Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-140-RC1, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "UK greenhouse gas measurements at two new tall towers for aiding emissions verification" *by* Ann R. Stavert et al.

Anonymous Referee #1

Received and published: 2 August 2018

Review Criteria: 1. Does the paper address relevant scientific questions within the scope of AMT? Yes, the paper provides a useful reference for greenhouse gas measurements at two sites and a discussion of systematic errors associated with water vapor that is important and timely.

2. Does the paper present novel concepts, ideas, tools, or data? It is worthwhile to document the analytical systems in use at these sites. Methods for dealing with water impacts on co2 and ch4 measurements are rapidly evolving and this group has worked hard to document and characterize systematic errors in their system. The paper describes a series of laboratory and field experiments that will provide useful information for researchers doing similar work.

3. Are substantial conclusions reached? The paper is overly long and the section

Printer-friendly version

Discussion paper

describing the experiments to characterize water vapor errors is difficult to follow. There are some significant conclusions, and the paper should be distilled to highlight those.

4. Are the scientific methods and assumptions valid and clearly outlined? The authors have worked hard to characterize their instrument, but substantial revision is needed to clarify the methods and results described in the paper. The statistical analysis of the data is flawed as described below.

5. Are the results sufficient to support the interpretations and conclusions? Generally yes, with some exceptions noted below.

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? Generally yes.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution? Generally yes, but a few key references are missing or incorrect as noted below.

8. Does the title clearly reflect the contents of the paper? yes

9. Does the abstract provide a concise and complete summary? yes

10. Is the overall presentation well-structured and clear? Improvement needed as described below.

11. Is the language fluent and precise? Improvement needed as described below.

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? Generally yes.

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated? See specific suggestions below.

14. Are the number and quality of references appropriate? yes

15. Is the amount and quality of supplementary material appropriate? The description

of the laboratory experiments in the text should be shortened and distilled. Many of the details could be moved to the supplement.

Major comments:

A full uncertainty analysis is evidently planned for a future paper, but more work is needed in the current manuscript to summarize the various contributions to the uncertainty of the datasets described here. This manuscript provides many details about systematic errors related to water vapor, but a concise summary of the impacts on the measurements is lacking. Ideally the dataset would include time and site dependent estimates of this uncertainty component depending on the method of drying/water correction employed. Other uncertainty components such as larger errors for measurements outside of the calibration range should also be reported for individual measurements.

Although some of the uncertainties discussed here are significant compared to the WMO internal reproducibility guidelines, these errors/uncertainties are likely very small relative to the so-called observation (model-data mismatch) errors assigned in the inverse modeling. Some context about how the measurement uncertainty compares with model-representation errors would be helpful. It would also be helpful to see how the measurement uncertainty compares to signals of emissions. The WMO extended measurement compatibility goals should also be noted (+/- 0.2 ppm for CO2, +/- 5 ppb for CH4).

The statistical analysis of the time series data needs major improvement. The Thoning fit is not a good choice for this dataset, as is clearly evident in Figure 5. Specific suggestions are provided below.

The inversion analysis as currently presented is not compelling due to only minor reported improvement in total uncertainty and apparent flaws in the inversion framework that cause uncertainties in some regions to increase with additional data. Since AMT is not suitable for a detailed discussion of the inverse modeling methodology perhaps better to omit. A simpler presentation of how the additional sites improve the sensitivity to

surface fluxes could be substituted (i.e. a map of the total surface sensitivity/footprints estimated by the NAME model showing the impact of the additional sites).

Discussion of the lab water vapor tests is very hard to follow and should be reorganized and significantly shortened. Specific suggestions are given below. It is not necessary to exhaustively present results from experiments that were inconclusive in the body of the paper. Although researchers who are struggling with similar issues might benefit from this information, it should be relegated to the Supplement or to an Appendix in order to simplify the main paper. Despite having direct experience with analyzing results from these types of water corrections, I found the presentation difficult to understand.

Specific comments:

Page 1, line 27/28: "...this error is mostly calibrated out" is the 0.02 umol per mol error the remaining error after applying the calibration? Or is the nafion-related error « 0.02 ppm after calibration? In either case 0.02 ppm is nearly negligible and likely smaller than the total measurement error of the analytical system, which has not been adequately characterized.

Page 2, line 25: does AMT allow references to a paper in prep?

Page 6, line 18: CRDS dwell times at each level are surprisingly long

Page 7: line 1, "This air is dried to <0.005%..." I think this refers to the counterpurge air but not totally clear at this stage if it might refer to the sample air. It would be useful to state what is level of drying that is accomplished with the nafion for the GC channel.

Page 7: droplet test has weaknesses due to rapid changes of humidity that are inadequately resolved. Potential mismatchs/lags among co2/ch4/h2o channels. Also, I am not sure that Yver Kwok et al. 2015 is the best reference for this. I quickly checked and did not see any discussion of the droplet test in that paper. Maybe it would be better to cite the Rella 2013 AMT paper which describes several implementations of the droplet method. A citation for the Rella 2013 paper is currently lacking.

Page 7, line 15: A cylinder of air was not humidified. Instead, air from a cylinder was humidified.

Page 8: data collected in the first 5 minutes following the injection was excluded. What is the maximum humidity sampled after these data have been excluded? (i.e. is the maximum h2o value included in the fits significantly lower than the 4.5% value mentioned on page 7, line 20?) It is frustrating & confusing that the water corrections are discussed in multiple sections of this paper (here on page 8 and again in section 3.3.1).

Page 8, line 22: What are the calibration gases spiked with?

Page 9, line 1: It would be helpful to specify what is the maximum systematic error due to differences between sample and standard isotopic composition, since spiking standards can result in isotopic compositions that are outside the ambient range.

Page 9, line 5: What is the uncertainty associated with the non-linearity correction? Dynamic dilution is a complex procedure and likely to have non-negligible uncertainties. Why not use a set of gravimetric standards instead?

Page 10, line 9: It is not clear how the long-term repeatability numbers here are being computed. The parenthetical description (xbar minus +/- one sigma) is not an adequate description. Are these numbers the mean standard deviation computed over all the tanks over a year? In any case, instead of “long-term repeatability” it would be better to report the “long-term reproducibility”, since a metric of the compatibility of the measurements over periods of months to years is needed. These terms are defined in the Guide to the expression of uncertainty in measurement: https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf The difference between repeatability and reproducibility has to do with whether the conditions of the measurement are changed, and over timescales of months to years, standards are changing, ambient conditions, humidity levels are changing, etc. It seems unlikely that one could confidently interpret differences in CO2 in measurements made months or years apart at the level of 0.018 or 0.013 ppm.

Page 10, line 21: How was the cycle time threshold of 8 sec determined? Are those data that are filtered based on cycle time obviously bad? Page 11, section 2.2.2: The Thoning fits are most appropriate for remote sites, and the method often produces spurious results when gaps or large pollution events are present. These effects can be clearly seen in Figure 5. Figure 5 does not add any value to the paper and should be removed. More careful analysis is needed if seasonal cycles and trends are to be reported for these sites.

Page 12, line 19: Are the HFD mean CO2 mole fraction of 407.5 ppm and the BSD mean of 404.7 the means over each entire record? If yes, then this quantity will seemingly be affected by the gaps and there is no use in reporting this quantity or interpreting the difference between the sites.

Page 12, line 25: The text implies that there are some high CO2 events at BSD that do not have associated high CO. If that is the case then any such events are not likely to result from biomass burning and must have a different source (e.g. power plant plume).

Page 13, line 5: It seems unlikely that multi-day CO2 enhancements resulting from pollution over London or Europe would not be associated with elevated CO. Could the elevated CO2 result from advection of air from higher latitudes?

Page 13, line 10: The typical diurnal variation of CO2 measured on tall towers is well understood, e.g., https://www.tandfonline.com/doi/pdf/10.3402/tellusb.v47i5.16070?needAccess=true

Page 14, Line 15: See comment above. The authors are correct that a major issue with the Thoning fitting routine is the underlying FFT, which requires interpolation across gaps. But since the fits are obviously flawed, why not use alternative methods to investigate the seasonal cycle and trends? A simple analysis using monthly mean or median values would be much more robust and simple to implement and explain. And/or box and whisker or fiddle plots could be used to describe the seasonal cycles and trends. Any months with significant gaps could/should be removed.

Page 14, line 20: It is not surprising that CO is decreasing. This has been widely reported. It would be interesting to compare the trends for all of these gases with background values from Mace Head.

Page 14, line 26: Again, box and whisker plots showing quantiles of the data would be helpful here to quantitatively describe any CO differences between sites. It is not useful to report he multi-year mean values without any uncertainties.

Page 15, N2O: There is a more recent Nevison et al paper that discusses N2O seasonality over the US (https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GB005759)

Page 16, Line 13: SF6 seasonality might be driven by atmospheric transport.

Page 16, Line 23: The additional sites appear to have negligible impact on the inversions. The estimate of total emissions does not change significantly, and the uncertainties are only slightly reduced.

Page 17, discussion of Figure 6: Perhaps additional panels could be included to show the posterior flux distribution for the GAUGE and/or DECC cases. Figure 6a could be revised to show the magnitude of the redistribution relative to the mean. Otherwise it is some work for the reader to understand whether the redistribution is significant.

How can the addition of more data make the inversion estimate more uncertain in some regions? I think this can only be the case if the inversion framework underestimates the uncertainty with the 4-tower case. Some additional explanation is needed. Perhaps it would help to indicate which regions are significantly constrained using a footprint/sensitivity map.

Page 19, Line 27: Apparent typo. . . .lasted between 2-5 (minutes?)

Page 26, discussion of figure S3: The average residuals given in Table 1 do not adequately describe the uncertainty indicated in these plots.

Page 26, line 13: The reported large difference between the humidity of samples and

standards is a bit concerning. Could this be mitigated by using a longer nafion drier? And/or a chiller could be used to remove the bulk of the sample humidity upstream of the nafion drier as is done in the NOAA tall tower systems.

Page 26, line 15: “However it is likely to be a systematic offset of the order of -0.05 to -0.1 ppm. . .” based on what evidence? The residuals in S3? Or the Reum tests? Or something else?

Page 26, last paragraph: Are these below ambient and above ambient results shown somewhere?

Page 27, line 10: It is surprising that the maximum humidity at these sites is <2%.

Page 27, line 12: I understand that when the Nafion drier was installed then many of the air samples have H2O < 0.1%, but what are the implications? This paragraph is discussing Figure 10 a & b, and this particular figure does not seem to show any troubling implications for samples with H2O < 0.1%.

Page 27, line 22: It seems very impractical to calibrate high-humidity sites weekly. Also, I am not convinced that the droplet test is accurate at very high humidity. The daily tests at U of Br in Figure 10e show extremely large variability at humidity > 2.5%.

Page 28, line 6: Why was this test not performed at H2O < 0.7% since that is where the droplet test is unreliable due to rapidly changing H2O?

Page 28, line 21: The range of humidity for HFD and BSD here is different than given on Page 27 line 10 (max of 2.5% instead of 2.0%). Meanwhile the Introduction states that the DECC/GAUGE network observes samples with humidity of up to 3.5%.

Page 29, line 8: Clarify 0.5 to 3.5% in the Wet experiment versus < 0.31% for the others.

Page 29 & 30, discussion of Figures 12 & 13 and table 2. This discussion is extremely hard to follow. Since it seems that the droplet tests are highly uncertain below about

<0.3%, then consider just omitting that discussion and sticking with the estimates from Reum. Or perhaps omit this discussion altogether. The tepid conclusion at the top of page 31 does not warrant the amount of discussion present in the current manuscript. If the experiments failed or were inconclusive then why include them? If you feel it is important for the community to understand the pitfalls of your attempt in order to avoid similar futile attempts, then perhaps relegate this discussion to the supplement.

Page 29, line 23: "Although small, these changes are an order of magnitude smaller. . ." This is puzzling. . .did you mean to write: ". . .these changes are an order of magnitude larger"?

Page 31, line 25: The measured difference of 0.02 ppm is practically negligible and unlikely to impact any conceivable scientific analysis, except perhaps analysis of spatial gradients in the high southern hemisphere.

Page 32, line 3: "These cylinders are very dry, H2O < 0.0001%..." But, above on page 26, it states that standards are significantly humidified by the nafion, and so the difference should be even smaller than the 0.005 ppm that you report.

Page 32, line 8: "This is not the case for the water correction bias, which varies with H2O (section 3.3.2)."

Page 32, line 22: Unfortunately, the inversion results as presented here indicate only marginal improvement in the flux estimates. However this is likely due to limitations of the inversion framework. In particular, it seems that the uncertainty estimated for the 4 tower network is too small.

Page 33, line 33: I think the should read 0.015 umol/mol instead of 0.15.

Table 1: Mean residual is not an adequate diagnostic of the uncertainty, since the residuals vary strongly as a function of H2O. Perhaps include an example plot for one of these calibration instances in the Supplement.

Table 2: ISO format for dates is YYYY-MM-DD

Figure 2: Include definition of TOC in the legend or the caption. It is mentioned in the text but should also be noted here.

Figure 8 & 9: Essentially no discussion of these figures is provided. Either describe these figures or omit.

Figure 11: Include the CO2 and CH4 values of the cylinders in the legend or the caption.

Table 2 & Table 3: What is the uncertainty of the assigned values for the calibration standards? The reproducibility of the NOAA scale for CO2 is estimated to be 0.03 ppm. But how well do the calibration centers propagate the scale?

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-140, 2018.

Printer-friendly version

Discussion paper