Authors reply to Referrees 1 & 2 and J. Laubach's reviews and comments on egusphere-2025-1605 manuscript :

« Aerodynamic gradient flux measurements of ammonia in intensively grazed grassland: temporal variations, environmental drivers, methodological challenges and uncertainties » by Abdulwahab et al.

Authors comments are indicated in blue.

Contents

****

1- General reply and opening remarks to all comments (RC1, RC2, CC1)

We thank the reviewers for their very positive and constructive reviews and comments. We are pleased that they all find the NH3 flux dataset valuable in the context of relatively poorly documented emissions in grazed grassland systems. There are some overarching comments, which are adressed below.

1. Both RC1 and CC1 comment on the length of the manuscript, but not RC2. RC1 and CC1 argue essentially that much of the materials and methods could be deleted, reduced or shifted to the Supplement for the sake of greater readability ; while by contrast RC2 welcomes the facts that « …*Data analysis is comprehensive* …» and « …*methodological challenges and its associated uncertainties*… » are discussed in detail.

   - We do acknowledge that the manuscript is extensive, but not by any means of a length unheard of in Biogeosciences. More importantly, we believe (like RC2) that the topics of field scale NH_3 flux measurements and related uncertainties warrant a very thorough description and critical assessment of measurement methods, post-processing, and careful interpretation of data. Ammonia is a very difficult gas to measure properly, and so many things can go wrong in the concentration/flux measurement, calculation, correction and filtering process. Hence the deliberately strong methodological slant we gave to the manuscript, to remind the BG readership of the importance of careful measurements and data treatment for NH3 fluxes, rather than simply reporting NH3 flux data and glossing over uncertainties. We (the authors) are only too aware of limitations of our data, and it is not in anyone's interest to pretend all is perfect.

   - Nonetheless, we do agree with RC1 and CC1 that it is in every reader's interest to remove from the main article any superfluous information, which just gets in the way and hinders the flow of the paper. We have gone through the paper and indeed we agree that the text can be streamlined in quite a few areas without siginificant loss of information. We will indicate below in the detailed reply to Referees exactly where space can be saved in the main article, mostly by shifting things to the Supplement.

   - In a preliminary revised version, we have thus reduced the total word count (including references and extras) from 16269 down to 14747 words (1522 words removed i.e. -9.3%). The reduction in word count excluding references and extras is from 12854 down to 11567 (1287 words removed, i.e. 10%). These are *net* overall numbers, accounting for the text we have *added* in parallel in response to reviewers requests for more information in some places.

2. Another comment in common by RC1 and CC1 is that the NH_3 flux data we present do not fundamentally add to our understanding of processes controlling the emissions (RC1 : « … *the analysis part of the article, which is rather descriptive/ phenomenological and lacks depth or real new insights on explanatory factors…* », and CC1 : « … *the full potential of the dataset has not been sufficiently explored…* », « … *NH_3 emissions vary enormously. It is a valid point to show this, but not a very novel one…* », « … *Are you planning to use this dataset for the testing of process-based models…* »), etc. They both argue that more should be done to better understand (i.e. model) the mechanisms underlying the emission process, and that the paper is too descriptive and fails to achieve this.

   - We fully understand the frustration of the reviewers in this regard, it is a very natural reaction. We carried out all those intensive field campaigns during 4 spring seasons in the field precisely with the ultimate objective to better our understanding of processes with the aid of process-based modelling. However, as pointed out earlier, the present manuscript is already long enough, and any further multi-variate analysis or modelling work was way beyond the scope of the present paper. This manuscript was aimed at 3 main things : i) to present a novel setup (lift system and associated data treatment) for AGM NH_3 flux measurements over pastures ; ii) to describe (and in the process make available to the community) a large flux dataset covering a wide range of conditions ; and iii) to discuss measurement challenges and uncertainties.
   - However, the next step is (and has been all along) a second paper on the adaptation and application of a process-based model (Generation of Ammonia by Grazing, GAG, Moring et al., Biogeosciences, 14, 4161–4193, https://doi.org/10.5194/bg-14-4161-2017). This second (modelling) paper is currently (as of 01.08.2025) at the draft stage, and is also intended for submission to Biogeosciences. « Samples » or « teasers » of this modelling paper can be made available to Reviewers and the Editor if need be.
   - Our original intention was to submit both papers at the same time as « twins » (along the lines of : Paper 1 : NH_3 flux measurements and uncertainties, and Paper 2 : Process-based modelling using GAG), similarly to what the second and last authors of the current manuscript have done in the past in Biogeosciences (e.g. https://bg.copernicus.org/articles/7/521/2010/ and https://bg.copernicus.org/articles/7/537/2010/). However, due to practical (time) constraints, we were not able to follow this plan, and we needed to submit the current manuscript as stand-alone.
   - The modelling (GAG) paper, to be submitted to BG towards the end of 2025, will adress the concerns of RC1 and CC1 with respect to the perceived lack of furthering mechanistic understanding. The GAG model incorporates the underlying thermodynamic, biogeochemical and micrometeorological mechanisms of NH_3 exchange, constrained by meteorology, soil-vegetation dynamics and pasture management and grazing livestock density.

   In our view, the current manuscript cannot be further extended to include multivariate or model-based analyses (out of scope and out of space here…), but the data description we provide makes it clear that the emission process is complex and multi-factorial, resulting in a very large variation in flux levels and dynamics, that can only be adressed through process-based modelling (in a subsequent publication).

   - To answer RC1's specific question on why no multivariate analysis was attempted in the current manuscript, the long and the short of it is that the most crucial driver of the multiple regression is missing in the available dataset from the field, namely : soil surface total ammoniacal nitrogen (TAN) concentration at sufficiently high temporal resolution. There was no way we could sample soil TAN during four spring seasons, at field scale (10-20 spatial reps), at half-hourly resolution (to match flux data) or even daily resolution ; the very few datapoints available in Fig. 6 do not even begin to describe the rapid (hour to hour, day to day) changes in urine-N deposition by many cows and the fast-changing urea and TAN concentration in soil. But that is what the GAG model simulates, and why we chose GAG to further interpret field data. Without a knowledge of soil surface TAN (and pH), the other controlling factors (temperature, VPD, turbulence) cannot constrain the magnitude of the flux.

3. Both RC1 and RC2 require clarification on the procedure for QCL concentration selection during each vertical mast ascent / sampling cycle, and subsequent averaging, detrending, etc, and the sensitivity of

calculated fluxes to the choices made in Methods. We are thankful for these comments, as they made us realize that this central part of the data treatment was not adequately described, and the procedures were not transparent. We will add a couple of sentences in the text, and a couple of figures in the Supplement (in order not to further add to the length of the main article) ; but it is crucial to realize that NH3 being a sticky/polar/hydrophillic molecule, there cannot be (with this sampling system) an instantaneous step change in concentration from one sampling height to the next, and much equilibration time is needed, which shortens the duration of available bias-free data for end-of-position-averaging and subsequent flux calculation. The bias introduced by sequential sampling in non-stationary conditions is a further complication. This will be further elaborated on below in reply to specific comments, as well as many other methodology-related questions, which highlight the importance of thorough and transparent method description.

2- **Reply to** **RC1**: 'Comment on egusphere-2025-1605', Anonymous Referee #1, 11 May 2025

General comments

The article 'Aerodynamic gradient flux measurements of ammonia in intensively grazed grassland: temporal variations, environmental drivers, methodological challenges and uncertainties' is a valuable contribution to the field of ammonia research. It presents an interesting dataset on ammonia fluxes in grazed grassland and the data treatment is conducted thoroughly and described in a transparent way.

One concern, however, is the length of the manuscript which is, based on the number of lines, roughly twice as long as is common. The length of an article has a strong influence on its accessibility. I, therefore, strongly recommend the authors to reduce the length of the manuscript by removing all lines which are not absolutely necessary. I understand that this leads to tough choices since none of the sentences are not worthy to read but all together the article length is way too long. Examples of lines which could be removed in my opinion are e.g. l35 'Globally – productivity', l66 'It -prevalent', l132 'The surface -profiles', the section describing the Delta denuder could be roughly halved etc.

A major emphasis of the paper was placed on methodological challenges and uncertainties in NH3 flux measurements, since ambient NH3 concentration and surface-atmosphere exchange of NH3 are notoriously difficult to measure. The difficulties arise not only from the sampling and bias-free determination of NH3 by analyzers in the field (adsorption/desorption effects, damping effects and loss of high frequencies, interferences by evaporating ammonium aerosols or gas-particle conversion, etc), but also, in the context of rotational grazing and relatively small paddocks, from footprint limitations, changes in wind direction, and potentially, interferences by other nearby NH3 sources.

Because of all these complications, we strove not only to describe the mesurement systems and the data treatment very thoroughly, but also to develop fairly innovative procedures to improve and control data quality. For example : i) a single sampling line moving up and down the mast to measure the gradient ; ii) the use of DELTA denuders to correct QCL NH3 concentrations ; iii) correction for footprint using background flux model ; iv) a new integrated flux classification index (qcA, qcB, qcC) to account for all of : turbulence status, stability, NH3 stationarity and footprint contribution ; and v) a critical assessment of a commonly used approach (mean diurnal variation) for gap-filling to derive cumulative flux estimates and emission factors. We believe that most of these innovations are worthy of description and might benefit the NH3 flux community, and we are therefore reluctant to downsize radically the Materials and Methods section (and subsequent discussions), because the innovative methodological aspects would be somewhat lost in the process.

That said, we definitely agree that the text can be significantly reduced in a number of places, as suggested by the referee. In the revised version we will :

- Delete around 200 words (and some associated literature references) from the introduction (l35-38, l46-50, l75-76, l81-83).
- Move section 2.3.5 to the Supplement (~320 words removed). The referee is right to point out that the DELTA system has been described many times elsewhere.
- Remove section 2.3.6 (~210 words saved), and put the corresponding information on instrumentation in a new table (as suggested by the referee) in the Supplement.
- Section 2.4.3, item (a) : remove most of the explanations on the qc 0-1-2 system by Mauder and Foken (2004), just keep the first sentence, mention the reference, and move the rest (l275-282, 140 words saved) to the Supplement. Like the DELTA description, this has been described extensively elsewhere.
- Figure 6 is not absolutely essential to this paper, and furthermore somewhat difficult to read, as commented by the reviewer. Fig.6 will be moved to the Supplement, side by side with the pH plots in Figure S9. (Subsequent figures 7 and 8 will be renamed 6 and 7).
- Section 2.6 : removed lines 349-368 ; the full technical description (including Eq. 12-13) of the mean diurnal cycle calculation for gap-filling is moved to Section S6 of the Supplement ; only a brief summary remains in 2.6 (opening 2 sentences), with a reference to S6. Around 280 words can be saved here, since the mean diurnal variation approach is not entirely novel and has been used and described previously.

- Small text segments were removed selectively throughout the discussion, which were a little repetitive or did not add to the argument. This saved some 250-300 words. A few references, used only once in these segments, were removed altogether.

My other major concern is related to the analysis part of the article, which is rather descriptive/ phenomenological and lacks depth or real new insights on explanatory factors related to emissions of grazing events. It must be possible to dig out more from such a valuable dataset. The authors explain that a multivariate regression analysis of the EF's didn't work out for various reasons. However, why wasn't an multivariate regression analysis of the fluxes attempted?

A multivariate regression analysis of the half-hourly fluxes requires that all major drivers are known and available as inputs to the regression, at a similar space and time scale as the measured flux data (i.e. field scale, half-hour). The data show clearly the effect of meteorology and turbulence (see e.g. Fig.7), which are available as drivers. However, the data also show clearly for all grazing events (Fig. 4) the strong temporal development of emissions over 1-2 weeks, starting from background (near-zero) levels before the arrival of the animals, increasing emissions over several days, a peak sometime near or just after the end of grazing, and a return to background over a week or so. This fairly systematic « background-increase-peak-decrease-background » pattern has little do to with weather or turbulence, but mostly to do with the gradual increase and then decrease in available mineral (ammoniacal) nitrogen in topsoil, driven by the arrival and departure of grazing animals (except in the case of applied fertilizer or slurry, but these data were excluded from the analysis).

Meteorology and turbulence matter, and do scale fluxes, but only as long as there is TAN available for exchange at the soil surface. We did not have the means/manpower/budget to sample TAN in topsoil at the field scale for every single hour or even every single day, for every grazing events ; thus no soil TAN data were available at high enough resolution to drive the regression at the required time scale (the data shown in Fig. 6 are actually very scant), and it is very hard to think of any other measured proxy for TAN to use in a predictive regression model.

In my opinion it also could be an option to divide the article in a part 1 and part 2. Part 1 consisting of the main part of the current manuscript extended with a sensitivity analysis of the choices made in deriving the fluxes . Part 2 could then consist of an elaborated version of the current results section (e.g. 3.4-3.6) extended with a statistical multivariate analysis of the fluxes; perhaps combined with a modelling part.

We explained our publication strategy in the introduction to this reply. The plan always was to have two papers, the first (this one) on flux measurements and methodological challenges and uncertainties, and the second one on process-based modelling. The second paper is currently at the draft stage, and should be submitted towards the end of 2025 ; the mechanisms underlying NH3 emissions will be investigated, with the multiple interactions (thermodynamics, biogeochemistry, turbulence, soil/grassland dynamics, herd management, stocking density) accounted for in a comprehensive modelling (GAG) framework.

Thus we propose to streamline the current paper, reducing content wherever possible to improve readability, and proceed as planned with the submission of the modelling and process interpretation paper in the next few months.

Sensitivity analysis of the choices made in deriving the fluxes (e.g the 5 sec length of the concentration time series at each height; how sensitive are the results when 10 secs are used?, the detrending procedure, the background flux correction and the gap filling procedure).

This is a very good point, and one that was in fact investigated in depth already at the beginning of the field flux campaigns, back in spring 2021, but not detailed in the current manuscript. In fact, the value of 5 seconds mentioned l202 was incorrect, and we apologise for the mistake and confusion. The averaging interval was always 10 seconds, centered on the last second of each sampling height (the time when the inlet starts moving to the next position), including the previous 5 seconds and the next 5 seconds. The extra 5 seconds (after the inlet has already started moving) were included to account for any time lag between inlet and QCL cell and to allow further equilibration time, based on observations that a further time delay was needed, in particular for the direct transit between top and bottom of the mast (start of a new cycle). To clarify the data treatment process we will add a new figure (S1, see below) to the Supplement and refer to the figure in the main text.

The topmost panel in Fig. S1 shows the continous 1-second NH_3 time series measured every second by the analyzer, for an example half-hour (nine consecutive lift ascents) with a strong emission gradient on 02 May 2023 12 :30. At each second the air sample inlet is located at a given height on the mast : either stationary at any one of the five designated sampling heights, indicated by a color from red (0.1m) to blue (2.1m) ; or moving from one position to the next (grey line between consecutive colors). The thick black horizontal lines indicate the time intervals during which the inlet was stationary at one height, indicated on the right axis. By contrast the longest time interval when the inlet was in motion was at the end of each cycle, from the top height (2.1m) all the way down to the bottom height (0.1m) without stopping, marking the beginning of a new cycle upon reaching 0.1m.

It is evident that in many cases for this example emission half-hour (and generally in the whole dataset), the NH_3 concentration continued increasing for quite a while after the bottom position had been reached, after the inlet had completed the move down from top to bottom. In such cases, it makes sense to wait as long as possible for the NH_3 signal to stabilize, before recording an average NH_3 value for this sampling height. (However, bear in mind that the lower two heights 0.1 and 0.2 m were not actually used for gradient determination). In the subsequent panels of Fig. S1, the averaged concentration and averaging interval are indicated by horizontal colored symbols (red to blue for the 5 heights), based on averaging times of 5 to 25 seconds (for testing purposes), with 10 seconds being the value used in all our calculations. It can readily be seen that an a longer averaging time of say 20-25 seconds would significantly reduce the magnitude of the gradient between top and bottom of the mast, due to carry over effects from one height to the next. Despite the high flow rate, heated and insulated PFA tubing for the sample line, the air sample was delayed and the high frequencies damped due to the stickiness of NH_3 and very likely the effect of the in-line filter located inside the QCL to preserve the cell mirror reflectivity.

Adding such a sensitivity analysis could provide valuable extra insight in the quality of the flux dataset; which is no luxury since each half hour flux value is only based on 45 sec of actually measurement data at each height. Although the data treatment appears to be carefully conducted the fact that only so few data points are used makes it hard to believe that the flux values are robust this being the last of my larger concerns); a sensitivity analysis could help to develop more trust in the flux values.

We will add a new figure (Fig. S2, see below) to the Supplement, showing flux results based on averaging times between 5 and 25 seconds and compared with the actual values based on 10 seconds. The results demonstrate that the flux results are not very sensitive to the averaging time (regression slopes very near 1 and near-zero intercepts), even though Fig. S1 showed significant differences in calculated mean NH_3 values for different averaging times. This is because the largest concentration equilibration effect, observed for the air inlet lift transit from the top to the bottom height, could not impact the calculated fluxes since the lowest two sampling heights (0.1 and 0.2m) were actually not used in flux calculations (section 2.3.4). Nevertheless, it is true that the shortest averaging time of 5 seconds did result in a greater scatter (see red triangles in Fig. S2). 10 seconds appears to be a reasonable compromise.

We will add a sentence to the main article, section 2.3.4 : « …See Supplement Fig. S1 and S2 for a comparison with other averaging times and a sensitivity analysis for flux calculations… »

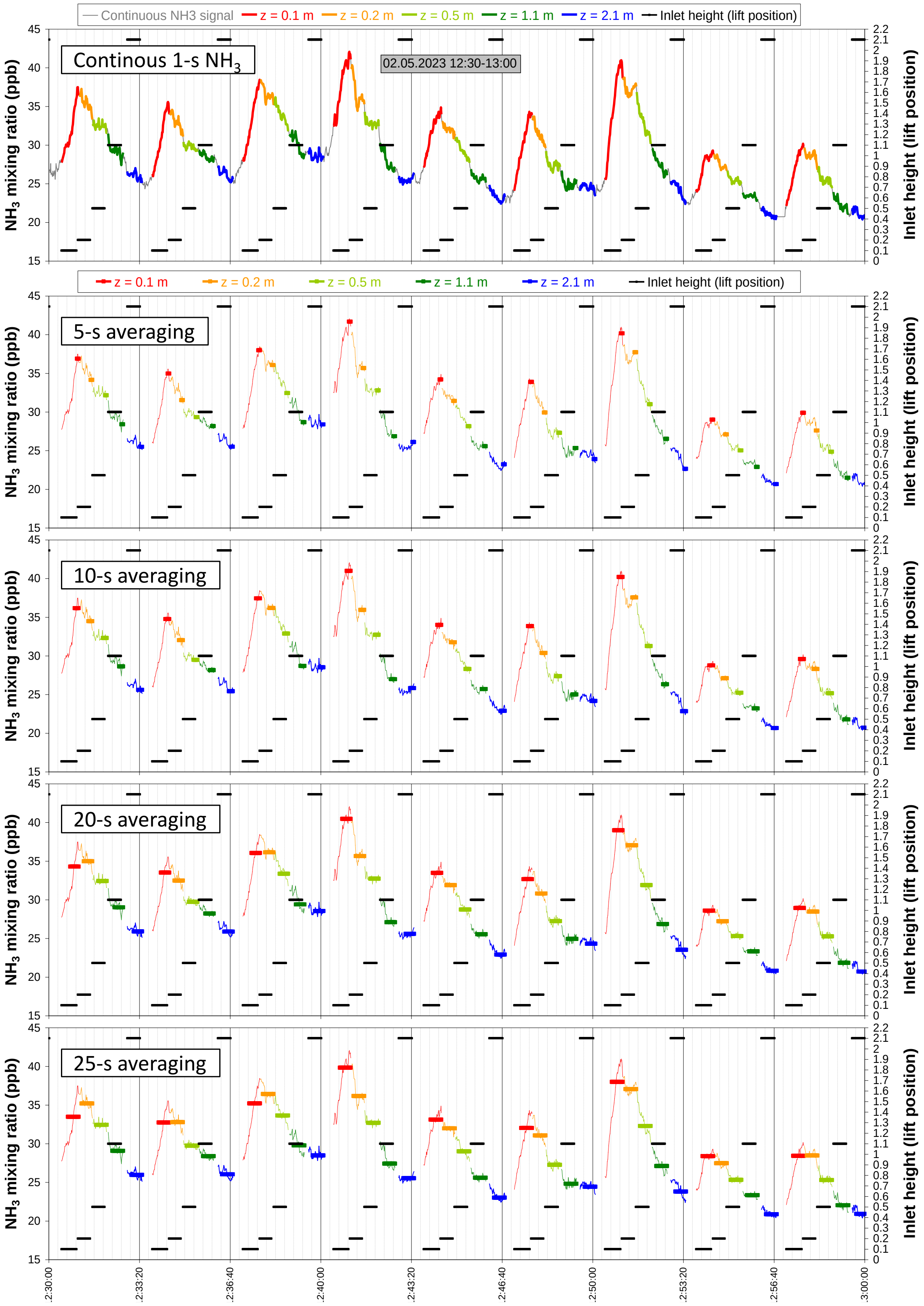


Figure S1 (added to Supplement): An example half-hour of high frequency (1-sec) gradient-lift data, showing nine consecutive mast ascents by the lift carrying the air inlet. The top panel shows the 1-sec continuous NH_3 time series, with concentrations measured at different heights when the inlet was stationary (colors red to blue) or moving between heights (grey line). All fluxes in the main paper were calculated using a 10-sec averaging time (5 seconds before + 5 seconds after end of a given position); other averaging times from 5 to 25 seconds are shown for comparison. The data show the need for sufficient equilibration time between heights in order to minimize carry-over effects, which inevitably reduce the magnitude of the gradient. The effect was naturally strongest for the lift descent from top (2.1 m) to bottom (0.1m) ; however, this did not affect flux calculations very strongly (see Fig. S2) since the lower two heights (0.1 and 0.2m) were not used in gradient calculations.

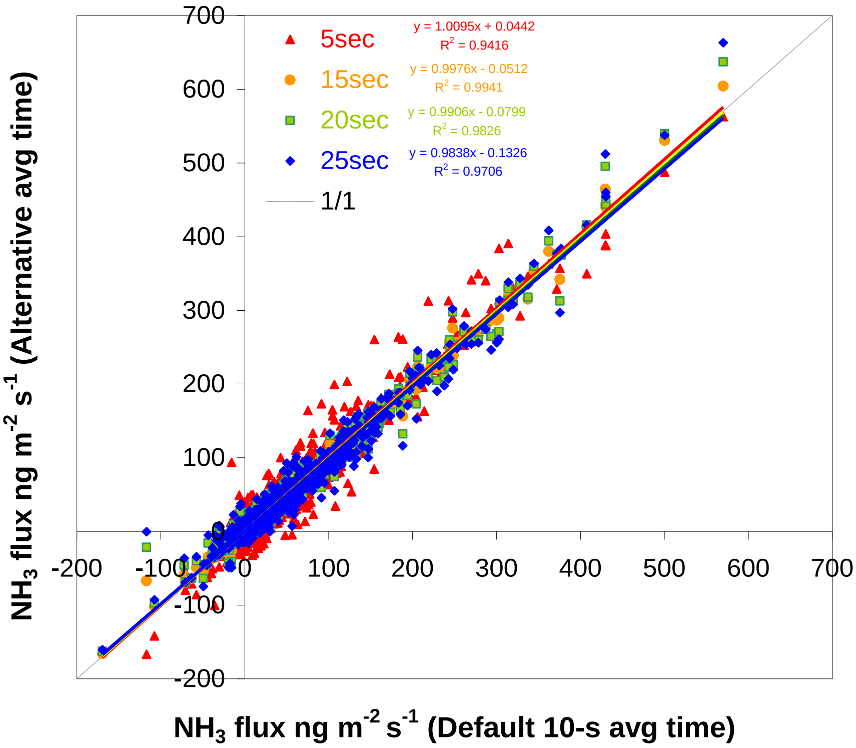


Figure S2 (added to Supplement): Sensitivity test of calculated fluxes (for 2023 data) to the averaging time chosen for high-frequency NH_3 concentrations at the end of each sampling height, following the scheme shown in Fig. S1, with 10-seconds being the averaging time used in all flux calculations shown in the main article. The calculated fluxes are not very sensitive to the averaging time (all slopes near 1 and offsets near 0). This is primarily because the 0.1 and 0.2m heights were excluded from the gradient calculations (see main article), while the longer equilibration time required was for the lift descent from the top (2.1 m) to the bottom (0.1 m) heights, as shown in Fig. S1.

Due to the length of the article itself I have not taken time to study systematically the supplementary material.

Specific comments

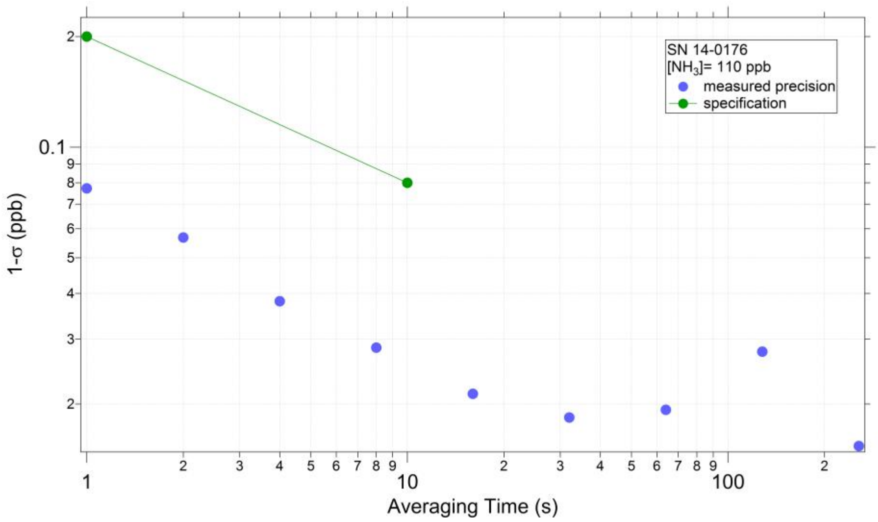
L75-l83: this paragraph is unclear to me, what is your main message? Agreed, this is unclear, or not especially relevant (in the case of bLS). We will delete l75-76 and l81-83.

L157: In neutral and stable conditions it is assumed… The same assumption is usually made for convective conditions as well. In many seminal publications (e.g. Webb, 1970 ; Dyer and Hicks, 1970 ; Thom, 1975, Monteith and Unsworth, 1990, see Reference list in the paper), the equality of eddy diffusivity for momentum, heat and trace gases is only assumed for neutral and stable conditions ; while for unstable conditions the different stability functions (for momentum and heat) lead to different eddy diffusivity values.

Comment : Sadly we missed a very good opportunity in this study to attach a fine thermocouple and a 2-D sonic anemometer to the lift system to measure sensible heat and momentum fluxes by the full gradient method, for comparison with the eddy covariance-derived H and τ at this site.

L191: what is meant with the phrase 'the nominal precision (1-sigma)'? Why (1-sigma)? We mean the specification provided by the manufacturer (Los Gatos Research) ; and we specify 1-sigma (1 std dev) because sometimes precision is expressed as 2-sigma. We will rephrase as follows : « *The nominal precision (1-std. dev.) of the QCL analyser, provided by the manufacturer, was 0.2 ppb at a 1-s integration time, and 0.08 ppb at 10-s integration time …* ». The value of 0.08 ppb is now also mentioned because we average concentrations over 10 seconds at the end of each height.

The figure below shows the Allan deviation plot provided by the manufacturer :



L199: In order to check whether I understand your procedure correctly: you apply linear detrending between two consecutive timeseries at the same height, hereby using the full 30-50 seconds of data?

No, not all the 30-50 seconds of data are used, for the reasons explained above in the discussion of the sensitivity to averaging times ; not all data can be used due to the need for an equilibration time every time a new height is reached. Only the calculated end-point 10-sec mean concentrations are used ; the rest of the dataset (inlet lift transit time between heights, and equilibration time at each height) is discarded. Here is the procedure :

1- The 10-second average concentrations NH3(z, i) are calculated for each of the N sampling heights (z), for each of the nine 200-sec lift ascent cycles (i=1..9) of each half-hour (as shown in the new Fig. S1, see above)
2- For all nine cycles of each half-hour, at each height (z) of the N-point gradient, a trend (slope) is calculated between cycle (i) and cycle (i+1) as follows :
   slope(z, i, i+1) = [NH3(z, i+1) – NH3(z, i)] / CycleDuration (slope unit : ppb s-1)

with CycleDuration = 200 sec

3- A mean slope across all N sampling heights is calculated as follows :

$$slope\_avg(i, i+1) = [slope(z1, i, i+1) + slope(z2, i, i+1) + \ldots + slope(zN, i, i+1)] / N$$

4- The middle height (z=3 in a 5-point gradient, or z=2 in a 3-point gradient) is taken as the reference (zRef) and is not corrected. The other heights (z=1,2,4,5 in a 5-point gradient ; or z=1,3 in a 3-point gradient) are corrected as follows :

$$NH3\_corr(z, i) = NH3(z, i) + slope\_avg(i, i+1) * \Delta t$$

with Δt being the time interval (in seconds) between the end of height z (when the 10-sec concentration is calculated) and the end of the reference (middle) height (zRef). Δt is positive (negative) if height z is sampled before (after) zRef. Thus, if there is an increasing trend over the interval cycle i to i+1, the heights sampled first (before zRef) are corrected upwards, and the heights sampled last (after zRef) are corrected downwards.

5- The half-hourly concentrations are calculated at each height as the average of the nine corrected concentrations :

$$NH3\_30min(z) = [NH3\_corr(z, 1) + NH3\_corr(z, 2) + \ldots + NH3\_corr(z, 9)] / 9$$

This procedure will be added to the Supplement (new section S2), and a reference to S2 inserted in the main article in section 2.3.4.

Using timeseries 1 and 2 and then 2 and 3 and then 3 and 4? Next you use the last 5 seconds of each of the 9 time series to obtain an average representative for 30 min, this is done for every height and thus obtained profile is used to determine the flux. Maybe it helps when you add words like 'first', 'second ', 'next' etc in the description. Or even better, add an conceptual figure on this crucial data treatment part. See above.

L208-214: Why is this calibration necessary when only Delta NH3 are needed for the calculation of the flux? Add one sentence explaining that it is worthwhile to have reliable estimates of the absolute NH3 concentration as well. We are not sure we understand the question. By « Delta NH3 », does the referee refer to the DELTA sampling system ? The comparison with the DELTA sampling system (regression equations in Fig.2) shows that the LGR/QCL concentrations are biased on both slope and intercept ; a correction (based on DELTA data) is necessary for the flux measurements, because an error in the slope affects the vertical gradient proportionately with the slope deviation from 1 (this is not the case for an offset). Further, a better calibration of the absolute NH3 concentration is needed for canopy compensation point modelling, where the concentration difference between soil/vegetation potentials and the atmosphere drive the magnitude and direction of the flux. An absolute concentration is also needed in case one calculates a deposition velocity (for dry deposition studies). We will add the sentence : « … *This correction of the absolute concentration is necessary because a slope deviation from 1 affects the vertical gradient (and flux) proportionately, and because accurate concentrations are required for compensation point or deposition velocity studies.* »

L229: 'near absolute mean' What is the uncertainty (nominal accuracy) in the concentration as measured by the Delta? We did not find any nominal accuracy or uncertainty value quoted in the literature for DELTA systems, because DELTA systems are samplers, not analyzers. Ultimately the accuracy of the NH3 concentration depends on the capture efficiency of the two dry denuders in series, the accuracy of the gas meters used in the system (which can drift over time), the accuracy of the NH_4^+ concentration of the denuder-extracted solutions, measured by colorimetry or ion chromatography in the laboratory, which means that the overall accuracy (or uncertainty) of a given DELTA sampler will depend on the procedures followed in the laboratory operating the DELTA system (as shown in the laboratory intercomparison exercise described by Tang et al., Atmos. Chem. Phys., 21, 875–914, 2021, https://doi.org/10.5194/acp-21-875-2021, in which our INRAE lab took part). We followed the rigorous protocol published by UKCEH, verified that the capture efficiency remained above 95%, checked the gas meters using high-precision rotameters, and performed regular NH_4^+ standard calibrations in the lab.

Section 2.3.6: for easy reference and in order to reduce the length of the article I suggest to put the information in this section in a Table. This is a good idea ; a new Table (S1) will be added to the Supplement.

L256 Storage flux correction: with a time scale of 30 minutes? Based on the measurements below z_mean?

Yes, the storage correction is calculated over a time scale of 30 minutes up to the mean measurement height.

Why is this discussed here? To me it appears more suitable as part of 2.5 (see also my comment on section 2.5 below) The storage correction is presented here as part of the section 2.4 on flux data processing and corrections ; while 2.5 deals with (random) uncertainty calculations.

L273-282: I suggest to shorten this section (once more to reduce the length of the article) only mentioning what you did and skipping the explanation of the flagging policy (just give a good reference for that).

Agreed. We have moved most of the section on momentum and heat screening to Section S4 of the Supplement, and added a reference to S4.

L311: small or near-zero emission or deposition: why do you assume this?

This is an assumption we make based on several thousands of half-hourly flux mesurements made at the site well outside of (in-between) grazing events, which have consistently shown small bi-directional exchange fluxes (such as the examples given in Supplement Fig. S14). But we do not make this assumption if we know that there has been a very recent grazing (or fertilization) event on the adjacent field ; in this case no footprint correction is applied (as explained on lines 323-326).

In figure 3 the time series of NH_3 concentration shows rather high values around 9/04, 20/04 and 22/04; fluxes around this time could be substantial (however are apparently filtered out in the quality control round) so how valid is the assumption that background fluxes are small?

Large ambient NH_3 concentration at this intensive farming location can be related to on-field emissions (if emission is strong and dispersion weak), and/or to other farm sources (farm buildings, manure storage, emissions from other nearby field), also if dispersion is weak and plumes are advected towards the measurement location. Large concentrations on those dates may or may not have resulted from nearby field emissions, or farmstead sources ; but what is certain is that large NH_3 levels were often observed at the site during calm/night-time/stable conditions, when AGM requirements in terms of turbulence statistics, stationarity, footprint, etc, were not met. For all the fields directly adjacent to the flux tower site, we made sure to document all spreading and grazing activities in order to avoid mis-interpretation of our data.

Eq 8 and 9: since 9 follows easily from 8 only one of the two equations needs to be presented (and when you really want to be lean both can be removed; l318-319 is clear enough about the procedure.) We have saved much space in many other spaces, the gain here would be minimal. This correction procedure being fairly novel in NH_3 emission studies, we believe it helps to be totally explicit as to how things are done.

Section 2.5 I miss a discussion on systematic errors here. Of course they are difficult to determine but a few can be discussed. For example the effect of a systematic error in the determination of the sampling heights (in a Delta z of 1 m an uncertainty of several centimeters already leads to a systematic bias of several percent) and displacement height. Also the (propagated) effect of an assumed but realistic systematic uncertainty in the concentration could be taken into account. Maybe the separate discussion of the theory on random uncertainties and the actual discussion in 3.5 on values of both random and systematic uncertainties doesn't work that well. I wouldn't mind when you combine the two and discuss them at one place.

Section 2.5 is part of Methods, in which we describe how things are done and calculated. We only have standard equations for random uncertainties, and none for systematic errors of course, because, as the referee says, and as has been demonstrated in other papers, no-one knows how to calculate systematic uncertainties. That does not mean we don't discuss them : we do, in Section 3.5 and 4.1. But we don't really have a proper method to describe here in M&Ms.

What we can do is rename *Section 2.5 – Random uncertainty analysis*, to be more accurate regarding the content.

L355: corrected -> you mean the correction for the background from section 2.4.5. No, we mean the fully corrected flux, including all types of correction : detrending, storage, DELTA, as well as the background term.

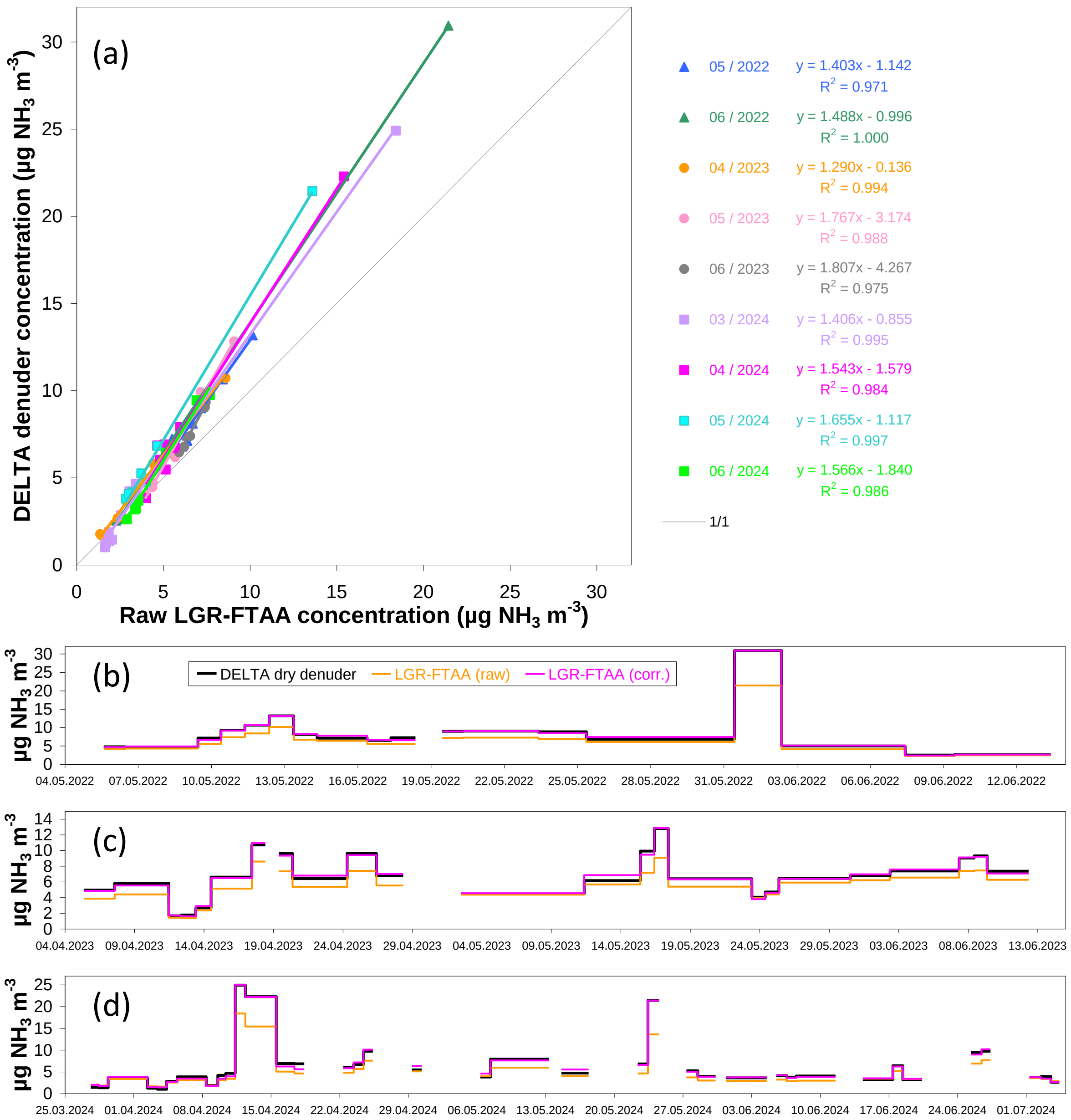
L366-368 I didn't check the supplementary material but these sentences are unclear without doing so: why are G9 and G10 different and do they need an modified gap-filling approach? As mentioned earlier in the reply, this section on mean diurnal calculations was mostly moved to the Supplement (S6). The difference between G9-G10 to the other grazing events was a near-total lack of data for a few days during grazing, so it's not just hours that needed gap-filling, but whole days. The procedure is now described in S6, we feel it would be superfluous to add more text here in the main article.

Section 3.1 How can you be sure that the results discussed in this section don't influence the flux measurements. E.g. doesn't the cleanliness of the mirror affect the uncertainty in the flux ?

We certainly do not think or claim that the correction of QCL data by comparison/regression with DELTA samples does not influence the calculation of the fluxes. Of course the correction of QCL concentrations propagates into the flux calculation, and it is intended to do so, because we believe i) the DELTA-based correction provides more accurate concentrations and vertical profiles, and ii) more accurate profiles lead to more accurate fluxes. And yes, we know from experience that the cleanliness of the mirrors does affect the concentration outputs by the QCL, that is precisely the reason why we carry out the correction using parallel DELTA samples, to provide a continous reference and correct for the analyzer's loss of signal. So we are not sure what the Referee means here ?

Figure 2b: to me it makes more sense to plot the results equidistant in time, just plot lines instead of bars and adjust the length of the line according to the delta time period it is represents.

We thank the referee for this suggestion. The results will be presented as suggested (see modified Fig. 2 below), but with stacked panels, one for each spring campaign, otherwise there would be large empty spaces in the figure between the spring periods. The modified figure also includes a re-drawing of the top panel as suggested by Referee 2, with a square (not rectangular) shape and colours.



Updated Figure 2, redrawn to accomodate changes requested by both Referees.

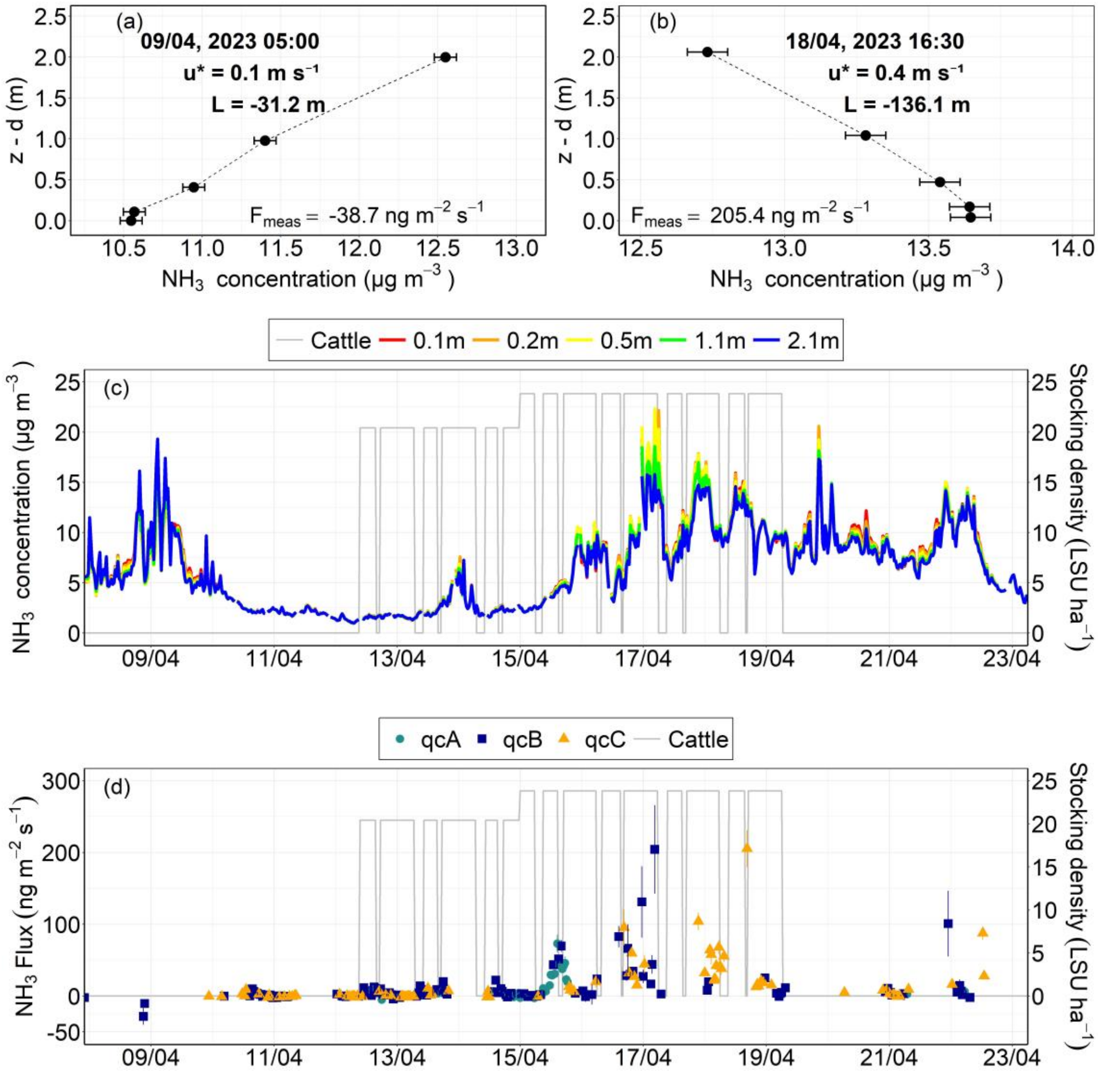
Figure 3 Please add uncertainty/error bars to the NH3 concentration values in a and b.

Uncertainty/error is ambiguous in this case ; we are not certain what the Referee means. First there are the precision (repeatability) and accuracy (closeness to absolute concentration level) at which the instrument measures NH3; then there is the natural variability of NH3 concentration within a half hour at a certain height, which is of course independent of the performance of the analyzer. The former two indicators wish to describe how well the instrument can resolve relatively small vertical gradients ; while the latter is a statistical descriptor of the natural variation around the mean NH3 value in a sampled population over a certain time interval (30 minutes). In field data the observed NH3 variability results from both types of errors.

We believe it is more useful to compare the nominal precision of the analyzer with the size of the gradient in Figure 3, rather than to show the standard deviation of all measured concentrations over a half-hour at a given height, since the latter is much larger (a few ppb in this case) but does not reflect on the measurement capability of the gradient system.

Thus we will add horizontal error bars corresponding to a 0.1 ppb precision (i.e. ~0.07 µg m^{-3} ; see updated figure below), since this is roughly the precision level specified in the Allan deviation plot (see earlier) for a 10-sec

integration/averaging time (the actual value is 0.08 ppb). This will be explained in the figure caption as follows :
« *... In panels (a) and (b) the given NH3 error bar is equivalent to +/- 0.1 ppb (~0.07 µg m*$^{-3}$*), which is roughly the manufacturer's specification for a 10-s integration time; note this does not reflect the NH3 variability (standard deviation) over the half-hour, which is much larger.* »



Updated Figure 3 including error bars on the NH3 concentrations in the example profile data.

Table 2 SE means standard error? « SE » will be changed to « *Std. Err.* » in the table

Figure 4 The visualization of the data in this figure doesn't work very well. Maybe leave out the information on the quality of the fluxes (the information in Table 2 is sufficient), plot the figures 10 by 1 instead of 5 by 2 so that the time axis can be extended and present the information on the cattle differently (only when it is present and not the variation in number since the variation is most of the times not very influential; just mention the milking times in the text and not visual in the figure since you conclude it has not real influence).

Fig. 4 will be redrawn as 10*1 (see below) instead of 5*2, as suggested by the Referee (also Fig. S5 in supplement). However, we strongly disagree with the suggestion of not displaying the info on the quality of the fluxes (qcA,B,C) ; to us it is paramount to show the flux quality classes in the time series here because it is a central point of the discussion on uncertainties, and also an innovative way of attributing integrated quality indices (see our opening remarks on methodology). The instantaneous values of stocking density dropping to zero during milking can be left in the figure ; they are real, objective data and serve to illustrate the dynamics of cattle movement on/off the field (even if, as the Referee argues, the instantaneous effect is not detected in measured fluxes).

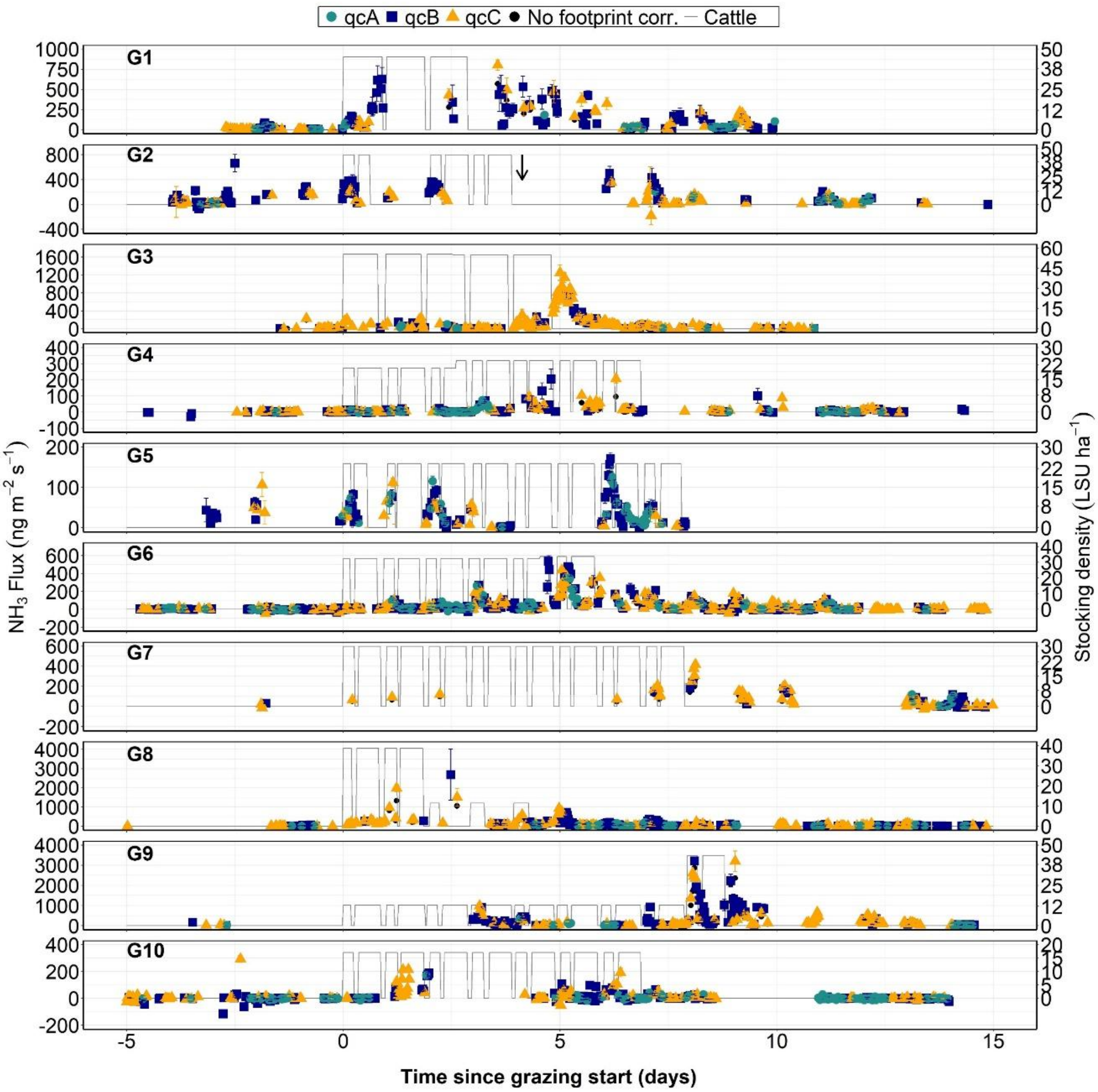


*Fig. 4 redrawn as 10*1.*

L442-446 Please refrase, your main message is difficult to grasp.

We agree the wording was confusing and the message unclear. The idea was that micrometeorologically valid fluxes, which also satisfied the 2/3 footprint requirement, but which, only for comparison purposes, were not footprint-corrected in Supplement Fig. S5, were still significantly larger than the fluxes which did not satisfy the 2/3 footprint requirement ; i.e., the field emission was clear even without FP correction. It was not the FP correction which made the fluxes much larger than background.

But since this is a somewhat confusing observation, this will be removed from the main text (to save a few dozen words), while the caption of S5 in the supplement, now renumbered S7, will be rephrased as : « …*Figure S7: Similar to Fig. 4 of the main article, but showing valid NH3 fluxes* <u>*without*</u> *footprint correction, and indicating in addition (grey crosses) fluxes that were micrometeorologically valid but for which the required footprint criteria (> 2/3) were not satisfied. The latter were almost systematically significantly lower than the former, highlighting the importance of footprint correction.* »

We will modify and simplify text in the main article as follows :

« *For comparison purposes, the flux timeries are also shown in Supplement Fig. S7 without footprint corrections, highlighting the importance of footprint in data interpretation. For some grazing events (e.g., G2, G7, G8) the valid flux data capture was patchy mostly because the wind was blowing from unsuitable directions; the flux data thus discarded due to footprint (shown as grey crosses in Fig. S7) may be fully representative of the adjacent plot (A or B, depending on which is the plot of interest), or a mixture of both plots, highlighting the difficulty of characterizing temporal flux patterns in rotational grazing with one single flux measurement setup located on the divide.* ».

L448-451 Is this a relevant observation (here)? Could perhaps skipped?

Agreed, this does not belong here in « Temporal variations ». The sentence will be moved to Section 3.5 on uncertainty, and merged with the content thereof.

Section 3.4 A nice description of observed correlations is given but as a whole this section is a little bit disappointing because it gets stuck at this level. In section 4.4 it is stated that no statistical multivariate analysis could be applied to derive the share of EF variance explained by variables but in my opinion it is a missed opportunity that no such analysis is applied on the fluxes! Why not try such analysis for each grazing event separately and one for all cases together in order to infer variables which explain specific events and which variables are important explaining generic behavior of emission events due to grazing. Now the reader is left with huge variability in emission over de various events and no hint at an explanation.

We explained earlier in this reply that the main reason for not trying multivariate analysis was a lack of data (or proxy) for the temporal variations in soil surface TAN content. The only way to deal with the absence of such data is a process-based model (such as GAG, https://doi.org/10.5194/bg-14-4161-2017), using cattle data (grazing density, urination count and volumes per day, urine N content, urea hydrolysis rate) to drive the hour to hour, day-to-day evolution of soil TAN and to explain the 1-2 weeks dynamics in the observed fluxes (Fig.4) . This will be tackled in the next (modelling) article.

Figure 5 Wouldn't be worthwhile to add an figure depicting the diurnal cycle over the ten grazing events? Either normalized or not before combining, depending on what is interesting to show?

The existing Fig.5 shows both the different magnitudes of the raw fluxes (between grazing events) and the normalized diurnal patterns for each event. The idea of showing data event by event was to demonstrate that there is a consistent shape of diurnal pattern across all events, with night-time minimum and daytime mid-afternoon maximum. It follows that over the whole dataset a similar pattern would indeed exist and could be aclculated (provided one normalized the fluxes first), but showing an additional figure would not add to the argument. Also, the reason for showing the mean normalized diurnal patterns for each event separately is that the mean diurnal cycle is actually used numerically for gap-filling each grazing event separately (we don't use a mean overall diurnal pattern over all grazing events for that).

Figure 6 Why not present these results as profiles with height labeled with time instead of height and time the other way around? I find this (and other comparable plots in the article) hard to interpret. How do the soil mineral nitrogen concentrations relate to precipitation?

We acknowledge that Figure 6 is somewhat difficult to read ; but more to the point Figure 6 is not absolutely essential to this paper. In the interests of reducing the size of the paper, Fig.6 will thus be moved to the Supplement, side by side with the pH plots (new Figure S9). It is impractical to change the figure in the way the Referee suggests (as vertical profiles), because not all depths were sampled systematically every time, and the soil layer depths sampled changed over time (sometimes 0-2, sometimes 0-5, sometimes 0-30 cm, etc).

The soil TAN information would of course be hugely valuable for multivariate analysis, as discussed earlier in the reply, if only soil samples had been taken every day during and after grazing events, for 2 weeks, for each grazing event, with enough replicates at field scale (10 ? 20 ?). Such an undertaking was naturally beyond our means. As it is, the available TAN data are far too sparse to describe temporal variations at the relevant time scales. But they were collected more with a view to calibrate or verify model (GAG) predictions.

Figure 7 Did you make plots with all grazing events (labeled by color and aggregated in periods of ‘days since grazing’) in one figure? Doesn’t this show a pattern?

These plots are available for all grazing events in Supplement Figs. 6-8 for each of air temperature, vapour pressure deficit and friction velocity. The point is that (once again) the whole flux data collection, either grazing event by grazing event, or with all G1-G10 data pooled, contains half-hour flux measurements made at different times since grazing started ; some of them at the beginning, some of them at the peak emission, some of them two weeks later, some of them in-between. Depending on nitrogen dynamics in the soil (in response to N addition by urination), the different fluxes are each related to a specific TAN content in soil at a certain point in time, and therefore the relationships to individual micro/meteorological drivers show very large scatter. This is the message we have tried to convey in Fig.6, in which the size of the symbol is scaled with the time elapsed since grazing started, but this remains an extremely poor proxy (and obviously not a linear one) for soil TAN over 2 weeks of grazing/post-grazing.

L511 Conversely- response This is rather an stand alone observation. It would be more interesting when information would be added whether this happened (albeit on a smaller scale) more often? E.g. during events G8 and G9??

We agree that the effect of rainfall is inconclusive, and the evidence confusing. Intuitively, light rainfall can trigger emissions in very dry conditions, or heavy rainfall can dilute and drain TAN deeper into the soil. But there simply aren’t enough data (and never any repetitions) to make a proper analysis, because soil/weather/grazing/grassland conditions for flux measurements are always unique in the field : nothing is controlled, flux drivers are always multi-factorial, and TAN information is always missing.

Figure 8 Isn’t this a somewhat worrisome result? The measurements are done in the constant flux layer so I would hope a smaller systematic bias was found (in absolute sense).

We actually don’t think this is a worrisome result, although obviously it highlights important considerations for flux-gradient measurements in managed/grazed grassland, which we wish to make clear for this methodology-oriented paper, for two reasons. First, the specialized literature (e.g. Garratt, J.R., 1992. The Atmospheric Boundary Layer. Cambridge University Press, Cambridge, UK, p. 316.) tells us that the traditional flux–gradient relationships of the Monin-Obukhov similarity theory tend to break down in the roughness sublayer (RSL), which is located in the lower part of the surface layer (below the inertial sub-layer) and represents the region immediately above an aerodynamically rough surface, such as a vegetation canopy. Observations suggest that the RSL extends from the vegetation canopy height (h) to approximately 2-2.5h (e.g. Tuovinen and Simpson, Atmospheric Environment 42 (2008) 8371–8381 ; see also references in the current paper). It is true that this is most often a concern discussed over tall rough vegetation such as forests, but the concept holds theoretically also for shorter vegetation. Clearly the measurement heights of 0.1 and 0.2 m were too close to the ground, being of the same order as canopy height itself (typically ~ 0.2m at the start of grazing, declining to ~0.1m at the end of grazing) and thus not clearing the top of the RSL a large fraction of the time.

The second reason was related to the presence of an elliptical enclosure around the instrumentation, to protect the equipment from grazing animals (this would not be needed in a fertilized cropped field). This was kept as small as possible, from 1 to ~3.5 meters between the gradient mast and the fence, depending on wind direction, and around 2m for prevailing wind from SW). Obviously no urine-nitrogen deposition took place within the enclosure, so the near-field footprint for the lower two heights was significantly affected by unfertilized/ungrazed vegetation within the fenced-in area. By contrat the 0.5m sampling height was not significantly influenced. This is discussed extensively in Section 4.1.3.

Did you check how choices made in the determination of the flux influenced this bias? Could this be used to determine the best choices regarding the flux derivation? E.g. the length of the period chosen (now 5 secs at each height)?

Fig. 8 actually shows that the sensitivity of calculated fluxes to the choice of the height range is rather small most of the time (7%, or slope of 1.07) for all fluxes below 2000 ng m^{-2} s^{-1}, when comparing the default case (0.5-2.1m) to the alternative case using the range 1.1-2.1m. The effect is very large obviously if one uses the lower heights, but as discussed above and in the text, there are very good, objective reasons for NOT using the lower two heights 0.1-0.2m in flux calculations. We believe this figure may be informative and useful for flux gradient setup design by other colleagues in the field.

We have previously addressed the issue of the averaging time (see above).

Table 3 How were the urinary N excreted values determined? Did I miss this somewhere or is this not explained?

We thank the Referee for pointing out that the information is indeed missing. We will add a Supplement Section « *S7 : Urinary nitrogen excretion during grazing* » to describe the protocole we used (with reference in Table 3).

S7 Urinary nitrogen excretion during grazing

Ammonia fluxes in grazing systems are strongly influenced by urinary N excretion, which varies in concentration, volume, and frequency based on N intake, diet composition, and water intake. Urinary N characteristics were determined by utilising an empirical calculator (UriNGraze; Edouard et al., 2024), based on a combination of tools and equations from previous work. Daily herbage intake was estimated using the HerbValo method (Delagarde et al., 2017, Wlodarski et al., 2024), considering pasture conditions, animal factors (body weight and milk production) and forage or concentrate supplementation. This enabled calculation of dietary N intake and diet crude protein (CP) content from both herbage and supplemental feeds (feed CP content = feed N content × 6.25). From the diet CP content, plasma urea N concentration was estimated using a regression derived from the CowNFlow database (Ferreira et al., 2021; Delagarde and Edouard, 2023):

$$Plasma\ Urea\ (mg\ N\ l^{-1}) = -19.7 + 0.2365 \times CP\ content\ in\ diet (g\ kg\ DM^{-1}) \quad (S8)$$

(Regression statistics: N = 114, $R^2 = 0.82$, sd = 3.81)

Plasma urea concentration was then used to determine urinary N excretion (g N cow^{-1} day^{-1}) using the following metamodel (based on the model presented in Rouillé et al., 2019; equation adapted from Edouard et al., unpublished, which combines linear and quadratic terms):

$$\begin{aligned} &\text{Urinary N excretion (g N cow}^{-1}\text{ day}^{-1}) = \ldots \\ &24 \times \{ 2.64 + \ldots \\ &\quad - 0.0517\ CONC + 0.00083\ CONC^2 \ldots \\ &\quad - 0.0552\ MPc + 0.00103\ MPc^2 \ldots \\ &\quad + 0.0196\ MY + 0.00063\ MY^2 \ldots \\ &\quad + 0.0609\ UREA + 0.000043\ UREA^2 \ldots \\ &\quad + 0.00021\ CONC.MPc + 0.000044\ CONC.MY - 0.00027\ CONC.UREA \ldots \\ &\quad - 0.00093\ MPc.MY - 0.00012\ MPc.UREA \ldots \\ &\quad - 0.00014\ MY.UREA \} \qquad (S9) \end{aligned}$$

where CONC is the concentrate proportion (%), MPc the milk protein content (g kg^{-1}), MY the milk yield (kg cow^{-1} day^{-1}), and UREA the plasma urea concentration (mg N l^{-1}).

The total N excreted per hectare during a grazing event was computed from daily urinary N excretion and total grazing days:

$$N_{event}(kg\ N\ ha^{-1}) = \frac{\text{Urinary N excretion (g N } LSU^{-1} day^{-1}) \times EGD(LSU\ ha^{-1} day)}{1000} \quad (S10)$$

where effective grazing days (EGD) were calculated as (see main article):

$$EGD = \text{Stocking Density (LSU ha}^{-1}) \times \text{Grazing Duration (days)} \quad (S11)$$

Line 646-654 Since you so nicely discuss this point it is good to add that another assumption is made. K for momentum is observed to behave differently than a K for heat or moisture. It is unknown whether it is allowed to assume that a K for ammonia behaves similarly as the K for heat.

See our earlier reply about the comment on similarity for momentum and heat. We have based our calculations on the classical assumptions by early authors (Thom, 1975, Monteith and Unsworth, 1990) that eddy diffusivity is the same for heat and trace gases, having no data available to the contrary.

L736-743: these lines repeat information of lines l705-712; I think L705-712 can be removed.

Actually, only lines 708-712 repeat the discussion line 736 and onwards. Their content will be deleted from line 708 onwards, and merged with the text starting line 736, as follows :

« …Nevertheless, several nighttime data points exhibited relatively high emission NH_3 gradients (sometimes up to tens of µg NH_3 m^{-3} difference between measurement heights, for example during G8), which, despite very low turbulence (u* < 0.05 m s-1), yield high emissions calculated from the flux-gradient equation. In such intermittent turbulence conditions, the uncertainty in the AGM flux is very large, even though they passed all criteria in the flux selection procedure (see Sect. 2.4.4), albeit with many data points assigned qcC (modest quality). Such data should not be over-interpreted, or even could be treated as outliers, but the data nonetheless suggest that under stable night conditions, NH_3 may accumulate near the surface during temperature inversions, trapping emissions closer to the ground due to reduced atmospheric mixing and limited vertical NH_3 transport… »

Make 'near the surface' more explicit; your measurement heights are already close to the surface so I guess you mean really close? A few centimeter?

This is a difficult question. « Near the surface » certainly means more than a few cm since we see a strong gradient over the depth of the profile mast, so probably more like the first few meters ? How far up can NH_3 diffuse upwards in very low turbulence ? Probably this is very variable and depends on the magnitude of the temperature inversion, on how very intermittent puffs might occasionally escape the surface. In such circumstances gradient-flux relationships break down completely, so we believe it is best not to provide any unfounded quantification.

L754 How was the estimated nitrogen input from grazing events determined?

This will be described in the new Supplement Section « *S7 : Urinary nitrogen excretion during grazing* », as described above in response to the Referee's question about urinary N excretion.

L823-825 Please rewrite, this sentence is difficult the grasp by first reading.

Agreed. The sentence was simplified as follows :

« …The data show that grazing in intensively managed grasslands can contribute larger cumulative seasonal fluxes than emissions from applied mineral or even organic fertilizers, depending on stocking density, vegetation characteristics, and meteorological conditions. »

Maybe I missed it but did you somewhere define the actual value of displacement height d? Based on figure 3 it appears to be ~ 10cm and time dependent?

We thank the referee for pointing this out. The displacement height was calculated as 2/3 * canopy height. We will add this sentence to Methods.

Technical corrections

L50: is 'crucial' the right term? Don't you mean 'significant'?

Agreed. We will rephrase as follows : « … Global NH_3 emissions from grazing are significant, though very uncertain, given the extensive grassland area, increasing livestock densities and excreta deposition on pastures. »

L93: intensively grazed grassland: LSU per ha mentioned in the table are at the lower-to middle end of the numbers presented. Is it really 'intensively' grazed?

The system is considered fairly intensive by French standards, and cannot by any standards be considered extensive. The grazing system is rotational, a frequently large stocking density (up to 55 LSU ha^{-1}) is applied for just a few days, whereby very little grass remains on the ground at the end of each grazing event, and the animals return after a month or so of re-growth. There is substantial addition of nitrogen as both cattle slurry and ammonium nitrate (altogether on average, around 150 kg N ha^{-1} added). It is true that larger stocking densities are reported in Table 4 at some of the NH3 flux measurement sites, for the duration of the grazing events investigated ; but it is not known whether this is also true for the annual average stocking density. One might wonder whether 120 or 150 LSU ha^{-1} reported at The Netherlands or New Zealand measurement sites represent usual practices.

But since the stocking density was sometimes less than 20 LSU ha^{-1} in our study, perhaps not all grazing events can be considered « intensive grazing » ; thus we propose to drop « intensive » from the paper's title.

L95: makes possible -> enables. OK

L174: for me the sentence is more clear when the first 'height' is removed. OK. We will rephrase as « … The sampling procedure followed a 200-s sequence from bottom to top… »

L198: if-profile: is this not repeating the argument? Could be removed? No, this is not repeating the argument, because in principle true stationarity really never happens ; however it only becomes an issue (for sequential sampling) when the concentration changes are significant, i.e. large.

L201: 'The concentration - averaging' can be removed without loss of information.

Yes, agreed, the sentence will be removed.

L214: Maybe it is good to already refer to later sections where you further discuss this choice.

Yes, agreed. We will refer here to Sections 3.5 and 4.1.3.

L452: It is also apparent that for some grazing events (e.g. …) the valid flux data capture was patchy, not necessarily – of interest (i.e. the wind was blowing from unsuitable directions). Please shorten to 'For some grazing events (e.g. …) the valid flux data capture was patchy due to the wind blowing from unsuitable directions.

Agreed. This section will been shortened and streamlined in response to earlier comments. The text has been modified as follows : « …*For comparison purposes, the flux timeries are also shown in Supplement Fig. S7 without footprint corrections, highlighting the importance of footprint in data interpretation. For some grazing events (e.g., G2, G7, G8) the valid flux data capture was patchy mostly because the wind was blowing from unsuitable directions; the flux data thus discarded due to footprint (shown as grey crosses in Fig. S7) may be fully representative of the adjacent plot (A or B, depending on which is the plot of interest), or a mixture of both plots, highlighting the difficulty of characterizing temporal flux patterns in rotational grazing with one single flux measurement setup located on the divide.* »

L630: 'with' missing in the second half? We are unsure what the Referee means here ? The sentence seems fine to us.

L675 'horizon' is maybe a correct (jargon) word? Or could it be replaced by layer?

Agreed. « Horizon » is inappropriate here. We will change to « layer ».

Citation: https://doi.org/10.5194/egusphere-2025-1605-RC1

3- Reply to RC2: 'Comment on egusphere-2025-1605', Anonymous Referee #2, 04 Jun 2025

Review of Abdulwahab et al. (2025): Aerodynamic gradient flux measurements of ammonia in intensively grazed grassland: temporal variations, environmental drivers, methodological challenges and uncertainties

Abdulwahab and colleagues present ammonia flux measurements in grazed grassland over four consecutive seasons. They use the aerodynamic gradient method with a custom-built sampling system. Data analysis is comprehensive and covers temporal flux variability under different environmental and management-driven controls, methodological challenges and its associated uncertainties as well as estimations of cumulative fluxes and emission factors. The latter include a well-justified application of a gap-filling approach making the study highly relevant for a wide range of readership.

Tackling ammonia exchange by direct flux measurements in the field under grazing management is challenging and every attempt using robust methodology is highly welcome. The authors did a great job in combining several well-established methods to be able to present a comprehensive assessment of grazing and fertilization-induced grassland NH_3 exchange.

We thank Referee #2 for the positive assessment of the manuscript, and for identifying the methodological aspects as key elements of the work.

I do not have any major concerns or major criticism, but a number of (rather minor) comments and suggestions for modifications, which should be addressed before publication. Given the challenges in ammonia flux measurements and data analysis, in my opinion, this work forms a milestone in methodology and verification of NH_3 emission factors for grazed grassland.

Line 126: Definition of LSU is needed. General comment: Could you briefly mention whether a correlation between F_{NH3} and LSU was found and elaborate a bit on herd movement, presence time in the footprint, etc.?

The definition of LSU will be added to the text at this point. There was no correlation of the half-hourly NH3 emission flux to the instantaneous grazing livestock density within each grazing event, since the stocking density was generally constant within each event (apart from G9). No decrease in flux was observed at milking times, when the herd was taken away from the field for a couple of hours. This is because the NH3 emission flux is driven by soil processes following urine deposition, not by standing or ruminating cattle themselves. The cows move around, but the urine patches stay in place. Thus there is no correlation to be sought between cattle movement in and out of footprint and the measured flux, as would be the case for CH4 emission directly from the animals. What would be needed is the detection and geo-referencing of all urination events in real time by each member of the herd across the field, to map out the spatial distribution of urine deposition, in order to analyse the correlation between measured flux and footprint weighted by urine deposition. Our phenocams took pictures of the field every 5 minutes (see example below, 16/04/2023 13 :20), and the images could be analyzed using shape recognition algorithms and coordinate triangulation within the field polygon, but still this could not tell us where urination events took place.

Line 126f: No fertilization at all in 2021, but two in 2022, is that correct?

There were applications of slurry and mineral fertilizer (ammonium nitrate) in 2021, but either earlier or later in the year, not at the time of the flux measurements. There was one application of ammonium nitrate, and one of cattle slurry, during the 2022 spring campaign (see Fig. S4), but also other applications earlier and later in the year. Not all N fertilizations events were mentioned in the paper if they took place a long time before or after the measurement campaigns.

Line 133: What does 'hybrid' mean in this context?

The full aerodynamic gradient method uses three types of vertical profiles for flux determination : vertical wind profile to determine u_*, vertical temperature profile to determine sensible heat flux H (together with u_*), and vertical gas (NH3) concentration profile. The method we used was hybrid in the sense that two methods were used : eddy covariance for u_* (from w'u') and for H (from w'T'), and gradient for NH3 (u* x dC/dz).

Text will be modified as follows : « … *By contrast to the original full AGM described in earlier studies (Fowler and Duyzer, 1989), which uses profiles of wind speed and temperature, as well as gas concentration, a 3-D ultrasonic anemometer was used here to measure friction velocity (u_*) and sensible heat flux (H) by eddy covariance.* »

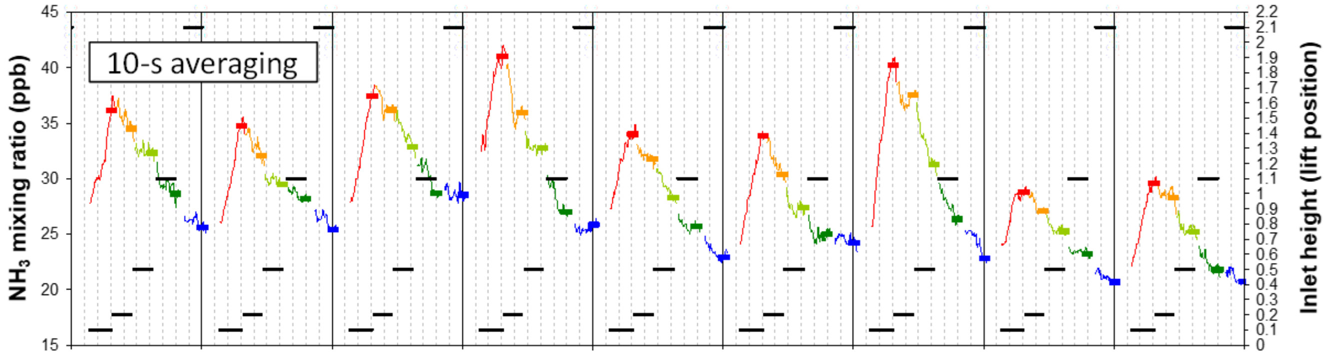
Line 135: Here and throughout the manuscript, also in some figures – the star in u_star is in the subscript, not in the superscript.

All occurrences of u* will be substituted by u_*.

Line 170: Refer to Section 2.3.3 after 'NH_3 analyser', otherwise people would expect more information here at this place. Agreed, this will be done.

Line 174f: I think the gradient lift is a well 'thought through' system for ammonia, particularly with regard to avoid several sampling lines. But what about highly dynamic situations? Did the concentrations not change significantly during the 200-s cycles (or those periods where data was finally taken from)? Can the authors show some raw data time series when the system was lifted to one level, stayed in a position and then lifted to the next level to illustrate/visualize the data selection and handling a bit better?

We have addressed a similar question in reply to Referee #1 (see earlier comments), and added a figure in the supplement (Fig. S1), which shows a half-hour of 1-sec data and illustrates how data are filtered out when the inlet is moving or still equilibrating at the new height, and how averages calculated for the last 10 seconds of each height. If we take a look at the third panel from top of Fig. S1 (see below), large fluctuations (factor of 2 in this example) of selected/averaged NH_3 concentrations can be seen between the nine consecutive lift ascent cycles, at all heights sampled. We have also added a section to the supplement (S2) to document explicitly how the detrending procedure is applied.

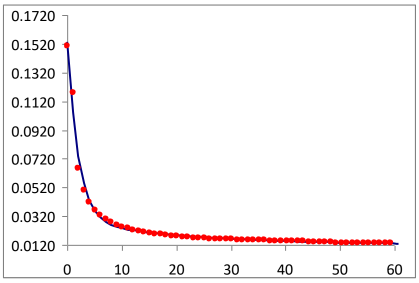

10-s averaging
NH3 mixing ratio (ppb)
Inlet height (lift position)
45
40
35
30
25
20
15
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Line 191: The nominal precision of the analyzer of 0.2 ppb at 1-s integration time: Is this number from the manufacturer's specifications or based on own tests? See also former and next comment to include some raw data examples for better illustration.

We have modified the text as follows (also in reply to Referee #1) : « ..The nominal precision (1-std. dev.) of the QCL analyser, provided by the manufacturer, was 0.2 ppb at a 1-s integration time, and 0.08 ppb at 10-s integration time… »

Lines 199-201: Have the authors done tests on step concentration changes?

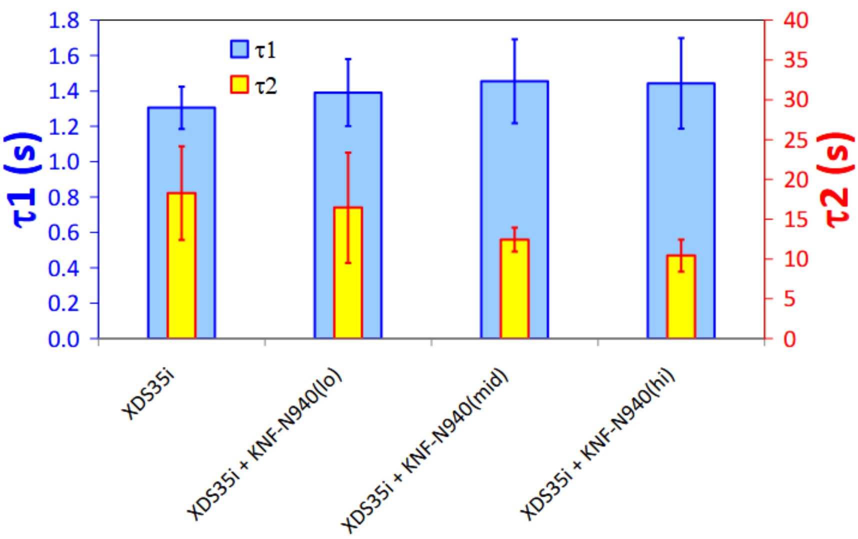
Yes, we have carried out a number of tests with our setup in the field, by which we placed a strong NH3 source (an open liquid ammonia bottle, or a dirty sweaty hand) just below the URG cyclone inlet, and removed the contamination source instantly from the inlet. We then analyzed the exponential decay of the NH3 signal back down to the ambient level :



Example time series showing the exponential decay versus time (sec) of the NH_3 signal (ppm), after removing (at t=0) an NH_3 contamination source from the inlet of the Los Gatos FTAA-QCL sampling system.

Each exponential time series was fitted with a double exponential function following (see similar earlier work by e.g. Ellis et al. 2010 (Characterizing a Quantum Cascade Tunable Infrared Laser Differential Absorption Spectrometer (QC-TILDAS) for measurements of atmospheric ammonia, Atmos.Meas. Tech., 3, 397–406, doi:10.5194/amt-3-397-2010, 2010), or Whitehead, et al. 2008 (Evaluation of Laser Absorption Spectroscopic Techniques for Eddy Covariance Flux Measurements of NH3, Environ. Sci. Technol., 42, 2041–2046, 2008).

The results of the tests we carried out using a range of configurations are shown in the figure below, with mean τ1 values of 1.4 sec and mean τ2 values of 10 sec for the setup used in this study. The τ1 values for our setup were roughly a factor of 3-4 longer than values found by Ellis et al. (2010) for an Aerodyne QCL (~0.4 sec), showing that our sampling setup using the LGR FTAA analyzer was not fast enough for eddy covariance, but adequate for profile sampling given enough stabilization time between heights.



Double exponential decay parameters τ1 and τ2 for a range of step concentration tests carried out in the field, with different sampling (pump) configurations.

We will add the following sentence in Section 2.3.4, referring to a new Supplement Section S3 which will be added (see below), summarizing these results : « ... *Step concentration change tests carried out both in the field and in the lab, fitted with a typical double exponential function versus time, indicated mean characteristic fast and slow time constants of 1.4 and 10 seconds, respectively (See Supplement section S3), showing that the stabilization times were long enough.* »

S3 Response time constants of the QCL sampling system

The combined inlet-spectrometer-pump sampling system has a relatively rapid response time, adequate for profile sampling (though not fast enough for eddy covariance), which reflects intrinsic analyser response time, internal volume turnover and tube/cell wall damping effects. Characteristic response times τ1 and τ2 for the measurement system were obtained by fitting double exponential decay curves to concentration time series following a step change in NH_3 concentration (see e.g. Ellis et al., 2010). The fitted fast time constant τ1 was 1.4sec, corresponding to the sample gas turnover (pumping) rate in the internal volume. The slow time constant τ2 was 10 sec, which is function of the damping/smoothing effect of internal surfaces (tube and cell walls, filter surfaces). See Bell (PhD thesis, 2017) for details.

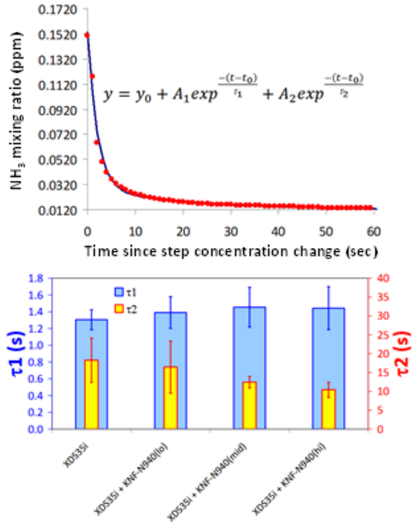
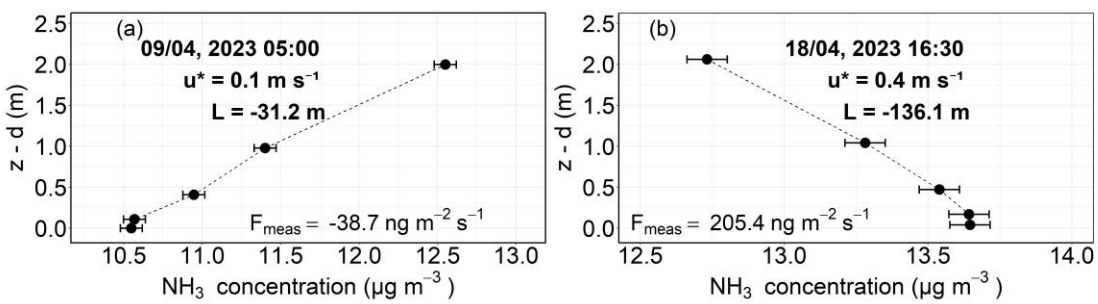


Fig. S17. Characterization of the fast (τ1) and slow (τ2) time constants of the LGR-FTAA-based NH_3 sampling system, obtained by fitting double exponential decay curves to concentration time series following a step change in NH_3 concentration (example time series shown in the top panel). The lower panel summarizes a series of decay experiments made using i) either the main Edwards XDS35i vacuum pump only, or ii) both Edwards XDS35i and the auxillary KNF-N940 pump, with the latter operating at low, intermediate and high flow rates. The auxillary KNF pump helps increasing the flow rate through the inlet line, from the aerosol cyclone to a T-piece at the back of the analyzer, thereby minimizing adsorption/desorption of NH_3 on tubing walls and thus decreasing the slow τ2 time constant to around 10 seconds.

For AGM the differences between heights (gradient) matter(s), but precision of the analyzer has an impact on uncertainty.

We concur. In reply to a comment by Referee #1 (see earlier in this document) we have added the nominal precision of the LGR FTAA analyzer (~0.1 ppb or 0.07 µg m^{-3} at 10-sec averaging time) as error bars to Fig. 3a and 3b, which we believe demonstrates that the measurement setup can detect typical emission and even deposition gradients :



Top two panels from the updated Figure 3, showing the nominal NH3 analyzer precision as horizontal error bars, to compare with the size of the vertical gradient to be measured.

Line 201: '…using half-hourly averaging': Can the authors clarify how exactly the 'half-hourly averaging' is meant? For each height? Despite a lot of information, it is unclear how this is handled.

We have clarified this point in reply to Referee #1, by adding a new Supplement section S2 ; the averaging is dealt with in Eq. S4, whereby the mean concentrations of the nine cycles for each half-hour are calculated in Eq. S3 :

S2 QCL NH_3 concentration gradient detrending procedure

To account for potential biases in vertical gradients arising from sequential sampling (non-simultaneous concentration measurements at the different heights) and non-stationary conditions, the QCL NH3 concentration gradient data were detrended, and the final half-hourly concentrations calculated, using the following procedure.

1- The 10-second average concentrations NH3(z, i) are calculated for each of the N sampling heights (z), for each of the nine 200-sec lift ascent cycles (i=1..9) of each half-hour (as shown in Fig. S1, see above)

2- For all nine cycles of each half-hour, at each height (z) of the N-point gradient, a trend (slope) is calculated between cycle (i) and cycle (i+1) as follows :
$$\text{slope}(z, i, i+1) = [\text{NH3}(z, i+1) - \text{NH3}(z, i)] / \text{CycleDuration} \quad \text{(S1)}$$
with CycleDuration = 200 sec (slope unit : ppb s-1).

3- A mean slope across all N sampling heights is calculated as follows :
$$\text{slope\_avg}(i, i+1) = [\text{slope}(z1, i, i+1) + \text{slope}(z2, i, i+1) + \ldots + \text{slope}(zN, i, i+1)] / N \quad \text{(S2)}$$

4- The middle height (z=3 in a 5-point gradient, or z=2 in a 3-point gradient) is taken as the reference (zRef) and is not corrected. The other heights (z=1,2,4,5 in a 5-point gradient ; or z=1,3 in a 3-point gradient) are corrected as follows:
$$\text{NH3\_corr}(z, i) = \text{NH3}(z, i) + \text{slope\_avg}(i, i+1) * \Delta t \quad \text{(S3)}$$
with Δt being the time interval (in seconds) between the end of height z (when the 10-sec concentration is calculated) and the end of the reference (middle) height (zRef). Δt is positive (*negative*) if height z is sampled before (*after*) zRef. Thus, for example, if there is an increasing trend over the interval of cycle (i) to (i+1), the heights sampled first (before zRef) are corrected upwards, and the heights sampled last (after zRef) are corrected downwards.

5- The half-hourly concentrations are calculated at each height as the average of the nine corrected concentrations within the half-hour:
$$\text{NH3\_30min}(z) = [\text{NH3\_corr}(z, 1) + \text{NH3\_corr}(z, 2) + \ldots + \text{NH3\_corr}(z, 9)] / 9 \quad \text{(S4)}$$

New Supplement S2 on NH_3 data selection and treatment from the raw 1-sec data to half-hourly averages.

Line 207: Add ', Panel (c)' after 'Fig. 3'. OK

Line 215f: The whole Section 2.3.5 is a chance of reducing the length of the main part of the paper as the DELTA method is described in detail elsewhere. Just mention the exposure times in the previous section and move the rest to the supplement.

Agreed. We have removed most of the DELTA method description to a new Supplement section S1 (see also our reply to Referee #1).

Lines 283-287: I'm confused now, mainly by the phrase 'to assess the stationarity of NH_3 concentrations over 30-minute intervals, …'. How is this exactly handled? Stationarity for each of the heights within a 30-minute window individually, i.e., just taking the respective chosen part of the 200-s interval? Otherwise, it is not clear how this is in accordance with earlier method description of the sampling scheme. Please consider rephrasing.

We believe these technical details will now be clearer with the new Supplement Section S2 (mentioned earlier), together with the new Supplement Figure S1 (also mentioned earlier). Once the nine mean 10-sec NH3 concentrations are calculated for the nine lift ascent cycles, for all heights (as described in S2), the mean half-hourly NH3 concentrations are calculated from the nine NH3 values, together with the standard deviations, and thus the CV%. We believe the slightly modified phrase describes this adequately (the text will also have been modified further up in M&Ms, as well as in the Supplement):

« …, based on the coefficient of variation (CV) of the nine NH_3 concentrations (averages of selected 10-second data) measured over the nine vertical lift ascents per half-hour… »

Further, did you also check for stationarity of consecutive half hours?

This is what we call the (change in) storage term, described in Eq. 7. We have already replied to a question by Referee #1 that the storage correction is indeed calculated over a time scale of 30 minutes, from the ground up to the mean measurement height. We will slightly rephrase the text to clarify this :

« … Nevertheless, the error in the flux due to storage (F_{sto}) was systematically calculated from the following equation, based on concentration changes from one half-hour to the next: »

$$F_{sto} = \int_0^{z_{mean}} \frac{\partial \chi}{\partial t} dz \quad (7)$$

Line 306f: Great approach for the correction of background flux interference!

We are glad the approach is received favourably. We believe such a correction is in principle necessary. It may be argued that the use of a (bi-directional exchange) model may add some uncertainty to measured fluxes, but the flux data measured in situ at this site during background conditions were used to parameterize/calibrate the model. The flux patterns observed during background (Supplement Fig. S14 in the original document) were convincing enough that the bi-directionality is real and that some small emissions can occur from plant or soil internal nitrogen cycling, outside of grazing or fertilisation events, not driven by TAN volatilization from hydrolyzed urea.

Line 328: Check bracket imbalance at '(SE(F_{meas})'. Yes, this will be fixed.

Line 368: Is the reference '(see Sect. S2 in the supplement)' correct? Shouldn't it be Figures S10-S12? Please check.

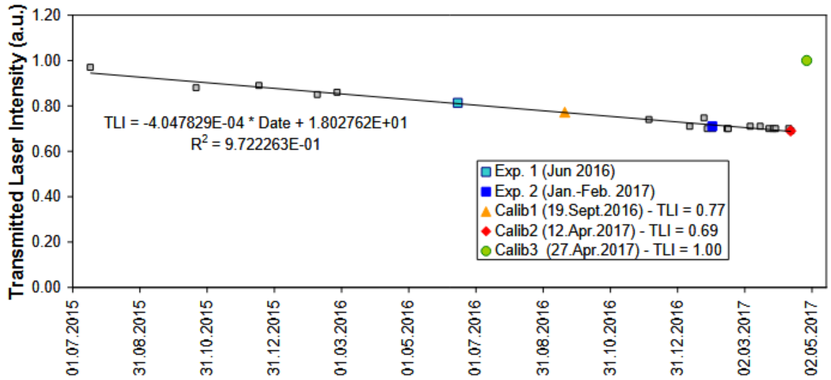
The original Section S2 referred to the special case for G9-G10, in case of large data gaps ; while Figs. S10-S12 were generic for all grazing events, so the reference was correct. Note that, to save space from the main article, the whole gap-filling section will be moved to a new Supplement section S6, covering all cases (see our reply to Referee #1).

Further, did the authors try to create artificial gaps in periods with high peaks to investigate whether or not the gap-filling method works?

No, we did not try to create further data gaps to test the gap-filling procedure, as the rigorous data selection already lead to a large loss (~50%) of flux data, for a large part due to footprint. This would be an interesting exercise to try at a different site where the flux tower has a larger data capture thanks to a near-360° « field of view », i.e. where the field is not divided into two paddocks, creating footprint issues such as encountered here.

Lines 375-376: 'The variability in regression slopes resulted from different degrees of cleanliness and reflectivity of mirrors…': How do you know that?

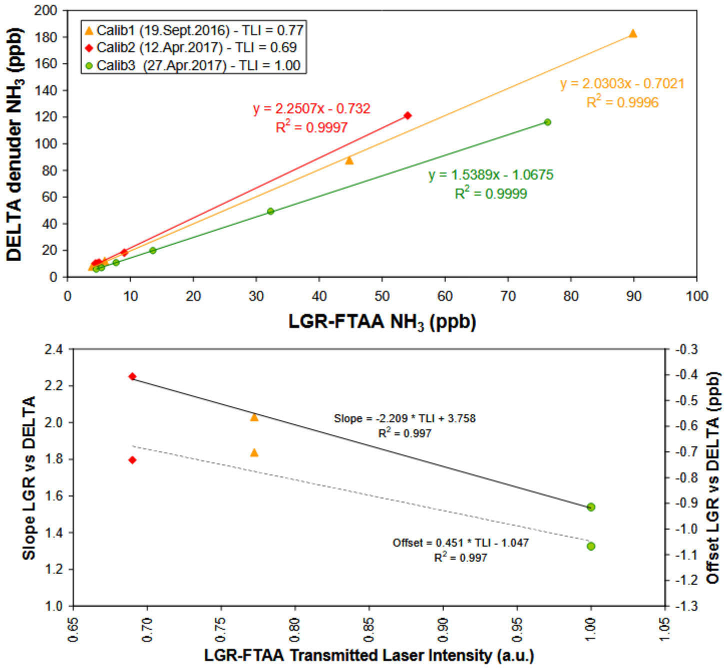
This is based on observations over 10 years of use of the QCL, related primarily to the observed gradual decrease in transmitted laser intensity (TLI) over several weeks and months of continuous high flow sampling. Despite the URG cyclone to remove aerosols larger than 1µm, and the unavoidable use of an internal 1.2µm teflon filter in the analyzer, the TLI inevitably decreases over time due to fine aerosol (<1 µm size) ingress, which are not filtered out by the cyclone and the teflon filter. Over 10 years the mirrors have been cleaned many times at the INRAE lab (LGR provided instructions and an acetone/methanol-based cleaning kit for the user) ; mirror cleaning restores the full TLI, and a new measurement cycle can begin with maximum signal strength.



Gradual decrease of transmitted laser intensity (arbitrary units) in the LGR QCL-FTAA over a period of 1.8 years, after the acquisition of the instrument in July 2015. The first-ever mirror cleaning took place at the end of April 2017, restoring TLI to the full signal potential (green circle). In subsequent years, mirror cleaning took place much more frequently, typically after 2 months of continuous use, or when TLI decreased below 0.75-0.8. Details available in: Bell, M. (2017). Emission, dispersion and deposition of ammonia from the plot to the landscape scale. PhD thesis, Agrocampus Ouest, Université Européenne de Bretagne, Rennes, France, 285 pp.

Over the 10 years of use of the analyzer we have also, prior to – and after – many mirror cleaning sessions, conducted lab-based experiments, in which we ran the instrument alongside a DELTA sampler for a few days, creating different levels of NH3 exposure in the room (typically between 5 and 100 ppb) by partially opening a liquid ammonia bottle (to « spike » ambient room air for a few hours), or by closing the NH3 bottle and opening windows and door (to lower the ambient level as much as possible). The idea was to generate different concentration levels in the room, loosely controlled but not necessarily stable, in order to plot a regression across a reasonable range between mean LGR measurements and DELTA denuders (similar to Fig. 2 of the paper). Thus a new set of dry denuders was put in place for every new concentration level. Every such experiment was made with a slightly different reduction in TLI (prior to mirror cleaning), compared with the full signal that resulted later from cleaning, depending on the recent sampling (and internal aerosol contamination) history of the analyzer.

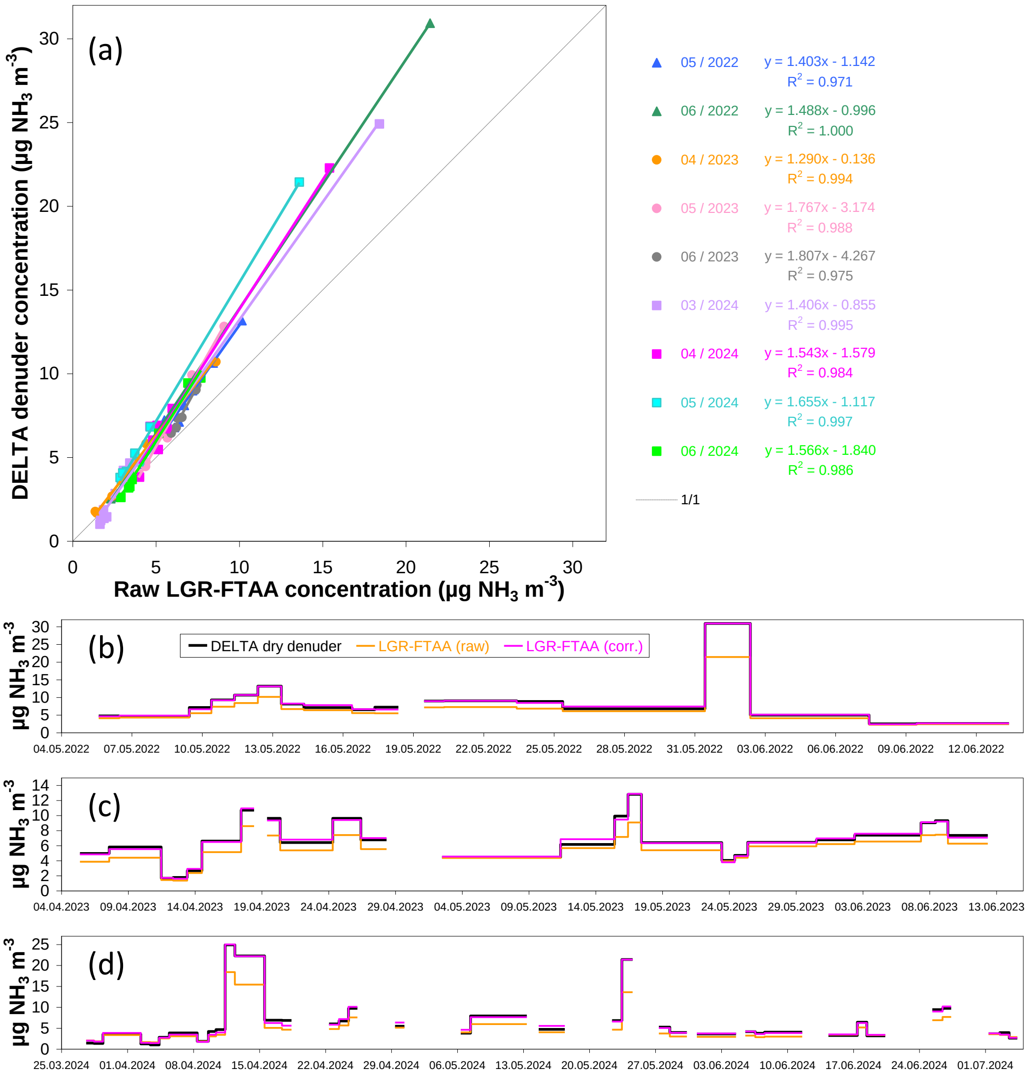
Thus we were able to compare the regression slopes of the DELTA vs LGR FTAA data for different levels of TLI reduction (say 70%, 75% , 80%, 85 % of full « clean » TLI) before cleaning, and for full potential (100%) TLI after cleaning, as shown in the example figure below. The data show that the regression slope increases with decreasing TLI, i.e. the magnitude of the correction required for LGR FTAA data is smaller for clean mirrors and larger for dirty mirrors.



Comparison of mean NH_3 concentrations co-sampled in ambient laboratory air by LGR QCL-FTAA and DELTA denuders (top) and regression slope and offset calculated for three levels of TLI (bottom). (from Bell, M. (2017). Emission, dispersion and deposition of ammonia from the plot to the landscape scale. PhD thesis, Agrocampus Ouest, Université Européenne de Bretagne, Rennes, France, 285 pp).

Figure 2: This figure is hard to grasp. The main message might be clear, but details are tough to extract. In general, x-y scatter plots with identical values on the axes and a 1:1 line should always come in a quadratic format so that an over- or underestimation of one variable against the other can be easily captured by the reader. Further, use color, a more readable legend and bigger symbols in Panel (a). I suggest replotting of Panel (b). The x-axis label is not readable. Try monthly boxplots or just plotting the differences against the DELTA numbers. I think that would make it visually more appealing without losing any main messages.

Agreed, the original figure was not optimal. Following a similar comment by Referee #1 (see above), we have completely re-arranged data display of Figure 2 (including a square shape for the top panel and colors), as follows :



Updated Figure 2.

Line 408: Can the authors say something about how they differentiate plumes from housings vs. plumes from grazing animals?

This is of course very difficult to do in the real world in a complex agricultural landscape with local sources. Figure 1 shows local farm buildings in two wind sectors : main farm to the SE, and a small old farm annex to the NW. The main farm to the SE was by far the larger NH_3 source, with the bulk of the animals and manure storage areas located there (350 m from flux tower to the main dairy barn, 430 m to the manure storage areas). However, the windrose in Fig. 1 shows that the frequency of SE winds was very small over the four spring campaigns.

The annex to the NW contained a small number of young animals (calves and heifers), thus a very small potential source, and only until 2023 and only in the building highlighted in red in the photo below (230 m from the flux tower) ; the rest of the buildings were either residential or disused.

We did not carry out any specific differentiation between plumes from field grazing and from farm buildings, considering that the impact of the farm buildings was small overall due to wind distribution (the main wind sectors were not strongly aligned with these two local sources), but also because it would be difficult to devise a reliable method to do so. Note that plumes advected from local (or other) sources will shift the vertical gradient at the flux tower towards deposition; while plumes from on-site grazing result in an emission gradient at the flux tower.

There were times, during calm, windless, stable nights, when large NH3 concentrations were observed, which could not be attributed to grazing animals (or fertilization) on the study field (because no animals were there, nor had been recently, nor any fertilizer application). For example (and this is a question raised by Referee #1), in Fig.3c, large concentrations are observed during the night of 08-09/Apr/2023. The plume does not originate on the study field, as there have not been any cows here since 28/Feb/2023. It may originate from nearby grazed fields (if any), or from the local farm buildings shown in Fig. 1, or even possibly from other farms further away (though the latter is less likely, considering distances). The point is, during that night and other typical nights at the site, wind speed was very low (0.5 m s^{-1}) and turbulence suppressed (u_* 0.05 m s^{-1}), such that wind and plumes hovers here and there randomly, wind direction is not reliable, and neither would dispersion modelling be, if spatial plume attribution was attempted. Not surprisingly, during the night of 08-09/Apr/2023, many half-hourly fluxes are discarded with a qc_µmet of 2. But the few datapoints that remain indicate deposition to the surface, which makes sense, as large NH3 concentrations advected from elsewhere and passing over the site will tend to produce a deposition gradient.

We will add the following sentence to the text in section 3.2: « … *Part of the deposited NH3 very likely originated from local farm sources, but we did not try to quantify the magnitude of the effect.* »

Figure 3: Mention in the caption what qcA, etc. means (or refer to description in text). OK, this will be done.

What is the reason for the high concentrations around 9th of April?

See our reply above in relation to plume differentiation. Without well developed turbulence, significant wind speed and reliable wind direction to drive a dispersion model, this is a very difficult question to answer. We can speculate that the plume came from the nearest farm, to the SE, as the recorded wind direction was in this sector some of the time, but the correlation between wind direction and NH3 concentration during that night was weak.

Line 431: Provide uncertainty ranges after flux numbers. On line 431 we provide the range, i.e. min and max of the whole dataset of calculated fluxes, not mean values, so confidence intervals are not relevant here ? If the referee means uncertainty for individual half-hourly fluxes (in this case the min and max of the dataset), this is something we are unable to provide if all sources of uncertainty must be accounted for. We have calculated and reported random errors (see section 2.5), but do not have any ready numbers for systematic errors, which are likely much larger but impossible to estimate with any confidence (see discussion in 3.5 and 4.1)

Table 2: I don't understand how to read the line 'LSU ha-1 (DG)'. Livestock unit per hectare is clear, but is the grazing duration in brackets given in days? If so, I can't really recognize the number for example in Figure 3, Panel (d) for the April 2023 period. Please check.

Error well spotted, thank you, and apologies. Indeed somehow the two numbers 5.6 and 4.3 were swapped between the two columns. Below is the corrected table. Now the calculation EGD = LSU * DG checks out for both 237 ~ 55*4.3 and 127 ~ 23*5.6. We have thoroughly checked the rest of the table.

Period	G1	G2	G3	G4	G
Date	09/04 – 19/04/22	15/04 – 19/04/22	20/05 – 31/05/22	12/04 – 28/04/23	1[illegible] 2[illegible]
plot	A	B	A	A	B
LSU ha^{-1} (DG)	45(2.7)	44 (2.1)	55 (4.3)	23 (5.6)	2[illegible]
EGD	122	93	237	127	1[illegible]

Line 602: Is there a part of the sentence missing?

Yes, apologies ; the last sentence of 4.1.1 on l601-602 (« Some systematic – relationships ») appears to be a left-over undeleted fragment from an earlier version. This will be removed.

Line 694f: Is there any information on how the farmer decides when and how much fertilizer is added after grazing? Well, slurry is probably due to seasonal storage issues, but why is so much mineral N added as well?

Slurry is often (though not systematically) spread at the end of winter, to empty storage as suggested by the Referee. Mineral nitrogen is then sometimes (not systematically) added after the first or second spring grazing phase to boost re-growth. All in all, the mean inter-annual N dose is large (~150 kg N ha^{-1} yr^{-1} over 2020-2024), but not that massive.

Line 709: Here another u-star appearance, see earlier comment. OK, this will be fixed.

Line 718: I don't get the logic here. Is the number of paddocks increased? If not, how can the N be more uniformly distributed? And how can NH3 emissions be mitigated under higher grazing densities?

Agreed. This sentence is ambiguous and confusing ; we will delete the whole sentence about mitigation options, since this is really not the focus of the paper.

Citation: https://doi.org/10.5194/egusphere-2025-1605-RC2

4- Reply to **CC1**: 'Comment on egusphere-2025-1605', Johannes Laubach, 04 Jun 2025

This is an interesting study, with a well-constructed measurement system and having collected a valuable dataset of NH3 emissions from a real-world grazing system. I would just like to make a couple of suggestions, because I feel the full potential of the dataset has not been sufficiently explored yet.

We thank J. Laubach for his interest in our study. We definitely agree that the full potential of the dataset has not been explored yet, because, as we explained at the beginning of this reply, two papers are intended. This first paper is measurement-methodology-uncertainty oriented ; the second paper (to be sumitted towards the end of 2025) will be more process-oriented with the use, application and adaptation of the Generation of Ammonia from Grazing (GAG) model by Moring et al. (2017).

1) In my view, the authors are too focused on individual grazing events. Because conditions during these are so variable, NH3 emissions vary enormously. It is a valid point to show this, but not a very novel one, and discussion of all the factors that might explain this, without being able to provide any insights, could be shorter.

We will reduce the discussion on flux variability to save a little space ; nonetheless we feel it is one of the strengths of our paper to present flux measurements over a long time period, covering a large range of meteorological conditions, at one single site.

2) Neither does it make much sense to me to then derive emission factors for individual grazings. They are short-term, highly variable, and defining the end of the emission response period to the fresh N input is rather arbitrary (N not volatilised quickly may remain available and could contribute to later bouts of emissions, be they triggered by weather changes or by fresh additional inputs).

We agree that it makes sense to provide an overall emission factor based on the sum of all fluxes over all grazing events, divided by the total number of effective grazing days. This is indeed likely a more robust mean EF, than the arithmetic mean of individual event-based values. Therefore we will add a line at the bottom of the revised Table 3 :

Table 3: Cumulative NH_3 emissions (g N ha^{-1}) and emission factors (per livestock unit and fraction of the excreted N) for eight grazing events. Cumulative fluxes were calculated using the DVmax and DVavg gap-filling methods based on mean diurnal variation (see Methods). Grazing events G2 and G7 were excluded from cumulative and EF estimation due to low data availability. The urinary N excretion was estimated following Supplement Section S7. $^{\Sigma}$: sum over grazing events; ^{A}: arithmetic mean over grazing events; $^{\#}$: calculated as the ratio of sum total emission to sum total EGD; §: calculated as the ratio of sum total emissions to sum total urinary N excretion

Period	EGD	Cumulative fluxes DVmax	Cumulative fluxes DVavg	Urinary N excreted Total herd	Urinary N excreted Per cow	EF per LSU DVmax	EF per LSU DVavg	EF relative to urinary N DVmax	EF relative to urinary N DVavg
		(g N ha^{-1})		(Kg N ha^{-1})	(g N LSU^{-1})	(g NH_3-N cow^{-1} grazing d^{-1})		(%)	
G1	122	1149	1637	13.7	112	9.4	13.4	8.4	11.9
G3	237	699	702	18.2	77	2.9	3.0	3.8	3.9
G4	127	174	179	16.8	132	1.4	1.4	1.0	1.1
G5	138	150	160	14.5	105	1.1	1.2	1.0	1.1
G6	150	411	438	15.8	105	2.7	2.9	2.6	2.8
G8	72	1482	1674	9.8	136	20.6	23.3	15.1	17.1
G9	118	2190	2322	17.2	146	18.5	19.6	12.7	13.5
G10	94	105	206	6.9	73	1.1	2.2	1.5	3.0
Overall	1058$^{\Sigma}$	6360$^{\Sigma}$	7318$^{\Sigma}$	113$^{\Sigma}$	111^{A}	6.0$^{\#}$	6.9$^{\#}$	5.6§	6.5§

The newly obtained mean « overall integrated » EF derived by this method (6.0-6.9 on the last line of Table 3) will be added to table 4 for the comparison to other studies.

The following sentence will be added to section 3.6 : « … *A robust mean overall EF of 6.0-6.9 g NH_3-N cow^{-1} grazing d^{-1} is calculated as the ratio of the sum of all emissions to the sum of all EGD cumulated over the available eight grazing events (see last line of Table 3).* »

3) Instead of defining emission factors per grazing event and per cow, my suggestion is to determine total NH3 emissions and total N inputs for each measurement season, and define emission factors as the ratio of the former to the latter. (Whole years would be even better, maybe that's an idea for the future.) That should make the numbers more easily comparable with those from elsewhere, and reduce the EF's uncertainty, too. A season's sequence of grazing events would represent the combination of management practice and climatic conditions better than individual grazing events. If you find the datasets too patchy for that, maybe consider making the quality control a little less rigid?

Since we have only 8 grazing events at our disposal, and thus not many per year/measurement season, we propose a mean overall EF calculated over the whole dataset G1-G10 (see above, modified Table 3).

And sadly, we only have flux data for spring (March-June) for 2021-2024; the work was fairly labour intensive and could not be carried out for four full years continuously, including summer and autumn grazing.

4) The flux-gradient method requires to approximate infinitesimal concentration gradients, d_C/d_z, with finite (measured) differences Delta_C/Delta_z, and to resolve Delta_C accurately, Delta_z cannot be too small. As concentration profiles near the ground tend to be curved, this approximation is likely to introduce bias. Some of the systematic differences between height pairs that you find may be due to this. A way to reduce this would be to only compare pairs of adjacent heights with each other, avoiding the larger Delta_z. Or, a possibly much neater way to use all heights would be to fit the profile shape for each averaging cycle and then determine its slope at the height where u* is measured (to match heights used for gradient and diffusivity).

We believe that the reason concentration profiles near the ground tend to be curved is twofold, as explained to the two Referees above, namely i) heights 0.1-0.2m are located in the roughness sublayer, in which classical K-theory gradient-flux relationships do not hold, and therefore the profile is not logarithmic, and ii) the effect of the ungrazed/unfertilized patch within the instrument enclosure, which strongly affects the footprint of the lower two heights. We would rather not use these two heights to fit the whole profile as suggested above, in particular since the lower two heights carry footprint information about the micro-ecosystem that is the instrument enclosure, which is not relevant for the grazed field of interest. Having removed the lower 2 heights, the study about differences between pairs 0.5-2.1 and 1.1-2.1 is already shown in Fig. 8, with a reassuring near-one slope of 1.07 in the range 0-2000 ng m^{-2} s^{-1}.

5) Please clarify if/how the differences between height pairs have been considered in the uncertainty estimates.

We believe this point is already adequately addressed in Fig.8 and the ensuing discussion.

6) I recommend considerable shortening of the sections on recalibration of C (slopes vs comparison instrument) and on quality control in general. State briefly that these were done and provide the details as supplementary material. Most readers would like to get to the interesting results quickly!

We have removed significant portions of the M&Ms section to the Supplement, including the section about the DELTA sampling system.

7) Title: consider a) swapping the words "gradient" and "flux" (many know the method as "flux-gradient"), b) dropping the words following the colon, c) including "over 3 spring seasons" (as a point of difference to other studies).

Including « over 4 spring seasons » is a good suggestion, but we wish to retain the methodological aspects in the title, which are central to the paper. We propose the following compromise, dropping « *intensive* » following RC1's comment and dropping « *temporal variations* » as the temporal dimension is now covered in « *four spring seasons* » :

« *Aerodynamic flux-gradient measurements of ammonia over four spring seasons in grazed grassland: environmental drivers, methodological challenges and uncertainties*»

8) The abstract states "2021 - 2024" but no data from 2021 are presented, unless I missed something.

The first flux measurements were made in spring 2021, as shown in Supplement Fig. S1, but without capturing a cattle grazing phase (and therefore no « G_X » grazing phase mentioned in Tables 2-3 for 2021). But the 2021 data were useful to characterize bi-directional background exchange in the absence of grazing or fertilization (see Supplement Fig. S14), needed to parameterize the Bi-Di flux model used for footprint flux correction.

9) Are you planning to use this dataset for the testing of process-based models?

Indeed yes, as we have mentioned a few times already in the course of this reply. We hope to submit our manuscript to BG towards the end of 2025.

Best regards

Johannes Laubach.

laubachj@landcareresearch.co.nz

Citation: https://doi.org/10.5194/egusphere-2025-1605-CC1