

Review of “Ammonia measurements over multiple fertilisation campaigns using eddy covariance and the integrated horizontal flux method” by Kukowski et al.

Summary:

This manuscript quantifies field-scale ammonia emissions from agricultural winter wheat following urea nitrogen fertilization across a three-year campaign. The authors present a detailed statistical analysis comparing concentrations and emission estimates from two distinct techniques: quantum cascade laser-based eddy covariance (QCL-EC) vertical fluxes with an integrated horizontal flux method applied to samples collected from Adapted Low-cost Passive High Absorption (ALPHA) samplers (ALPHA-IHF). This work aims to establish whether emissions estimates differ from these different measurement techniques and flux calculation methodologies, especially where the measurement techniques differ greatly in approach (offline passive vs. real-time optical), cost, and temporal resolution.

This analysis provides a valuable and timely contribution to the peer-reviewed literature given current EU legislative efforts to curb agricultural ammonia emissions. Agricultural ammonia emissions remain difficult to quantify, and few multi-year studies are available for contrasting measurement and analysis approaches across varying environmental conditions and known fertilization events. This direct comparison of eddy covariance vertical fluxes and integrated horizontal fluxes provide essential insights that should be carefully considered when planning future field studies.

Overall, I recommend this manuscript for publication following some revisions and clarifications.

General Comments:

Additional context regarding individual measurement techniques for both measurements (e.g., inherent precision, limit of detection, accuracy, and drift) should be included in the Methods section and propagated throughout the Results and Discussion sections for context.

An uncertainty analysis for endpoint concentration and flux estimates across the independent measurement methods is currently lacking. Could uncertainties for each individual measurement technique and flux calculation method be propagated through for comparison with the combined statistical approaches?

The theoretical background in the Supplementary Information is cumbersome. Can some parts be streamlined to focus on essential support for the main conclusions? For instance, given that the equal-precision scenario is ultimately rejected in favor of a weighted-mean approach in the Bland-Altman analysis, consider asserting directly that the equal-precision assumption is unrealistic and limiting detailed methodology to the weighted-mean implementation for this approach.

Including a few basic supplementary figures—such as a time series of the measurements with propagated uncertainties, diurnal profiles of the concentration data and meteorological parameters, and box or violin plots showing data variability—could provide valuable context for the analysis (particularly for Section 3).

Specific Comments:

P2, L64: Were any passive samples lost due to preparation, handling and analysis in this work?

P3, L111: Did soil type ultimately influence the observations (e.g., resulting in higher concentrations or fluxes for specific soil types)? If a future campaign included sufficient statistical sampling of paired measurements across different soil types, could soil type be incorporated into a covariate analysis?

P3, L114: To clarify, did all of the topsoils have the same pH, C-content, cation exchange, and urease activity even though the specific sites and their soil types differed?

P4, L118: For readers unfamiliar with the area, what does “local best management practices” mean in this region?

P4, L121: Did measurements only occur after fertilization or did you also capture some pre-fertilization time periods as well. If the latter, the fertilizer application times should be marked in Figures 2 and 3.

P4, L122: The time periods for each campaign, which ranged from 6 to 25 days, are quite different. Did it rain or do you know the irrigation schedules during these time periods? Could also be added to Figures 2 and 3.

P6, L156: Roscioli et al, 2015 used 1H,1H-perfluorooctylamine to improve the ammonia time response. Is this a typo or did you passivate with a different perfluorinated compound? If yes (and it worked) this would be a really exciting addition to the literature in the measurement details. In either case, how much passivant was used over what period of time and at what flow rate. Was the improvement in time response as good as that reported in Roscioli?

P6, L160: Was an external calibration source applied to the QCL at any time before, during, or after the study? Were routine zeros or span calibrations performed during the campaigns to assess drift? Zero drift, even with temperature stabilization, is a particularly important consideration for the QCL instrument. This section would also be an appropriate place to add more information with references about the QCL instrument’s precision, detection limit, time response, and accuracy according to the manufacturer’s specifications and for the configuration used in this study.

P6, L174: What does poor performance mean? And how were those periods identified? What criteria were used for manual inspection?

P7, L193: Was frequency locking used for the QCL? If so, what was the frequency and duration of the locking routine, and were gaps filled using linear interpolation or ignored? If the frequency lock routine was not used, can you estimate how much variability there was in the NH₃ concentrations due to spectral drift?

Section 2.3.1: Please add more information about the inherent precision, detection limit, and uncertainties for the passive measurement technique with references.

P9, L251: What was the source of the elevated background concentrations? Are there other agricultural sources upwind? How would an upwind source confound this analysis? How often did this background exceed the measurements of interest? What percentage of the full sampling period was retained v. eliminated for this reason?

P10, L301: It is unclear what “the more precise method” refers to in this sentence.

Table 1 and Section 3:

-Please clarify how many usable pairs there are per campaign year. If I have this correct, 12 pairs are excluded leaving 9 remaining usable pairs for 2021. Or possibly fewer as indicated in Figure 2 and Table 2. Is this enough pairs for effective statistical analysis for campaign I?

-It is interesting that the largest fertilization event, in 2021, led to the shortest duration of campaign days. Was that expected? Can you discuss why that might be the case?

-Was there something unique about 2021 compared to the other campaign years? Would the conclusions of this paper change if all 2021 pairs were eliminated from further statistical analysis?

-Please also address the difference in 8-20 days compared to an earlier section noting 6-25 days.

P14, L350: To clarify, is this related to the inherent detection limit of the passive samplers or the propagated uncertainties of the IHF method?

Figure 2: Can you add the standard deviations on the mean for each data point in the figure to get a better visual of the day-to-day variability? Also, measurements were conducted continuously from March to July of each year, yet the paired data are only shown for campaign periods <20 days. In which months were these campaigns conducted? Does collection during different seasons impact this analysis, or is this sufficiently accounted for in the covariates analysis of Temp and RH?

P18, L390: Would orthogonal distance regression (ODR) or weighted orthogonal distance regression (WODR) have been a better approach to the initial linear regression analysis (e.g.,

<https://amt.copernicus.org/articles/11/1233/2018/>)? In this case, a specific measurement would not need to be selected for the x-axis.

Figure 4: how do the fits change if ODR or WODR is used instead of OLS?

P27, L543: Do you have soil moisture measurements or model-based estimates? If yes, could this be included in the covariates analysis?

P28, L458: What were your criteria for the manual decisions and visual inspections?

Conclusions: Based on these findings, can you expand your recommendations to address broader field deployment strategies? Specifically, despite their inherent differences and constraints, should future studies adopt a hybrid approach—such as pairing a high-precision QCL instrument with a distributed network of lower-cost ALPHA sensor towers to better constrain the fetch mass balance? Or would the additional methodological complexity outweigh the benefits?