

Author's response to review 1

This paper by Reum et al., presents a new atmospheric observatory for CO₂ and CH₄ in the Siberian coastline. The presence of this new research infrastructure is a very valuable opportunity for the atmospheric science community, especially taking into account the possible role that carbon stocks in the arctic region (i.e. permafrost melting) can play in the next decades under the current climate change.

Performing such high level quality measurements in remote regions is not an easy task and strong scientific and technical skills are necessary to obtain reliable data with dense time coverage (as needed to perform inversion for GHG emission studies)

The authors describe with good details the experimental set-up adopted for CO₂ and CH₄ measurements as well as methods for data correction and data screening. Very basic analysis of the first months of data are provided.

Even if the methods adopted in this paper are not innovative, I think that the availability of this new station (and related data-sets) is a matter of interest for the atmospheric community.

Personally, I have some concerns about the design of the gas handling system and the data screening. For these reasons, I ask the authors for providing more explanation or details for some specific points (listed in the following) before publication.

SPECIFIC COMMENTS

"2.3 Gas Handling" Any kind of rain guard was mounted on the air inlet?

Yes, the air inlets used for our station come with a rain guard. They are shown pictographically in Fig. 3, and we add “Air inlets *with rain guards*” to the text.

The air flow diagram presents a very complicate system, with a number of connections and valves which increase the possibility of leaks and dead volumes. Even if I do not see anything wrong in this set-up, nonetheless I'm wondering why a so-complex system was adopted.

The basic concept behind this arguably complicated setup was to be less dependent on the rotary valve. More details on this choice are given three comments below.

For this reason, I'm wondering if the authors performed specific leak test on the system. If yes, what kinds of test have been carried out? Are these tests repeated routinely?

Avoiding leakage was a top priority during assembly of the gas handling system. Leak tests have been carried out, but were omitted in the manuscript for the sake of brevity. In the revised manuscript, we include them as part of Sect. 2.3 (“Gas handling”). In short, leak tests were performed by evacuating the gas handling system and observing the pressure increase over several hours. Leak rates determined this way were found to be negligible. During later maintenance visits, only simple breathing tests were performed to avoid opening tubing connections.

By using two flushing pump for each sampling lines you would avoid the complex switching

system downstream of the antiparticulate filters F4-F3.

We use only one flushing pump to minimize power consumption of the measurement system, which is an important consideration at this remote site.

I'm also wondering why you didn't use a rotary valve with more inputs to manage also the ambient air: this would have the advantage of simplifying the system (less possibility of leaks) and use a larger part of sampling circuit for both ambient air and calibration/target gas measurements with clear advantages and effectiveness for application of calibration and evaluation of target gas results.

This setup would in principle be possible and would indeed have the advantages that the reviewer pointed out. However, our setup minimizes the workload put on the Valco valve. This is an advantage because the rotor of the Valco valve is a consumable that, for top performance, has to be exchanged regularly, particularly if switched frequently. This is not a trivial task that only well-trained personnel can perform. One may alternatively exchange the Valco valve completely, but this would introduce additional risk of leaks, as all connections on the Valco would have to be opened. In our setup, the Valco valve switches position about twice per day. In the setup suggested by the reviewer, the Valco valve would, in addition, be used for switching between the two inlet lines from the tower, i.e. six times per hour. For the above reasons, we chose to avoid this workload.

Please provide the residence time of sampling within the system.

The full residence time is on the order of 12 sec. We add this information to the Sect. 2.3 ("Gas handling").

No water traps are used along the ambient inlet lines. In the paragraph 3.2 you mentioned that "longer probing time of the first tank serves to flush residual water out of the tubing". Do you mean that water are present in the tubing? Is this due to condensation or drizzle sampling? In both case this can represent a problem since the presence of liquid water can create artifact in the measurement. Please explain and comment.

This statement refers to residual water *vapor* due to washing out water molecules adhering to the tubing wall. We clarify the sentence to avoid misunderstandings.

"3.1 Water correction"

It is possible to add in the supplementary material more info about the water vapour test? E.g. plot of Concentration(wet)/Concentration(dry) ratios plotted as function of water vapour level for CO₂ and CH₄ or time series of CO₂wet, CH₄wet during the water correction experiment. I'm wondering which is the absolute difference (in ppm and ppb for CO₂ and CH₄) if the "classical" water droplet experiment is used instead the Reum et al. (20q8) procedure. I'm pretty sure that this difference is well lower than the WMO compatibility goals.

First a clarification: we originally cited the discussion paper Reum et al. (2018). In the meantime, the final revised paper was published, so we update the reference throughout this response and in the revised manuscript.

Differences between the water correction methods were documented in Reum et al. (2019), where they were up to 50 % of the WMO inter-laboratory compatibility goal for CO₂ and up to 80 % that goal for CH₄. Thus, the differences were smaller than the WMO goals, as suggested by the reviewer. However, the value of the new method is due to the fact that the WMO goals refer to overall compatibility, which suffers from other errors.

We think that the differences between the methods are documented in enough detail in Reum et al. (2019). However, perhaps we have not communicated clearly enough that data from Ambarchik were used therein: the gas washing bottle experiments in 2015 and 2017 were analyzed in Reum et al. (2019) to evaluate the new method, and the CRDS analyzer in Ambarchik is the one labeled “Picarro #5” in Reum et al. (2019). In the revised manuscript, we add a note to the main text and Appendix B about this correspondence, so that the interested reader can find additional details in Reum et al. (2019). As documented in the manuscript, a direct comparison of the two experimental methods (gas washing bottle vs droplet) with the Ambarchik system may be conflated with drift, because droplet measurements were only done in February and July 2014, whereas the gas washing bottle method was used in 2015. A more direct comparison of the methods based on data from a different Picarro analyzer can be found in Reum et al. (2019).

The requested plot of wet air- over dry air mole fractions is given below (Fig. 1). However, as the reviewer pointed out, the differences between the methods are small and hardly visible in this type of plot. Because of this, and since the gas washing bottle data were presented in more detail in Reum et al. (2019), we chose to present the differences between the experiments with the Ambarchik system as in Fig. 4 in our manuscript. The plot below is not suitable to present the differences between the methods. Therefore, we prefer not to include it in the manuscript.

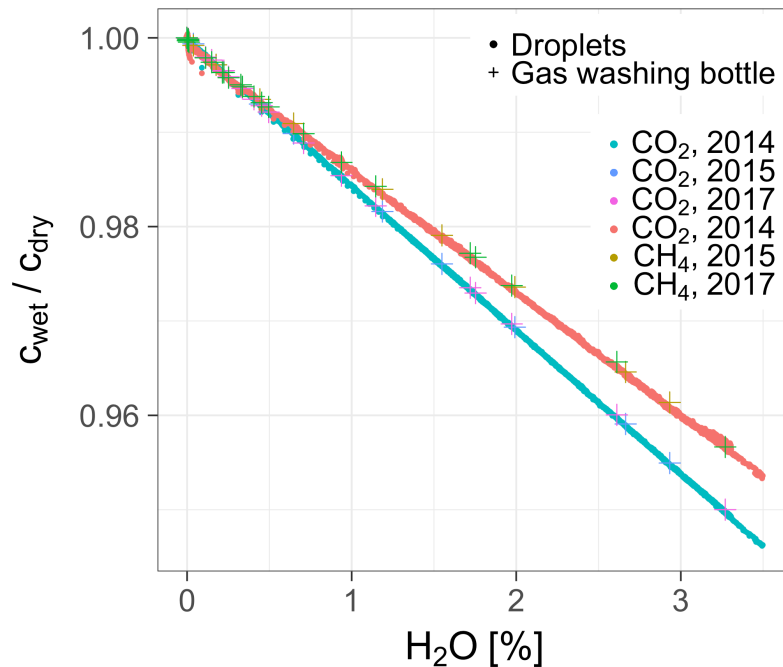
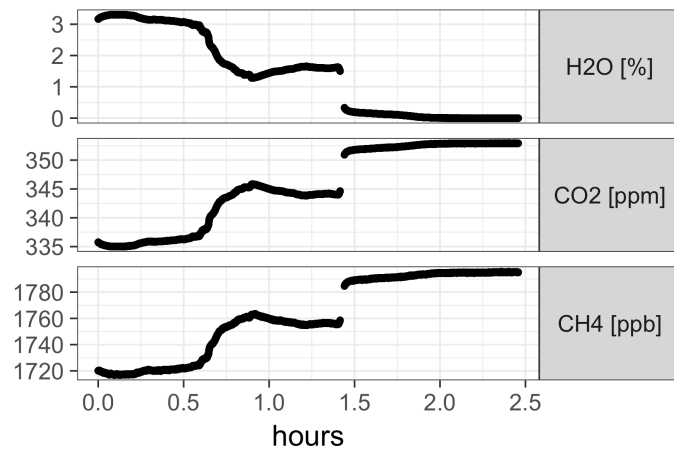
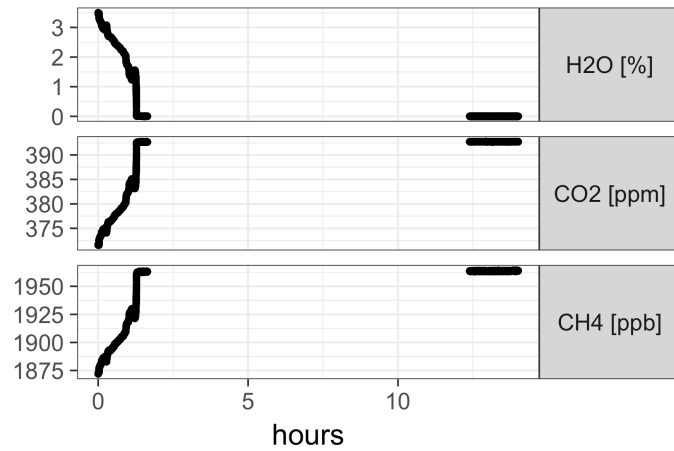
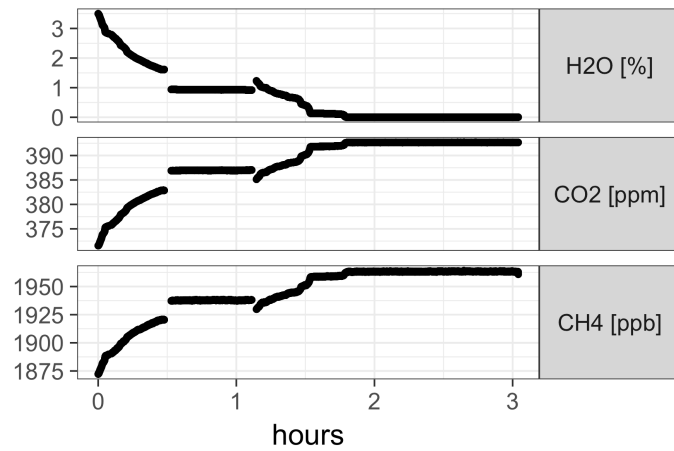


Fig. 1: Wet air mole fractions vs water vapor from the water correction experiments presented in the main text of the manuscript. The data from the three droplet experiments are shown with the same color.

The requested time series are plotted below (Fig. 2). As above, we think that this material is too much to be included in the manuscript.



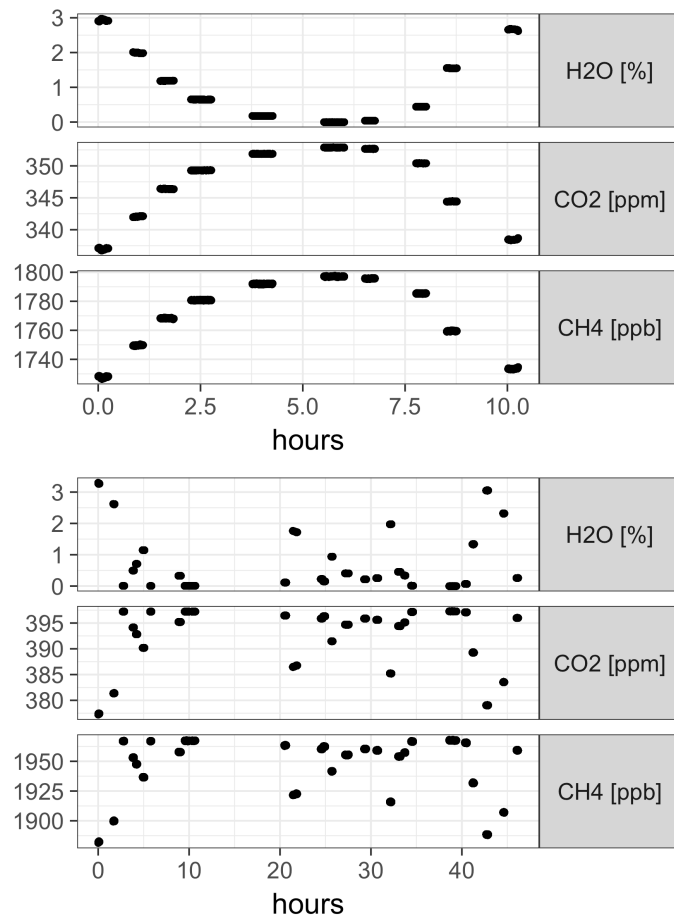


Fig. 2: Time series of water vapor, CO₂ and CH₄ mole fractions during all water correction experiments. Only data used for inferring water correction functions are shown. Panels 1-3: droplet experiments conducted in 2014. Panels 4-5: Gas washing bottle experiments conducted in 2015 and 2017, respectively.

Line 19: Did you apply the same correction for all the data series by considering an average value of the correction coefficients derived during each single experiment? This approach will become unfeasible when many years of measurements will be available, I suppose.

Yes, as documented in the manuscript, an average correction was applied because the data so far do not allow deriving trends in the water correction. For the future, we plan to include the influence of potential trends, but this can only be implemented when more experiments become available. As we see it, this strategy is the most practical solution at this point.

How much would change CO₂ and CH₄ correction if results from single experiments are used?

The variability in the output based on single experiments was shown in Fig. 4 in the manuscript. Differences were up to 0.14 ppm CO₂ and 1.2 ppb CH₄. These maximum differences occur at 3 % H₂O and 2.3 % H₂O for CO₂ and CH₄, respectively. Note that ambient H₂O mole fractions in Ambarchik rarely exceed 1.5 % H₂O. In this domain, the differences between the water correction experiments are smaller.

Figure 4 and Figure 5: the WMO goals are wrong in these figures. They are +/- 0.1 ppm for CO₂

and ± 2 ppb for CH₄.

In the revised manuscript, we adopt the captions by Reum et al. (2019). In short, these should be thought of as the WMO internal reproducibility goals. The WMO inter-laboratory compatibility goals as cited by the reviewer refer to differences between data from different laboratories. However, keeping the accuracy with respect to a common calibration scale, which is what plots depict, within these thresholds does not ensure achieving this goal. Consider CO₂ data from two stations. One has a positive bias of +0.1 ppm CO₂, while the other has a negative bias of -0.1 ppm CO₂ with respect to their common calibration scale. Thus, the bias between these two stations is 0.2 ppm CO₂, exceeding the interlaboratory compatibility goal. However, keeping the accuracy with respect to the calibration scale within half of the WMO inter-laboratory compatibility goal ensures that biases between stations do not exceed them. Therefore, we chose these goals as context for the accuracy of our measurements with respect to the WMO calibration scales (i.e. water correction and calibration, Fig. 4 and 5, respectively). The WMO refers to these goals as the “internal reproducibility goals” (WMO, 2016).

We make this point clearer by adding a short version of this paragraph to the text where Fig. 4 is introduced, and by modifying the labels in the figures from “Range covering WMO goal” to “WMO internal reproducibility goal”.

"3.2 Calibration"

How many measurement cycles are carried out during each calibration event?

One calibration event consists of one cycle of three tank measurements, in order High – Middle – Low. We clarify this in the revised manuscript. Note that data are calibrated based on a weighted average of about 25 such events. Note also that it's important to assess whether the order of gas tank measurements affects the results. We discussed tests with reversed order of the tanks in Appendix E.3 (E.4 in the revised manuscript). Based on these data, we could not rule out the presence of small biases, but determined that their impact on calibrated data would be small. Therefore, we consider our calibration strategy to be adequate. As stated in the Appendix, we nonetheless plan to rule out these small potential errors with additional experiments in the future.

Did you apply any metric to evaluate the success of the calibration (e.g. standard deviation of single injection or data coverage).

In our opinion, the scatter of raw CO₂ and CH₄ data observed during calibrations as well as of the derived fitted coefficients is the best indicator of the precision of the calibration procedure. These data are also manually checked for outliers.

Did you consider stabilization time after starting of the single cylinder injection?

Yes, we investigated whether the gas tank measurements are sufficiently stabilized based on two methods. One was to fit an exponential stabilization function to the average of all measurements of each tank doing this with individual measurements was not robust because the signal was. We then compared the fitted equilibrium mole fractions to the average mole fractions of the last two minutes per tank (i.e., the values that were used to calibrated data). As reported in Appendix E.3 (E.4 in the revised manuscript), the differences were negligible. The other method was to

measure the tanks on site for up to two hours. As reported in Appendix E.3 (E.4 in the revised manuscript), there were small variations of CO₂ and CH₄, but no consistent drifts over the full span of the experiments. Therefore, we think that the worst-case numbers given in the manuscript, which are well below 0.1 ppm CO₂ and 2 ppb CH₄, do not indicate insufficient stabilization. As stated in the Appendix, we nonetheless plan to rule these small potential errors out with additional experiments in the future.

How do you handle the fitting of calibration parameters when discontinuity of data appear (e.g. instrument switch off/on).

The regular operation period of the instrument without switching it on/off is usually very long (several months) since both on-site maintenance and power supply are very reliable under normal conditions. However, particularly during the start-up period of the site, when minor flaws in the setup still needed to be straightened out, the system sometimes needed to be restarted more than once a week. Even then, discontinuities in instrument drift (Fig. 5) were not observed, so we are confident that the record is not affected by discontinuities. Thus, discontinuities like analyzer restarts are not explicitly accounted for.

Can you provide the time series of standard deviation (based on 1-minute averages) of single target measurements (a measure for the CMR) and the time series of the standard deviation over 72 hours of the target gas injection means (LTR)?

We provide the requested quantities and plots here. However, in our opinion, they express similar uncertainties as already accounted for and quantified in our uncertainty estimates. In addition, they reflect neither our calibration strategy nor the averaging strategy (i.e. hourly averages). Therefore, we prefer not to include them in the revised manuscript.

In Yver Kwok et al. (2015), CMR and LTR were defined as average values, not as timeseries. Here, we first provide values according to those definitions, and then time series as requested by the reviewer. We did not perform measurements that follow precisely the protocol in Yver Kwok et al. (2015). However, a similar value as LTR is already incorporated in our uncertainty estimates: as stated in Appendix E, analyzer drift and precision ($\sqrt{u_p^2 + u_b^2}$) were jointly estimated from a dry air measurement in the lab prior to field deployment that lasted 12 days. More precisely, analyzer drift and precision of hourly values were calculated as the standard deviation of hourly averages over this 12-day measurement and was reported in the manuscript (0.013 ppm CO₂, 0.25 ppb CH₄). This is somewhat similar to the LTR by Yver Kwok et al. (2015), which was defined as the standard deviation of 10-min averages over a period of 3 days. Alternatively, one may use the regular target gas injections for computing a measure for LTR. To obtain a meaningful standard deviation, more than 72h have to be considered, since measurements take place only every 29h. Therefore, we show the timeseries of the standard deviation over 13 days, which corresponds to 11 regular Target tank measurements. We use the averages of the final 2 min of each target tank measurement that were used to calibrate data (Fig. 3). The average LTR estimated this way is 0.018 ppm CO₂ and 0.25 ppb CH₄.

From the same 12-day measurement used to estimate analyzer drift and precision, CMR may be calculated as standard deviation of raw data within 1-minute intervals. In 30-h intervals, they were on average 0.015 ppm CO₂ and 0.17 ppb CH₄. This is shown in Fig. 4 for one 30-h interval.

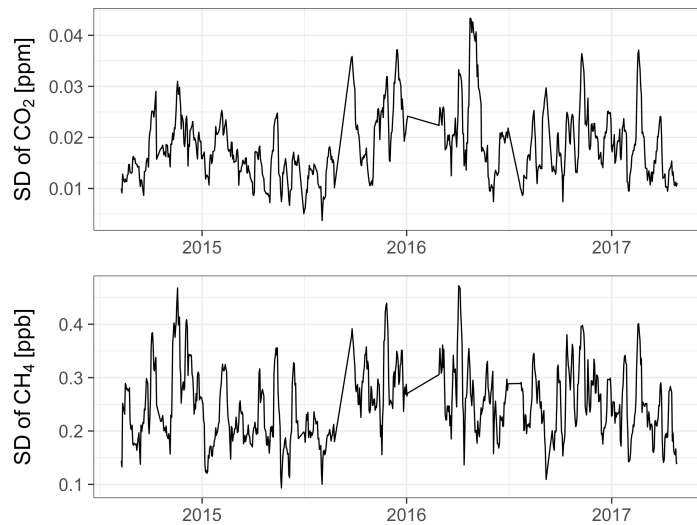


Fig. 3: Running standard deviations of Target tank data averaged over 13 days (used for estimating LTR).

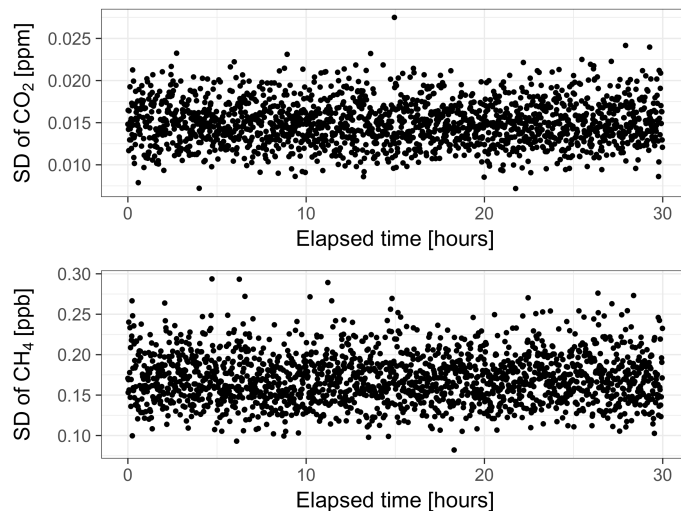


Fig. 4: Timeseries of standard deviation of 1-min averages of a dry gas measurement over a 30 hour period (used for estimating CMR). The experiment was performed prior to field deployment.

Fig. C1: please express CH4 in ppb.

Ok.

The spread of intercept looks pretty high (for both CO2 and CH4). Please can you provide the time series of measurement results (expressed as average value of CO2 and CH4) for each single tank during the calibration events?

The spread of the intercept represents the uncertainty for theoretical measurement values of 0 ppm CO2 and 0 ppb CH4. The uncertainty in the range covered by the calibration tanks (and

thus, typical ambient mole fractions) is much smaller and quantified by se_{fit} , which, if computed for individual calibration events, would on average be 0.047 ppm CO₂ and 0.11 ppb CH₄. For more details, see also the difference between σ_b and $\sigma_{b,min}$ explained in Andrews et al. (2014). The requested plot is shown in Fig. 5 below.

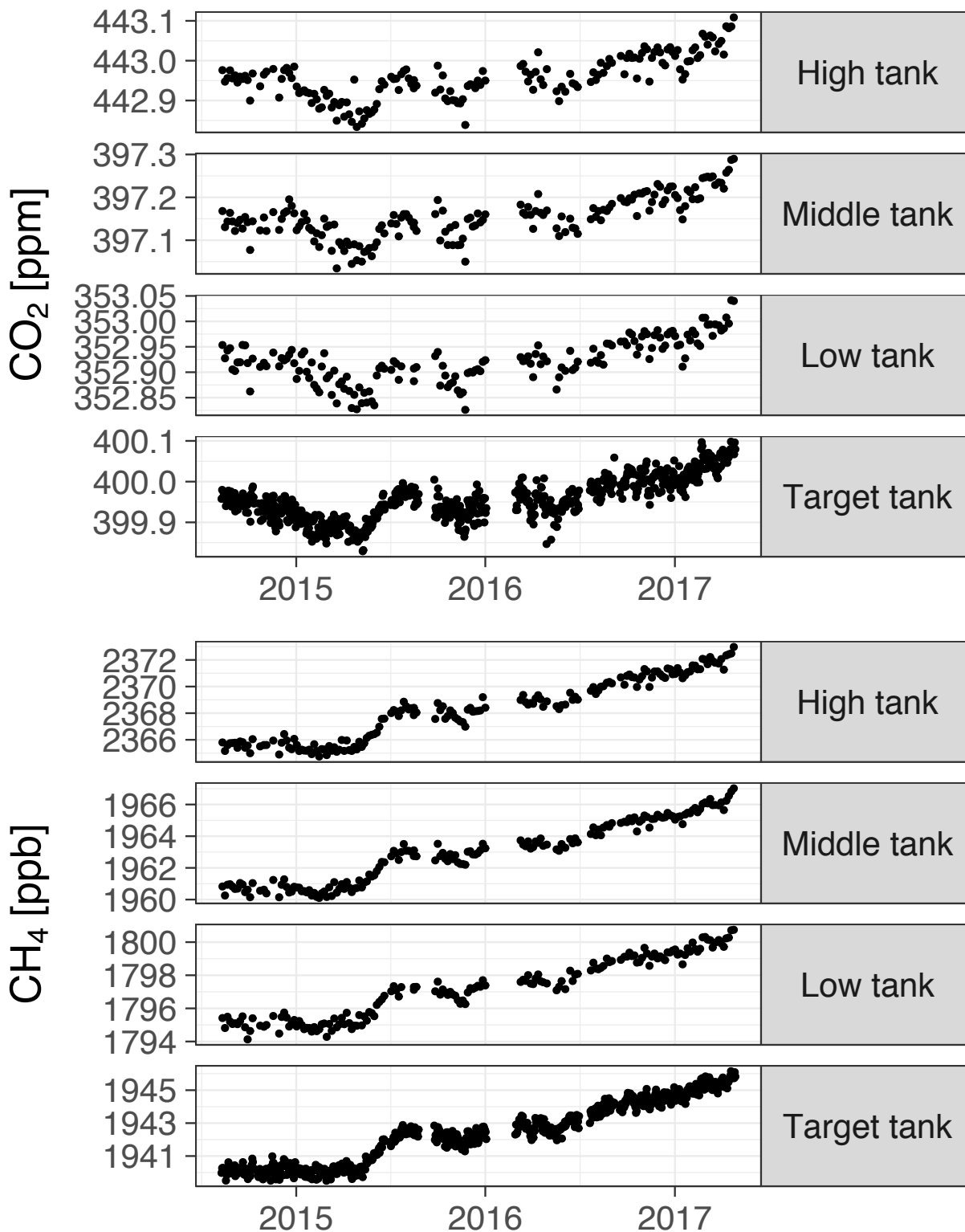


Fig. 5: Timeseries of CO₂ and CH₄ values from gas tank measurements. Shown here are the last two minutes of each gas tank measurement, i.e. the values that were used to derive calibration coefficients.

Which is the typical value of H₂O during the calibration for each tank?

The residual H₂O mole fraction is well below 0.01 %. We add this information to the main text.

"3.4 Data screening"

It is not clear if this check are performed automatically or manually. Please provide more details about the screening procedures here adopted (e.g. threshold values, which kind of air pollutants are considered,. . .)

Erroneous measurements (Sect. 3.4.1, 3.4.2, 3.4.4) are removed automatically. For other procedures (listed in Table 1), flagged data are supplied and it is up to the user to use them as they see fit for their application. We clarify this in the introduction to Sect. 3.4. The thresholds given for these parameters were merely examples used in this paper, and it is up to the user to choose thresholds best suited for their application.

-3.4.1 Analyser status diagnostic I'm rather surprised that the OUTLETVALVE parameter is not mentioned in this screening. In my experience this is a pivotal parameters to check the presence of obstruction (e.g. filter) in the system.

We monitor obstructions and other problems based on the flow meters and the pressure sensor in the sampling system. The outlet valve provides redundant information.

-3.4.2 Flushing of the measurement 30 sec is not sufficient as stabilization time. I think that a few minutes is more suitable.

To avoid misunderstandings, the 30 seconds were only used for switching between ambient air inlets. After switching from tanks to ambient air inlets, 5 minutes of flushing were used to account for the much larger differences in CO₂, CH₄ and H₂O mole fractions between gas tank and ambient measurements. We agree with the reviewer that longer flushing of the lines reduces chances of cross-contamination between air sources. However, the purpose of switching between the two inlets is to be able to filter for situations where differences between the two inlets are large. In cases when the differences are large, the timing of trace gas changes when switching between inlets can be observed and is roughly as follows (an example is given in Fig. 6): after ~8 seconds, recorded values start changing rapidly (i.e. this is the residence time between V1 and Picarro cavity). This rapid change lasts for roughly ~10 seconds. Afterwards, changes are small. Certainly not all cross-influence will be removed by then, but since our intention is to detect cases of large differences between the inlets for the purpose of filtering them out, we think that 30 seconds of flushing are sufficient.

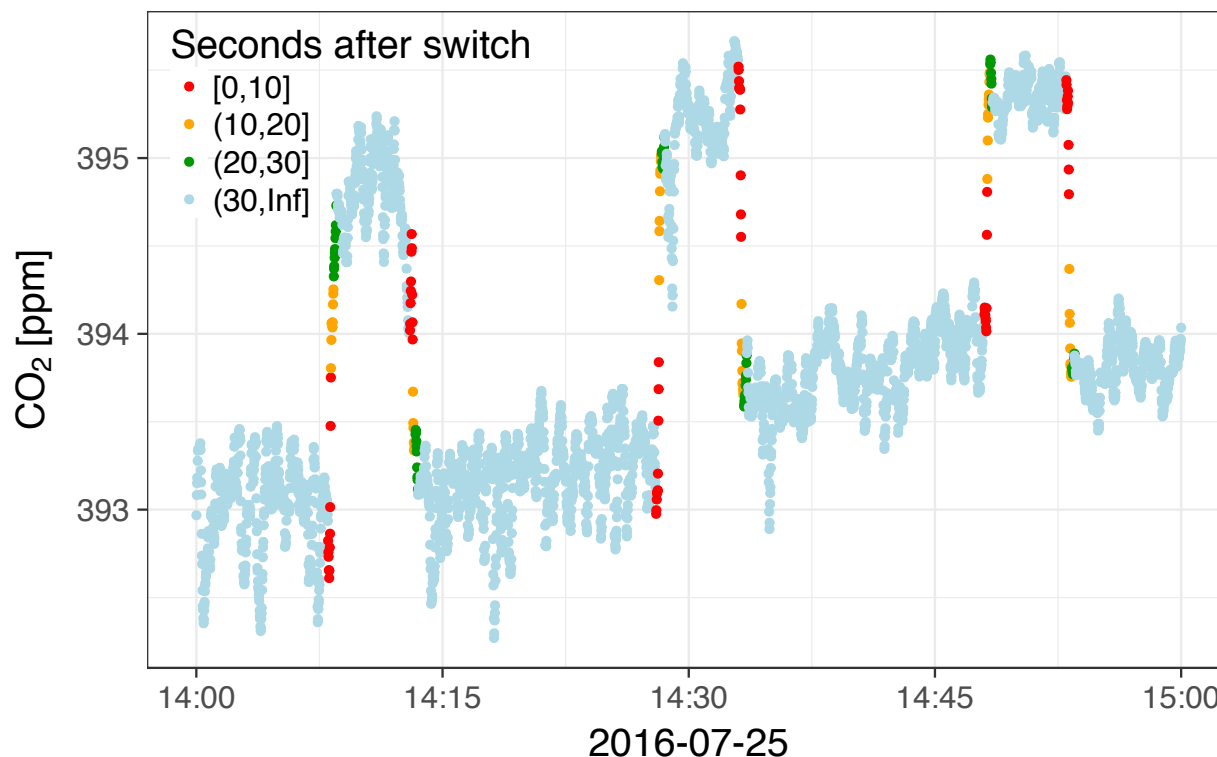


Fig. 6: An example of ambient data of both air inlets illustrating the flushing of the tubing common to both tower inlets. For clarity, we chose an example where differences between the inlets were large compared to the intra-hour variability of each inlet. Colors indicate the time that passed after the inlet was switched by valve V1. Data are used for analysis starting 30 seconds after switching, i.e. the data points in light blue.

-3.4.3 Contamination from local polluters I think that also CH₄ need a proper spike detection. What about biological waste management of the base?

We agree that our approach may, in rare cases, lead to CH₄ spikes remaining undetected. As described in the manuscript, the major pollutant sources to consider are emissions by the meteorological station itself. Contamination due to waste may be possible in summer only because during other seasons it is mostly frozen. However, the application of the CO₂ filter to CH₄ data was based on the assessment that, if CH₄ spikes were present, they often coincided with CO₂ spikes. At the same time, large CO₂ spikes are much more frequent than large CH₄ spikes at Ambarchik. Thus, given the small impact of the CO₂ filter (Table 3 in the manuscript), we think that contamination of the CH₄ signal independent from CO₂ is a negligible source of bias. Furthermore, every filter based on signal variability is somewhat subjective and bears the risk of removing natural signals. This is particularly problematic in the case of CH₄ due to the high variability of its natural emissions. Therefore, we believe that, in case of the Ambarchik station, separate spike detection for CH₄ rather holds the potential to reduce the data quality, and we decided that a common filter based on the CO₂ time series works best. Note also that CH₄ contamination from local sources may be filtered out by other criteria made available to the user. In particular, intra-hour variability is directly affected by potential undetected spikes, and independently provided for CO₂ and CH₄.

We include these motivations in Sect. 3.4.3 of the revised manuscript.

Table 3: the statistic is referring to all data or the 1-4 PM selection? Looking to Fig. 6, it seems that WD has a strong seasonal variability. How the fraction of flagged data is shared among the different months of the year?

Table 3 refers to all data (we add that to the caption of the table). The annual variation of the impact of the spike filter depends on whether filters are applied (Fig. 7 below).

Here, we present seasonal variations of the fraction of flagged data. We consider the case of applying the temperature gradient (“T”) and wind speed (“wv”) filter to remove the seasonally varying impact of temperature inversions and wind speed variations. The fraction of affected data follows indeed the wind speed pattern (Fig. 7): more data are affected during the period March–September than October–February. This roughly corresponds to the period when the prevailing wind direction is Northeast, where the inhabited building, and thus contamination sources are located.

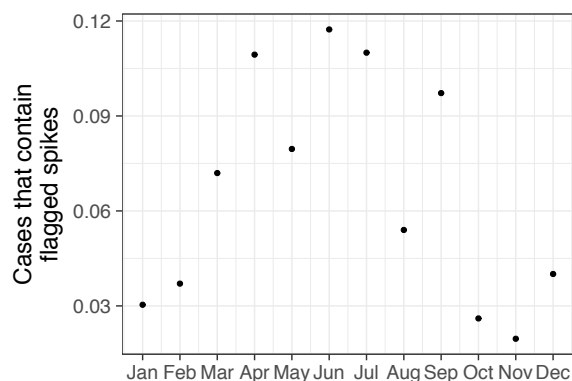


Fig. 7: Fraction of data affected by the CO2 spike flagging. Shown here are data that pass the temperature gradient and wind speed filter from Table 3 of the manuscript.

I would suggest to implement as soon as possible measurements for the monitoring of pollution emissions (CO, NOx or aerosol particle) to consolidate the detection of local pollution influence.

The authors agree with the reviewer that monitoring pollutants continuously would aid the quality control of the greenhouse gas data. Monitoring additional gas species would also significantly enhance the impact of the Ambarchik station for the pan-Arctic GHG monitoring network. Unfortunately, additional continuous sampling is currently not available due to lack of funding. However, we originally planned to extend the observing system in Ambarchik by an ICOS-style automated flask sampler in summer 2019. Due to customs problems, we had to postpone this installation until summer 2020. Drawing air samples at regular intervals, but also considering the targeted sampling of specific emission events, including spikes, this new data source will allow in-depth interpretations of the existing, continuous monitoring program presented within this manuscript.

"Section 5.1."

Line 12. Please do not use “trend” for this short time period. Use “tendencies” or (when appropriate) “growth rate” (the same for CH4)

We replace the term “trend” by “average growth rate over the analyzed period”.

Line 26: more than this “trend” along the whole measurement period, a discussion of the annual growth rates could be more interesting.

We attempted to quantify annual growth rates based on averaging springtime measurements, but the results depended on the averaging period. The curve fitting procedure applied here, consisting of 4 harmonics plus a linear trend, may also be used to infer annual growth rates, but in light of the short data coverage period, the present analysis was chosen because it appeared most robust.

I do not agree that the attribution of the very high values of CO₂ in December 2016 are outside the scope of the paper. They can indicate analytical/experimental problems or interesting phenomena can be investigated at the station. I strongly suggest to provide some sounding explanations.

We are highly confident that this signal is not due to measurement errors, because it was also observed by the gas analyzers of an eddy-covariance station operated by MPI-BGC near Chersky, approximately 100 km to the south of Ambarchik. Even though these analyzers (LosGatos FGGR) are not as well calibrated as the Picarro instrument at Ambarchik, their data quality is clearly good enough to observe such a pronounced signal.

At the same time, it is obvious that the detailed interpretation of such a signal, including an attribution to specific emission processes and/or source regions, would require an extended analysis that is clearly beyond the scope of the presented paper. With the signal being detected also at other sites, it is clear that we either see a large-scale anomaly in surface-to-atmosphere emissions, or the effect of an unusual atmospheric transport pattern. To differentiate between both, data from many more sites within the Arctic domain would be required, including also a reliable dataset on the variability of the background signal entering this domain. All of this must therefore be referred to a follow-up paper.

"Section 5.2"

This section is really basic. The related goal is not clear to me as well as the method for deriving the background values of CO₂ and CH₄. Please explain better. No explanation or discussion are provided for the results from wind analysis in section 5.2.1 and 5.2.2.

The analyses shown in Section 5.2 provide hints on source/sink regions for the signals detected at Ambarchik. This is of particular interest at Ambarchik because it is located at a junction of several different ecoregions (e.g. land/ocean). Thus, to first order, we expect that differences in the signals by wind direction hint at differences between these regions. To clarify our intention, we add this motivation to Sect. 5.2. The results demonstrate that there is indeed an angular dependence in the observations that hint at terrestrial regions, as opposed to the ocean, as the dominant contributor to regional CO₂ and CH₄ anomalies captured at Ambarchik. This demonstrates the value of sampling at this location for insights into regional carbon cycle processes. We add this consideration to the conclusions section.

Our computation of background values follows a standard method in regional inverse modeling of atmospheric tracer transport, meaning that the background corresponds to the contribution of

CO₂/CH₄ transported into the examined domain. By subtracting this signal from the observations, only the signature of sources and sinks inside the domain remains. We reformulate the section to clarify this procedure.

A conclusion drawn from results presented in Sect. 5.2 was already given in Sect. 6, i.e. that larger CO₂ and CH₄ signals appear to be of terrestrial rather than oceanic origin. In the revised manuscript, we slightly expand this by highlighting the added value of the unique station location as a reference to the newly added motivation of Sect. 5.2.

"Section 6"

Line 14. I do not think that the WMO compatibility goal and your total uncertainty can be directly compared. Instead, the “compatibility goal” is not (better: may not be) your achievable total uncertainty but a specific value within which your measurements must agree (see GAW Report No. 206).

Correct. We delete the sentence in question.