

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 #1'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*.

---

## Response to Reviewer 1

This article discusses the calibration via colocation experiments and five different calibration models, deployment, stability, and relocation/post deployment check of low-cost air quality sensors (specifically ARISENSE measuring PM, O<sub>3</sub>, NO<sub>x</sub>, and CO) in Malawi. Low cost sensors are increasing in use, specifically in areas with less infrastructure or resources for reference air quality monitoring stations, and long-term deployments are useful to see how they perform. Comparing sensor calibration robustness over time is also useful. I recommend eventual publication. However, there are a few large issues with this paper that must be addressed first. General comments first, then specific comments.

### General Comments

This paper feels like an add-on/afterthought to another paper (on the air quality findings by the low-cost sensors in Malawi, which is mentioned by the authors as a separate publication in prep). That's an efficient data use strategy, but means that this study was not designed to optimize it's stated main goal of sensor calibration scheme comparison and robustness. There are no reference monitors available for comparison during deployment, so the ability of the scientists to really understand how their calibrations performed over time and how performance changed when exposed to different conditions, beyond knowing when the sensors are returning non-physical data (e.g., negative values), is unclear.

- **We did not intend to portray the main goal of this paper as an optimized sensor calibration scheme comparison. Our overarching goal is to get data in never-studied regions using affordable LCS technology. The main goal of this paper is to fully address the methodology of deploying this specific technology in our target deployment site and to provide examples for approaches others may use in the situation where no *in situ* reference monitors are available. More plainly, our goal was to determine and share how to use and infer results from the equipment we deployed.**
- **To clearly establish the goal of 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.”*

Further, to establish this papers’ independence from the forthcoming paper, we reframed how the objectives were both defined (Section 1) and addressed (Section 4). In the Conclusions (Section 4), we comment on the viability of using such a technology for our goal and provide insight on how to improve future efforts.

- To improve readability, we also reorganized sections to improve the flow of the paper. 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.”*

- We realize the approach we had to take to calibrate and assess the data was not ideal, but that was inevitable, given the lack of reference monitors in the region. We believe we have reframed the paper to more clearly show why we took this path and to clearly state when we are using quantitative vs. qualitative inference. Finally, we believe we sufficiently acknowledge the shortcomings of our methodology, and as a result, we express to the extent possible the uncertainty associated with quantitative values.

A comparison of sensor data to satellite data is done; however, the authors themselves acknowledge past work that shows that, in Africa, comparison of satellite to ground data is non-ideal.

- For clarity, we changed the name of Section 2.7 from “Remote sensing data” to “Remote sensing and reanalysis data”. MERRA-2 is a “reanalysis” data product, so it is the output of an atmospheric chemistry model that has “assimilated” other data (including but not limited to satellite data) in its estimations.
- We believe this comparison is adding to the literature on surface-to-satellite comparisons over Africa. More broadly, comparison of satellite data to single-point ground measurements is never “ideal” but it is done all the time. We believe this comparison adds value to the discussion and literature on such comparisons. For example, this information may help inform decisions about where and why

permanent ground-stations should be added in Africa. We added a new opening to the first paragraph of this section to emphasize our purpose for this comparison and the main outcome:

*“In the absence of in-situ surface data, we rely on satellites and models to estimate surface air quality. To contribute to the literature on surface-to-satellite comparisons over Africa, we compared calibrated ARISense CO observations to a satellite observation (MOPITT) and a model estimate (MERRA-2) for the study region. We confirmed that all three data sets reported similar annual qualitative trends, although they disagreed in magnitude.”*

- In the absence of "ideal" data from well-maintained regulatory-grade surface monitors, we are using what we can (low-cost sensor data), while being honest about the limitations. In the absence of any in-situ data, we only have satellites and models; this analysis compared with both and confirmed they are all seeing the same trends, although they disagree on magnitudes.
- We believe that we have correctly pointed out the limitations of this comparisons and shown the qualitative agreements, which are useful. We have added to the text in Section 3.3 “Comparison of ARISense CO to remote sensing data” to make this clearer for the reader:

*“Monthly mean CO ARISense values were 2 to 4 times higher than those reported by MOPITT and MERRA-2. We found differences of 175 to 200% between the annual mean CO concentration from ARISense and MOPITT, depending on the site, and even larger differences (up to 360%) with MERRA-2. Differences between MOPITT and MERRA-2 were smaller (30 to 35%). There are few comparable studies available to explain these differences, which are greater than previously reported in the literature available for SSA. One study in South Africa reported relative differences of  $\pm 40\%$  between ground-based CO measurements and Aura satellite observations at Cape Point station (Tohir et al., 2015). Many studies found good agreement (within 10-20% bias) between ground measurements and MOPITT observations, but this was for Total Column CO, and the observations were not limited to comparisons over Africa (Buchholz et al., 2017; Emmons et al., 2009, 2004; Yurganov et al., 2008, 2010). However, these studies found negative satellite bias when intense biomass plumes affected observations, when CO levels were low in the Southern Hemisphere, or when atmospheric CO levels changed rapidly (Buchholz et al., 2017; Emmons et al., 2004; Yurganov et al., 2008, 2010). Each of these conditions could reasonably be expected to occur in the southern Africa troposphere, potentially explaining differences observed between the ARISense and remote sensing observations in this study.*

*This comparison of low-cost sensor surface data, satellite observations, and model estimates in Malawi suggests each of these resources can give consistent information on qualitative, long-term trends in a region without ground-based reference monitoring. However, because of inherent differences in spatial and temporal resolution, each observation will likely disagree in magnitude. Satellite retrievals and real-time surface measurements do not result in directly comparable quantities. Satellite data are generally collected as a once-daily flyover observation, averaged over a  $\sim 12,000$  square kilometer area (corresponding to  $1^\circ$  spatial resolution). In contrast, the ARISense data were 1 min resolution, fixed-site, long-term point measurements at the surface. Further, the ARISense data were collected near visually identified biomass emission sources and were not representative of background conditions. Meanwhile, the satellite observations provide an*

*estimate of regional background conditions. Despite these differences, the MOPITT, MERRA-2 and ARISense data sets agreed on the long-term seasonal trends present in this region, and even corroborated site-to-site differences (e.g., higher mean CO at University compared to Village Mean site). These findings suggest the ARISense captured synoptic-scale variation in CO, but comparison to remote sensing data does not allow for a quantitative assessment of data collected at higher temporal resolutions.”*

There are mentions of ‘known emissions’ near each site, but no clear definition and numbering of these emissions or discussion of their distance/location from the site.

- We added markers to indicate the location of visually identified biomass cookstoves in Fig. S12 and S13 (see Supplement .pdf attached to this response).
- We also added the following context to Section 2.6:

*“ARI013 (“Village 2” site) and ARI014 (“Village 1” site) were deployed < 5 km apart in two rural residential villages in Mulanje, Malawi, adjacent to many households. Almost all rural households in Malawi (99.7%) use solid fuels (e.g., firewood, charcoal) for cooking (National Statistics Office, 2017). Emissions from widespread biomass cookstove use are known to impact local ambient air quality (Aung et al., 2016; Zhou et al., 2011; Amegah and Agyei-Mensah, 2017). Homes regularly using biomass cookstoves within 50 m of the monitoring sites were visually identified at the onset of the study (shown as red cross-hatches on Fig. S12 and S13).”*

If the goal of the study was really to compare long-term performances of these models, this study was not ideal for that.

- We agree with the reviewer that this study is not ideal for that goal, and this is one of several objectives of the study. We did not intend to portray this paper as a comparison of the long-term performance of the calibration models. The goal of the study is to determine if measurements made during the 1-year deployment were reliable and consistent. Characterizing and comparing the long-term performance of the models was a method used to support that goal. We hope that we resolved this issue with changes made in our first response statement above.

Perhaps this could still be done in part—could you also pretend the initial colocation was a full experiment, .

- We used the initial colocation data in N.C. to develop and quantitatively assess the performance of the models before deployment to Malawi. We cross-validated the models by separating our full data set into separate testing and training data sets and evaluating the performance using EPA-recommended assessment metrics.
- Because of the lack of reference monitors in Malawi, we were unable to quantitatively assess model performance during the deployment to Malawi. We qualitatively assessed the performance by comparing the diurnal trends predicted by each model.
- We cannot use the post collocation data collected in N.C. to quantitatively assess the long-term performance of the models, because the sensors were degraded during the high-concentration biomass burning experiments conducted in Malawi after the 1-year deployment. Unfortunately, we cannot assess long term model

performance in any other way than our 1-year in situ comparison discussed in Sect. 3.7.

- We believe we sufficiently address the limitations and shortcomings of our approach in the final three paragraphs of Section 3.7. To make this clearer to the reader, we added text in Section 3.7:

*“Ultimately, the O<sub>x</sub>, NO, NO<sub>2</sub> sensors were permanently altered by the biomass burning emission experiments in Malawi, leading to the poor performance during post-deployment collocations with reference instruments in NC. Given these dramatic changes in sensor responses, the models were unable to generate reasonable concentration values from sensor signals and consequently, we were unable to use the post-deployment collocation data set to quantitatively assess long-term model performance.”*

I acknowledge that this is hard in this particular region; however, another major issue is that there is no one from the region involved as an author on this study.

- We would like to clarify that although no one from the region is listed as an author on this specific manuscript, the larger study involved scientists and civilians from the study region. We do not wish to under-emphasize the role that our local collaborators played and so we mention, by name, the many persons that contributed to this work in the Acknowledgements section.
- For clarity, we explain how each author contributed to the manuscript in the ‘Author contribution’ section of the manuscript. In conferring authorship, we adhered to the criteria given in the Vancouver recommendation (<https://i.ntnu.no/wiki/-/wiki/English/Co-authorship>):

*“The person in question must have made a substantial contribution to the conception or design of the work; or to the acquisition, analysis, or interpretation of data for the work.*

- 1. She or he must have been involved in drafting the work or revising it critically for important intellectual content.*
- 2. She or he must have approved the version of the manuscript to be published.*
- 3. She or he must agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. In addition to being accountable for the parts of the work he or she has done, an author should be able to identify which co-authors are responsible for specific other parts of the work.”*

- Our local contacts at the University of Lilongwe Centre for Agricultural Research (CARD) are in a different research discipline and as such were not considered to have contributed in the ways outlined above in the development of this paper. We attempted and failed to identify and involve interested students at CARD, possibly because air quality science is not included in/or a focus of the CARD curriculum. A few years into this project, we established a collaboration with, and eventually a monitoring site at, an engineering-oriented university in Malawi, recently renamed the Malawi University of Business and Applied Sciences (MUBAS), but they were not involved with collecting or interpreting any of the data presented here. We will engage them further in the analysis and documentation of the full data set. Given the content of this initial paper in the two-paper series,



we decided to limit authorship to those involved in the establishment of the sampling sites (AG, EL, and AB), developers of the technology (EC, DH) and developers of the calibration tools used (CM, DH, EC). Our unique contributions are further described in the “Author Contribution” section at the end of the manuscript.

Involving local scientists would have greatly increased the scientific merit of this paper in a few ways.

- We agree and we plan to engage with researchers at MUBAS and invite them to participate in the authoring of the second paper in this two-paper series.

#### 1. Better understanding of conditions on the ground and local context

- Three authors (AG, EL and AB) identified and established the monitoring sites through on-the-ground visitations and discussions with local residents. AG and AB maintained contact with local staff throughout the duration of the study, listened to their comments and concerns, and passed information on the equipment and progress of the study to staff and project participants. AG and AB annually visited the sites.
- At the onset of the study in July 2017, AG, EL, AB and CARD faculty participated in village-wide meetings at Village 1 and Village 2, attended by the village chiefs, their families, and many residents. Residents shared their comments and questions, and some volunteered to be study participants. During this meeting, Dominic, a resident in Village 2 (mentioned in the Acknowledgements) expressed his interest in helping our team execute the study. He was hired as our interpreter, guide, person of contact, and assistant for the remainder of the study (July 2017 to March 2021).

#### 2. Better data capture by regular maintenance

- We hired two regular maintenance staff for the duration of the study: One at the Village 2 site (Dominic Raphael) and one at the University site (Misheck Mtaya). We originally identified and hired an additional maintenance assistant at the Village 1 site but failed to retain them. After their resignation, our Village 2 assistant took on the role of maintenance assistant for both Village sites.
- In all cases, field staff had limited technical background, and so maintenance (of which there is relatively little for these sensor units) was not a major role. Instead, they assisted with data download, checking on continued operation, and with some troubleshooting via remote communication with us when the monitors failed to operate.

#### 3. Better understanding of how these sensors might be used

- The target user for this technology is a government or private institution. The technology is not intended for citizen science purposes. We held a collaborative planning meeting with government officials in July 2017 in Lilongwe, Malawi before the beginning of the study. Another collaborative, reflection workshop was planned in summer 2020 to mark the conclusion of the study, but it has been postponed due to the on-going COVID-19 pandemic.

4. Better understanding of how well corrections factors on data might be applied. (e.g., are the correction factors easy to apply with limited computing power and limited software?).

- **In the development of this study, it was determined that the calibration of the sensor packages would be the combined responsibility of the product developer and researchers. The end user (e.g., a national research institution or governing body) was not expected to be involved during that stage of product development. Therefore, we did not consider the ease of application or computing power/software limitations.**
- **While during this study, the post-processing of data was required (developing and validating the models was a key component of this work), the future application of these models/calibration factors in real-time is feasible. The models used are light-weight, pre-trained models that run server-side and are thus not limited by computational power on the edge device itself.**

Understanding the actual use, by people in the region, of these sensors could have made up for some of the lack of reference data comparisons by discussing another essential facet of low-cost sensor use (the people using them).

- **We agree with the reviewer that documenting use cases/how these sensor packages will be used would be a great next step. However, this will require integration of efforts with national or regional-scale regulatory or research institutions. We have continued these conversations with representatives of such organizations in Malawi (e.g., via our link with MUBAS) but this is a longer-term effort and is outside the scope of this focused paper.**

Specific comments

Table 1: define QR in table caption

- **Original caption:** “Description of the five calibration modelling approaches and data inputs for each gas sensor and model combination (CO = carbon monoxide, NO = nitrogen oxide, NO<sub>2</sub> = nitrogen dioxide, O<sub>3</sub> = ozone).”
- **Revised caption:** “Calibration modelling inputs for each gas sensor (CO = carbon monoxide, NO = nitrogen oxide, NO<sub>2</sub> = nitrogen dioxide, O<sub>x</sub> = oxidants) and model combination (‘All’ indicates k-nearest neighbor (kNN) hybrid, random forest (RF) hybrid, high-dimensional model representation (HDMR), quadratic regression (QR), and multi-linear regression (MLR)).”

How reasonable is developing individual models for each sensor and each gas component?

- **The different gas sensor types (CO, NO, NO<sub>2</sub> and O<sub>x</sub>) require separate and individual calibration approaches because the physical basis of their response and cross-sensitivity to environmental conditions can and does vary.**
- **The different sensor packages (ARI013, ARI014, and ARI015) could potentially be calibrated by a ‘group’ model. However, inter-unit variation among the sensor packages (Figure S3) motivated the use of individual models. Further, we determined that the small size of this monitoring network (N=3) made this option feasible. Other studies have investigated the tradeoffs of different approaches using larger networks (Malings et al., 2019).**

- In any case, we posit that it is reasonable and common to develop and use calibration factors unique to an analytical instrument. In the case of the U.S. Federal Reference Methods, it is a requirement.

60 minute averaged data: will miss ‘events’ potentially like cooking/agricultural and trash burning—can you speak to the significance of these types of events?

- We agree with the reviewer that 60-min averaged data is unsuitable to visualize trends in short/variable emission events like trash burning. We acknowledge that nearby emission events can impact the representativeness of ambient, background measurements and we feel we communicate this throughout the paper. We plan to quantitatively explore the impact that source emissions have on background levels in the second paper of this two-paper series. However, we clearly see the influences of village- and household-scale cooking activity, even at 60-min averaging intervals, at least in the aggregated CO and PM<sub>2.5</sub> diurnal trend data (Fig. 3 and Fig. 7) for the villages.

Figure 1: I found this confusing. And hard to tell if the different models worked better or worse on the different sensors, or there was consistent agreement from looking. Are they the same graph just color-coded differently? If so you could make sure no overlaps in color on the two graphs. I think you are trying to convey a bit too much information on one graph for ease of interpretation. A table, or clearer labeling with different colors, or an extra panel, might work better to convey your message, which is a bit lost right now. I think a table (like those in the supplementary nfo) or adding some numbers of relative performance in your text could make the determination of which model worked best clearer for the reader.

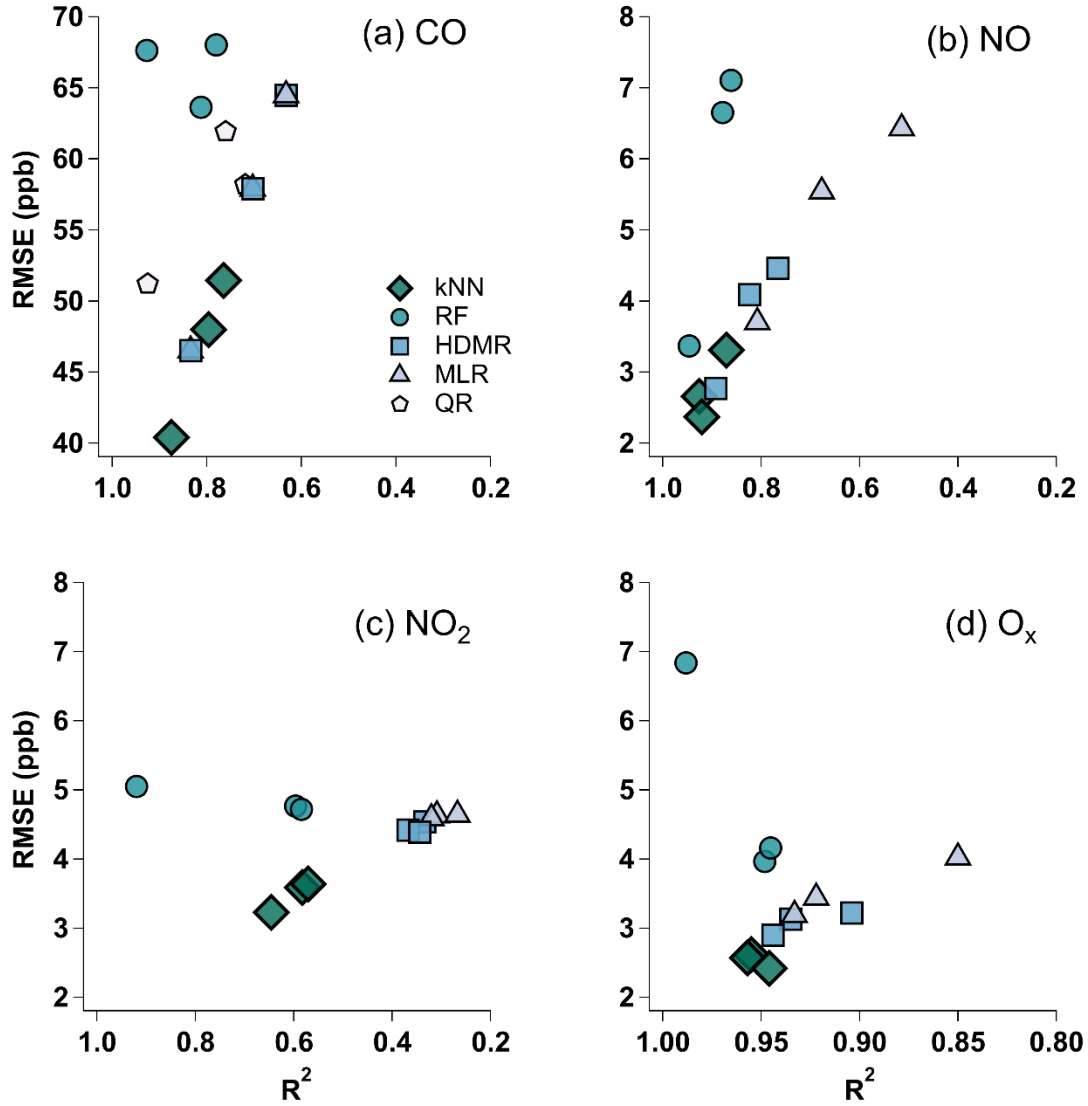
- Based on the reviewer’s feedback, we decided to change the way information was presented in Figure 1 (revised Fig. 1 below). We removed panel (b) and used four panels, one for each gas sensor. We selected new color-blind friendly colors.
- Based on Reviewer #2’s feedback, we decided to use the EPA-recommended RMSE metric to assess error, rather than MAE, which was previously shown on the y-axis of Figure 1. The RMSE and MAE values were similar, however, and did not change the interpretation of results or the conclusions of the paper.
- To emphasize that the kNN model performed the best for all gas sensors, we used a darker color and increased the marker size and the border weight compared to the markers used to represent the other model types.
- We also added the following discussion about the relative performance of the models in this section of the results:

*“In almost all cases, the kNN hybrid model returned higher  $R^2$  values and lower RMSE values than any other model. The RF hybrid model attained similar, and occasionally higher  $R^2$  values than the kNN hybrid, but it had higher (and therefore worse) RMSE values by comparison. Further, the kNN hybrid model showed the least inter-monitor variation in performance. In Fig. 1b-d, the kNN hybrid points are closely clustered together, suggesting that this model was able to attain approximately the same performance for each of the three ARISense. Conversely, the other models, in particular the RF hybrid and MLR, showed a wide range in performance across the three ARISense. Even if another model was able to attain performance metrics higher than the kNN hybrid (e.g., HDMR and MLR CO models in Fig. 1a) it was only for one of the three ARISense monitors, never all three. Given that we seek an approach uniformly applicable to all the gas sensors and*



*all three ARISense, any model other than the kNN hybrid was unsuitable. Additionally, the MLR failed to meet target values for some ARISense-gas sensor combinations (Fig. 1a-b)."*

- We also plan to add a summary table of the performance metrics for all models to the supplementary information for the revised submission.



**Figure 1 (revised):** Performance comparison of gas sensors (a) CO, (b) NO, (c) NO<sub>2</sub>, and (d) O<sub>x</sub> as calibrated by the five types of calibration models adopted for this study (kNN hybrid, RF hybrid, HDMR, MLR, QR). The model type is indicated by color and marker shape. An individual data point represents the paired metrics (MAE and R<sup>2</sup>) for one ARISense monitor. Since there are three ARISense (ARI013, ARI014, ARI015) monitors, there are three markers for each gas sensor-model combination. MAE is mean absolute error. R<sup>2</sup> is the coefficient of determination ( $-\infty \leq R^2 \leq 1$ ). The lower left corner region of each panel indicates the highest performance based on these metrics.

For the atmospheric pressure versus sensor performance issues: are there any studies you could quickly compare your work to? Maybe some done in Boulder, for extrapolation? If not, I don't think that limits this work at all, as the difference in elevation is not huge.

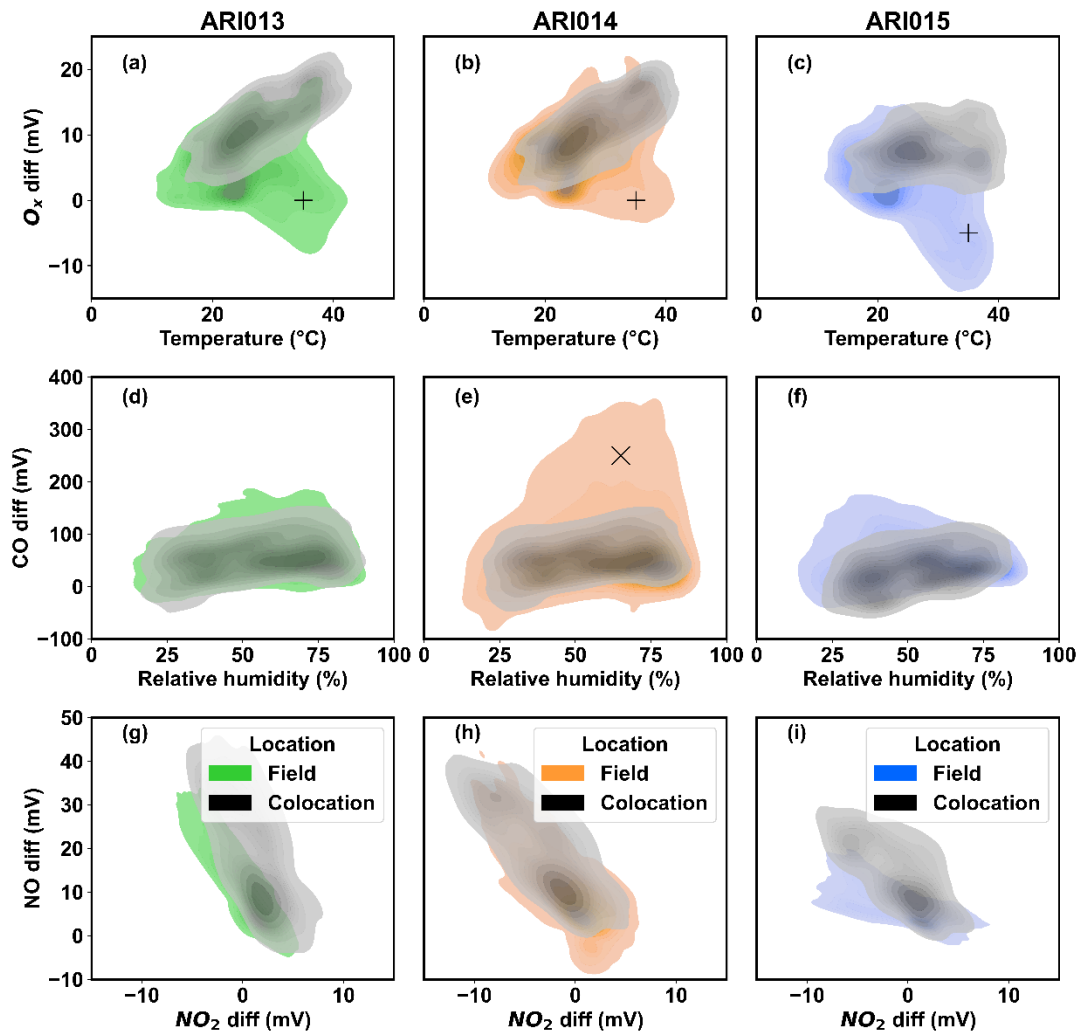
- **We added an additional reference to the literature to strengthen our argument:**  
*“Further, others have shown no statistically significant change in electrochemical sensor sensitivity due to changes in pressure (Popoola et al., 2016).”*

Figure 2: These plots are somewhat confusing and hard to interpret, in my opinion. I would at least add a color scale bar. I think there's likely an easier way to display this info. Also I searched your text and I don't see V1,V2, Uni defined anywhere? I assume it is village 1 village 2 university, but I would repeat that info in the graph caption and maybe in the text when introducing the sites define that V1=village 1 so readers just browsing the graphs can figure out the meaning more quickly.

- **We removed all instances of V1, V2, and Uni and only use ‘Village 1’, ‘Village 2’ and ‘University’ throughout the manuscript. However, to aid in interpreting this figure, we changed “NC” to “Colocation”, and each deployment site to “Field” to emphasize the point of this figure. Displaying the colocation and field data as bivariate histograms allows us to visualize regimes where our models are poorly constrained (i.e., regions where the colocation and field ‘blobs’ do not overlap). We added additional text to the caption (**shown in red below**) for clarification. We plan to choose a color blind friendly color scheme to denote the three sites in the revised manuscript.**
- **The colorscale denotes the number of data points in that location, or the “density”. Because of data gaps throughout the field data, the quantitative density values are different for each panel, therefore a single colorscale bar would not accurately represent all panels. Further, we think that adding two colorscale bars (colocation and field) for each panel (a-i) would only further complicate this figure, especially since the actual density values are not relevant to our message. We only want to communicate data density qualitatively, and we feel the explanation in the caption conveys this sufficiently to the reader: “Density is reflected in the color scheme; Darker colors indicate more data points in that region”.**
- **We display this information using simpler, traditional histograms in the supplementary information, but that presentation prevents us from identifying important interactions between the variables (e.g., O<sub>x</sub> voltage and T) that might impact model performance. To further emphasize the focus of this figure, we added markers (‘x’ and ‘o’) to identify the regions discussed extensively in the text (revised figure below). This allows the reader to synthesize information between the text and figure and find the exact region that we are referring to more easily. We added to the text in the section (**shown in red**) to point the reader to the exact region we are referring to:**

*“Figure 2e shows the maximum ARI014 CO differential voltage in Malawi (350 mV) was 3 times higher than the maximum voltage registered in NC (100 mV). **This high CO regime is denoted by an ‘x’ on Fig 2e.** This difference was aligned with observations of nearby sources (Fig. S12 and S13). We expected higher CO in Malawi than in NC, where biomass burning is less common and emissions from other sources (e.g., vehicles) are controlled by strict federal regulation. ARI014 was deployed in a densely populated village, adjacent to more biomass cookstove activity than ARI013 or ARI015 (Fig. S13).*

The  $O_x$  differential voltage ranges were the most dissimilar between the collocation and deployment environments (denoted by a '+' on Fig. 2a-c). The most frequent regimes, the heaviest shaded regions in Fig. 2a-c, did not overlap for any of the ARISense. In NC, the relationship between the  $O_x$  sensor voltage and ambient temperature was positive and approximately monotonic. Generally, higher temperatures facilitate ozone production, therefore this relationship fit our expectation for an urban site in a single season. However, the positive relationship between  $O_x$  sensor voltage and temperature did not always hold in the deployment sites. Figure 2a-c shows a high temperature-low ozone regime in Malawi that was not present in the NC data (region denoted by a '+' marker)."



**Figure 2 (revised):** Bivariate distributions of gas sensor calibration model data inputs (RH, T, and  $O_x$ , CO, NO, and  $NO_2$  differential voltage) for each ARISense monitor using kernel density estimation. Density is reflected in the color scheme; Darker colors indicate more data points in that region. Training data collected during collocation in North Carolina are shown in grey; data collected during field deployment to Malawi are shown in color. Regions where the field deployment distributions overlap with the N.C. collocation distributions indicate the regimes for which the calibration models were trained. Regions where the deployment location distributions extend beyond the N.C. collocation distributions indicate regimes where the calibration models extrapolated to estimate pollutant concentrations. These regions are indicated by overlaid markers 'x' and '+'. ".

Line 279 : would recommend you to define more what you mean by 'well controlled'

- **This statement was rephrased as:** *"We expected higher CO in Malawi than in NC, where biomass burning is less common and emissions from other sources (e.g., vehicles) are controlled by strict federal regulation."*

Line 288 can you expound a bit on the 'difference in ozone precursor regimes?' I know a separate study is likely coming out about the measurements themselves, and this will be expounded on, but this line as it is just is hanging there begging for a bit more info.

- **We agree with the reviewer and in response we decided to remove discussion of this topic. It is not necessary to understand the main point (that we lack representative collocation data, particularly for the oxidant sensor).**
- **We plan to address this topic in the second paper of the two-paper series, where we can do a first-order estimate of ozone precursor regimes at the collocation and deployment sites (using O3 to NOx ratios).**

Line 295 discussions: is calculating a diurnal trend really a way to see if the models are transferable, if you don't have any nearby ground-based air quality data? How did you get your local knowledge of air quality? (anything to cite?)

- **We posit that calculating the diurnal trends is a qualitative approach to assess if the models are transferable, since the deployment site does not have ground-based air quality data for quantitative comparison.**
- **The diurnal data show we're in the right ballpark in accordance with what we know about atmospheric chemistry (based on our expertise as air quality scientists), these specific deployment sites (based on our firsthand observations of diurnally variable emission sources), consistency between combustion tracer diurnal trends (CO and PM), and what other studies in Africa have shown. While this doesn't necessarily show the models are transferable perfectly, it does suggest they likely capture some of the major features of diurnal trends at these sites.**
- **To better support our approach in this section, we rephrased the text and added references:**

*"Since the deployment site does not have reference data for quantitative comparison, we calculated and compared the annual mean diurnal trends of each pollutant, at each site, as predicted by the five models to qualitatively assess the transferability of the calibration models to Malawi. Our definition of a transferable model required that it produce: (a) non-negative concentration values and (b) diurnal trends consistent with our first-hand observations of nearby emission sources and their timing, our knowledge of the ambient trends in regions with widespread biomass cookstove use (Dionisio et al., 2010; McFarlane et al., 2021; Subramanian et al., 2020) and atmospheric chemistry."*

*"This analysis can contribute to our confidence in the estimated concentration values and trends, but ultimately cannot address or estimate the quantitative error."*

Line 319: sensor has less RH/T interference at higher concentration—can you provide a reference for this? It makes some logical sense but would be nice to expand this thought.

➤ **This statement was rephrased to:**

*“RH/T interference induced on the CO-B4 sensor, ~0.2 mV/ppb (Lewis et al., 2016) has relatively less influence on overall sensor response in the higher voltage (i.e., concentration) regime.”*

Also, are there any chemical species co-emitted with CO that, at higher levels of CO, would also influence CO?

- For the major emission sources near the monitoring sites (biomass cookstoves), CO is co-emitted with fine particulate matter (PM). We do not expect co-emitted PM to influence the formation of CO (chemically). Photochemical oxidation of VOCs (also co-emitted during biomass burning) can form CO, but this is on much longer time scales (hours-days), so we expect this to be a background contribution and not one that influences diurnal trends in the way that local emissions do.
- But it is possible, and at extremely high concentrations probable, that co-emitted pollutants (e.g., gas phase hydrocarbons, other reactive gases, fine particulates, etc.) can influence the response of the CO sensor due to cross-sensitivity. For example, ethylene oxide and carbon monoxide sensors share the same catalyst and structure (but have different pre-filters), therefore small VOCs could interfere with CO sensor performance. At the low ambient concentrations in Malawi, we do not expect this to be an issue because VOC concentrations are typically 2-5 orders of magnitude lower than CO in biomass burning emissions (Akagi et al., 2011) and the cross sensitivity of the CO-B4 sensor to an example hydrocarbon (acetylene) is less than 1% (<https://www.alphasense.com/wp-content/uploads/2019/09/CO-B4.pdf>).
- However, in wildfire-like conditions, or during the high-concentration biomass burning fenceline experiments we conducted after the ambient deployments, certain sensors (even high-quality reference instruments like the 2BTech 202 Ozone analyzer) can be rendered essentially useless due to cross-sensitivities (Long et al., 2021).

How did you characterize the known CO sources?

- We visually identified nearby biomass cookstoves through on-the-ground site walkthroughs and by spending full days in the village sites over multiple weeks.
- We characterized a sample of the known CO sources visually and with emission measurement equipment (Section 10 of SI - this is a separate paper and research study which will be published later).

The grid averaged for CO surface concentration is 12,000 km<sup>2</sup>, right?. How many different CO sources are within an area of that size that would cause heterogenous CO measurements?

- The spatial resolution is 1° (corresponding to ~12,000 km<sup>2</sup>) for the MOPITT CO surface concentration (Table S5) and 0.5° x 0.625 for the MERRA-2 product (Table S6). Every combustion site is a potential CO source, and there are quite a lot of them contained in each grid cell at those spatial scales. We do not know how many different CO sources are within each grid cell nor do we try to estimate it.



Further, we postulate that the number of CO sources is important, but so is the relative proximity of the point measurement to the source. The ARISense were placed in the middle of a cluster of CO sources (biomass cookstoves in the village) while the satellite is seeing sources from those two villages, plus many other villages, plus the less-populated areas between them that make up the majority of the land cover.

- We recognize that estimates from a satellite observation and measurements collected by a single surface instrument are not directly comparable quantities. We emphasize that we do not expect them to perfectly agree, since they are not measuring the same thing, but we can still draw qualitative conclusions by comparing them. And although an area-average and a point value are not directly comparable, they are the only measures available for comparison in this region, so when making this comparison, we offer educated guesses about the reasons for observed differences (i.e., the proximity of the ARISense to a cluster of CO sources – the biomass cookstoves in the villages).

Figure 4: Label individual village sites, be consistent with naming of your sites.

- Data from the individual village sites are not shown in this figure. Data from the individual village sites was averaged in the data points shown. To make this point clearer to the reader, the term “Villages” was renamed to “Village Mean” (defined in Section 2.7).
- Previously in Fig. 4, the ‘Village Mean’ color was identical to the Village 1 color; this was misleading the reader. Now, a different color scheme has been selected to differentiate the Village 1, Village 2, and Village Mean data. Additionally, the naming convention now appears consistently throughout the paper (i.e., Village 1, Village 2, and Village Mean).

Line 651: I assume the ‘relatively short’ and ‘100 hours’ is 100 hours for the whole year? Or is that per biomass burning episode?

- **Approximately 100 hours was the total duration of all biomass burning emission experiments conducted at the conclusion of the 1-year period of ambient sampling.**
- **Original:** “Operation beyond specified conditions, combined with repeated, although relatively short (< 100 hours), exposure to high concentration gases during the post-deployment emissions monitoring experiments, made the three less robust sensors unsuitable for future use.”
- **Revised:** “Operation beyond specified conditions, combined with ~100 hours of exposure to high concentration gases during the post-deployment emissions monitoring experiments, apparently damaged the three less robust sensors (NO, NO<sub>2</sub>, O<sub>x</sub>) and made them unsuitable for future use.”

Also, are you certain that O<sub>3</sub>, NO, and NO<sub>2</sub> came from ‘fresh’ biomass burning emissions, and not other sources (e.g., NO<sub>x</sub> from diesel trucks with poor emission controls)? O<sub>3</sub> from biomass burning not so straightforward and a result of aging of biomass burning emissions.

- We did not sample near diesel vehicles. We agree that, as a secondary pollutant, O<sub>3</sub> was not emitted from the ‘fresh’ (primary) biomass burning emissions.
- The O<sub>x</sub>, NO, and NO<sub>2</sub> sensors were likely indirectly affected by high concentrations of PM and volatile gases (hydrocarbons, formaldehyde, etc.) emitted during the

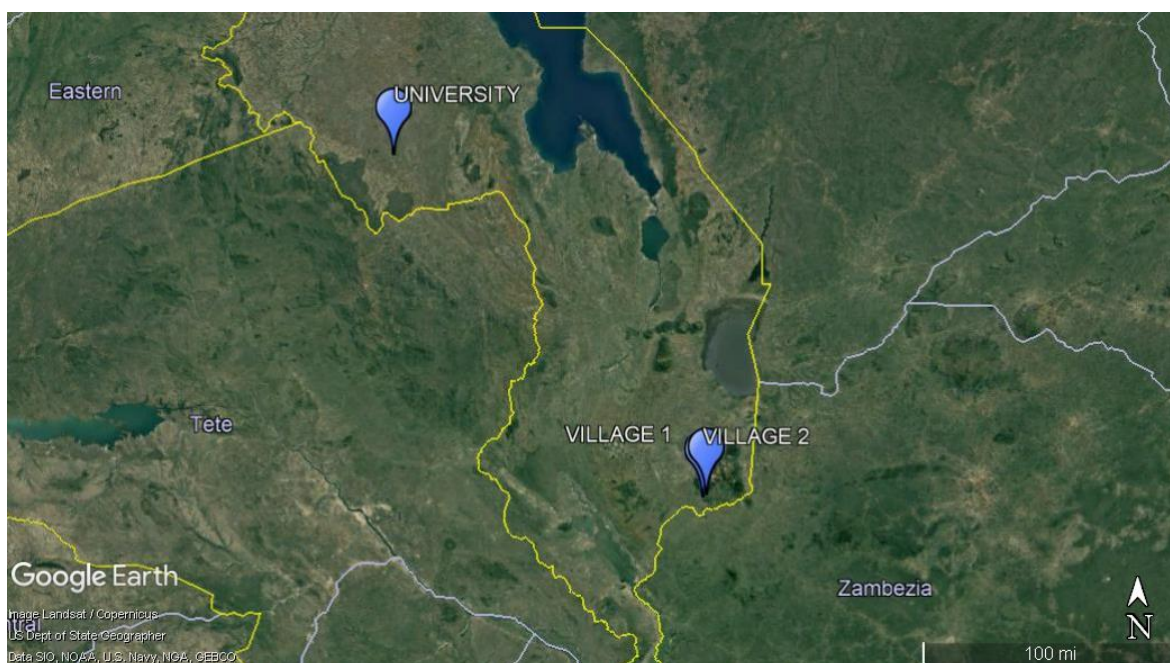
biomass burning experiments, not necessarily high concentrations of the target species ( $O_3$ ,  $NO_x$ ) themselves. Very high concentrations of emissions can chemically degrade or contaminate the sensors, for example, the catalyst or electrolyte can be affected or depleted by repeated interactions with high concentrations of non-target species emissions. Further, if there are high concentrations of fine PM permeating the inlet and flow line, it can condense and block or attenuate the sample flow rate.

- We made this point clearer to the reader by adding more explicit text in Section 3.7:

*“These ambient sensors (except for possibly the CO sensor) were likely affected by high concentrations of PM and volatile gases (e.g., hydrocarbons, formaldehyde, etc.) co-emitted during the biomass burning experiments. Very high concentrations of emissions can chemically degrade or contaminate the sensors, for example, the catalyst or electrolyte can be affected or depleted by repeated interactions with high concentrations of non-target species emissions.”*

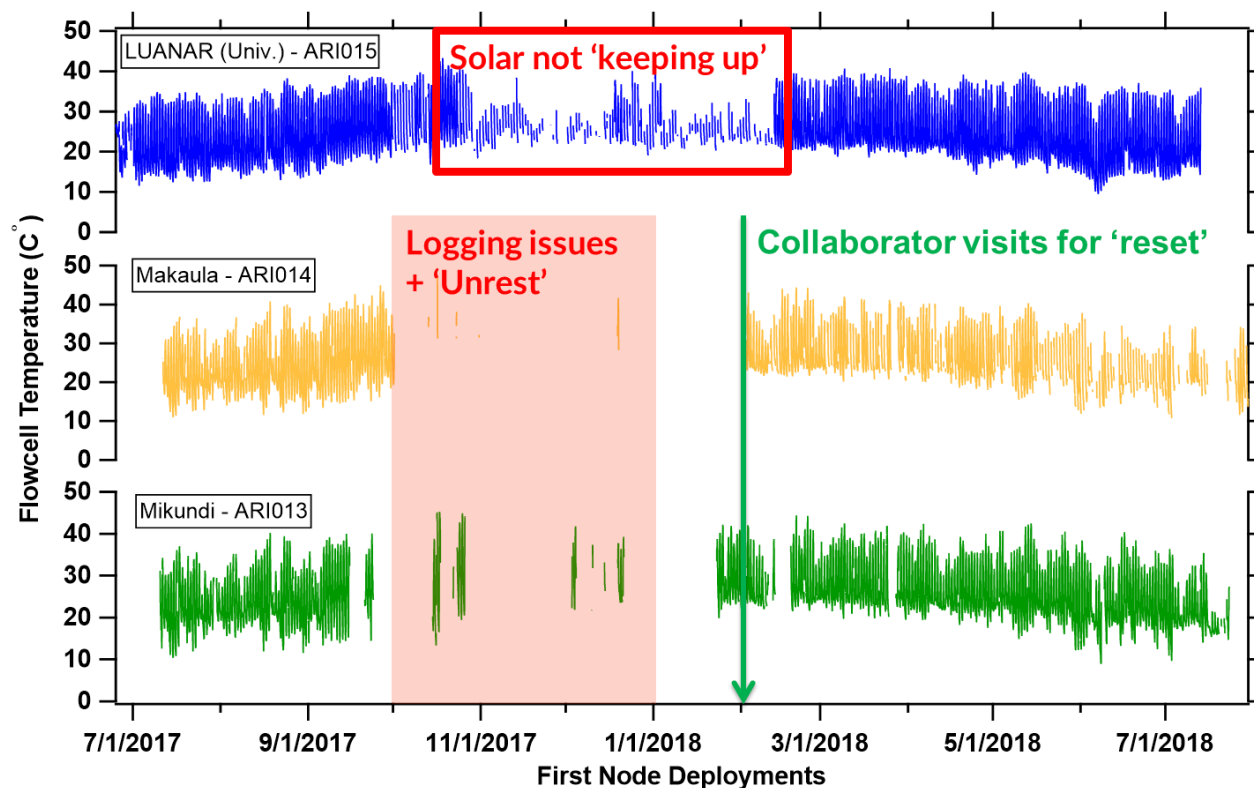
Would be helpful to have a map of the stations & a figure showing when sensors were collocated with and when the instruments weren't collecting data. There are some maps in the supplement, but a nice and simplified summary graphic in the main paper would be good.

- We agree with reviewer and propose adding an additional Figure (Figure 0 below), which would become the 'new' Figure 1 in the revised manuscript. This figure was Fig. S10 in the Supplementary information, however Reviewer #2 suggested we move this map to the main manuscript. Further, we provide a timeline of the project, indicating which ARISense were deployed to which sites, and when they were collocated, deployed to the field, and used for an emissions monitoring campaign.
- We plan to add timeseries of the temperature data from the full year, example given below (Fig. S0), to show when the instruments weren't collecting data.



	May - June 2017	July 2017 - July 2018	July - Aug 2018	Aug 2018 - Mar 2019
<b>ARI013</b>	Colocation (N.C., USA)	Field deployment (Village 2)	Emissions monitoring (Village 2)	Colocation (N.C., USA)
<b>ARI014</b>	Colocation (N.C., USA)	Field deployment (Village 1)	Emissions monitoring (Village 2)	Colocation (N.C., USA)
<b>ARI015</b>	Colocation (N.C., USA)	Field deployment (University)	Emissions monitoring (Village 2)	Colocation (N.C., USA)
<b>ARI023 (OPC-N2)</b>	n/a	n/a	Colocation w/ MicroPEM (Village 2)	n/a

**Figure 0 (previously Fig. S10):** Satellite map of Malawi (top), blue markers indicate ARISense monitoring sites and project timeline (bottom). The timeline indicates the activity conducted for each time period, and the location of that activity is given in parenthesis. Image source: Google Earth Pro Version 7.3.4.8248. University, Village 1, and Village 2, Malawi, Southeastern Africa. Borders and labels layer. Accessed: June 5, 2020.



**Figure S0:** Timeseries of temperature data from ARI015 (top), ARI014 (middle), and ARI013 (bottom) from the full 1-year pilot deployment in Malawi. Gaps in the timeseries indicate periods when the ARISense were not collecting data. Text labels indicate the causes of data loss: ‘solar not keeping up’ refers to insufficient solar power in the winter months; ‘logging issues and unrest’ refer to the combination of corrupted USB devices which failed to log data, and a period of social unrest in the southern region of the country which created unsafe conditions for our assistant to visit the monitors; ‘collaborator visits for reset’ indicate when a collaborator visited the village locations to replace the USB devices and update the firmware.

If possible, a map with direction of known emitters could help (would assist with the discussion in change in sensor performance with wind direction). As it is, emission sources are mentioned but are unclear where and what they are, and how information that they even existed was obtained.

- We revised Fig. S12 and S13 (shown below) to show the known emitters.
- We added the following context (shown in red) to Section 2.6:

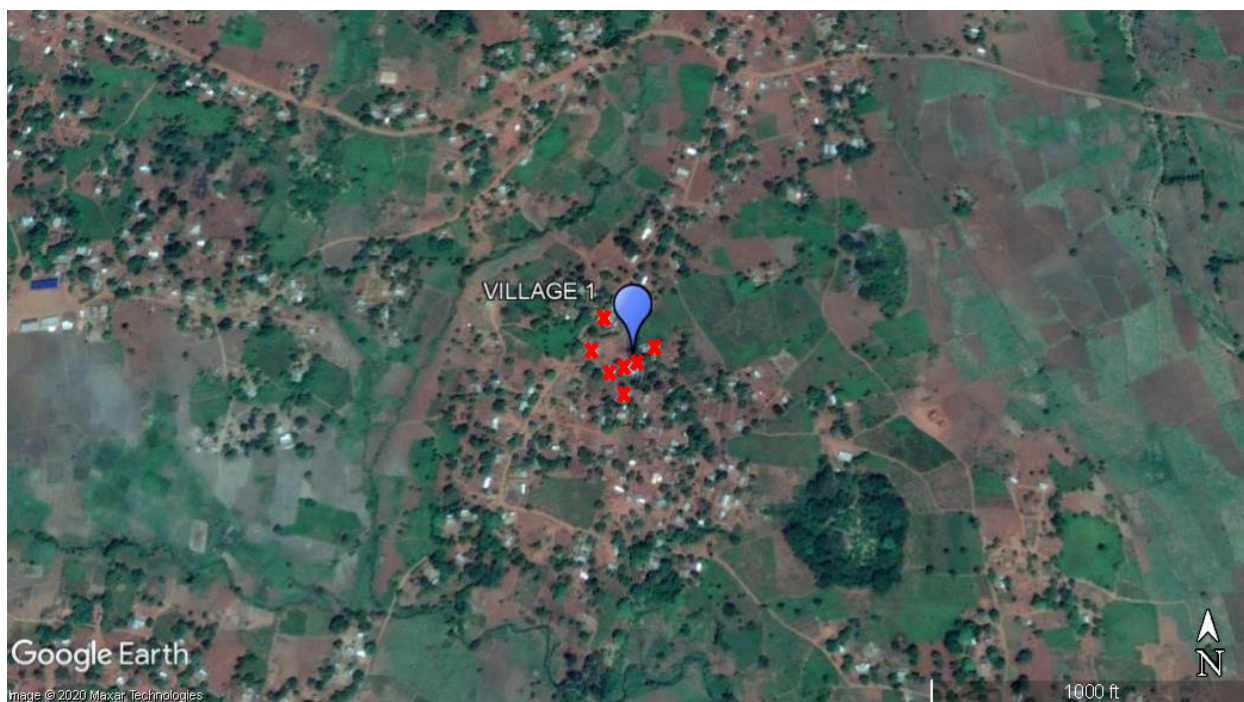
*“ARI013 (“Village 2” site) and ARI014 (“Village 1” site) were deployed < 5 km apart in two rural residential villages in Mulanje, Malawi, adjacent to many households. Almost all rural households in Malawi (99.7%) use solid fuels (e.g., firewood, charcoal) for cooking (National Statistics Office, 2017). Emissions from widespread biomass cookstove use are known to impact local ambient air quality (Aung et al., 2016; Zhou et al., 2011; Amegah and Agyei-Mensah, 2017). Homes regularly using biomass cookstoves within 50 m of the monitoring sites were visually identified at the onset of the study (shown as red cross-hatches on Fig. S12 and S13).”*





**Figure S3:** Satellite image of Village 2 (1000ft scale), blue markers indicate ARISense monitoring sites (ARI013), red crosshatches indicate nearby biomass cookstoves. In addition to the explicitly marked cookstoves, each white square is a household, likely with its own cookstove. ARI013 was deployed to the Village 2 site and was mounted on the roof of the residence of the village chief (4 m above ground) in the Mikundi village of Mulanje, Malawi for 382 days from 6 July 2017 to 23 July 2018. Image source: Google Earth Pro Version 7.3.4.8248. Mikundi village, Mulanje, Malawi. 36.056°S, 35.535°E, eye elevation 626 m. Borders and labels layer. Accessed: June 5, 2020.





**Figure S4:** Satellite image of “Village 1” (1000ft scale); blue markers indicate ARISense monitoring sites (ARI014), red crosshatches indicate nearby biomass cookstoves. In addition to the explicitly marked cookstoves, each white square is a household, likely with its own cookstove. ARI014 was deployed to Village 1 site and was mounted on the roof of the residence of the village chief (4 m above ground) in the Makaula village of Mulanje, Malawi for 384 days from 11 July 2018 to 30 July 2018. Image source: Google Earth Pro Version 7.3.4.8248. Makaula village, Mulanje, Malawi. 16.045°S, 35.555°E, eye elevation 645 m. Borders and labels layer. Accessed: June 5, 2020.

The conclusions could be tightened up. What are the absolute main points your study, in particular, found? The novel work here, as stated at least in the introduction and abstract, is the comparison of different correction models over time. This gets a bit lost in the weeds.

- We thank the reviewer for pointing out that we missed an opportunity to reinforce the main messages of our paper. We restructured and rewrote the conclusions in a way that better connects the objectives to our main findings and takeaways (shown in bolded text):

**“Our experience showed that LCS networks are a viable method to collect novel surface AQ data in regions without reference equipment, but this approach requires strict data quality procedures to ensure the conclusions drawn from the resulting data are valid.** Performance assessment in NC suggested that the ARISense sensor package (excluding the NO<sub>2</sub> sensor) calibrated by four of the five calibration models (excluding MLR) would be appropriate for supplemental monitoring based on U.S. EPA guidelines. However, performance during the pre-deployment NC assessment did not reflect performance in Malawi. For this deployment site, **we found that detailed information about nearby sources and their diurnal emission patterns, ambient meteorological data, and a familiarity with air pollutant behavior were helpful when qualitatively assessing LCS performance in a region where quantitative**

**assessment was not an option.** A lack of coherency in diurnal trends between calibration model predictions and frequent non-physical concentration values (Fig. 3) showed that LCS measurements made in deployment environments different from the collocation environment can be unreliable and may lead to biased information about the deployment environment. For example, although the  $O_x$  sensors showed the highest performance of all sensor types during collocation and the measured RH, temperature, and  $O_x$  voltage ranges were similar in the collocation and deployment environments, the calibrated  $O_3$  data in Malawi were unreliable. The collocation data were collected in an urban area near a highway and the deployment data were collected in a rural area heavily impacted by biomass burning emissions. This difference in ozone precursor emissions could have contributed to the poor performance of the calibration models in the deployment environment. We expect our experience in Malawi may generalize to other regions, suggesting that additional research is needed to address the issue of LCS calibration for secondary pollutants.

**Ultimately, we found that the kNN hybrid modelling approach performed the best in the U.S. and when applied to data collected in Malawi.** However, the general lack of standardization in LCS calibration and assessment approaches complicated and extended this process for our study. Although there have been advancements in calibration methods, the difficulty of identifying and applying a singular best in-field calibration model remains a common issue among LCS users (Topalović et al., 2019; Lewis and Edwards, 2016; Giordano et al., 2021). From an end user perspective, the burden of calibration easily becomes overwhelming; there is presently no clear guidance on which model would be appropriate for which sensor under which circumstances. This limits the potential user base of LCS technologies, complicates our ability to generalize findings across different studies, and may even lead to poor quality measurements. Given the wide range in potential LCS technologies and deployment conditions, it is not possible to fully generalize the viability and sensitivity of the ARISense to another LCS package deployed in a different area. Nonetheless, we surmise LCS are most useful when they are carefully selected and calibrated for a single purpose and location, for which the environmental and pollutant conditions are at least partially characterized.

**This pilot deployment also provided lessons regarding the design and deployment of low-cost AQ monitoring systems for off-grid applications.** The ARISense packages survived the 1 year deployment to Malawi and enabled collection of a large, novel dataset, however they suffered individual sensor failures and frequent power losses. Given that 20 to 50% of the deployment data were lost due to insufficient power and corrupt data storage systems, for future solar-powered deployment efforts we suggest that the power system be designed to allow for primary and secondary data recovery goals (i.e., a back-up plan to prioritize the most desirable data in the event of insufficient power). Further, we were frequently restricted in troubleshooting and repair operations by spotty cellular connection, limited human resources, and our inability to remotely locate and procure appropriate equipment. A repair kit with basic equipment (e.g., pre-programmed USB devices, alternate SIM cards, hand tools with attachments specific to each LCS, etc.) stored in a nearby, secure location would have allowed for quicker troubleshooting and repair. We suggest that in addition to solar power limitations, other potential confounding factors like extreme weather and limited technical capacity and assistance availability be considered before deployment to remote locations. We found that the more closely located the monitor was to a trained local assistant, the lower the overall data losses were.

**The responses of the LCS were not remarkably different after 1 year in the field (Fig. S29-30), assuming actual concentrations did not vary significantly from 2017 to 2018.** However, except for CO, repeated exposure to high-concentration biomass emissions completely degraded the sensors. Key manufacturer specifications indicated that the CO sensor was the most robust. the CO sensor exposure limit was 40 times higher than that of the O<sub>x</sub>, NO, and NO<sub>2</sub> sensors. Further, the maximum temperature and RH range for the CO sensor was 50°C and 90%, respectively, and 40°C and 85% for the O<sub>x</sub>, NO, and NO<sub>2</sub> sensors. During deployment, the maximum ranges were occasionally exceeded for every sensor except CO. **Operation beyond specified conditions, combined with ~100 hours of exposure to high concentration gases during the post-deployment emissions monitoring experiments, apparently damaged the three less robust sensors (NO, NO<sub>2</sub>, O<sub>x</sub>) and made them unsuitable for future use.** We caution end users to carefully select an appropriate sensor package given pilot information about the emission sources in their target site.

A growing body of literature highlights the potential value of LCS technologies for Sub-Saharan Africa and other low-resource settings (Subramanian and Garland, 2021; Wernecke and Wright, 2021; Rahal, 2020; Sewor et al., 2021; Awokola et al., 2020). **We found that our LCS surface observations were largely consistent with the only other available data sources in this region (remote sensing data and model products) and data from similar studies across SSA. This suggests LCS have a key role to play in providing reliable information on general air quality conditions and trends in regions without a historical record.** Advancements in machine learning techniques show how LCS can be used for source identification and attribution in regions where little quantitative information currently exists on dominant emission sources (Hagan et al., 2019; Thorson et al., 2019). While LCS in SSA show promise, many of the issues experienced in this study stemmed from a lack of in situ reference monitors. Additional reference grade monitors throughout the region may help circumvent some issues related to calibration modelling and quality assurance. A regional, shared facility would enable periodic, regionally representative collocations without requiring every country to establish its own regulatory network. Recent research has improved our ability to synthesize data from networks of LCS through computational calibration solutions which minimize the need to transport and collocate each individual monitor separately and increase the spatiotemporal resolution beyond that of reference networks (Buehler et al., 2021; Malings et al., 2019a; Kelly et al., 2021; Considine et al., 2021; Sahu et al., 2021). Concurrently, policy-focused researchers are helping to bridge the gap between governments and AQ scientists by creating comprehensive frameworks which provide systematic procedures to establish AQ monitoring networks in low and middle income countries (Gulia et al., 2020; Pinder et al., 2019). In the meantime, we found support from local universities, which helped maintain the pilot deployment of this LCS network. We expect that any AQ program in SSA will benefit from building long-term, local capacity and knowledge transfer systems for training on-site staff and for receiving their feedback and guidance.”



## References

- Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, *Atmos. Chem. Phys.*, 11, 4039–4072, <https://doi.org/10.5194/acp-11-4039-2011>, 2011.
- Amegah, A. K. and Agyei-Mensah, S.: Urban air pollution in Sub-Saharan Africa: Time for action, *Environmental Pollution*, 220, 738–743, <https://doi.org/10.1016/j.envpol.2016.09.042>, 2017.
- Aung, T. W., Jain, G., Sethuraman, K., Baumgartner, J., Reynolds, C. C., Grieshop, A. P., Marshall, J. D., and Brauer, M.: Health and Climate-Relevant Pollutant Concentrations from a Carbon-Finance Approved Cookstove Intervention in Rural India, *Environ. Sci. Technol.*, 50, 7228–7238, <https://doi.org/10.1021/acs.est.5b06208>, 2016.
- Buchholz, R., Deeter, M., Worden, H., Gille, J., Edwards, D., Hannigan, J., Jones, N., Paton-Walsh, C., Griffith, D., Smale, D., Robinson, J., Strong, K., Conway, S., Sussmann, R., Hase, F., Blumenstock, T., Mahieu, E., and Langerock, B.: Validation of MOPITT carbon monoxide using ground-based Fourier transform infrared spectrometer data from NDACC, 1927–1956, <https://doi.org/10.5194/amt-10-1927-2017>, 2017.
- Dionisio, K. L., Arku, R. E., Hughes, A. F., Vallarino, J., Carmichael, H., Spengler, J. D., Agyei-Mensah, S., and Ezzati, M.: Air Pollution in Accra Neighborhoods: Spatial, Socioeconomic, and Temporal Patterns, 44, 2270–2276, <https://doi.org/10.1021/es903276s>, 2010.
- Emmons, L. K., Deeter, M. N., Gille, J. C., Edwards, D. P., Attié, J.-L., Warner, J., Ziskin, D., Francis, G., Khattatov, B., Yudin, V., Lamarque, J.-F., Ho, S.-P., Mao, D., Chen, J. S., Drummond, J., Novelli, P., Sachse, G., Coffey, M. T., Hannigan, J. W., Gerbig, C., Kawakami, S., Kondo, Y., Takegawa, N., Schlager, H., Baehr, J., and Ziereis, H.: Validation of Measurements of Pollution in the Troposphere (MOPITT) CO retrievals with aircraft in situ profiles, 109, <https://doi.org/10.1029/2003JD004101>, 2004.
- Emmons, L. K., Edwards, D. P., Deeter, M. N., Gille, J. C., Campos, T., Nédélec, P., Novelli, P., and Sachse, G.: Measurements of Pollution In The Troposphere (MOPITT) validation through 2006, 9, 1795–1803, <https://doi.org/10.5194/acp-9-1795-2009>, 2009.
- Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., Smith, K. R., Buckley, J. W., Ellis, M., Gillot, S. R., and White, A.: Evaluating the performance of low cost chemical sensors for air pollution research, *Faraday Discuss.*, 189, 85–103, <https://doi.org/10.1039/C5FD00201J>, 2016.
- Long, R. W., Whitehill, A., Habel, A., Urbanski, S., Halliday, H., Colón, M., Kaushik, S., and Landis, M. S.: Comparison of ozone measurement methods in biomass burning smoke: an evaluation under field and laboratory conditions, 14, 1783–1800, <https://doi.org/10.5194/amt-14-1783-2021>, 2021.
- 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, 12, 903–920, <https://doi.org/10.5194/amt-12-903-2019>, 2019.

McFarlane, C., Isevulambire, P. K., Lumbuenamo, R. S., Ndinga, A. M. E., Dhammapala, R., Jin, X., McNeill, V. F., Malings, C., Subramanian, R., and Westervelt, D. M.: First Measurements of Ambient PM<sub>2.5</sub> in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-calibrated Low-cost Sensors, *Aerosol Air Qual. Res.*, 21, 200619–200619, <https://doi.org/10.4209/aaqr.200619>, 2021.

National Statistics Office: Household Socio-economic Characteristics Report, Republic of Malawi, 2017.

Popoola, O. A. M., Stewart, G. B., Mead, M. I., and Jones, R. L.: Development of a baseline-temperature correction methodology for electrochemical sensors and its implications for long-term stability, *Atmospheric Environment*, 147, 330–343, <https://doi.org/10.1016/j.atmosenv.2016.10.024>, 2016.

Subramanian, R., Kagabo, A. S., Baharane, V., Guhirwa, S., Sindayigaya, C., Malings, C., Williams, N. J., Kalisa, E., Li, H., Adams, P., Robinson, A. L., DeWitt, H. L., Gasore, J., and Jaramillo, P.: Air pollution in Kigali, Rwanda: spatial and temporal variability, source contributions, and the impact of car-free Sundays, 30, <https://doi.org/10.17159/caj/2020/30/2.8023>, 2020.

Tohir, A. M., Venkataraman, S., Mbatha, N., Sangeetha, S. K., Bencherif, H., Brunke, E.-G., and Labuschagne, C.: Studies on CO variation and trends over South Africa and the Indian Ocean using TES satellite data, 111, 1–9, 2015.

Yurganov, L., McMillan, W., Grechko, E., and Dzhola, A.: Analysis of global and regional CO burdens measured from space between 2000 and 2009 and validated by ground-based solar tracking spectrometers, 10, 3479–3494, <https://doi.org/10.5194/acp-10-3479-2010>, 2010.

Yurganov, L. N., McMillan, W. W., Dzhola, A. V., Grechko, E. I., Jones, N. B., and Werf, G. R. van der: Global AIRS and MOPITT CO measurements: Validation, comparison, and links to biomass burning variations and carbon cycle, 113, <https://doi.org/10.1029/2007JD009229>, 2008.

Zhou, Z., Dionisio, K. L., Arku, R. E., Quaye, A., Hughes, A. F., Vallarino, J., Spengler, J. D., Hill, A., Agyei-Mensah, S., and Ezzati, M.: Household and community poverty, biomass use, and air pollution in Accra, Ghana, *PNAS*, 108, 11028–11033, <https://doi.org/10.1073/pnas.1019183108>, 2011.