Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-325-RC2, 2018

Atmospheric Measurement Techniques Discussions
Open Access

Interactive comment on "Accurate measurements of atmospheric carbon dioxide and methane mole fractions at the Siberian coastal site Ambarchik" *by* Friedemann Reum et al.

Anonymous Referee #2

Received and published: 26 November 2018

The paper describes a unique and valuable dataset collected in a harsh environment in an under-sampled region. These data will be valuable for inverse modeling to estimate emissions and removals of CO2 and CH4. Arctic data such as these are particularly needed, since release of carbon from permafrost is an expected outcome from warming temperatures and current estimates of Arctic fluxes vary widely. The authors provide a useful and complete description of challenges of operating in the Arctic and their strategies for maintaining continuous operations and filtering data to remove local effects. The description of the configuration is comprehensive and clear. The authors have provided quantitative and time-varying uncertainty estimates and a clear description of how the uncertainty was estimated.

Printer-friendly version

Discussion paper

A concern is that the data is available "on request" rather than readily available for download (e.g. from the WMO Global Atmosphere Watch World Data Center for Greenhouse Gases or these data could be included in the GLOBALVIEW+ ObsPack product compiled by NOAA). The value of these data will only be realized when combined with other datasets from the global community.

Also, the spike detection algorithm seems to be highly tuned and somewhat arbitrary (but to be fair data from many sites are manually flagged, which relies on expert judgment that is arguably even more arbitrary). Please see specific comments about making the flagging criteria explicitly available so that users have enough information to develop their own filtering scheme.

~~~~~~

Review Criteria for AMT:

~~~~~~

Does the paper address relevant scientific questions within the scope of AMT?
yes
Does the paper present novel concepts, ideas, tools, or data?
yes the data from this new Arctic site are novel and uniquely valuable for tracking possible release of CO2 or CH4 from permafrost
Are substantial conclusions reached?
yes in the sense that 2+ years of data are presented along with an assessment of enhancements over background presented versus wind direction and season
Are the scientific methods and assumptions valid and clearly outlined?
yes
Are the results sufficient to support the interpretations and conclusions?
yes
Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

yes with some minor requests for clarification below
Do the authors give proper credit to related work and clearly indicate their own new/original contribution?
yes
Does the title clearly reflect the contents of the paper?
yes
Does the abstract provide a concise and complete summary?
yes
Is the overall presentation well structured and clear?
yes
Is the language fluent and precise?
yes
Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?
yes
Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
no, the paper is of appropriate length and detail
Are the number and quality of references appropriate?
yes
Is the amount and quality of supplementary material appropriate?
yes

~~~~~~

Specific Comments:

~~~~~~

page 6 what is the flow rate through the analyzer and what is the purge flow rate?

page 9, line 25: Is there any indication if the time synching with the GPS fails?

page 10 line 20: State that "synthesis" function is defined in Appendix B.

page 10 lin3 16: Variability of water correction experiments discussed by Stavert et al., AMTD, 2018 (https://www.atmos-meas-tech-discuss.net/amt-2018-140/) and could be referenced here. They found that short-term repeatability of water corrections was similar to long-term repeatability.

page 12: what is the expected lifetime of each calibration cylinder?

page 13: it would be useful to describe the stochastic and non-random components of the estimated measurement uncertainty (i.e. to what extent does the uncertainty improve with averaging). The text states that the uncertainty is dominated by the water correction, which is not going to improve with averaging. But perhaps also include a statement about the short-term precision of the analyzer for each gas (i.e. what is the standard deviation on each 10-minute calibration after the gas has equilibrated).What is the typical standard error of the residuals?

page 16: description of data filtering algorithms is useful and the results shown in Table 3 demonstrate that impact is practically negligible.

page 16: description of water vapor spikes is interesting, and the explanation seems plausible

page 17: it would be useful to see how the virtual potential temperature threshold corresponds to other indicators of difficult-to-model observations. For example, are hourly standard deviations typically higher than during well-mixed conditions? What is

Printer-friendly version

Discussion paper

the duration of a typical inversion (i.e. how many consecutive hours of data are typically flagged)? Can these events be reliably screened based on something like enhancement above a smoothed background? This type of information could be helpful for developing filters for other sites (particularly Arctic sites) where virtual potential temperature information is lacking.

page 18, line 5: what is the duration of the back trajectories (i.e how many hours or days backward in time)?

page 20, line 6: How are Barrow data selected for this comparison. State clearly that you are including Barrow data that has not received a first column flag if that is the case. Can you speculate about why the virtual potential temperature filter would remove such a large fraction of the data at Barrow? Is there some obvious difference in the meteorological conditions at the two sites? Does this result have implications for interpreting the Barrow data?

page 15: regarding amplitude estimation, maybe it would be better to use the curve including residuals and then estimate the amplitude based on the difference between the min max smooth curve values (and you could just compute the average for all the consecutive min-max or max-min pairs). Then you could do same with to ensure apples to apples comparison. Otherwise when you compare Barrow and Ambarchik are the amplitudes different because of different time periods?

page 20, line 22: are you sure that the smaller variability at Barrow was real and not due to differences in screening for the two sites?

page 22: Were the trajectory endpoints the actual endpoints for the entire Arctic WRF domain? Or did you define a subdomain? It would be useful to provide some information about the locations of the endpoints (such as vertical and lat lon distributions by season or for some typical examples).

page 24 line 16: instead of "exceeded the goal" perhaps say "did not meet the goal" (although I am not sure the uncertainty estimate is accurate to 0.01 ppm, so maybe you could say instead something like "meeting the goal to within our ability to estimate the uncertainty). Certainly you are doing as well as any other group in the world, and better than most at documenting the uncertainties.

page 29, line 6: differences among sequential individual co2 measurements?

page 29, line 11: it's not clear how "cases when all CO_2 data in the interval have rather uniform variations" are identified so that they can be unflagged

page 29, step 3: why is it not desirable to also flag short-duration spikes? Couldn't these originate from a very local source, such as a generator?

page 30, line 2: why choose a threshold of 8 std deviations? this seems arbitrary

page 31, Figure D.1: This figure shows the utility of using an algorithm to remove spikes and it does seem to work reasonably well for this case. But the complexity of the strategy is concerning. When the data is distributed, it would be best if the flagging for spike-detection is reported separately from other types of flagging (e.g. flagging after transitions, flagging for maintenance) so that the end user can consider alternative strategies.

page 32, E.1 It would be useful to describe the Allan variance of the analyzer and to distinguish between random error that reduces with averaging versus uncertainties that result from systematic errors that cannot be reduced by averaging. Specifically, if laboratory tests or field calibration data can be used to estimate the random

component at the native frequency of the measurement and for hourly averages, then that would allow the user to determine when atmospheric variability exceeds the random noise of the Picarro analyzer. This can help with data selection and weighting in inverse modeling. See the discussion of "sensor precision and atmospheric variability" in the recently released GGMT 2017 meeting report (GAW Report 242). A related question is whether the standard error of the fit takes into account the 120 day smoothing of the coefficients. For a simple case with a uniform (boxcar) 120 day weighting, there would be approximately 24 separate calibration episodes = 70 degrees of freedom. The standard error is substantially reduced compared to a single calibration episode. An example with realistic values and errors is given in the attachment (coded in R) and improvement in the fit coef uncertainties and the overall residual standard error of the fit is evident when multiple calibrations are combined. Here I neglected noise on the assigned values. It should be straightforward to adapt the equations from Andrews et al., 2014 Appendix D to account for the tricubic kernel weighting if that method is demonstrably superior to simple boxcar smoothing. And/or you could use the "residual standard error" of the fit to find the optimal averaging window and weighting strategy. The "sigma prime y" term in E.4 will also be affected by analyzer noise, and may be smaller for an hourly average value than for a single calibration episode. In any case, it is important to describ
the random error characteristics of the analyzer and the individual calibration episodes.

Please also note the supplement to this comment:
https://www.atmos-meas-tech-discuss.net/amt-2018-325/amt-2018-325-RC2-supplement.pdf

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-325, 2018.