Reply to the comments of reviewer #1 of "UK greenhouse gas measurements at two new tall towers for aiding emissions verification"

Ann R. Stavert[1], Simon O'Doherty[2], Kieran Stanley[2], Dickon Young[2], Alistair J. Manning[3], Mark F. Lunt[5], Christopher Rennick[4] and Tim Arnold[4,5].

[1] Climate Science Centre, CSIRO Oceans & Atmosphere, Aspendale, VIC, 3195, Australia
[2] School of Chemistry, University of Bristol, Bristol, BS8 1TS, United Kingdom
[3] Met Office, Exeter, Devon, EX1 3PB, United Kingdom
[4] National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom
[5] School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3FF, UK

General 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 uncer- tainty 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 es- timates of this uncertainty component depending on the method of drying/water correc- tion 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 in- verse 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 mea- surement compatibility goals should also be noted (+/- 0.2 ppm for CO_2, +/- 5 ppb for CH_4).

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 re- ported 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 bet- ter 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.

The authors would like to thank reviewer 1 for their helpful comments. We have endeavoured to incorporate all the suggestions that they have made. The locations of changes to the paper are given as Page Number, Line number of the new paper while the reviewers comments contain the locations in the original paper.

As requested by the reviewer a summary of the errors associated with the drying and water correction methods used at the Heathfield and Bilsdale sites is now included (see Table 5). Estimates of the error associated with making observations outside the range of the calibration suite has also been included (Page 14, lines 13 – 23). Discussion of these errors relatively to the WMO extended compatibility goals, model-data mismatch error and the magnitude of emissions signals have also been included (Page 28, lines 20 - 26).

As advised by the reviewer we have significantly altered the statistical analysis, no longer using the Thoning et al. approach and instead using box plots of linearly detrended data. References to specific changes are included in the replies to the specific comments below.

Also, as suggested by the reviewers, due to the measurement focused scope of AMT we have removed details of the inversion modelling and its results. These will be discussed more fully in a later paper.

Following the advice of both reviewers we have removed a large section of the text relating to the first set of Nafion® drying tests and the accompanying figures. As such some technical corrections are no longer applicable, these have been noted as such. We have also simplified and reordered the discussion of the water correction work. The paper has been significantly restructured with the Nafion® and water correction work no longer a separate section. Instead they are incorporated in to the main experimental and discussion sections.

Even with the inclusion of the extra information requested by the reviewers these changes have reduced the length main body of the paper by 6 pages and removed 4 figures.

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.

The 0.02ppm error is the error prior to carrying out the calibration. We expect the error following the calibration to be ≪0.01ppm. This has been clarified in the text. It is important to note that this error is only removed through calibration because the calibration cylinders are also passed through the Nafion® which may not be the case for all monitoring sites. As such we feel it is important to highlight and quantify this possible source of error.

We would politely disagree with the reviewer, and suggest that a 0.02ppm bias is far from negligible - a bias of this magnitude represents 40% of the WMO Southern Hemisphere inter-laboratory compatibility guideline. As the reviewer correctly notes, other measurement errors associated with an analytical system are most likely to be larger, hence, in conjunction with this bias they could well lead to a significant error within a data set.

As outlined in the text, a separate paper currently in preparation will characterise the measurement error of the analytical system of these two new sites along with the other four sites within the UK GHG monitoring network.

Page 2, line 25: does AMT allow references to a paper in prep?
We have removed this reference.

Page 6, line 18: CRDS dwell times at each level are surprisingly long
The length of the CRDS sampling period at each height was designed to obtain as long as possible period of continuous data at each height during an hour. That is, for the site with three intakes 20 minutes was spent at each height during each hour. Such a sampling strategy is not unusual for greenhouse gases, for example the INFLUX experiment measured from some intakes for up to 40 minutes of every hour (Richardson et al., 2017).

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.
We have clarified and added extra information to the text.
Page 6, line 21 – 23, "The counter purge is dried to < 0.005 % H_2O by the compressor (50 PLUS M, EKOM, Slovak republic) and a gas generator for total organic carbon systems (TOC-1250, Parker Balston, USA). Previous examinations of this drying method have found that samples are dried to < 0.0002 % H_2O (Young, 2007)."

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.
We thank the reviewer for highlighting this oversight. We have replaced the reference to Yver Kwok et al. 2015 with a reference to Rella 2013 (Page 7, line 20).

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

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).
The text has been clarified to include
Page 8, line 8, "This 5-minute cut-off reduced the maximum H_2O value included in the fit to 4 % H_2O."
Following the restructuring of the paper, the water correction work is now split into two components, an experimental (Sections 2.3.4, 2.3.5 and 2.3.6) which outlines how the water correction tests were conducted and results & discussion (Sections 3.3 and 3.4) which details the results of the experiments and their implications.

Page 8, line 22: What are the calibration gases spiked with?
The below has been added to the text.

Page 15, lines 4 -10, “The cylinder spiking and filling techniques of the calibration cylinders varied. The Heathfield calibration suite and the second Bilsdale calibration suite were filled at GasLab MPI-BGC Jena and consisted of natural air spiked using a combination of pure CO_2 and a commercial mixture of 2.5 % CH_4 and 0.5 % CO in synthetic air. The “high” calibration cylinder of the first calibration suite used at the Bilsdale site was filled with peak-hour ambient air at EMPA, Dübendorf, Switzerland while the “low” and “mid” cylinders were based on Mace Head air, in the case of the “low” this was diluted with scrubbed natural air.”

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.
We cannot comment directly on our measurement error as we do not have measurements of the isotopic composition for our calibration gases. However, we refer readers to other papers who investigate this topic in detail and give an estimate based on the literature.
Page 15, lines 13-19, “The effect of an isotopic mismatch between the calibration standards and the sample has been examined in detail by Flores et al. (2017), Griffith (2018) and Tans et al. (2017). With Griffith (2018) showing that, for a sample of 400 $\mu mol\ mol^{-1}$ CO_2 and 2000 $nmol\ mol^{-1}$ CH_4, the error will range between 0.001 – 0.155 $\mu mol\ mol^{-1}$ CO_2 and 0.1 – 0.7 $nmol\ mol^{-1}$ CH_4 depending on the magnitude of the sample to standard mismatch. As such, we expect a worst-case scenario estimate of the error associated with our measurements < 0.04 % for both CO_2 and CH_4.”

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?
These non-linearity corrections are typically exceptionally stable with Hammer et al. (2008) finding differences of < 0.2 $nmol\ mol^{-1}$ N_2O between non-linearity corrections determined 10 years apart and over the range of interest (326 to 340 $nmol\ mol^{-1}$ N_2O) highly linear. While, studies including Hall et al. (2011) and van der Laan et al. (2009) have found SF_6 to be very linear at mole fractions > 4 $pmol\ mol^{-1}$ and for the instrumental response function to be very stable over time. As such, we expect the uncertainty of the non-linearity correction to be small. Although not discussed in detail in this paper this uncertainty will be addressed in the associated uncertainty analysis paper as outlined in the introduction.
Page 2, lines 23 – 26, “A further paper, currently in preparation, will discuss the integration of these new sites with the existing UK Deriving Emissions linked to Climate Change (DECC) network (Stanley et al., 2018) funded by the UK Department of Business, Energy and Industrial Strategy (BEIS) and provide a full uncertainty analysis for data collected at all the DECC/GAUGE sites.”
The UoB group, along with other groups (e.g. the Schauinsland GAW station) have a history of successfully using dynamic dilution for the quantification of the N_2O and SF_6 non-linearity. The laboratories at the BSD and HFD sites are physically small making it difficult to store or use a set of dedicated gravimetric N_2O and SF_6 standards at each site and shipping a set of cylinders between the sites would be logistically difficult considering the remote locations of and limited access to the sites. In contrast, the dynamic dilution method is able to generate multiple calibration points using just two cylinders and is logistically simple to deploy and ship between sites. As such, the decision was made to use this approach.
We have also added the information below

Page 15, line 23 – Page 16, line 3, "This approach, dynamic dilution, has a history of use in similar field locations (Hammer et al., 2008) and is able to generate multiple calibration points using just two cylinders. This greatly reduces the number of cylinders needed, a key concern for space-limited locations like BSD and HFD. An assessment of the uncertainty associated with this non-linearity approach will be included in a future paper currently in preparation. However, previous studies (Hall et al., 2011;van der Laan et al., 2009;Hammer et al., 2008) have found the ECD detector response to be extremely stable over time and very linear for both SF_6 and N_2O in the mole fraction range typical of the HFD and BSD stations. As such, we expect the uncertainty of the nonlinearity correction to be very small."

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.
We thank the reviewer for identifying the incorrect use of repeatability. We have now corrected this to reproducibility throughout the text.
A reproducibility of minute mean values at the level of 0.02 μmol mol^{-1} CO_2 over a period of years would indeed be unlikely however we are reporting the reproducibility of a 20-minute mean value which is much more stable. We have more clearly defined how these numbers have been calculated. See
Page 17, lines 3 – 4, "The long-term reproducibility of a 20-minute mean was estimated as the mean standard deviation of the daily 20-minute measurements of 4 standard cylinders at each site."

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?
The 8-second cycle time threshold was determined after a close examination of the typical cycle times and their relationship to other key parameters. Longer cycle times were typically related to periods of more variable cavity pressure and outlet valve parameters. Measurements of cylinders with long cycle times were also typically significantly different, between 0.5 and 50 μmol mol^{-1} CO_2 from the mean cylinder value.

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.

The Thoning fits and Figure 5 have been removed from the paper and replaced with a seasonal cycle analysis using box plots and simple trend estimate based on a linear fit to the data.

Page 18, Lines 2-6, "The long term trend in mole fraction at each site was estimated as the mean linear trend in the minute mean data over the period 2014-2017, inclusive. Seasonal and diurnal trends in the data were assessed using monthly and hour-of-day box plots of hourly means of detrended minute-mean data developed using the Python matplotlib.boxplot package. Here the long-term trend was removed by using a least-squares fit between a quadratic and the minute mean data."

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.

We thank the reviewer for noting this. References to mean mole fractions calculated in this manner have been removed.

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).

We thank the reviewer for this observation and have included it in the text.

Page 18, Lines 17 - 19, "While events that do not show corresponding high CO mole fractions, the majority of which occur in the higher two intakes, are likely to be driven by less local CO_2 sources, for example power plants."

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?

We thank the reviewer for this comment as it demonstrates that our text was not written clearly. We had meant to say that the difference in pollution patterns, short sporadic events vs. prolonged periods of increased background mole fraction, was not common to all gases between the two sites rather than not common between the gases at an individual site. In fact, coincident 2-3 day periods of increased CO, CH_4, N_2O and SF_6 were observed at HFD. This has been clarified in the text.

Page 18, Lines 20 - 26, "HFD is located in southern England, just south of London (Figure 1). Here, high CO_2 events are typically longer — 2 to 3 days — and coincide with elevated CH_4, CO, N_2O and SF_6. Rather than appearing as peaks superimposed on a background value these periods have a positive shift in the entire diurnal cycle, suggesting a change in the background mole fraction. Air histories, based on the output of the Numerical Atmospheric dispersion Modelling Environment (NAME) Lagrangian dispersion model, outlined in Manning et al. (2011), for these periods of elevated CO_2 typically show the source of the air to be from over London or Europe."

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

We have clarified our text and included a reference to this paper.

Page 19, Lines 1 -6, “Both sites show a clear relationship between CO_2 mole fraction and intake height with the lowest height generally having the most elevated mole fractions, followed by the higher heights (Figures 3a & 4a). This trend, also apparent for CH_4 and CO (Figures 3c & e and Figure 4c & e), is typical of tall tower measurements and is driven by proximity to surface sources (Bakwin et al., 1998). The gradient in CO_2 and CH_4 mole fraction is most apparent during the early hours of the morning (see Figure 3b & d and Figure 4b & d) when the boundary layer is the lowest.”

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.
All references to the Thoning fit have been removed and replaced with a discussion of seasonal cycles based on box plots.

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.
The -7 nmol mol^{-1} yr^{-1} trend observed in HFD and BSD CO is slightly larger than the -5 nmol mol^{-1} yr^{-1} trend apparent over the same period in the CO data collected at Mace Head. However, as HFD and BSD are on a different calibration scale to the MHD it is not possible to determine whether this is a real difference or a calibration artefact. As such, we decided not to include this in the text of the paper. However, as all three sites use common calibration scales for CO_2, CH_4, N_2O and SF_6 comparisons between the growth rates of these gases have been added to the text.
Page 20, Lines 1-7, “Of the 5 gases measured at HFD and BSD, CO is the only gas to show a decrease in mole fraction between 2013 and 2017, roughly -7 nmol mol^{-1} yr^{-1}. In contrast, the CO_2 and CH_4 data increase by 2-3 µmol mol^{-1} yr^{-1} and 5-9 nmol mol^{-1} yr^{-1} respectively, varying on the intake height. These agree well with the ~2 µmol CO_2 mol^{-1} yr^{-1} and ~8 nmol CH_4 mol^{-1} yr^{-1} trends observed at Mace Head (MHD, 53.327 °N, -9.904 °E, Figure 1), a remote site within the UK DECC network located on the west coast of Ireland. The CO data collected at MHD is not on the NOAA x2014 CO calibration scale making direct comparisons between growth rates at the three sites meaningless.”
Page 20, Lines 20-22, “The long-term trend, ~0.8 nmol N_2O mol^{-1} yr^{-1} (calculated using data from the 108 m and 100 m intakes at BSD and HFD over the period of coincident data collection, 2014 to mid 2016) also agrees well between the two sites and with MHD, also ~0.8 nmol N_2O mol^{-1} yr^{-1}.”
Page 21, Linese 17-18, “The long-term trend in the SF_6 mole fraction at BSD and HFD shows a gradual increase of 0.3 pmol mol^{-1} yr^{-1} again agreeing well with MHD which showed an identical growth rate.”

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 the multi-year mean values without any uncertainties.

This has been corrected and a discussion based on box plots included.
Page 20, Lines 8-10, "While the range of minute mean CO mole fractions was significantly larger at BSD, 63 to 9500 nmol mol^{-1} than HFD, 60 to 4850 nmol mol^{-1}, the high CO values observed at BSD were relatively rare. This is reflected in the smaller quantile spread of the BSD data compared with the HFD data (Figure 7e & f)."

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)
We thank the reviewer for highlighting this new Nevison paper. However, as our paper, and the original Nevison reference, focuses on the UK while the new paper examines the USA we feel the original paper is the more appropriate reference.

Page 16, Line 13: SF6 seasonality might be driven by atmospheric transport.
This increase in pollution events of SF6 during the winter months is seen across the combined DECC/GAUGE network (4 sites across the UK but not MHD) and is very intriguing. However, the air history maps for these events do not consistently indicate any one area in the UK or the continent suggesting that this is an emission rather than transport related increase. This has triggered further investigations in the inventory community but no conclusions have yet been made.
Page 21, Lines 21 -25, "This seasonal shift occurs across the wider DECC-GAUGE network and air history maps suggest that it is not associated with an obvious UK or continental region. As such, instead of an atmospheric transport driven shift we believe this to be a true change in emissions and hypothesise that this may be due to increased load on, and hence increased failure of, the electrical switchgear during the colder months."

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.
Section removed.

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.
Following the comments of both reviewers the decision was made to remove the discussion of the effect of the two towers on UK flux estimates. This will be discussed in a separate paper.

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

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

Table 1 now includes the interquartile range which better expresses the range of the uncertainties as indicated in plot S3.

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.
Yes, the calibration gases could be further humidified using a longer Nafion® or the samples further dried using a chiller system. This suggestion has been incorporated into the conclusions and future work section (Page 28, Lines 15 - 16).

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?
This estimate was based on the residuals shown in Figure S3 (now Figure S5). This has been clarified in the text.
Page 23, Lines 17-20, "However, for BSD and HFD, assuming that the residuals of the droplet water tests are an accurate reflection of the likely error (Figure S3, we expect there to be a systematic offset of the order of -0.05 to -0.1 µmol mol^{-1} CO_2 and -1 to -2 nmol mol^{-1} CH_4."

Page 26, last paragraph: Are these below ambient and above ambient results shown somewhere?
The coefficients for the above and below ambient tests are shown in Table 1 and plots of the residuals are included in the Supplement, Figure S6.

Page 27, line 10: It is surprising that the maximum humidity at these sites is <2%.
This is a typo. It should have been 2.5%. This has been corrected (Page 25, Line 11).

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%.
This paragraph was meant to convey that in light of the variability of the water correction over time, particularly at higher water contents, using a Nafion dryer to obtain a relatively low and stable sample water content was an advantage. This has been clarified.
Page 24, Lines 14 – 16, "In light of the temporal variability of the water correction over time, particularly at higher water contents, using a Nafion® dryer or alternative drying method to obtain a relatively low and stable sample water content would be an advantage."

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%.
As the reviewer has noted, performing weekly water calibrations would be a burdensome task and the droplet test can be unreliable at high humidities. This has been included in the text.
Page 24, Lines 24 – 27, "The impracticality of such a frequent testing regime along with the apparent unreliability of the droplet test at H_2O > 2.5 % (See Figure S3 g) mean that an

alternative method, possibly partial drying, or a higher level of uncertainty may need be applied to measurements made at higher water contents."

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?
We used the lowest stable set point of the dewpoint generator, 2.5°C, which equated to a water content of 0.6-0.7 %. As such, we were unable to use the dew point generator to examine water contents < 0.7 %.

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%.
The values reported on Page 27, line 10 was a typo. This has been corrected. Other sites within the DECC/GAUGE network, specifically the Angus site in Scotland, experienced humidities > 2.5%.

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

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.
As suggested by the reviewer we have removed this section and the accompanying figures and table.

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"?
As suggested by the reviewer we have removed this section and the accompanying figures and table.

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.
Yes, as the reviewer has stated these differences are small, particularly in the light of the magnitude of the signals observed in the UK regional network. However, small errors can compound and individually may be a cause of concern at "clean air" locations as such we feel it is important to highlight them.

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.

The relationship between water content and CO_2 exchange discussed in this section and plotted on the X axis of Figure 9 is in reference to the water content of the sample (or standard) prior to it entering the Nafion®. Yes, the standards are humidified by the Nafion® but the CO_2 exchange occurs prior/during this humidification process.

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

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.
At the suggestion of the reviewers references to the inversion study have been removed from the paper.

Page 33, line 33: I think this should read 0.015 umol/mol instead of 0.15.
Yes, this was incorrect. We thank the reviewer for noting this. It has been corrected Page 28, Line 6.

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.
The plots of all of these calibration instances are included in the supplement – see Figures S5 and S6. In an effort to encompass the variability of the residual Table 1 has also been altered to include not just the mean residual but the 25th and 75th percentiles.

Table 2: ISO format for dates is YYYY-MM-DD
The date format in Table 2 (now Table 5) has been corrected.

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

Figure 8 & 9: Essentially no discussion of these figures is provided. Either describe these figures or omit.
Figure 8 has been removed. Figure 9 has been moved to the supplement (now Figure S4) and referred to specifically Page 11, Lines 22, 23, 28, Page 12, Line 4 and Page 13, Line7.

Figure 11: Include the CO2 and CH4 values of the cylinders in the legend or the caption.
These values are now included in the caption of Figure 9 (Now Figure 11). "The (a) CO_2 and (b) CH_4 change in the Wet – Dry sample treatment difference with sample water content for cylinders UoB-04 (515.3 µmol mol^{-1} CO_2 and 2585 nmol mol^{-1} CH_4), H-296 (406.6 µmol mol^{-1} CO_2 and 1947 nmol mol^{-1} CH_4), UoB-06 (384.8 µmol mol^{-1} CO_2 and 1975 nmol mol^{-1} CH_4) and H-306 (372.5 µmol mol^{-1} CO_2 and 1776 nmol mol^{-1} CH_4). Error bars are the larger of either the standard deviation of the mean difference or the uncertainties of the two sample types added together in quadrature."

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 centres propagate the scale?

The uncertainties provided by the calibration centres have been added to Table 3. The values reported in Table 4 were made at the AGAGE Mace Head laboratory or the University of Bristol laboratory. An uncertainty assessment for scale propagation in this manner is outside the scope of this paper but will be included in the future uncertainty paper.

References

Bakwin, P. S., Tans, P. P., Hurst, D. F., and Zhao, C.: Measurements of carbon dioxide on very tall towers: results of the NOAA/CMDL program, Tellus B, 50, 401-415, doi:10.1034/j.1600-0889.1998.t01-4-00001.x, 1998.

Flores, E., Viallon, J., Moussay, P., Griffith, D. W. T., and Wielgosz, R. I.: Calibration Strategies for FT-IR and Other Isotope Ratio Infrared Spectrometer Instruments for Accurate δ13C and δ18O Measurements of CO2 in Air, Analytical Chemistry, 89, 3648-3655, 10.1021/acs.analchem.6b05063, 2017.

Griffith, D. W. T.: Calibration of isotopologue-specific optical trace gas analysers: A practical guide, Atmos. Meas. Tech. Discuss., 2018, 1-19, 10.5194/amt-2018-187, 2018.

Hall, B. D., Dutton, G. S., Mondeel, D. J., Nance, J. D., Rigby, M., Butler, J. H., Moore, F. L., Hurst, D. F., and Elkins, J. W.: Improving measurements of SF6 for the study of atmospheric transport and emissions, Atmos. Meas. Tech., 4, 2441-2451, 10.5194/amt-4-2441-2011, 2011.

A gas chromatographic system for high-precision quasi-continuous atmospheric measurements of CO2, CH4, N2O, SF6, CO and H2: https://www.researchgate.net/publication/255704060_A_gas_chromatographic_system_for_high-precision_quasi-continuous_atmospheric_measurements_of_CO2_CH4_N2O_SF6_CO_and_H2, 2008.

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., and Derwent, R. G.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, Journal of Geophysical Research: Atmospheres, 116, n/a-n/a, 10.1029/2010JD014763, 2011.

Richardson, S. J., Miles, N. L., Davis, K. J., Lauvaux, T., Martins, D. K., Turnbull, J. C., McKain, K., Sweeney, C., and Cambaliza, M. O. L.: Tower measurement network of in-situ CO2, CH4, and CO in support of the Indianapolis FLUX (INFLUX) Experiment, Elementa-Sci. Anthrop., 5, 14, 10.1525/elementa.140, 2017.

Stanley, K. M., Grant, A., O'Doherty, S., Young, D., Manning, A. J., Stavert, A. R., Spain, T. G., Salameh, P. K., Harth, C. M., Simmonds, P. G., Sturges, W. T., Oram, D. E., and Derwent, R. G.: Greenhouse gas measurements from a UK network of tall towers: technical description and first results, Atmos. Meas. Tech., 11, 1437-1458, 10.5194/amt-11-1437-2018, 2018.

Tans, P. P., Crotwell, A. M., and Thoning, K. W.: Abundances of isotopologues and calibration of CO2 greenhouse gas measurements, Atmos. Meas. Tech., 10, 2669-2685, 10.5194/amt-10-2669-2017, 2017.

van der Laan, S., Neubert, R. E. M., and Meijer, H. A. J.: A single gas chromatograph for accurate atmospheric mixing ratio measurements of CO2, CH4, N2O, SF6 and CO, Atmos. Meas. Tech., 2, 549-559, 10.5194/amt-2-549-2009, 2009.

Young, D.: Development of a Pre-Concentration System for the Determination of Atmospheric Oxygenated Volatile Organic Compounds, Doctor of Philosphy, School of Chemistry, Faculty of Science, University of Bristol, Bristol, 2007.