Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-402-AC1, 2020

Atmospheric Measurement Techniques
Discussions

Interactive comment on "Long-term reliability of the Figaro TGS 2600 solid-state methane sensor under low Arctic conditions at Toolik lake, Alaska" *by* Werner Eugster et al.

Werner Eugster et al.

werner.eugster@usys.ethz.ch

Received and published: 16 February 2020

Response to Reviewer #1

Reviewer feedback in copied in with black text, our response, how we plan to revise our manuscript, is given in indented blue text.

1 General comments

The manuscript summarises the findings from a 6-year field deployment of a small, low-cost methane sensor under low Arctic conditions. Given the interest of the community in small, low-cost sensors on the one hand and the measurement of atmospheric methane on the other hand, this work is of high relevance to the readers of Atmospheric Measurement Techniques.

The manuscript is written and structured well. The methods are described in appropriate detail. However, I see some shortcomings in the data analysis and presentation of results, detailed below, that should be addressed before publication.

> Thank you for your very valuable and detailed assessment, which definitely helps us improving our paper.

2 Specific comments

It is unclear which quantity the Authors use when they report the abundance of methane. The sentence 'We report all gas concentrations in mixing ratios by volume (ppm or ppb dry mole fractions)' (ll. 78–79) is contradictory, as it uses three distinct quantities as if they were synonymous. In a mixture of two components A and B, concentration c is defined as $c = qA/(VA + VB)$, where V is volume and q one of the quantities mass, amount, volume or number concentration [1]. Mixing ratio by volume r is rather uncommon and defined as $r = VA/VB$ [2]. Mole fraction x (IUPAC recommends y for gaseous mixtures, but this is not common in atmospheric science) finally is defined as $x = nA/(nA + nB)$, where n is amount of substance [3]. Given that the WMO scale for methane abundance is a mole fraction scale [4], the reporting of mole fractions would be desirable. While the use of the term ‘concentration’ for mole fraction is accepted for communication with the general public [4], a publication in a scientific journal should in my eyes favour exact terminology. In any case the Authors must make clear which quantity is reported.

> We apologize for the units confusion. We measured the dry mole fraction of methane (confirmed with our associate Dr. Martin Steinbacher from the Swiss Federal Labs for Materials Science and Technology, Empa), and we have replaced all other terms used to express methane units in the manuscript with “mole fraction” (defined in the first instance as dry mole fraction). We also now define ppm explicitly as μmol mol^{-1} for clarity.

Two TGS 2600 sensors were deployed at the site, referenced to as #1 and #2. However, in several instances in the manuscript ‘TGS 2600’ appears without a number when I think TGS 2600 #1 is meant. Also, ‘sensor’, ‘TGS’ and ‘TGS2600’ are used. This should be made more consistent. Results from TGS 2600 #2 are presented exclusively in l. 104, ll. 185–186 and Fig. 9. Explaining the minor role of TGS 2600 #2 around l. 59, l. 104 or l. 162 might prevent confusion of the reader.

> The paper is revised to be more explicit about the minor role of TGS 2600 #2. With only one sensor in place, an acceptable critique would have been that one sensor could be far off in its readings. Hence the second sensor #2 was only there for redundancy, and using this sensor actually showed that sensor #1 on which our analysis is primarily based, could be backed up by another (randomly selected) sensor of the same type.

ll. 74–77 How often and to which scale were the reference analysers calibrated?

> Note that this is not a GAW reference station, but a seasonal research camp where we were able to check the LGR every summer season. We’ll add this information and the numbers of the NOAA reference gas bottles we used. Before we had NOAA reference gases available it turned out that the accuracy of LGR was better than the specification of the available reference gases for calibration of the research station.

l. 153 ‘[...] relative humidity (which is a ratio and not a physical variable of atmospheric water content)’ – I think I have an idea of what the authors mean, but I find the wording not quite right. Would the authors say that the refractive index of a material is not a physical variable because it is a ratio? In general, I miss some thoughts about the temperature dependence of the quantity used for expressing humidity. The Figaro TGS 2600 has a heated sensing element, so the relative humidity at the sensing surface is different from the relative humidity in the environment. The temperature dependence of relative humidity makes this quantity a less than ideal choice for this type of correction. Unfortunately, both alternative quantities chosen by the authors to express water vapour content depend on temperature as well. Mixing ratio (by mass) or specific humidity would be temperature-independent alternatives [6]. Using the ideal gas equation, the terms in Eq. 2 that contain the product Ta Âův would also cancel out the temperature dependence of absolute humidity if Ta was absolute temperature (in K) – but in the manuscript a Celsius temperature is used. Hence, my suggestion to the authors is to try out either mixing ratio by mass or specific humidity as an independent variable in Eq. 2. Using absolute instead of Celsius temperature might be advisable as well.

> We see the point, but one challenge is that such low-cost sensors are closer to a black-box system than to a fully controllable physical sensor. In addition, the manufacturer does not reveal all details of the sensor operation. In a conversation with Dr. Nick Martin, National Physics Laboratories, Lon-

> don, we learned that his lab has also explored the validity of using water concentration instead of relative humidity – in their laboratory tests they found an almost linear relationship between the two. We tried absolute humidity, vapor pressure, and relative humidity in our calculations, but absolute humidity was always performing slightly better than the other two. We considered specific humidity and mixing ratio to be too close to absolute humidity and thus did not test this so far. Further tests could be run, but our approach is to adopt the same method as was used by previous authors (Collier-Oxandale et al., 2018; lines 258–260 in the discussion paper version).
>
> We also understand the point that if we mix units of K or °C in the various equations (due to how the temperature dependence is expressed), we might introduce a constant bias (or at best just make the units messy). We can empirically test the question of whether temperature in K or °C gives the better results in our prediction models, and select the one that performs better for the revised version.

ll. 161–162 Using the entire dataset for estimating the parameters of Eq. 2 is a comprehensive test of how well the model can describe the dataset, but is of limited relevance for field deployments where calibrations are performed during limited periods of time and the main interest is in the uncertainty of independent measurements. For this reason, splitting the dataset into a calibration and a validation part yields important insights. The caption of Table 1 explains that the authors have in fact performed analyses of a split data set. This fact should also be mentioned in the main text around the lines given.

> This point was also addressed by Reviewer 2. We will thus recalculate the ANN with the same selection for training vs. validation as we used for the linear model.

l. 167 For the reasons given before, the results presented in the columns 'Linear Model - Calibration' and 'Linear Model - Validation' in Table 1 should be discussed here, even more so because the results for the validation period are substantially worse than for the period used for calibration.

> This will be done.

ll. 193–194 Is there any conclusion that can be drawn from this finding of a -1:1 relationship?

> The conclusion should be that both variability and stability can be improved at the same time because there is no tradeoff visible in Fig. 11. We will add text about this in the revised manuscript version.

Sect. 3.4 The discussion in Sect. 3.1 leans heavily on the coefficient of determination R2. In Table 1, each R2 for the ANN approach is higher than the corresponding R2 for the Linear Model. Considering just the validation period, the ANN approach outperforms the Linear Model by a factor of 3–10 by this measure. Similarly, in Fig. 6 and 8 the ANN approach outperforms the Linear Model (comparing R2 of 'ANN' and 'c/v', 'all' is irrelevant in this respect); in Fig. 5 and 7 they perform nearly equally well. None of these comparisons is made here. Instead, the authors state that the root mean square error (RMSE) does not improve substantially with the ANN approach. While I generally appreciate the reporting of RMSE together with R2, its interpretation here is questionable.

> The reviewer is right, R^2 in an ANN is not the same as R2 in a linear model and thus comparing the numbers may be misleading. We will revise this and provide a goodness-of-fit statistic that is comparable between a linear

> model and ANN. RMSE is of course the better comparison because it has the same meaning in cases where either linear or nonlinear fits are used and evaluated.

On the one hand, RMSE is reported in Table 1 with one significant digit only, potentially masking up to ∼30% differences for an RMSE of 0.03 ppm (0.025 ppm vs. 0.0349 ppm). On the other hand, the RMSE should be seen in the context of the variability of the data, specifically the root mean square difference between the reference measurements and their mean value over the whole dataset, which is not stated. Overall, the discussion in this section appears negatively biased with regard to the ANN approach. This also manifests in the last paragraph of this section, where 'understand[ing] the physical response of TGS sensors' is prioritised over 'technically nicer fits to data', a stark contrast to the lack of a physical interpretation of the terms in the empirical model (Eq. 2). Section 3.4 must be revised to reach the level of neutrality expected from a scientific publication.

> Thank you for these suggestions; we will remove the wording used in the conclusion about neural networks, and rethink how we can best compare the two models both with unbiased estimators of their fit to the data and with their usefulness in interpreting the data in mind. We will also adjust the significant digits of the measures we use (e.g., RMSE) to make them more comparable and useful.

ll. 229–231 To make such an argument, the reader must be informed about the amplitude of all input variables, especially SC.

> In light of the comments above we will reword or delete this statement.

l. 237 I might be mistaken, but as far as I understand the term homoscedasticity it would in this case mean that the variance of the deviation in CH4 abundance is the same for every temperature bin. The authors do not report variances, but both interquartile range and 95% confidence interval suggest that the variance is higher at low temperatures than at high temperatures, i.e. heteroscedasticity.

> What we wanted to say is that the variances (shown with the 95% CI) are quite constant over a wide range of temperatures and absolute humidity. We of course agree that if one focuses at the difference between one bin and another there are many comparisons where the statement is not 100% correct, even when the Bonferroni correction for multiple testing is considered. In comparisons that span a difference of more than 10–15 °C, the within-bin variances are of course not always homoscedasticity. We will therefore reword our description of this point to avoid this confusion.

ll. 253–257 'laboratory conditions simplify the real world too much' – What could be the simplification that makes laboratory calibrations problematic? The input variables used in the empirical model (Eq. 2) can – practical difficulties taken aside – be controlled in the lab. Any other variable that might prevent transfer of lab results to field conditions is not included in the empirical model, so the problem would not be a simplification of the lab environment but a model deficiency. The last sentence of the paragraph seems to go in this direction ('relevant factors'), but is unclear. Please explain better or leave out.

> This statement also caught the attention of Reviewer #2, and given both reviewer comments we will revise the text. As the reviewer says, there are several ways in which laboratory and real world conditions apply to sensor calibrations. For example, in our revision we will emphasize that testing in the temperature and moisture conditions that exist in the range

> from –40 °C to 0 °C is not easy in a laboratory environment. In addition, we have discussed this general point with Dr. Nick Martin, National Physics Laboratories, London, and will adopt some of his perspective and advice in our revised text.

ll. 258–261 Suggesting to move the first sentence to l. 150 and to remove the other one (repetition).

> This will be done.

Fig. 5–8 The graphs are squeezed in horizontal direction, making comparisons between the lines difficult. A shorter period, e.g. 14 days, would give more insight.

> We agree that the graphs are squeezed, but we are concerned that expanding them horizontally would invite comparisons over shorter time periods that are not the focus of the study. For example, on lines 179–181 we wrote "The TGS 2600 is not expected to provide short-term accuracy comparable to high-quality instrumentation (see also Lewis et al., 2018). However, Eugster and Kling (2012) argued that such measurements still may provide additional insights as compared to the passive samplers described by Godbout et al. (2006a, b), integrating over longer time frames". Given that, we tried to draw the attention of the reader to the fact that such a low-cost sensor is not expected to provide high accuracy over the short-term. Zooming in would thus highlight the opposite of what we consider meaningful in this study. However, we will publish the original data along with our paper (we will however select a different repository than the original mentioned under "Data availability"), which will allow the interested reader to download the data and zoom in as desired.

Fig. 5 and 7 The collected in 2012 and 2015 are both part of the calibration period, not the validation period, which is important to know for the reader to correctly interpret 'TGS 2600 c/v' and 'ANN'. I therefore strongly suggest a note in the figure caption.

> This information will be added.

Fig. 5 Suggesting to replace '(TGS 2600 - Reference)' with '(TGS 2600 all - Reference)' in the caption

> This will be done.

Fig. 9 A plot of the difference of the methane abundance calculated from the measurements of the two sensors would be of high interest for the readers. With such a new panel it is also important to state if the parameters derived for TGS 2600 #1 have been used when applying Eq. 2 to the measurements of TGS 2600 #2. In my opinion the new panel could replace panel (b), as the signal difference seems of little relevance.

> We will replace panel (b) that currently shows the relative differences with a new panel showing the absolute differences as suggested by the reviewer. And we will more clearly state that we used the parameters derived for sensor #1 also for sensor #2 (as correctly interpreted by this reviewer).

Fig. 10 If the main text in ll. 189–190 is correct, 'and the reference' is missing at the end of the first sentence of the figure caption.

> Yes, this will be added.

3 Technical corrections

l. 1 Suggesting to remove “weak” to avoid misunderstanding. Alternatively, it could be written in parentheses like it the conclusions.
l. 8 Insert a space between value and unit of temperature. This correction is necessary wherever ‘°C’ is used [5].
l. 76 replace**d**
l. 140 typeset ‘Ta’ as Ta
l. 305 ‘cross-sensitivities [...] **are** of no concern’
Fig. 2 There seems to be a non-displayable glyph at the beginning of the label for the vertical axis, possibly a Δ. This is also the case in Fig. 3, 5 and 12. ‘CH4’ specifies a substance, not a quantity. Use ‘x_{CH4}’ or another appropriate quantity symbol. The same applies to Fig. 3 through 9 and Fig. 12.

> These will be incorporated in the revised manuscript, thank you for highlighting these items.

4 References

[1] IUPAC Gold Book, term ‘concentration’, https://goldbook.iupac.org/terms/view/C01222
[2] IUPAC Gold Book, term ’mixing ratio’, https://goldbook.iupac.org/terms/view/M03948
[3] IUPAC Gold Book, term ’amount fraction’, https://goldbook.iupac.org/terms/view/A00296
[4] GAW Report No. 242 19th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Measurement Techniques (GGMT-2017), https://library.wmo.int/doc_num.php?explnum_id=5456
[5] SI Brochure: The International System of Units (SI), https://www.bipm.org/en/publications/si-brochure/

[6] WMO Guide to Meteorological Instruments and Methods of Observation, https://library.wmo.int/doc_num.php?explnum_id=4147

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-402, 2019.