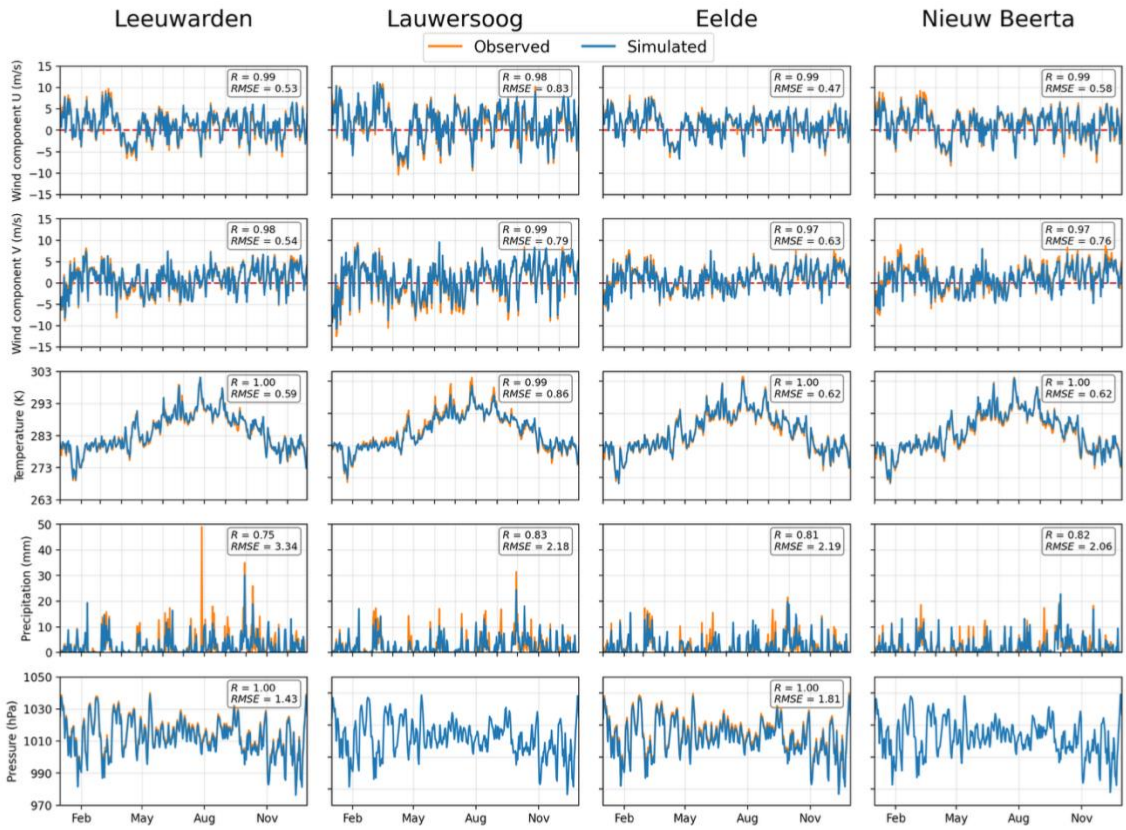


Figure 12 Horizontal components of 10~m wind (U and V), precipitation, temperature, and pressure obtained from simulations and observations, 2019

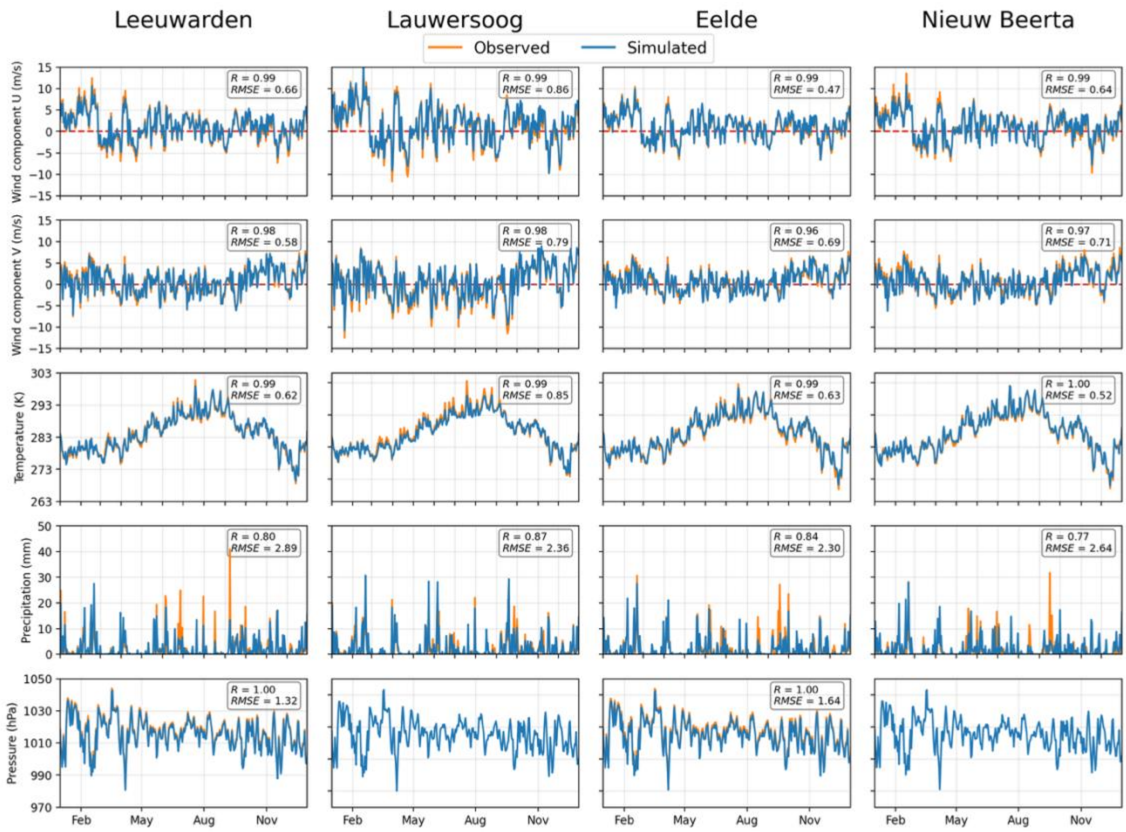


Figure 13 Horizontal components of 10~m wind (U and V), precipitation, temperature, and pressure obtained from simulations and observations, 2022

4) Inversion evaluation

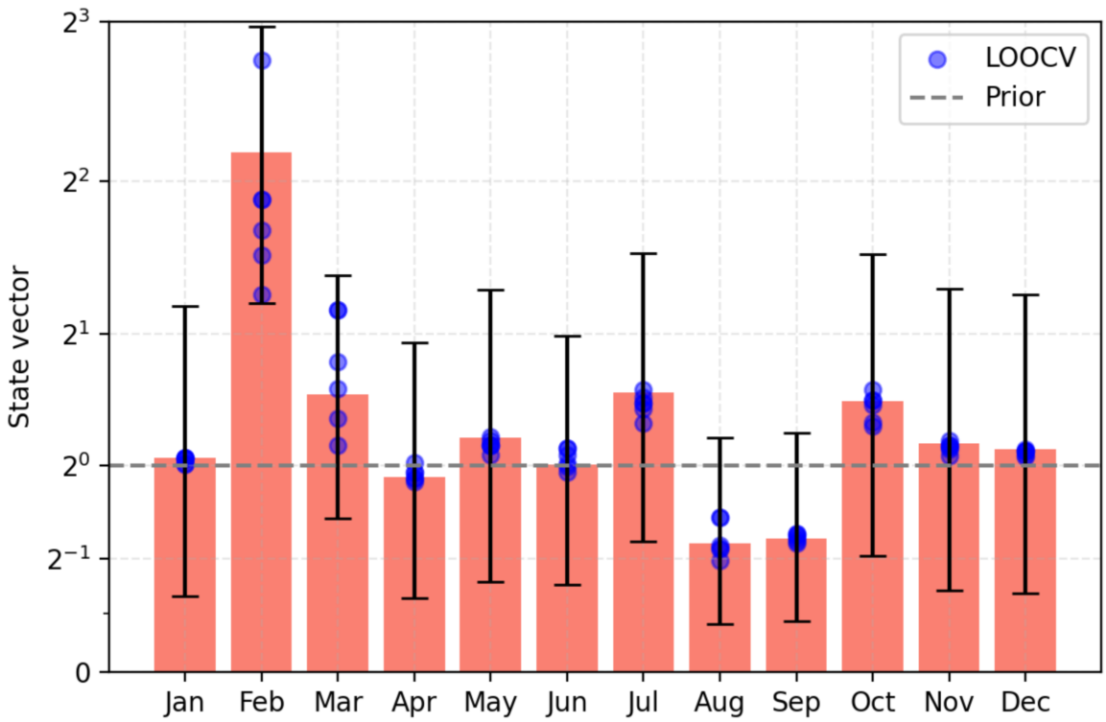
They claim the monthly inversion produces "unrealistically high" values in the spring (Line 358), but how was this assessed? Evaluation against independent observations or k-fold cross validation are two common approaches for evaluating inversions, but it does not seem like the authors have done that.

We have conducted the leave-one-out cross validation (LOOCV). The current result falls within the validated bounds, however the value in February exhibited 4 times higher comparing the prior, which is much higher as expected.

We have rephrased the paragraph in the draft (Sect. 3.3, line 370-379):

Particularly, the posterior value of February exhibited more than four times the prior value, although the result falls within the leave-one-out cross-validation bounds (see the figure below). This increase likely reflects the onset of manure application season in February, when agricultural ammonia emissions typically peak due to fertilizer spreading on farmland. While both February and April fall within the spring period, in 2022, February experienced much stronger wind speeds (Fig.~S5b), enhancing transport and dispersion. In contrast, April had lower wind speeds, which reduced the spread of ammonia and increased sensitivity to local sources.

Additionally, in months like April, August, and September, prevailing north winds placed most observation sites on the leeward side of the source, reducing their sensitivity to local emissions and thus weakening the inversion constraint. In other words, the low results from the inversion may not be due to actually low emissions but rather to the measurements of those months that failed to capture and represent local emissions adequately.



5) Inversion discussion

The entirety of the discussion of the real inversion is half a page. It seems odd to spend 17 pages discussing work and methods to estimate emissions of ammonia and then only discuss the results for half a page.

We have rewritten and re-examined our results. See also our response to reviewer 1.

Specific comments

Abstract: *Why is it broken into 4 short paragraphs? The whole abstract is 10 sentences, why is it broken into 4 paragraphs? Standard practice would typically be to have one paragraph. Also a lot of superfluous intro. Nearly half is intro/background (4/10 sentences).*

We have revised the abstract:

Quantifying real-world emission reductions is a core goal of atmospheric inversion methods, yet direct validation against known events remains rare, especially for reactive species like ammonia. In this study, we have applied local-scale Bayesian inversions using ground-based measurements and the LOTOS-EUROS air quality model, with high-resolution emission inventories as prior input, not to explore a theoretical scenario, but to evaluate a documented emission reduction. On the island of Schiermonnikoog in the Netherlands, where GVE (grazing livestock units) decreased from 639 to 541, with a particularly notable reduction in dairy cattle, ammonia emissions are expected a 23% reduction between 2019 and 2022. Our inversion captured a similar trend, estimating a 51% decrease, which may be overestimated, largely attributed to uncertainties in the 2019 posterior emissions. The posterior for 2022 shows consistency with the validation and indicates a 27\% reduction compared with the prior emissions of 2019. The associated uncertainty, derived from the posterior error covariance, highlights both the potential of the method and its limitations for policy verification. Moreover, we developed a method to assess the usefulness of individual observations and propose that adding a single high-quality continuous measurement in a strategically chosen location can significantly enhance the inversion performance. This strengthens the observational constraint and enhances the system's ability to resolve temporal variations in emissions.

Figure 1: *How big is this island? Some sort of scale would be important*

We have added a figure (Fig. 2b) indicating the Dutch provinces and the size of Schiermonnikoog.

Line 145: *How many grid cells? Maybe add to table 1*

We have added the grid configuration of the finest domain to Table 1 (174 * 160 grid cells).

Line 175: *Related, how is* ***x*** *defined? Is it each pixel, time-dependent, etc? Ah, see its defined on Line 208, would be good to explicitly state in a table somewhere or make sure its apparent. Is it time-dependent or are you assuming temporally constant?*

We have added this definition explicitly at line 160. For the yearly emission inversion, the state vector is time-independent; for the monthly emission inversion, the state vector is time-dependent with a block diagonal Jacobian.

Lines 235+: *I'm confused why they add to* $\boldsymbol{S_o}$*. If anything, having off-diagonal correlations in So would be more appropriate because the model could easily introduce correlated errors due to transport, the PBL, etc*

Following the previous discussion, we elaborate on the reason why we neglect the off-diagonal terms and further discuss the possible limitations in the draft (Sect. 4.2, line 483-487):

In this study we neglected off-diagonal terms in the observational error covariance matrix, effectively assuming errors are uncorrelated between sites. This assumption is supported by the dominance of random measurement errors and by the relatively small representation errors observed between nearby stations. Nevertheless, neglecting correlations means that potential systematic errors, such as seasonal biases in the model, are not explicitly accounted for. While this simplification is unlikely to strongly affect monthly inversions, correlated errors could become more relevant for annual-scale inversions.

Table 2 and Figs 4+: *are these using the prior? It's a bit confusing because you describe the inversion and then start talking about the model performance but not clear if its prior or posterior. It would be helpful to clarify that in the figure captions.*

Yes, these are the results using the prior. We have also revised the captions of figures (Fig. 4, 5, and 6) and Table 2.