We would like to thank both reviewers for their helpful feedback. The revisions involve clarifications to the text, reordering of the sections to improve flow, and changes to the way the data are presented in Figures 1, 2 and 6.

Below we respond to the reviewer #2's comments in detail. Reviewer comments are in black text while author responses are in **bold purple text.** Excerpts from the original manuscript are shown in *italicized purple text.* Revisions to the manuscript are shown in *italicized red text.*

Note that reviewer #2 commented directly on a PDF version of the manuscript. To respond to these specific comments, we excised them and listed them as bullet points under the corresponding Line number.

Improved semi-conductor technology has made possible the significant evolution witnessed around the development and use of low cost air quality sensor in the last couple of years. The increased use of low-cost air quality (LCAQ) sensors in Africa, especially Sub-Saharan Africa has been brought about by its affordability and relative ease of deployment. However, the accuracy and quality of the measurement have been questionable and hugely debated in the scientific community. If the quality of the data from the sensors are improved, LCAQ sensors would bring a great revolution to air quality monitoring globally and enhance the our understanding of the problem, especially in low and middle income countries (LMICs) where the problem of air quality is endemic but reference grade instrument are not available.

General Comments

The study discusses an approach to enhance the quality of data obtained from units of LCAQ sensors (O_X, NO, NO_2, CO and PM) in Malawi by calibration using pre- and post-deployment collocations and five model approaches. The structure and layout of the manuscript makes it very difficult to follow through and understand.

- **To improve readability, we reorganized the subsections in the Results to improve the flow:**
  - **3.1 Gas sensor performance during colocation**
  - **3.2 OPC-N2 performance during colocation**
  - **3.3 Gas sensor performance during deployment**
  - **3.4 OPC-N2 performance during deployment**
  - **3.5 Comparison of ARISense CO to remote sensing data**
  - **3.6 Comparison to other ambient measurements in SSA**
  - **3.7 Performance of ARISense sensor packages over time**
- **Further, to clearly establish the goal and objective of this paper, we added the following statement in the Introduction:**

  "*Our overarching goal was to assess the viability of establishing and maintaining a small, temporary network of LCS monitors in Malawi until a more formal governmental regulatory monitoring system can be established. Given that comparison to regulatory grade equipment in Malawi was not possible, the objective of this work was to devise an* *alternative methodology to evaluate the ARISense technology for accuracy, precision, and stability over the 1-year pilot deployment*."

- **We also added an overview statement of the paper's organization in the Introduction. The last paragraph of Section 1 was rewritten as:**

  "*In Section 2.3 and 2.4, we discuss separate collocations of the gas sensors (in North Carolina, USA) and particle sensor (in Mulanje, Malawi) with reference or semi-reference instruments (Section 2.2). We use the collocation data and quantitative assessment metrics (Section 2.5) to compare the performance of five modelling approaches to calibrate the gas sensors (Section 3.1) and estimate error in the particle sensor data (Section 3.2). After deployment to Malawi (Section 2.6), we qualitatively assess how the ARISense performed in the field using contextual information about nearby emission sources, diurnal trend data, and an inter-comparison of calibrated gas model observations (Section 3.3 and 3.4). In Section 3.5 and 3.6, we compare the results to remote sensing and reanalysis data products and surface measurements from similar environments in SSA. Finally, in Section 3.7, we qualitatively assess the long-term stability of the sensor readings and calibration models in Malawi by comparing seasonally similar ambient data collected 1-year apart at the same location. In concluding (Section 4), we use evidence from this pilot study to characterize the benefits, limitations, and robustness of this technology and methodology for our application: collecting AQ data in under-studied regions. Additionally, we offer guidance on considerations to improve future remote deployment efforts.*"

The language also needs adequate tone up to enhance the flow.

- **We improved grammar and sentence structure throughout the paper. Importantly, we removed the excessive number of semicolons.**

The materials in the manuscript should be arranged as much as possible in the way they are referred to in the manuscript. This makes it easy for the reader to follow the manuscript and supplementary material together without having to flip through pages of the supplementary material haphazardly.

- **In the final manuscript, after revisions are finalized, we will reorder the the supplementary information to ensure that it chronologically aligns with references made in the main manuscript as best as possible.**

Some of the figure (I have indicated these in the annotated version of the manuscript) are difficult to understand in their present form.

- **We redesigned Figures 1, 2 and 6 to improve their clarity. Revisions to Figure 6 are shown in the supplement to this response. We invite Reviewer 2 to see the supplement attached to our response to Reviewer 1. That file contains revised versions of Figure 1 and 2.**

Some sub-sections could be further divided into sub-sub-section to improve the organization and readability of the manuscript. Overall, the manuscript needs serious reorganization, restructuring and editing to enhance its understanding.

- **We respond to the reviewer's specific comment about subdividing the sections in "Specific comments" below. We hope our above response detailing the reorganization of the sections addresses the remaining portion of this comment.**

Specific comments

I have included the specific comments in the annotated version of the manuscript.

Line 17: What are the basis for selecting these five models?

- **We added the following text to Section 2.3 to better motivate our selections:**

  "*The five models were selected based on their performance in previous studies. The kNN hybrid model was found to enable accurate measurements even when pollutant levels were higher than encountered during calibration (Hagan et al., 2018). Given that we expected pollution levels to be higher in Malawi than during calibration in N.C., we expected kNN hybrid models to perform well for our unique data set. Further, the authors indicated that the kNN hybrid approach was expected to be widely applicable to a range of pollutants, sensors, and environments (Hagan et al., 2018). In a calibration and validation study conducted by Malings et al. (2019a), RF hybrid models were recommended for any low-cost monitor using electrochemical sensors similar to their sensor package, the Real-time Affordable Multi-Pollutant (RAMP) monitor. Given that the RAMP and ARISense monitors use the same electrochemical sensors and have similar integrated designs, we expected RF hybrid models to perform well for our dataset. HDMR models were found to effectively model interference effects derived from the variable ambient gas concentration mix and changing environmental conditions over three seasons for the sensor types used in the ARISense package (Cross et al., 2017). Finally, MLR and QR are some of simplest and most popular calibration approaches and they were included in this study for that reason."*

Line 34-35: Sources of air pollution in SSA are expected to increase over time given the regional growth in population and energy demand, a biomass fuel dominated energy mix, and slash and burn agricultural practices.

- Source or level?
- ?????
- **This statement was rephrased to "***Air pollution in SSA is expected to increase over time given regional growth in population and energy demand combined with a biomass fuel dominated energy mix.***"**

Line 50: Given the potential applications, LCS deployments are increasingly common (Giordano et al., 2021).

- ???? do you mean becoming common?
- **This statement was rephrased to "***becoming common***".**

Line 96-97: The ARISense sensors were collocated with reference equipment in North Carolina (NC) before and after deployment to Malawi. One OPC-N2, the ARISense particle sensor, was collocated with a semi-reference instrument at a field site in Malawi.

- Is the OPC-N2 part of the ARISense sensor package or is it a separate unit ? From Section 2.1, I think is a part of the ARISense package. If so, reframe this sentence.
- **This statement was rephrased to "***The ARISense were collocated with reference instruments in North Carolina (NC) before and after deployment to Malawi. One ARISense was collocated with a semi-reference PM instrument at a field site in Malawi to assess the performance of the integrated OPC-N2.***"**

Line 126-127: Previous validation studies found the MicroPEM performed well across a wide range of ambient PM concentrations and the real-time nephelometer, after gravimetric correction, agreed with fixed-site reference monitors.

- This sentence is confusing in its present state
- **This statement was rephrased to "***In previous evaluation studies, after gravimetric correction, the MicroPEM real-time nephelometer agreed with fixed-site reference monitors across a wide range of ambient PM concentrations (Du et al., 2019; Williams et al., 2014a).***"**

Line 132: 2.3 Gas sensor colocation and calibration

- There should be a subsection here that presents the component of the sensor units and the source of power. It should include a picture showing parts of the unit.
- **Information on the sensors and power source is currently presented in "Section 2.1 ARISense sensor packages". External and internal photographs of the ARISense are given in Sect 1. of Supplementary Information.**

Line 174: 2.5 Assessment metrics

- There should be a sub-section after this to discuss each of these metrics.
- **Quantitative descriptions for each metric are given in Sect 4. of the Supplementary Information. To make this easier for the reader to find, this statement: "***Quantitative descriptions for each metric are given in Sect. 4 of the Supplementary Information***." was added immediately following the mention of the assessment metrics in Section 2.5.**

Line 177-178: Instead of the EPA recommended Root Mean Square Error (RMSE) metric, the Mean Absolute Error (MAE) was used to assess error in the estimated measurements compared to the true values.

- Is there any reason for opting for MAE instead of RMSE proposed by US EPA
- **We originally preferred MAE to RMSE as MAE does not weight outlying points as heavily as RMSE. However, after considering the reviewer's comment, we believe that adhering exactly to the EPA recommended metrics is a better approach. Therefore, we began replacing MAE with RMSE throughout the paper, and plan to only use RMSE in the revised and final manuscript. In our data sets, RMSE is generally <10% different from MAE, and this replacement does not change the interpretation of main results or the conclusions of the paper.**

Line 191: Supplementary Information, Sect. 5.

- The Google map in Figure S10 should be in the manuscript.
- **We agree with both reviewers who requested a map in the main manuscript. We plan to add a new figure to main paper (which would be Figure #1 in the revised**

manuscript) showing a map of Malawi, denoting the monitoring sites, and a timeline of the project. A mock-up of this figure, based on Fig. S10, is shown in the supplement to our response to Reviewer #1.

Line 212: RF hybrid model

- The authors need to present and discuss in details in section 2 the five models used in the calibration. You are presenting the results from the models without proper introduction to the models themselves.
- **We agree with the reviewer. We hope we resolved this issue with our response to the first specific comment above.**

Line 213: sensors

- The authors need to reframe this paragraph and find a way to get rid of the semi-colons.
- **This section was revised to remove semi-colons.**

Line 214: similarly

- Similarly? How do you mean?
- **"***performed similarly***" was replaced by "***returned similar performance metrics***"**

Line 214: MA

- What is MAE? Discuss the statistical tools used for your analyses before presenting the results.
- **Mean absolute error (MAE) is introduced as a statistical tool in "Section 2.5 Assessment Metrics". The formula is given in Sect. 4 of the Supplementary Information. Note that we will update the formulas to include Root Mean Square Error (RMSE) in the final manuscript.**

Line 215: The NO and CO sensors performed similarly, considering MAE values compared to the typical ambient concentration ranges; ambient CO concentrations are generally 1-2 orders of magnitude larger than NOx.

- This sentence is confusing. Be explicit.
- **This sentence was rephrased to "***The NO and CO sensors returned similar performance metrics, considering ambient CO concentrations are generally 1-2 orders of magnitude higher than NOx concentrations. The CO RMSE values (40-70 ppb) were correspondingly 1 order of magnitude larger than NOx RMSE values (2-7 ppb).***"**

Line 235: 3.2 Gas sensor performance during deployment

- It should be tidier and more readable if you could breakdown the results and discussion for CO, NOx, Ox into 3.2.1, 3.2.2, and 3.2.3.
- **We intend to divide this section (which will be Section 3.3 in the revised manuscript) into two sub-sub sections (***3.3.1 Bivariate histograms* and *3.3.2 Diurnal trends***). We believe that if we divide these sub-sections further by gas type, there will be too many sub-sections. We believe we cannot divide only by gas type, given that the discussion in this section often compares across gas sensor types.**

Line 246: ;

- ????
- **Extraneous use of semicolons was addressed throughout the paper.**

Line 221: O_3, Line 282: Ox, & Line 335: O3

- Have you resolved OX to get O3? Are you using Ox and O3 interchangeably? Or did you estimate O3 for Ox? O3 or Ox?
- **We do not intend to use 'Ox' and 'O3' interchangeably. 'Ox' is used when referencing the sensor itself or the raw data voltage readings. 'O3' is used when referencing the calibrated Ox measurements. All references to "'O3' indicate a data+model product. All instances of O3 and Ox were corrected throughout the paper to match this definition.**
- **Further, the following clarifying statement was added to Sect. 2.3: "***Note that references to 'O3' indicate estimates made from calibrating the Ox data. References to 'Ox' indicate raw voltage measurements from the total oxidant sensor. 'Ozone' is used when referring to the gaseous air pollutant.***"**

Line 291: annual mean diurnal trends

- A comparison of the observation with models on season basis could produce better results. Malawi has two main seasons - the cool dry season between May and October with mean temperatures of around 13C and the hot season between November and April with temperatures between 30-35 C.
- **We agree with the reviewer, and we note that the diurnal trends do vary by season. However, we think this is beyond the scope of this specific paper, as we are only using the diurnal trend to assess coherency between the models. We postulate that we would come to the same conclusion (that the kNN hybrid and RF hybrid perform the best) even if we looked at the seasonal differences. Further, since we do not plan to use different models for different seasons (we will select one model to apply to all our field data), we believe that evaluation of annual trends is sufficient for this analysis.**
- **We do plan to explore the impact of seasons on our findings in the second paper of this two-paper series.**

Line 339: LT

- What is LT?
- **'***LT***' was changed to "***local time***"**

Line 381: higher

- Slightly higher
- **'***higher***' was changed to '***slightly higher'*

Line 394: was highest during the burning season (Figure 4).

- This is only obvious for the university site and not the village site. I suggest you include "especially at the university site".
- **"***especially at the university site***" was added to the existing sentence.**

Line 406: Total Column CO

- Can you also use total column CO? Just in case you will get better agreement.
- **Technically, we could use Total Column CO for comparison, but the surface CO mixing ratio value is most closely comparable to our ARISense surface observations. Both quantities are mixing ratios with the same units (ppb), allowing us to perform qualitative and quantitative comparisons. If we used the Total Column CO, we would be limited to doing a qualitative comparison (such as correlation).**
- **Since we are not aiming to assess the quality of the MOPITT surface CO product in Africa, and only using it as a benchmark to compare the trends in our surface observations with, we think the surface CO product is the best choice.**
- **Further, in absence of in-situ data, air quality managers would likely first use the surface mixing ratio satellite product to inform on-the-ground conditions, rather than the Total Column, so this makes it the logical dataset to compare with.**

Line 410: Presently,

- How do you mean by presently? Do you mean more recent studies?
- **"***Presently,"* **was changed to "***Existing***"**

Line 425: biased high compared

- ????
- **This statement was changed to "***ARI013 $PM_{2.5}$ mass concentration measurements were higher than measurements made by ARI014 and ARI015 (slope > 1), despite all ARISense being in the same location***."**

Line 437: five suggested EPA target values (m, b, MAE)

- A background for this should have been set in section 2.
- **These metrics were introduced in Section 2.5 Assessment metrics. Their formulas and the recommended target values are given in Table S4 of Sect. 4 of the Supplementary Information**.

Line 481: Figure 6

- This figure is very confusing in its present state. It is really difficult to interpret this figure and understand the message you intend to put across to the reader.
- **We made a few changes to this figure (revised version below) to aid interpretation: 1) removed black borders on points and histogram blocks to eliminate an unnecessary visual barrier, 2) colored the text labels to facilitate matching between the histogram bin, its corresponding data point, and the text label (when applicable), 3) selected a sequential colorblind friendly palette, which creates a more natural relationship between color and increasing value range (i.e., on panel (c), redder colors indicate drier ambient conditions, bluer colors indicate wetter ambient conditions).Finally, we bolded the text labels, the panel labels and the figure text title. We would like to note that in the revised and final manuscript, we plan to use RMSE on the y-axis instead of MAE, to address reviewer #2's earlier comment. Preliminary analysis suggests the difference between the 2 metrics is**

around 10%, and thus we expect the main findings of the paper to remain unchanged.

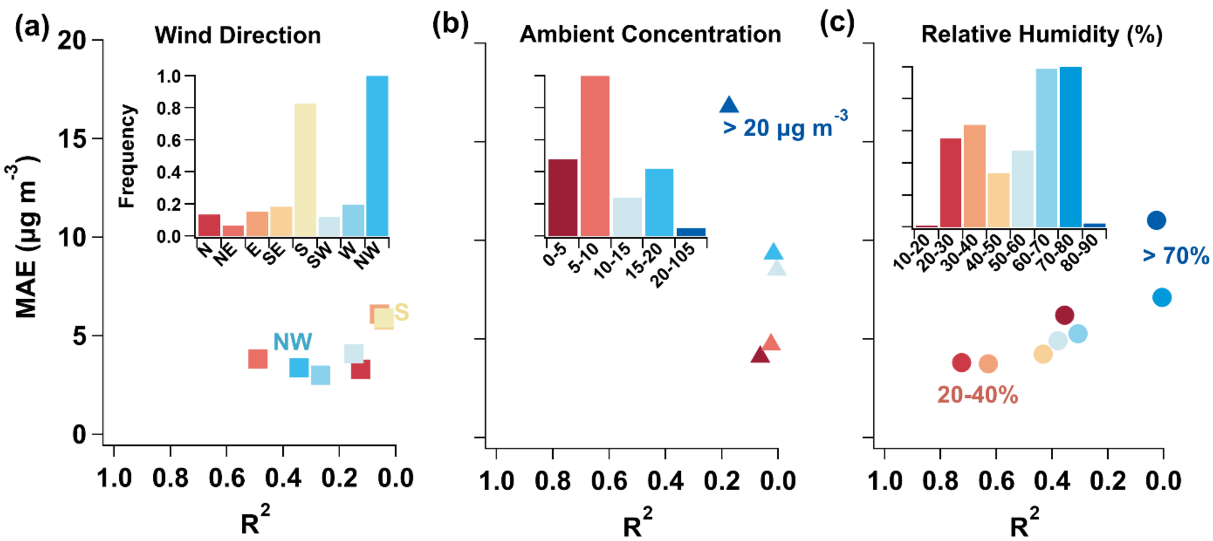

Figure 1 (revised): Performance comparison of the RH-corrected Alphasense OPC-N2 compared to the MicroPEM under different environmental conditions: (a) wind direction, (b) ambient concentration, and (c) relative humidity during collocation at the Village 2 site in Mulanje, Malawi. An individual data point represents the paired metrics (MAE and R^2) for the OPC-N2 for a specific range of each condition. The histograms (inset) show the normalized frequency distributions for the ranges of each condition recorded during the collocation period. The colored markers in each panel correspond to the colored histogram bins. The metrics were calculated from 60-min averaged RH-corrected OPC-N2 $PM_{2.5}$ concentrations compared to the MicroPEM mass-corrected nephelometer. MAE is mean absolute error, assuming the MicroPEM concentrations as the true values; R^2 is the coefficient of determination. The lower left corner region of each panel indicates the highest performance based on these metrics.

References

Du, Y., Wang, Q., Sun, Q., Zhang, T., Li, T., and Yan, B.: Assessment of PM2.5 monitoring using MicroPEM: A validation study in a city with elevated PM2.5 levels, Ecotox. Environ. Safe., 171, 518–522, https://doi.org/10.1016/j.ecoenv.2019.01.002, 2019.

Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., Worsnop, D. R., and Jayne, J. T.: Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements, Atmos. Meas. Tech., 10, 3575–3588, https://doi.org/10.5194/amt-10-3575-2017, 2017.

Hagan, D. H., Isaacman-VanWertz, G., Franklin, J. P., Wallace, L. M. M., Kocar, B. D., Heald, C. L., and Kroll, J. H.: Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments, Atmos. Meas. Tech., 11, 315–328, https://doi.org/10.5194/amt-11-315-2018, 2018.

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S. P. N., Zimmerman, N., Kara, L. B., Presto, A. A., and R. Subramanian: Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring, Atmos. Meas. Tech., 12, 903–920, https://doi.org/10.5194/amt-12-903-2019, 2019a.

Williams, R., Kaufman, A., Hanley, T., Rice, J., and Garvey, S.: Evaluation of Field-deployed Low Cost PM Sensors, U.S. Environmental Protection Agency, Washington, DC, 2014a.