Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2022-24-RC2, 2022

Comment on amt-2022-24

Anonymous Referee #2

Referee comment on "Field inter-comparison of low-cost sensors for monitoring methane emissions from oil and gas production operations" by Vincent M. Torres et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-24-RC2, 2022

This work attempts to address an important issue in relation to the large methane emissions from oil production, in which low-cost sensors (LCS) are evaluated as a possible alternative to expensive measurement instrumentation. However, the way the article is organized makes it difficult to follow, the objectives are vague, the methodology is not clearly outlined -which undermines the reproducibility of the results-, and the discussion of the findings and their potential impact is absent in the text. It is the consideration of this reviewer that this manuscript requires further maturation, therefore, major changes are suggested.

Three things seem to be crucial and require better description, further development, and deeper analysis: the dispersion model (and on which some of the subsequent efforts depend), the devices under evaluation (which must meet certain criteria to be considered fit for the purpose) and the last is the colocation experiment (which allows their evaluation).

Since the use of the Calpuff model is crucial for the definition of the assessment criteria as is presented here, is essential to describe the details of the modelling exercise. What configuration was used? What other inputs did it require? What are the most relevant uncertainties regarding the problem in question? What is the spatial and temporal distribution predicted by the model? Are there similar works in the literature, in which a dispersion model has contributed to defining this type of criteria? The Calpuff model has a 3 km spatial resolution, so if the model was run at this standard resolution, how were the results interpreted considering that the sources are located just a few meters from the sensors? Or is it perhaps that the code was modified to improve its resolution? Using only a few weeks' worths of weather may also contribute to limiting the scope of the results, and perhaps this deserves further attention. It is also not clear how meteorology of 1 min time res. was obtained (the work mentions that hourly data was interpolated, which seems wrong. See specific comments below).

As for the low-cost sensors (words that appear only once in the entire text), important issues are not considered by the authors. For example, interferences and drift that this cheaper hardware usually suffers from is something that needs to be considered. It would then be important to discuss the different techniques used by these devices and what are the potential problems that they may experience in relation to the measurement technique they use. Furthermore, if these systems have already been used in other scientific studies and/or evaluated, it may also be useful to understand other potential

limitations. This discussion is absent in the manuscript and is essential to understand the potential uses and limitations in this oil industry scenario.
The colocation experiment used to determine performance also requires a deeper description and analysis. This is perhaps would be the richest aspect of the manuscript and needs to be exploited. Something that catches my attention is that the Tildas is said to be operated continuously, however, the results of the comparisons seem to leave out some elements that could contribute to understanding the usefulness of the LCS in this scenario. The figures showed (and almost no discussion) seem to represent only bits of the colocation study (supposedly 9 months long). Although direct comparisons between sensor "x" and the reference instrument have been intermittent in those 9 months, it would be important to show and analyze the results throughout the entire period and if there are changes in performance (which is common in LCS) try to identify the possible causes or at least the observable impacts. Furthermore, if the reference instrument is continuously present at the site, it could also be used to test the initially assumed capabilities of the dispersion model and redefine (or not) the evaluation criteria.
Complimentary to this certified gas challenges are used to compare LCS and Tildas instrument response. However, a final report (University of Texas, 2021) is cited repeatedly, but when this reviewer accessed it almost no details are shown (in contrast to what is said in the manuscript). This in other circumstances might not be relevant, but as the text presents it as an important source of information and on which this work seems to rest, it would be very important to be better describe how the experiment configuration was.
As a final consideration, I think that the work also needs to be improved regarding how it is structured and how the results are presented and analyzed. The figures presented are sometimes unclear and little explanatory, which hinders the reading quite a bit. The quality of the figures and tables must be improved throughout the text, as well as enriching the text with the analysis of the graphic results. The discussion and conclusions, almost absent in the manuscript, need to be greatly improved.

Specific comments
Title: the low-cost sensors are not mentioned elsewhere
Abstract: It is not clear what the contribution of the work is. The results highlighted here are somewhat sparse. Something that catches my attention is that the data capture is detailed, but results that could be more relevant are not considered.
Introduction: Although the use of LCS may be relevant, the motivation behind this study is not clear. This may be related to the fact that the reviewed literature is scarce. Regarding the potential of using dispersion models as an independent source of information, it is not reviewed either, which seems to have an impact on the methodology.
Line 31-34: it mentions the interest in using networks of sensors, which seems to be outside the scope of this study more focused on the evaluation of independent instruments (and not as part of a network).
Line 23-24: citation of own works seems somewhat excessive, even more so if one considers that the literature review referred to the topic that the paper tries to address is scarce. Here are some papers that could be considered in this matter:
https://www.sciencedirect.com/science/article/pii/S0048969721012614
https://www.sciencedirect.com/science/article/pii/S1352231020301771
https://www.sciencedirect.com/science/article/pii/S1352231018306241
Line 55: “The emission rate (5-10 kg/hr detectable at a distance of 50-100 m) and temporal resolution requirements (one-minute resolution) for sensors were converted into requirements for precision using dispersion modelling.” This requires a more detailed explanation. What are the logical steps?
line 59-66 (plus figures): “Results and discussion” seems to be a more appropriate place for this. It is also important to include the location of the mentioned station. How far is it from the site? Is it representative of the study site? What are the risks and uncertainties

associated with considering only 4 one-week periods?
Figure 1 and 2 could be merged
line 78: were the emissions modelled simultaneously or each emission source separately? constant or variable over time?
Figure 3: the details are described above but also below the fig (it should be only below). The quality of the image needs to be improved. I suggest combining it with figure 4 or be sent to the supplementary. They both add very little info to the manuscript. In the text that accompanies the figure, potential locations for the sensors are discussed. What were the criteria for deciding the final location? What are the distances to the sources?
Line 89: It is suggested to create a new section to describe the model.
line 93: it says that the meteorology of 1-hour resolution was interpolated to one minute. It is not clear from the manuscript how this was done.
Line 99: check the two opening sentences, they seem to say similar things.
Table 1. The text is not explanatory. This table is not discussed/analyzed. In that case, it is suggested to be sent to the supplementary. Regarding the information contained, it is heterogeneous and due to the diversity of units used, comparisons cannot be made on the potential use of these instruments. Also, in the description below (line 107-119) some bits are repeated so it is suggested to harmonize this.
Line 122: The description of the site is probably better if it is included at the beginning of the methodology
Line 132 says that the Tildas was used to confirm the certified gas. How was this instrument calibrated? Does it suffer from specific interferences?
Line 147-150: Is this for the entire 9-month period? How often was this done?
line 154-156: this seems to fit better at the beginning of the methodology.
line 159: This part of the methodological design is important, and its description should be improved. Including a fuller explanation of how these experiments were performed would help replicate these experiments. Even including diagrams and schematics would be of great help.
Figure 5: improved quality is needed (labels and contrast). Please add the dates and show the 9 months of colocation. Descriptive statistics would also offer great inside.
Line 195: the title seems to be wrong. Perhaps the model results could be included here?
Lines 220 to 235: the metrics' names used to assess the performance of the sensors (Rise Speed, Decline Speed, Percent of Target Concentration, and Integrated response Relative to QC-TILDAS) differ from those defined in Table 2. I suggest using the same names.
Figure 3. Improve the quality of the figure (labels). I suggest including the sensor's response to the challenges plots.
Table 2 shows the concentrations used in the challenges (10 and 100 ppm), which exclude the expected concentrations according to the model (3-4 ppm). Furthermore, it is not clear when these experiments were performed. Were they one after the other? At different times of the year? At the beginning of the lifespan of the sensors? Did this behaviour change in time? How?
Linea 251: "dispersion modelling reported in Section 3.2 suggested that the ability to detect mixing ratio enhancements of ~1 ppm". Perhaps the Tildas measurements could be compared to the model results (here an example of polar plots that may help https://www.sciencedirect.com/science/article/pii/S1352231016307166)
Linea 253: "Representative results for one month of data are shown in Figure 8". Please show the full comparison time series and show the regression plots.
Line 256-260: the slopes and R2 are mentioned, but the time series and regression plots are not shown, which are super useful to understand sensors' performance. Including other metrics like RMSE and MAE can also provide useful information.
Line 258: R2 is not a correlation coefficient, it is the Coefficient of Determination.
Line 262: "relative to baseline" is twice.
Figure 7: improved quality is needed (labels, labels, fonts, etc. include date and time). Extend it to the entire comparison period.
Figure 8: improved quality is needed (labels, fonts, etc.) Merge the Tildas and Sensor x layouts into a single panel to compare the histograms easily. Use the entire comparison

period.
Figure 9: improved quality is needed (labels, fonts, etc.). It is not clear which panel corresponds to which sensor.
Figure 10: improved quality is needed (labels, labels, fonts, etc. include date and time). Use the entire comparison period.
Table 3: Consider moving to supplementary. R2 is not a correlation coefficient.
Table 4: Consider moving to supplementary.