Response to Anonymous Referee #2 on "In-situ measurements of NH3: instrument performance and applicability" by Marsailidh M. Twigg et al., Atmos. Meas. Tech. Discuss.,https://doi.org/10.5194/amt-2022-107-RC2, 2022

The authors thank the referee for the careful and constructive consideration of this manuscript. The answer is structured as follows: the comments from referee #2 are marked in black and the authors' response and changes the to the manuscript are written in blue.

GENERAL COMMENTS

The manuscript presented by Twigg et al. represents an important study towards improving the quality of in-situ ammonia (NH_3) measurements. Significant challenges still exist in obtaining accurate and precise NH_3 measurements, specifically in field observations where a large range of ambient conditions impact the performance of instrumentation. So far, there have been only a limited number of studies comparing different NH_3 measurement methodologies under field conditions. The authors present a comparison of 13 online instruments, spanning over the most important measurement techniques, and 1 type of passive samplers, making it to my knowledge the broadest NH_3 intercomparison study conducted up to date. The authors use different statistical methods to compare the instruments determining their precision and accuracy. The comparison of closed-path systems with the open-path miniDOAS as well as employing some instruments of the same type with different inlet systems, allows the authors to investigate dampening effects caused by inlet tubing surfaces. Finally, the authors also give advice on what future measurement setup should consider such as regular calibrations (which were missing in this study). The setup of the study, analysis of data and presentation of results was performed with great care, considering the major challenges of each instrument. Given the importance and uniqueness of the study, I suggest it to be published in AMT after addressing the comments below. The major parts where the manuscript in my view needs improvement is (1) the determination of the residence time and (2) the description of inlet tubing effects and the derivation of time constants.

We thank the reviewer for taking the time to review this study and providing the constructive feedback of the manuscript.

SPECIFIC COMMENTS

P. 1, L. 35: The CEN EN 17346 reference protocol for passive sampling is from the year 2020, therefore, I suggest to write "first developed in the 1990s" or similar.

We have updated the text as suggested.

P. 2, L. 13/14: Other studies, which investigate ammonia emissions on PM2.5 formation on a global scale include Gu et al. (2021) and Pozzer et al. (2017)

We added Gu *et al.* (2021) and Pozzer *et al.* (2017) to the references listed in line 23/24. We have now updated the text to include the following to the manuscript:

"*The same conclusions have been made by (1) and (2) on of mitigating PM$_{2.5}$ across the world.*".

P. 4, L. 19 ff (2.2. Instrumentation): I suggest to give similar details in the instrument description (where possible). Sometimes performance indicators (like detection limits) are given. 2.2.7 is the only

section where the instrument operators are named. Providing wavenumbers for the spectroscopic techniques in 2.2.4 and 2.2.5 would be beneficial.

To maintain consistency the text outlining the operators of the mini DOAS has been modified, as the purpose of the text was to highlight that are differences between the two instruments. As rightly pointed out we do not state anywhere else the other operators. The revised text in Section 2.2.7 states:

"The two systems taking part in this campaign were miniDOAS #1 developed by the Bern University of Applied Sciences, Switzerland, in collaboration with Neftel Research Expertise and the miniDOAS # 2 developed by the Dutch National Institute for Public Health and the Environment (RIVM), Netherlands."

P. 4, L. 23: Do you mean ½" OD? In Table 2, the tubing diameter is given as 9 mm, which is less than ½". In Table 2, I suggest to specify that the value represents the inner diameter.

Apologies there was an error in the table. We have now updated the table to provide the inner diameter and have also updated all the other calculations in this table due to the error of using the outer diameter for some instruments.

P. 4, L. 24 ff & Table 2: When calculating the residence time from the flow rate and tubing dimensions, the pressure drop along the tubing and respective volume flow increase needs to be taken into account. Looking at some of the values in Table 2, is seems a constant volume flow was used for the calculation, is that correct? If so, values would need to be adjusted for the pressure drop, which would lead to faster residence times (assuming that given flow rates are at STP).

Most inlets did not have a substantial pressure drop in the inlet and therefore did not require the pressure to be taken into account for calculation of residence time, however there was an error in the diameter used which has now been corrected in table 2. However, the reviewer is correct to point this out as we neglected to take this effect into consideration for the inlet of the QCLAS which operates with a much reduced pressure over most of the inlet length. Therefore the residence time of the QCLAS has been recalculated and updated in Table 2.

The residence time of the manifold (+inlet) is given as 1.8 s, however, in Table 2 both add up to 1.62 s. Is that correct? In Table 2, it should be specified if flow rate was at STP or different temp/pressure conditions.

Apologies this is an error the text should read 1.62s and not 1.8 s. All measurements were reported in STP.

P. 13, L. 13: I suggest to state the sampling interval (of the 282 ppb) to be clear that this is based on the original sampling interval and not on the 1 hr resolution (as in Table S1).

Thank you for highlighting the fact that this is not clear. We have updated the text as follows to ensure to make sure the reader understands this is the instruments reporting resolution:

"The LSE instrument reported the highest concentration with a maximum of 282 ppb (1 minute average)."

P. 17, Figure 6 caption: NH3 mixing ratios is the median value (as reported above)? If so,it should be clarified.

In Figure 6, it is the average concentration (ppb) that is reported as this value is used in the calculation of the CV as stated in the caption.

P. 17, L. 21 ff: The response time due to inlet effects was determined by low pass filtering the respective instrument time series to match the unattenuated miniDOAS time series. While this approach seems sound, the retrieved response times seem very high (e-folding time ranging to more than 2 hrs). Another way to determine the response time is by investigating the exponential decay/rise during calibrations (when zero or span is applied). From Figure 13 a) the OGS (i.e. Picarro #2), seems to have a much lower e-folding time. E.g. the 63% (1-1/e) increase at the step change from 0 to 10 ppb (which would represent ambient conditions) seem to be reached within 1 min as opposed to the 49.5 min with the method used here. The dampening effects of ammonia in inlet systems is better described by a double exponential function with a fast time constant that represents the air exchange in the inlet tubing and a slower time constant that describes the adsorption and desorption effects (Ellis et al., 2010). E.g. Moravek et al. (2019) showed the evolution of the dampening over time using the double exponential function. Even if a single exponential decay/increase is used (-> the low pass filtering method yields a single time constant), the authors should include the time constant values from the calibration measurements, also as comparison with other studies. Also, it should be explained why the time constants of both methods (i.e. low pass filtering and visual fitting to miniDOAS time series) would be so different.

We agree with reviewer #2 that the ideally a double exponential function should be used to describe the time response of instruments, as the response observed is a function of volume and then adsorption/desorption effects. However, we cannot fit two time constants in our methods where we deteriorate a fast-response time-series to match each instrument. This would be an under-constrained mathematical problem.

We could indeed be able to fit a double exponential if we had controlled rise / decay measurements with the setups used for the ambient measurements, under ambient conditions, but unfortunately we do not.

The response observed in Figure 13 is indeed much faster. But here the instruments are used with different inlets, with a much drier calibration gas (which probably accounts for most of the better time response) and, for some instruments, with a modified flow rate (to match the limited output of the ReGAS). Therefore we conclude that the suggested approach by Reviewer #2, although in good in principle, is unfortunately not applicable in this case.

"The instruments that were evaluated against the ReGaS (i.e. LSE, Picarro #2, LGR #1, LGR #2 and Tiger Optics) were transferred from the Pyrex manifold to the Teflon manifold for this purpose. Due to the maximum flowrate of the ReGaS1 (5 l min-1) the LGR #2 did not use its external pump but was reliant on the internal pump of the instrument, so had a flow rate of 0.25 l min-1 which equates to a

residence time of 11.0 s for the inlet, which is slower than LGR #1. The system was set for the following concentrations in sequence for the duration of 31 minutes each: 0 ppb, 9.98 ppb, 24.39 ppb, 39.71 ppb, 2.95 ppb and 1.02 ppb. Unfortunately, the data of following instrument was excluded from the analysis; the LGR #1 concentrations remained low even at elevated concentrations indicative of a fault and the Tiger Optics reported 0 ppb as it could not detect concentrations below the 10 ppb detection limit. As a result in this study only information from the OGS, LSE and LGR#2 are evaluated against the ReGAaS."

P. 18, L. 10-12: Why was it discounted or what is your underlying hypothesis regarding the influence of a turbulent flow regime? I would have thought the opposite: if laminar flow conditions are increasing the dampening (due to segregation along the tubing cross section), this would explain why the Picarro #1 does not perform as well as the Picarro #2 (a similar statement is expressed by the surface/volume ratio).

It is agreed that a laminar flow regime will smear out high temporal variations concentrations but when reporting an averaged over a seconds to an hour this loss is not considered to be great. In general it is assumed that in using a turbulent flow this will lead to increase in wall interactions leading to further adsorption. We have revised the text as follows to highlight this discussion:

In section 3.4*: "This is not, however, the only controlling factor for the response of an instrument, as the Picarro #1 inlet is calculated to have a residence time for air of 2.9 s compared to Picarro #2 that has a residence time of 4.7 s (including the manifold inlet and manifold), but it still appears that the Picarro #2 performs better. It is postulated that as the surface area/ volume ratio for the Picarro #1 is two times the surface area/volume ratio of Picarro #2 (Table 2), resulting in more molecules interacting with the inlet walls leading to the observed a smoothed feature. It was discounted that turbulent flow was a controlling factor in the response time, as it would be expected that wall interactions would increase under a turbulent regime leading to greater losses (Table 2).*
."

P. 21, Figure 8: The y-axis offsets from the linear regressions are used to describe the accuracy (i.e. over- or underestimation) of the respective instruments at low mixing ratios. However, the offset may also result in uncertainties in the linear fit (lowest ensemble median mixing ratios are just under 2 ppb). Was this taken into account?

We agree that the y-axis offset is a result of uncertainties in the linear fit and not necessarily representative of the performance at low concentration. Interpretation of the intercept here is limited in order to understand the relationship between predicted NH_3 (from the ensemble median) and the concentration response of the instruments. Contamination, inlet losses, limits of detection and non-linear instrument response are the major issues, which will lead to linear slopes with significant offsets. We have added the text below to note that the discussion is indicative and that in each case reasons for the offset should be investigated, i.e. if an instrument is being deployed to make quantitative measurements.

"Most instruments (Figure 8) had a slope less than 1 with the exceptions of the AP$_2$E, Picarro #1 and the LSE. The largest slope reported was from the AP$_2$E (1.47) and it had the largest negative offset of -1.39 ppb. The y-axis offset is a result of uncertainties in the linear fit, and contamination/losses of NH$_3$ in inlet or the instrument. Interpretation of the intercept is here limited in order to hypothesise regarding the relationship between predicted NH$_3$ (from the ensemble median) and the concentration response of the instruments. Contamination, inlet losses, limits of detection and non-linear instrument response are the major issues which will lead to linear slopes with significant offsets. Negative intercepts are often indicative of losses of NH$_3$ either to the inlet or the instrument, however the large slope and high scatter (r^2=0.76) would also be contributing to the offset value. The instrument with the smallest offset is the QCLAS, which had an offset of 0.05 ppb but had a slope of 0.82 compared to the ensemble median. The largest positive offsets are seen in the Picarro #1 (with an offset of 1.05 ppb), miniDOAS #1 (0.74 ppb), LGR #1 (2.11 ppb), LGR #2 (0.65 ppb) and the AiRRmonia #2 (0.75 ppb). Working with the assumption that within the uncertainty of the regression, the positive offsets are real, the positive offsets in this case could be attributed to contamination in the inlet or in the case of the CRDs on the inline filters."

P. 32, L. 24 ff: The authors use the results of the linear regression between instruments to described the precision. While the R2 value is influenced by the instruments' precisions, the slope would rather indicate the accuracy of an instrument in comparison to the ensemble median. I suggest to make clear that the precision only refers to the R2 value and not the slope.

We fully agree but may not have been clear in the text. We have now updated the text to highlight we are referring to the R^2 and not the slope.

"In this study we assessed the precision by comparing the inter-variability between instruments and the variability against the ensemble median (R^2)."

P. 34, L. 25 ff: Next to the tubing material, contamination of the tubing surface over time can influence the time response significantly (e.g. Moravek et al., 2019). Although the experiment was probably not long enough for it to have a major influence, this point may be included in the discussion.

We have now expanded the discussion on the impact of contamination of the inlet on the time response. The text now reads as follows:

"Not considered in this study is the contamination of the inlet, which is likely to occur over time and has been discussed previously in the literature, though there still no recommendations for frequency of either cleaning or replacing inlets. Moravek et al. (2019), for example, demonstrated that for the QCLAS time response degrades with age (based over a 5 month study) due to contamination of the inlet; but even after cleaning the response time did not always return. As a result, some network instruments have already started to frequently replace their inlets. Twigg et al. (2015) replace their inlet at quarterly intervals for the MARGA currently operated in the UK to try to minimise contamination. Though not studied here, it would therefore also be recommended to consider the frequency that an inlet is either cleaned or replaced to account for potential loss of precision."

In addition we have expanded the discussion in section 4.1 with regards to artefacts from filters with the following text:

"It is however noted that the average concentration reported when compared to the ALPHAs (Table S1) that the instruments with filters tended to report higher concentrations compared to filter free methods (Table S1), supporting the suggestion that filters introduces an artefact. There is recent evidence that frequent filter changes are starting to be considered by network operators to limit artefacts in measurements. For example, (3) reported changing filters a frequency of between 2 weeks to monthly, dependent on atmospheric conditions for a CRD instrument."

P. 34, L. 34 ff: Next to avoiding condensation, inlet line heating was shown to improve the time response in previous studies (e.g. Ellis et al., 2010). This should be mentioned here as well.

Thank you for recommending the Ellis et al. (2010) paper. We have added the following text to manuscript:

"A previous study by Norman et al. (2009) demonstrated the importance of condensation on inlet lines and that care needed to be taken to ensure that condensation did not occur in the inlet. They recommended an optimal design might therefore include thermal insulation and, if possible, keeping inlets heated a few degrees above the ambient temperature, particularly also any sections that run within air conditioned measurement cabins. Ellis et al. (2010) also evaluated using a heated inlet, they found that heating the inlet line led to an improvement in the time response of a QCLAS. During this study, only the LGR#2 and the QCLAS used heated inlets. Caution however is required when heating an inlet, as if the temperature is too high, this will lead to the dissociation of NH_4NO_3 leading to an artefact."

P. 35, L. 13 ff: Ideally, the humidity in the zero air would match the ambient air humidity levels. One way to produce zero air is by removing NH_3 from ambient air through a heating catalyst (without a drying cartridge or similar). Assuming that the water vapor is conserved, the zero air would then have similar humidity levels than the ambient air. Next to its influence on spectroscopy, humidity levels can also affect the adsorption/desorption processes in the inlet line. Therefore, having humidified zero air (at least to some degree) would be a beneficial for all instruments with an inlet line.

Many thanks to the reviewer for their comment. We have now reviewed and updated the progress towards a standard section of the manuscript. Please see below the revision to the text.

"However, preparing humid gas samples with accurately characterised NH_3 concentrations in the ambient concentration range is challenging and work is required to develop standard methodologies to produce a humidified gas standard such as using a scrubber or heated catalysts. Pollack et al. (2019) provides a valuable study in evaluating these approaches. It would be advisable that a standard is also used on a frequent basis to determine the contamination of the set-up, as previously demonstrated by Ellis et al. (2010) and Pollack et al. (2019), who observed that the calculated time response alters due to contamination. In addition, a standard design of inlet needs to be agreed (where applicable), as evidence from the Picarro and AiRRmonia set-ups in this study (Figure 14), would suggest that this can lead to losses of information of the temporal pattern of NH_3. Consideration is also required to determine if passivation of the inlet is valuable to routine air quality monitoring, as there is evidence this limits the interactions of the NH3 with the inlet walls, as discussed in Roscioli et al. (2016). Open path techniques, such as DOAS, will benefit from the availability of zero-air facilities, where instruments can check their zero level on ammonia free air. Work on such a facility is ongoing."

TECHNICAL COMMENTS

P. 7, Table 2: Typo in "AiRRmonia #1" & "Operated with a filter" -> "No" Corrected

P. 7, L. 14: "are sampled" instead of "is sampled" Corrected

P. 7, L. 15: do you mean "through which air is drawn and …"? Yes, may thanks for the correction Corrected

P. 7, L. 27: ("Erisman, 2001)" Corrected

P. 22, L. 7: Insert space before "ppb". Corrected

P. 25, Figure 10: "a)" and "b)" missing Corrected

P. 33, L. 13: Move comma to at the end of subclause (after parentheses in L. 14). Corrected

REFERENCES

Ellis, R. a., Murphy, J. G., Pattey, E., van Haarlem, R., O'Brien, J. M., and Herndon, S. C.: Characterizing a Quantum Cascade Tunable Infrared Laser Differential Absorption Spectrometer (QC-TILDAS) for measurements of atmospheric ammonia, Atmos. Meas. Tech., 3, 397–406, https://doi.org/10.5194/amt-3-397-2010, 2010.

Gu, B., Zhang, L., Dingenen, R. Van, Vieno, M., Grinsven, H. J. Van, Zhang, X., Zhang, S., Chen, Y., Wang, S., Ren, C., Rao, S., Holland, M., Winiwarter, W., Chen, D., Xu, J., and Sutton, M. A.: Abating ammonia is more cost-effective than nitrogen oxides for mitigating PM2.5 air pollution, Science (80-.)., 374, 758–762, https://doi.org/10.1126/science.abf8623, 2021.

Moravek, A., Singh, S., Pattey, E., Pelletier, L., and Murphy, J. G.: Measurements and quality control of ammonia eddy covariance fluxes: a new strategy for high-frequency attenuation correction, Atmos. Meas. Tech., 12, 6059–6078, c, 2019.

Pozzer, A., Tsimpidi, A. P., Karydis, V. a., de Meij, A., and Lelieveld, J.: Impact of agricultural emission reductions on fine-particulate matter and public health, Atmos. Chem. Phys., 17, 12813–12826, https://doi.org/10.5194/acp-17-12813-2017, 2017.