We appreciate the reviewer's thoughtful evaluation and critique of our paper. We generally agree with the reviewer's comments, and feel that the suggested revisions improve the clarity of the manuscript. We reply to the specific points below: the reviewer's comments are italicized and our responses are given in ordinary text.

1. *Line 14: 'moderate performance' could be expressed more quantitatively.*

   Revised to add an RMSE value of < 0.6 ppm.

2. *Line 19-20: It seems that the lack of high-resolution data and low-cost network have few causations. In addition, could you provide some stronger evidence to support the lack of high-resolution data?*

   Reviewer 1 commented on this as well; although we believe this to be true from observation, we agree that the assertion is not well supported and have decided to remove this statement.

3. *Line 75-76: Based on the previous experiments, it is a good idea to exclude the influence of environmental factors through the response difference between the two TGS sensors, thus enabling the monitoring of methane in low concentrations (1-10ppm). However, it is well known that TGS2600 is sensitive to hydrogen, ethanol, iso-butane, carbon monoxide (CO), methane, relative humidity (RH) and temperature, while TGS2611-E00 is only sensitive to methane, hydrogen, ethanol, RH and temperature owing to its additional filter covers. (see FIGARO TGS 2611-E00&2600 product information, the above factors are ranked in order of influence.).*
   And therefore, does the effect of hydrogen, ethanol, isobutane, and CO be taken into account in evaluation of the application potential of the TGS2611-E00 in low concentration situation? Especially in the indoor food experiment room. The differences in principle caused by physical between the two sensors and the response characteristics of both two resistance should be presented here.

This is an excellent point. We have added discussion of the specific gasses the two sensors are sensitive to, and explained how including both could help to account for some interfering gasses (hydrogen, ethanol), while other interfering gasses may cause a response in TGS2600 but not TGS2611-E00. It is also unclear to us which gasses were tested by the manufacturer, and accordingly if there are undescribed interfering gasses that might occur in our scenarios.

4. *Line 115: In Fig 4B, where is the indoor sample gas inlet of CH_4 reference instrument? It may affect time delay.*

   We added clarification in the main text that the reference analyzer was drawing air from directly next to the sensor node for the indoor site.

5. *Line 119: As mentioned above, TGS2600 is also responsible for hydrogen, ethanol, isobutane, and carbon monoxide, and therefore nearby sources of relevant gas emissions or its surrounding environment are needed to be additionally stated.*

   We agree; we have added that we are unaware of nearby sources of interfering gasses, in addition to methane.

6. *Line 122: Why these data are averaged at 10 mins scale rather other time scale? Is this a universal practice? or is it tested with some experiment then after 10 mins is selected? In addition, what do you means of 10 mins? Average all the data over 10 mins after then record a valid value?*

   We have expanded our mention of this towards the end of section 2.3 to explain that we averaged to 10 minutes to reduce the effect of lag for our passive sensor node, and to smooth out short spikes which might not diffuse quickly into our device. We choose the 10-minute scale at the beginning of our analysis as a

duration long enough to smooth out any very short transients, but short enough to have real-world utility. We have also added an explanation that the 10-minute average is consecutive averages of the collected data.

7. *Line 132: In the title of Sect 2.2.1, the author reports the situation of increasing background concentration, and how many ppm the concentration is it?*

   We have added detail at the end of the section about the enhancements resulting from our controlled releases, specifically with concentrations measured by the reference analyzer. The highest 10-minute averaged value from a release was 5.8 ppm, with most releases falling between 3 and 4.5 ppm.

8. *Line 134-135: A brief description of laboratory-generated gases with potential effects on resistance is needed.*

   We agree; we have added context that we expected VOCs, hydrogen sulfide, nitrous oxide, and other unknown gases. We have also added that the nearby lab conducts a range of bioprocessing projects, including manure processing and other fermentation studies.

9. *Line 173: 'time parameter'. Is the time parameter taken for the fit calculated from the start time of the per interval? In fact, within a 10-day window, environmental factors do not increase or decrease linearly but time does. Adding the time factor is a good attempt, but what is the contribution or significance of adding this factor?*

   We agree this is ambiguous. We have added context that the time variable is the total time the sensor was powered on since the beginning of the experiment, which would account for sensor aging as well as environmental factors.

10. *Line 182: Is '2.3 ppm' the optimal selection after lots of tests? Or was it chosen randomly by the author? It would be an important impact factor on fitting result.*

    We did not want to bias the selection by trying values and then choosing the best, and so we chose 2.3 ppm a priori based on our previous work as a threshold these sensors would likely not be able to differentiate from baseline. We have added text to this effect.

11. *Line 185: I have seen lots of fitting functions in Appendix B, but can you explain why log is used rather than others? Has any other research used a similar fitting formula before? Or is the log more appropriate after statistics? Some explanation is needed here.*

    We have added brief additional discussion to Appendix B to note that we chose the log transform as a common statistical transformation. The regression residuals did not show clear bias, and so we did not investigate other transformations further.

12. *Line 212: Authors report that a diurnal cycle during rainy week is caused by soil process, please provide some solid evidence for this diurnal cycle driven by temperature.*

    We have removed the mention of soil processes as speculative and have added an example of this diurnal pattern as an appendix.

13. *Line 277: Statistical indicators are recommended to be presented on the figure, and therefore looks more intuitive. In addition, the numbers of data on each sub-figure need to be represented. The first impression is that the numbers of data in ABCD is different from EFGH.*

    We appreciate the suggestions on figure clarity. We agree with adding statistical results, and will label the panels more obviously.

14. *Line 281-283: 'Even with the additional sensor term, the accuracy of the regressions varies with time period, as can be seen in the coloring of Fig. 4. For example, the baseline at the beginning of the inside experiment in Fig. 4E has a worse fit than the baseline closer to the end of the experiment.'.*
    In Figure 2-A2, from early Apr to the end May, in this period temperature and H_2O fluctuate over a wide range. While before 1 Apr, temperature and H_2O changes obviously smaller, especially temperature. When the author only used only one piecewise to fit the whole study period, and the fitting result will obviously be affected by the number of samples. The former months are about 3 times than the latter months. Thus, it can be expected that the fit will be significantly worse in April and May, i.e. bring greater RMSE. And the data prefer to attribute this result to its representative rather than time.

    In addition, in both Figures 4A and 4B, even if the author uses piecewise fitting in 4B, it can be clearly seen that the RMSE of the yellow points in both 4A and 4B is larger than the RMSE of the green points. Compared Fig-2A to Fig-2B, temperature and H_2O in former months (Jan to Apr) also well show larger variation than latter months (after Apr) and therefore imply different possibility with author.

    This is an excellent point, and we agree that the increased fluctuation in environmental conditions is worth considering. However, the early period shows worse performance in Fig. 4E than the later months, which is the opposite of what we would expect if the increased environmental variability were solely responsible. As you say, in Fig. 4A the late period has a worse fit, as expected; so, it seems that some of the benefit of adding TGS2600 is improved tracking of environmental factors (as you have noted elsewhere). But, since the non-piecewise Equation 2 performs worse at the beginning of the experiment than at the end, despite the increased sample size of the relatively stable period of the experiment, we believe something else is going on.

15. *Line 293-295: I agree with baseline needs to be regression and is also a good experiment and perspectives. But the drive factor should be clarified.*

    To give a better sense of the importance of the different terms, we have added a sentence about statistical significance; the time term was non-significant in two of the ten piecewise subsets, and the intercept was non-significant in one, but all terms were otherwise significant in all subsets with $p \leq 0.001$.

16. *Line 320: How about the averaged change speed of CH_4? Actually, I see lots of black point (high change speed points) on the red line.*

    We have added corresponding data on rates of change for the full dataset as follows: "…40% of the outliers show a rate of change greater than 1 ppm per 10 minutes; 31% exceed 2 ppm per 10 minutes; and 6% exceeded 5 ppm per 10 minutes. For the full dataset, 5.1%, 1.7%, and 0.16% show the same rates of change respectively, suggesting that the outliers are considerably more likely to occur during periods of rapidly changing concentrations than are the other data."

17. *Line 354-358: From Fig 2, the temperature and H_2O in outside vary great in a day. And from Fig 3 data also Line 368 of MS show that resistance is sensitive to temperature and H_2 Why not have a try on constructing a fitting baseline by interval of temperature and H_2O? It might get more interested conclusion than a 10 days interval in the outside.*

    This is an interesting and reasonable suggestion; we have also found in our previous work that the sensor response varies by temperature and H_2O bands (such as in Fig. 7 in https://amt.copernicus.org/articles/15/5117/2022/). However, the outside dataset in this manuscript has an extremely small methane range – almost all the data is 2 to 2.5 ppm – and we are unaware of plausible previous work suggesting the sensors will respond with sufficiently small error for this range. As our baseline fit for the outside data is already quite strong, we believe it is unlikely that an improved baseline fit will yield a better methane response; the fit for methane does not show any obvious trend, and so we conclude that the sensors

are simply unsuited for this low range without further breakthroughs, which we think are unlikely to be algorithmic.

18. *Line 361: How about the statistical indicators in per sub-figure? From Fig 7, the effect of piecewise fitting is significantly better than full fitting. The input of TGS2600 can be used to eliminate some predicted extreme outliers, but methane fitting results is not improved significant, especially at low concentrations.*
    In Fig 4 data also show that a better resistance fit with the TGS2600 as baseline results with a resistance from 20 to 70. And it also just well confirms that the two TGS have similar changing characteristics potentially caused by temperature and H_2O. Based on the very weak methane fitting improvement performance after the input of TGS2600, the response of TGS2600 under low concentration conditions cannot be ruled out. Therefore, the input of TGS2600 is only more conducive to predicting the resistance of TGS2611, because their resistance changes in principle are relatively similar. Therefore, the experiment almost failed to achieve the purpose of improving methane retrieval by inputting TGS2600 to weaken the influence of other factors.

    In consideration of this comment and comments from reviewer 1 we have added RMSE and R^2 values for the other three cases; we had previously omitted these as we were unsure if negative R^2 values (indicating performance worse than always predicting the data mean) would confuse the reader. All of the fits other than Fig. 7D produced negative R^2 values, although all clearly captured some methane trend. The fit using Equation 1, without TGS2600, appears to perform significantly worse than with TGS2600 included. It is unclear to us if this improvement is due to better tracking of T and H_2O, as you reasonably suggest, or if interfering gasses or some other environmental condition is responsible. Picking apart these different factors will require further research, and appears to us to be crucial for using these sensors in a low concentration range.

19. *Line 371: Why not make the x and y axes change in the same range? This seems more intuitive.*

    We agree this will make the figure clearer, we will revise accordingly.

20. *Line 457, 'we did not find TGS2600 to respond to methane in the studied 2 to 10 ppm range.'. If this MS aimed to conclude this, and it requires more stronger evidence.*
21. *Line 458-459: 'We suggestion….'. As comment 20 mentioned above, the response of TGS2600 to methane is not excluded in this MS and is therefore not recommend appear in here.*

    20 and 21: We agree with both you and reviewer 1 that our statement was stronger than the evidence supports. We have decided to remove the sentence beginning with "We suggest", and clarified that we did not find TGS2600 to respond in the range and with the algorithms we applied.

22. *Line 461-463: The indoor and outdoor conclusions obtained by the author are not fair. The temperature and water changes are very small in the indoor experiment, which is a relatively ideal condition compared to the outdoor experiment. This ideal condition can significantly reduce the uncertainty caused by diurnal water and temperature in the baseline fitting (see comment 14). Moreover, correlation analysis also shows that resistance is related to temperature and water intensity. This implies that such networks have high challenging at outdoor application.*

    We agree; we have added a mention of the wider temperature and humidity range and noted that this is typical in outdoor settings in many climates.