Performance Characterization of Low-cost Air Quality Sensors for Off-grid Deployment in Rural Malawi

Ashley S. Bittner[1], Eben S. Cross[2], David H. Hagan[2], Carl Malings[3], Eric Lipsky[4], and Andrew Grieshop[1]

[1]Department of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC 27606, USA
[2]QuantAQ, Inc., Somerville, MA 02143, USA
[3]NASA Postdoctoral Program Fellow, Goddard Space Flight Center, Greenbelt, MD 20771, USA
[4]Department of Energy Engineering, Penn State Greater Allegheny University, McKeesport, PA 15132, USA

Correspondence to: Andrew Grieshop (apgriesh@ncsu.edu)

Abstract. Low-cost gas and particulate sensor packages offer a compact, lightweight, and easily transportable solution to address global gaps in air quality (AQ) observations. However, regions that would benefit most from widespread deployment of low-cost AQ monitors often lack the reference grade equipment required to reliably calibrate and validate them. In this study, we explore approaches to calibrating and validating three integrated sensor packages before a one~~1~~ year deployment to rural Malawi using collocation data collected at a regulatory site in North Carolina, USA. We compare the performance of five computational modelling approaches to calibrate the electrochemical gas sensors: k-Nearest Neighbor (kNN) hybrid, random forest (RF) hybrid, high-dimensional model representation (HDMR), multilinear regression (MLR), and quadratic regression (QR). For the CO, O_x, NO, and NO_2 sensors, we found that kNN hybrid models returned the highest coefficients of determination and lowest error metrics when validated.~~;~~ Hybrid models~~they~~ also appeared to be the most transferable approach when applied to deployment~~field~~ data collected in Malawi. We compared kNN-hybrid calibrated CO observations from two regions in Malawi to remote sensing data ~~in two regions in Malawi~~ and found qualitative agreement in spatial and annual trends. However, ARISense~~the~~ monthly mean surface observations were 2 to 4 times higher than the remote sensing data, possibly due to proximity to residential biomass combustion~~small scale combustion~~ activity not resolved by satellite imaging. We also compared the performance of the integrated Alphasense OPC-N2 optical particle counter to a filter-corrected nephelometer using collocation data collected at one of our deployment sites in Malawi. We found the performance of the OPC-N2 varied widely with environmental conditions, with the worst performance associated with high relative humidity (RH > 70%) conditions and influence from emissions from nearby residential biomass combustion~~cookstoves~~. We did not find obvious evidence of systematic sensor performance decay after the one~~1~~ year deployment to Malawi.~~;~~ ~~however,~~ O~~o~~verall data recovery was limited by insufficient power and access to technical resources at deployment sites. Future low-cost sensor deployments to rural Sub-Saharan Africa would benefit from adaptable power

systems, standardized sensor calibration methodologies, and increased ~~regulatory grade~~ regional regulatory grade monitoring infrastructure.

1 Introduction

Ambient air pollution is a leading cause of morbidity and premature mortality in Sub-Saharan Africa (SSA) (Murray et al., 2020). ~~Sources of~~ Aair pollution in SSA is~~are~~ expected to increase over time given ~~the~~ regional growth in population and energy demand combined with~~,~~ a biomass fuel dominated energy mix~~, and slash and burn agricultural practices~~ (Shikwambana and Tsoeleng, 2020; Stevens and Madani, 2016; Liousse et al., 2014; Amegah and Agyei-Mensah, 2017). However, regulatory air quality (AQ) monitoring is uncommon in many SSA countries, partially due to the high cost of reference grade equipment (Amegah, 2018; Petkova et al., 2013). Remote sensing is a valuable tool to address these data gaps, but satellite observations alone have various shortcomings relative to in situ measurements (Martin et al., 2019). Additional validation with reliable surface measurements is required, particularly in SSA (Malings et al., 2020). In the meantime, low-cost gas and particulate sensor packages provide an affordable, compact, and easily transportable approach to supplement air quality networks in regions where reference grade instrumentation is not accessible. Malawi, located in southeastern Africa, provides a relevant context to investigate how low-cost sensors (LCS) can be used to address the global dearth of AQ observations. The Malawi Bureau of Standards published ambient air quality limits based on World Health Organization guidelines in 2005 (Mapoma and Xie, 2013; MBS, 2005), but there is no regulatory air quality monitoring program in the country to date. Previous studies of~~n~~ AQ in Malawi have primarily focused on indoor air quality or were unable to capture long-term trends (Fullerton et al., 2009, 2011; Jary et al., 2017; Mapoma and Xie, 2013). A reliable and~~,~~ affordable LCS monitoring network in Malawi could provide data to monitor the evolution of air quality and establish baselines for future AQ management.

Given the potential applications, LCS deployments are becoming ~~increasingly~~ common (Giordano et al., 2021). However, as the cost of LCS decreases, so may the selectivity, linearity, and accuracy. Electrochemical gas sensors are prone to interference and cross-sensitivities.~~;~~ Interference occurs when sensors ~~they may~~ respond to changes in temperature (T) and relative humidity (RH). ~~or~~ Cross-sensitivities occur when sensors respond to the presence of gases other than the target analyte (Lewis et al., 2016; Mead et al., 2013). Failure to properly account for these during calibration can result in substantial measurement error under ambient conditions (Lewis et al., 2016; Cross et al., 2017; Castell et al., 2017; Mead et al., 2013). The calibration and application of LCS technologies to augment existing regulatory monitoring networks has been widely explored (Cross et al., 2017; Hagan et al., 2018; Malings et al., 2019a, b; Mead et al., 2013; Zimmerman et al., 2018; Li et al., 2021), but historically there has been little standardization in calibration approach or performance evaluation (Castell et al., 2017; Duvall et al., 2021; Morawska et al., 2018; Rai et al., 2017). In response to this, the U.S. Environmental Protection Agency (EPA) recently released two reports outlining testing protocols, metrics, and target values to evaluate the

performance of ozone and fine particulate matter ($PM_{2.5}$) sensors for non-regulatory supplemental and informational monitoring applications in the U.S. (Duvall et al., 2021a, b). Unfortunately, there is no similar guidance for validating LCS for deployments in settings without in situ regulatory monitors. The deployment and evaluation ~~of LCS~~of LCS packages in areas without existing AQ monitoring infrastructure is a growing research area ~~limited, but is becoming increasingly common~~ (Chatzidiakou et al., 2019; Hagan et al., 2019; Subramanian et al., 2020, 2018). A lack of in situ regulatory monitors requires collocation, calibration, and validation at another site, potentially under a set of environmental conditions different from those of the target deployment environment. Advancements in laboratory chamber calibration may help resolve this issue.~~;~~ ~~studies have shown that~~ In a controlled environment, gas sensors can be exposed to and calibrated for a range of environmental conditions (i.e., gas concentration, RH, T, pressure, etc.), which may allow LCS cross-sensitivity and interference to be measured and controlled for before deployment (Williams et al., 2014b; Spinelle et al., 2016; Lewis et al., 2016; Spinelle et al., 2015). However, studies of low-cost particle sensors have observed better performance under laboratory versus field conditions (Rai et al., 2017). For example, previous long-term field assessments of the Alphasense OPC-N2 optical particle counter have observed large variability with changing seasons, environmental conditions, and background pollution levels (Bulot et al., 2019; Rai et al., 2017; Sousan et al., 2016). Low cost optical particle sensors can systematically overestimate mass concentrations under high RH (>75%) conditions due to hygroscopic growth of the particles (Crilley et al., 2018; Di Antonio et al., 2018), with errors ranging from 100 to 500% depending on ~~the~~ aerosol hygroscopicity (Hagan and Kroll, 2020). Further, the complex chemical, physical, and optical properties of aerosol can complicate the field evaluation of low-cost particle sensors. For the Alphasense OPC-N2, particle composition may impact the sensor output by as much as a factor of 30 (Rai et al., 2017; Sousan et al., 2016). A recent modelling effort by Hagan and Kroll (2020) found that the optical properties and particle size distribution of the source aerosol can result in errors of up to 100% and 90%, respectively, in mass measurements made by low-cost optical particle sensors. Measurement errors were highest for strongly absorbing aerosol dominated by small (< 300 nm) particles. These traits can be characteristic of aerosol emitted by biomass-burning (Reid et al., 2005), a dominant source of ambient PM throughout SSA (Marais and Wiedinmyer, 2016; Queface et al., 2011; Liousse et al., 2014). Therefore, s~~S~~tringent quality assurance is necessary to ensure the validity of LCS particle measurements in this environment.

In this study, we calibrated and evaluated the "ARISense", a moderate-cost, integrated gas, particle, and meteorological sensor package (Aerodyne, Inc.) for long-term field deployment to Malawi. Our overarching goal was to assess the viability of augmenting and maintaining a small, temporary network of LCS monitors, until a more formal governmental regulatory monitoring system can be established. ~~where~~ Given that comparison to regulatory grade equipment in Malawi was ~~is not currently~~ not possible, the objective of this work was to devise an alternative methodology to evaluate the ARISense technology (Section 2.1) for accuracy, precision, and stability over the 1-year pilot deployment. ~~.~~In Section 2.3 and 2.4, we describe collocations of the gas sensors (in North Carolina, USA) and particle sensor (in Mulanje, Malawi) with reference or semi-reference instruments (described in Section 2.2). We use collocation data and quantitative assessment metrics

(described in 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 (described in Section 2.6), we qualitatively assess how the ARISense performed in the field using contextual information about nearby emission sources, diurnal trends, 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 deployment results to remote sensing and reanalysis data products and to 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 one year apart at the same location. In concluding (Section 4), we draw on these pilot results to characterize the benefits, limitations, and robustness of this technology and methodology for our application: collecting AQ data in under-studied and -resourced regions. Additionally, we offer guidance on considerations to improve future remote deployment efforts. ~~Our objectives are to 1) evaluate various modelling approaches to calibrate the gas sensors using field collocation data from North Carolina, USA., 2) identify one modelling approach that is best suited to interpret 1 year of deployment data collected in Malawi, 3) estimate error in the particle sensor using a mass-corrected nephelometer, 4) compare calibrated observations in Malawi to remote sensing data products and surface measurements from similar environments, 5) characterize the stability and longevity of the sensors and calibration models after 1 year in the field and after repeated exposure to high concentration emissions, and 6) provide contextual evidence of the benefits, limitations, and durability of this technology and methodology for our application. Lastly, we provide guidance on considerations to improve future remote deployment efforts.~~ Detailed analysis and discussion of several~~the full~~ years of data ~~measurements~~ collected in Malawi will be presented in a forthcoming complementary publication.

2 Methods

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 deployment site in Malawi to assess the performance of the integrated OPC-N2. ~~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.~~ Instrumentation, collocation, and calibration are covered in Sect. 2.1 – 2.4. Performance assessment metrics are given in Sect. 2.5. Calibrated ARISense were deployed to Malawi (Sect. 2.6) and compared to remote sensing data products (Sect. 2.7).

2.1 ARISense sensor packages

The ARISense package (Fig. S1) integrated the following sensors from Alphasense Ltd., UK: carbon monoxide (CO-B4), nitric oxide (NO-B4), nitrogen dioxide (NO2-B43F), total oxidants (Ox-B421), and the OPC-N2 optical particle counter. The ARISense reported voltage readings from electrochemical gas sensor working electrodes (WE) and auxiliary electrodes (AE).~~:~~ ~~the~~ ~~S~~sensor differential voltage (ΔV) was calculated as WE – AE. The Alphasense OPC-N2 recorded counts in 16

size bins spanning particle diameters from 0.38 to 17.5 µm, mainly coarse (> 2 µm) and some accumulation mode (0.1 to 2 µm) aerosols (Badura et al., 2018; Crilley et al., 2018; Sousan et al., 2016). Although the OPC-N2 has~~d~~ embedded algorithms to convert count measurements into mass concentrations of $PM_{1.0}$, $PM_{2.5}$ and PM_{10} (particulate matter with aerodynamic diameters less than 1.0, 2.5, and 10 µm, respectively), ~~here~~ the bin count data were manually integrated, converted to number concentration (cm^{-3}) assuming unity measurement efficiency across the bin range, and then to mass concentration assuming spherical particles with uniform density (1.65 g cm^{-3}). The values reported for $PM_{2.5}$ are $PM_{2.5}$~~.~~; ~~The~~ The location of the adjacent bin separations ~~(~~at 2.0 and 2.99 µm~~)~~ did not allow for direct estimates of $PM_{2.5}$. However, this was only one of many contributing sources of error in approximating true mass concentration with the Alphasense OPC-N2. Given the minimum cut-off diameter, we were unable to measure (nor did we try to estimate) the mass from particles smaller than 0.38 µm.

We used four ARISense monitors in this study: serial numbers ARI013, ARI014, ARI015 (Version 1.0, 2017), and ARI023 (Version 2.0, 2018). The monitors were powered by solar panels charging external batteries and recorded data to an internal USB device. Details and images are provided in Sect. 1 of Supplementary Information. ~~Supplementary Information, Sect. 1.~~ Additional environmental and meteorological sensors (i.e., T, RH, pressure, solar intensity, and noise) and system design are described in Cross et al. (2017).

2.2 Reference instrumentation

Gas concentration measurements for NO_x/NO/NO_2 (Teledyne Model T200UP), CO (Thermo Scientific Model 48i-TLE), and Ozone (Ecotech Federal Equivalent Method instrument) were obtained from reference instruments operated by the North Carolina Department of Environmental Quality (NC-DEQ) and the Environmental Protection Agency (EPA).

The semi-reference MicroPEM (RTI International) instrument was used to assess the performance of the OPC-N2 in Malawi. The MicroPEM, equipped with T and RH sensors, sampled (0.50 L/min, 100% duty cycle) via a $PM_{2.5}$ inlet into a nephelometer (0.1 Hz) and 25 mm PTFE filter. 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 ~~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~~ (Du et al., 2019; Williams et al., 2014a). However, deployments observed baseline (zero) drift and poor performance at RH conditions above 94% (Williams et al., 2014a; Zhang et al., 2018). To account for baseline drift, the MicroPEM was zeroed before each deployment using a HEPA filter. Additional details on the MicroPEM sensor, filter analysis, and quality assurance are provided in Sect. 1 of Supplementary Information.~~in Supplementary Information, Sect. 1.~~

2.3 Gas sensor collocation and calibration

Before deployment to Malawi, ARI013, ARI014, and ARI015 were collocated with EPA and NC-DEQ reference instruments (Fig. S2) at a near-highway site near Durham, North Carolina, USA (35.865°N, 78.820°W) between 29 May and 15 June 2017 (boreal summer – warm, mild season). ARI013 and ARI014 were collocated for 17 days. ARI015 was collocated for only 8 days due to a defect identified early in the deployment. All data were recorded at 1 minute resolution. Collocation site details are provided in Sect. 2 of Supplementary Information.

~~After their return from Malawi, ARI013 and ARI014 were collocated at the same site from 22 August 2018 and 20 March 2019. The Ecotech Ozone monitor at this site was no longer operational. As a proxy, we obtained ozone data (Thermo Environmental Instruments, Inc. Model 49i) from the Millbrook NC-DEQ site (35.856°N, 78.574°W), located < 20 km away. ARI015 remained in Malawi and was relocated to a new site. Collocation details are provided in Supplementary Information, Sect. 2.~~

The pre-deployment collocation data were used to train, test, and compare the performance of five calibration models to convert the raw voltage data to concentration units and to account for sensor interference and cross-sensitivities. Outlying data points in the raw ARISense gas sensor voltage data due to noise and power cycling were visually identified and removed.~~;~~ R~~r~~aw NO sensor data collected within 8 hours of a power cycle were also removed due to the extended warmup time of the NO-B4 sensor. ARISense data were time aligned with the reference data and both datasets were averaged to 5-min resolution. A random 70% of the collocation data were used for model training and the remaining 30% were withheld for testing. Performance a~~A~~ssessment metrics were calculated only for the withheld data.

~~**Table 1:** Description of the five calibration modelling approaches and data inputs for each gas sensor and model combination (CO = carbon monoxide, NO = nitrogen oxide, NO_2 = nitrogen dioxide, O_3 = ozone). ΔV is the voltage difference between the working electrode (WE) voltage and the auxiliary electrode (AE) voltage measured by each electrochemical gas sensor; RH = relative humidity, T = temperature, DP = dew point.~~

~~Gas Sensor~~	~~Data Inputs to Model~~	~~Model(s) applied~~
~~CO~~	~~CO ΔV, RH, T, & DP~~	~~All~~
~~NO~~	~~NO ΔV, RH, T, DP, & NO WE[a]~~	~~All except QR~~
~~NO_2~~	~~NO_2 ΔV, RH, T, & DP~~	~~All except QR~~
~~O_3~~	~~O_x ΔV, DP, & NO2 ΔV[b]~~	~~All except QR~~

~~[a]kNN hybrid only~~

[b]RF hybrid only

Individual calibration models were built for each gas sensor (O_{x3}, NO, NO_2, CO) in each monitor (ARI013, ARI014, ARI015) using five modelling approaches: k-Nearest Neighbor (kNN) hybrid (Hagan et al., 2018), Random Forest (RF) hybrid (Malings et al., 2019a), High-Dimensional Model Representation (HDMR) (Cross et al., 2017), quadratic regression (QR) (Malings et al., 2019a), and multi-linear regression (MLR). The five models were selected for consideration 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 levels of some pollutants to be higher in Malawi than during calibration in NC, we expected kNN hybrid models to be well suited for our application. Further, the kNN hybrid approach is 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. (2019), 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.

Table 1: Calibration modelling inputs for each gas sensor (CO = carbon monoxide, NO = nitrogen oxide, NO_2 = nitrogen dioxide, O_x = oxidants) and model combination ('All' indicates k-nearest neighbor (kNN) hybrid, random forest (RF) hybrid, high-dimensional model representation (HDMR), multi-linear regression (MLR), and quadratic regression (QR)."ΔV is the voltage difference between the working electrode (WE) voltage and the auxiliary electrode (AE) voltage measured by each electrochemical gas sensor, RH = relative humidity, T = temperature, DP = dew point.

Gas Sensor	Data Inputs to Model	Model(s) applied
CO	CO ΔV, RH, T, & DP	All
NO	NO ΔV, RH, T, DP, & NO WE[a]	All except QR
NO_2	NO_2 ΔV, RH, T, & DP	All except QR
O_x	O_x ΔV, DP, & NO2 ΔV [b]	All except QR

[a]kNN hybrid only

[b]RF hybrid only

The modelling ~~approaches~~ inputs are summarized in ~~Table 1~~. O_3 models were designed to account for sensor cross-sensitivity to NO_2 (Cross et al., 2017). Note that references to 'O_3' indicate estimates made from calibrating the O_x sensor data. References to 'O_x' indicate raw voltage measurements from the total oxidant sensor. 'Ozone' is used when referring to the gaseous air pollutant. For our study, the CO HDMR models were set to allow only first-dimensional interactions, as second-order interactions led to spurious results for data collected outside the bounds of training data (see Sect. 3.3 - on deployment conditions). For the CO sensors, this effectively made the HDMR model equivalent to the MLR model. Therefore, the statistical metrics achieved by both models were identical and are shown as overlaid points on Fig. 2a.

2.4 OPC-N2 collocation and calibration

ARI023 was collocated with a MicroPEM in an ambient, combustion source-influenced environment on a house rooftop (4 m above ground level) in Mikundi village in Mulanje District, Malawi (16.056°S, 35.535°E) between 25 July 2018 and 7 August 2018 (austral winter – cool, dry season). We collected 130 hours of collocation data over three multi-day collection periods (i.e., 3 PTFE filters). A 75% completeness requirement was applied before the raw 1 min data were averaged to 1 h and 24 h intervals. Sub-daily averaging intervals were used to assess the OPC-N2 for near real-time (1 min) and diurnal trend (1 h) monitoring applications. A bin-wise RH-correction algorithm based on κ-Köhler theory was applied to correct for hygroscopic growth under high RH conditions, initially assuming particle density (ρ) equal to 1.65 g cm^{-3} and aerosol hygroscopicity (κ) of 0.6 (Di Antonio et al., 2018). To observe sensitivity of this correction to the assumed hygroscopicity, the density was held constant at 1.65 g cm^{-3} and the κ value was varied (κ = 0.15, 0.6, and 1). To observe variability due to the assumed source of the aerosol, the density and hygroscopicity were varied to approximate ammonium nitrate, dust, wildfire, and background aerosols. Aerosol property assumptions (κ and density) are based on Hagan and Kroll (2020) & Petters and Kreidenweis (2007). ~~More detail is available in Supplementary Information Sect. 3.~~

2.5 Assessment metrics

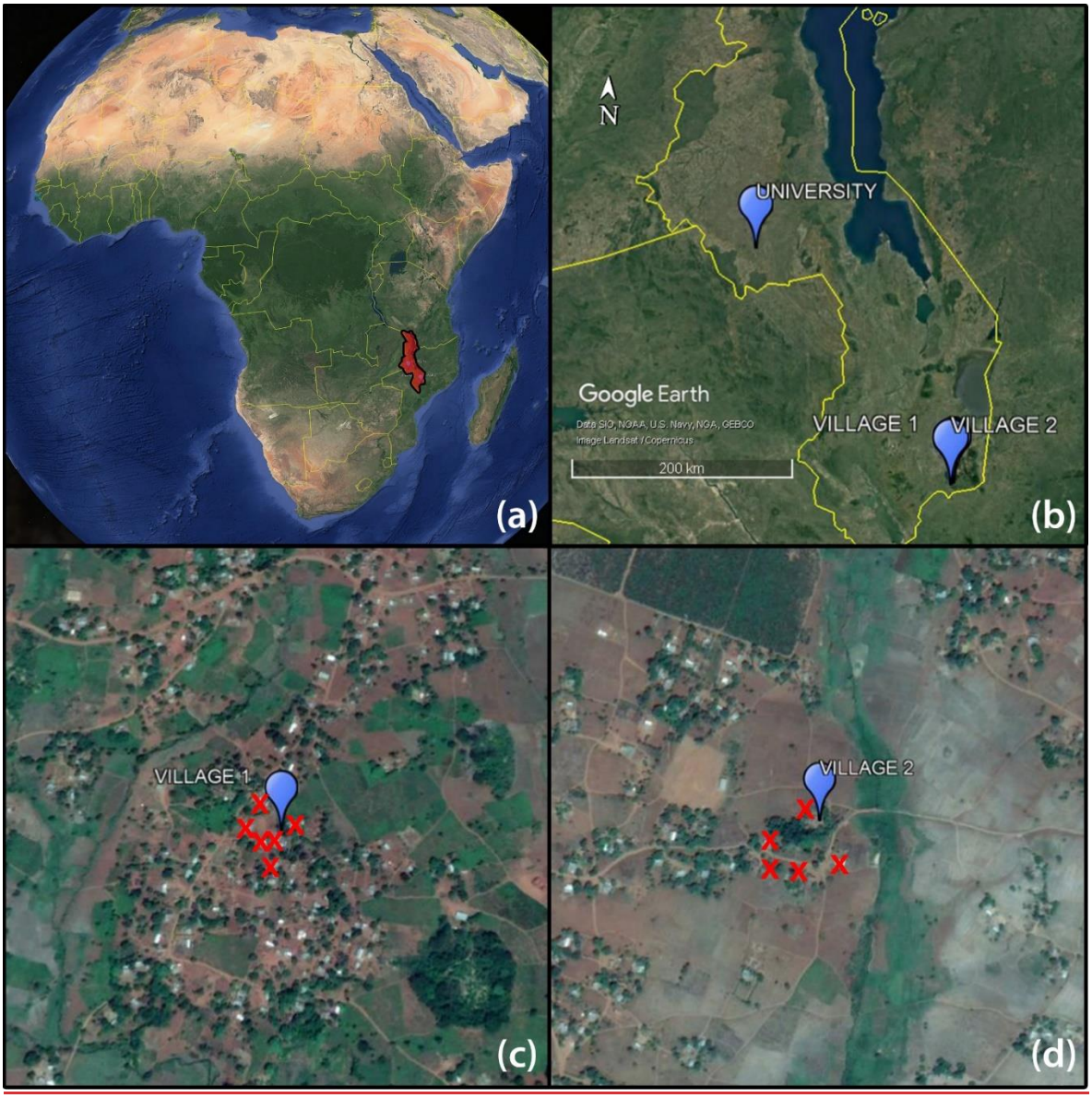
We adapted performance m~~M~~etrics and target values ~~adapted~~ from recently published U.S. EPA guidelines (Duvall et al., 2021a, b) ~~were used~~ to assess ARISense performance (Table S14). The EPA guidelines suggest using linearity, bias, precision, and error metrics to assess air sensor performance and they offer target values for each. We use the U.S. EPA target values as a quantitative marker to indicate satisfactory or unsatisfactory sensor performance, however given the differences in our study compared to the U.S. EPA methodology, we do not consider these categorizations to be definitive. Further, we emphasize that even if a sensor meets, or surpasses, the performance target values for each metric, this does not constitute endorsement by the U.S. EPA. Their guidelines were developed for O_x and $PM_{2.5}$ air sensors, and we used these to assess the ARISense Ox-B421 and OPC-N2 sensors, respectively. Although there are no formal guidelines for CO, NO, and NO_2 sensors at the time of writing, for coherency, we opt to assess those sensors using a similar approach. ~~To assess linearity, the correlation between estimated and true concentrations was calculated using~~

Tthe coefficient of determination (R^2), an indicator of the correlation between estimated and true concentrations, was used to assess linearity. The Instead of the EPA recommended rRoot mMean sSquare eError (RMSE) metric, the Mean Absolute Error (MAE) was used to assess error in the estimated measurements compared to the true values. The coefficient of variation (cV) was used to assess precision. Finally, tTo assess bias, a linear regression model (y = mx + b) was fit using the ARISense measurements as the dependent variable (y) and the reference measurements as the input variable (x), , and the resulting slope (*m*) and intercept (*b*) were calculated. Quantitative descriptions for each metric are given in Sect. 3 of the Supplementary Information.

The same metrics were used to assess the OPC-N2 sensor. In addition, tTo estimate confidence in- hourly averaged OPC-N2 measurements made under characteristic fielddeployment conditions in Malawi, we calculated 68% (1-sigma) prediction intervals using the Malawi collocation data set (Fig. S49). The 60-min averaged observations were used to fit a linear model, which required a Box-Cox transformation (Box and Cox, 1964) to obtain normally distributed residuals (Fig. S310). Details and quantitative descriptions are given in Sect. 3 of the Supplementary Information., Sect. 4.

2.6 Deployment to Malawi

ARI013, ARI014, and ARI015 were deployed to their respective monitoring locations in Malawi from July 2017 to July 2018 (shown as blue markers on Fig. 1). The three locations were selected to provide measures of regional variation and replicates in two paired village sites. ARI013 ("Village 2" site) and ARI014 ("Village 1" site) were deployed < 5 km apart in two nearby rural villages in Mulanje, Malawi. ARI015 ("University" site) was deployed >375 km northwest of the village sites at a rural university campus ~30 km from the capital city (Lilongwe).ARI013 ("Village 2" site) and ARI014 ("Village 1" site) were deployed < 5 km apart in two rural residential villages in Mulanje, Malawi, adjacent to private residences (Fig. S5). ARI015 ("University" site) was deployed >375 km northwest of the village sites at a rural university campus ~30 km from the capital city (Fig. S6). 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 with red 'X's on Fig. 1c-d). -Additional sSatellite images , maps, and site characteristics are given in Sect. 4 of Supplementary Information., Sect. 5.



310

Figure 1: (a) Satellite map of Malawi in southeast Africa, (b) ARISense monitoring locations in Malawi, (c) satellite map of Village 1, and (d) satellite map of Village 2. Blue markers indicate ARISense monitoring sites. Red 'X's indicate the location of known biomass cookstoves within 50 m of the monitoring site. 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. © Google Earth 2021.

315

A timeline of the ARISense collocations and deployments is given in Table 2. After the one year ambient deployment was completed, the ARISense were used for high-concentration emissions monitoring experiments in rural Malawi in July and August 2018. The details of those experiments (i.e., number of experiments, duration, approximate CO concentrations) are discussed in Sect. 5 of the Supplementary Information. We explore the impact of these experiments on sensor operation, but we do not discuss the data itself in this paper.

Table 2: Project timeline of collocations, deployment, and emissions monitoring experiments. The description under each period indicates the activity conducted during that timeframe. The location of the activity is given in parenthesis.

ARISense	May - June 2017	July 2017 - July 2018	July - Aug 2018	Aug 2018 - Mar 2019
ARI013	Collocation (NC)	Deployment (Village 2)	Emissions monitoring (Village 2)[a]	Collocation (NC)
ARI014	Collocation (NC)	Deployment (Village 1)	Emissions monitoring (Village 2)[a]	Collocation (NC)
ARI015	Collocation (NC)	Deployment (University)	Emissions monitoring (Village 2)[a]	n/a
ARI023	n/a	n/a	OPC-N2 collocation (Village 2)	n/a

[a]Data from emissions monitoring experiments not discussed in this paper

At the conclusion of the emissions monitoring experiments, ARI013 and ARI014 were returned to NC and were collocated with reference instruments at the near-highway NC-DEQ site used in the pre-deployment collocation (described in Sect. 2.2). ARI015 was relocated to a new monitoring site in Malawi.

2.7 Remote sensing and reanalysis data

Two publicly available NASA data products were obtained from the Goddard Earth Sciences Data and Information Services Center (GES-DISC) Interactive Online Visualization and Analysis Infrastructure (GIOVANNI): 1) area-averaged, monthly Multispectral CO Surface Mixing Ratio (Daytime/Descending) from MOPITT ~~(a satellite observation),~~ and 2) CO Surface Concentration - ENSEMBLE from MERRA-2 ~~(a global reanalysis product)~~; henceforth referred to as "MOPITT" and "MERRA-2", respectively. MOPITT is a calibrated satellite observation and MERRA-2 is a global reanalysis data product. MERRA-2 is the output of an atmospheric chemistry model that has "assimilated" other data, including but not limited to satellite data, in making its estimations. Monthly averaged MOPITT and MERRA-2 observations were compared to ARISense CO surface data collected at the Village and University locations. Given the physical proximity of Village 1 and Village 2, and the similarity in monthly mean CO concentration at each site (Fig. S7~~17~~), the average of the data sets ("Village Mean") was used. Additional details are given in Sect. 6 of the Supplementary Information~~, Sect. 6.~~

3 Results and discussion

3.1 Gas sensor performance during collocation

Raw gas sensor voltages (5-min averaged data) from all three ARISense monitors (ARI013, ARI014, ARI015), excluding the O_x sensor in ARI015, were highly correlated ($R^2 > 0.8$) during the pre-deployment collocation, suggesting changes in sensor response were due to environmental changes, not sensor-to-sensor variability (Fig. S93). The sensors in ARI013 and ARI014 were most closely correlated ($R^2 > 0.9$). The raw ARI015 O_x sensor data showed weaker temperature dependence and the lowest correlation ($R^2 < 0.6$) with O_{x} sensors in ARI013 and ARI014.

Figure 2 shows two performance metrics representing each sensor-model combination for the three ARISense. Results from all ARISense-sensor-model combinations for all five performance metrics are given in Tables S4-6. Data points toward the lower left corner of each Fig. 2 panel indicate better performance. We found that performance varied by ARISense monitor, but none of the ARISense consistently performed better than the others. Overall performance varied by gas sensor type and modelling approach. The calibrated NO_2 sensors in all three ARISense were the least correlated with reference measurements compared to the other gas sensors. Only the ARI015 NO_2 sensor, calibrated by the RF hybrid model, surpassed the target value for the linearity metric ($R^2 > 0.8$). Further, no NO_2 sensor-model combination met the bias target values for slope and intercept. For all three ARISense, the calibrated NO_2 sensors considerably underestimated the true concentration compared to the reference ($0.26 < m < 0.71$). However, all NO_2 sensor-model combinations met the error target and approached or surpassed the target for precision.

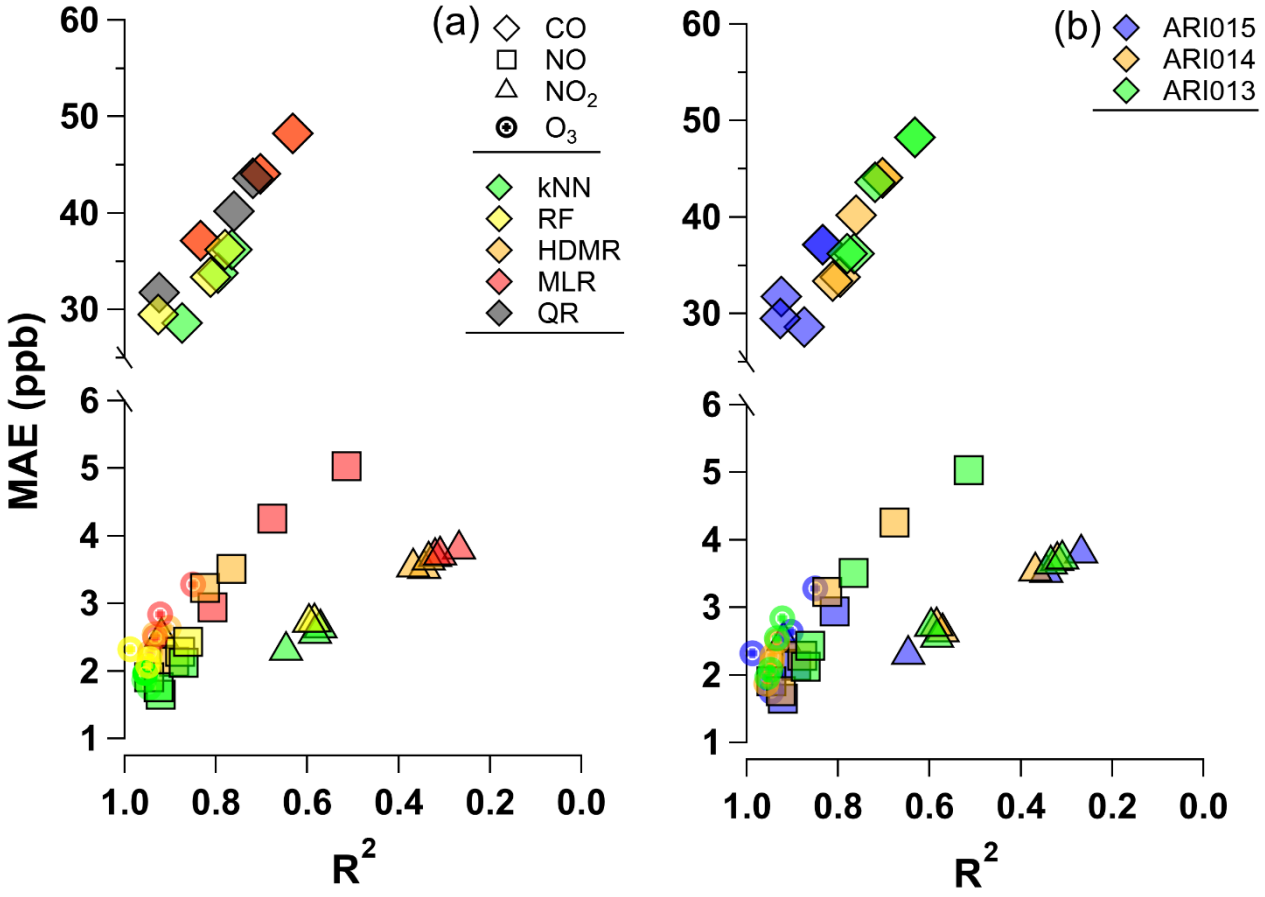
At the other end of the performance spectrum, the calibrated O_3 sensors performed the best compared to the other gas sensors during pre-collocation. Nearly all O_3 sensor-modelling approaches attained similar linearity and error metrics ($0.85 < R^2 < 0.98$ and $21.5 < RMSE < 53.5$ ppb), well within the target values. Only the ARI015 O_x sensor calibrated by the RF model failed to meet the RMSE target value, yet it returned the highest R^2 value compared to the other models. Additionally, all O_x sensor-model combinations met the slope and intercept target values for bias. For the kNN hybrid model, the calibrated O_3 observations had a slope approaching 1 ($m > 0.98$) compared to the reference. Only the precision values ($37\% < cV < 54\%$) were outside the EPA guideline target range ($cV < 30\%$).

Most NO sensor-model combinations met the target value for the bias, error, and linearity metrics, but precision was low for all combinations tested, with most cV values > 100%. The MLR model showed the worst performance for all three ARISense compared to the other models. However, for ARI015, all NO sensor-model combinations surpassed the target for every metric except precision. Again, the ARI015 NO sensor-RF hybrid model combination was the outlier compared to the

ARI013 and ARI014 sensor-model combinations (Table S6). We hypothesize that the shorter collocation period of ARI015 (8 days compared to 17 days of collocation for ARI013 and ARI014) led some of the sensor-model combinations to be overfit or poorly constrained.

Most CO sensor-model combinations met the target values for bias, linearity, and precision. The O_x target values for these three indicators can reasonably be used to compare against the CO sensor values to approximate performance, but we surmise the error target value (RMSE ≤ 5 ppb) cannot. The U.S. EPA National Ambient Air Quality Standards suggest CO concentrations are generally 1-2 orders of magnitude larger than ambient ozone or NO_x concentrations. By extension, we posit that a reasonable error target value for the CO sensor is 50 ppb. Except for the CO-kNN hybrid model combination, most CO sensor-model combinations did not meet our adapted error target value. However, considering the magnitude differences, the CO sensor-model combinations performed similarly to the NO, NO_2 and O_x sensors, in terms of error. The CO RMSE values (40-70 ppb) were correspondingly 1 order of magnitude larger than NO, NO_2, and O_3 RMSE values (2-7 ppb).

~~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 NO_x.~~


(a)
CO
NO
NO2
O3
kNN
RF
HDMR
MLR
QR
(b)
ARI015
ARI014
ARI013
MAE (ppb)
R2
60
50
40
30
6
5
4
3
2
1
1.0
0.8
0.6
0.4
0.2
0.0

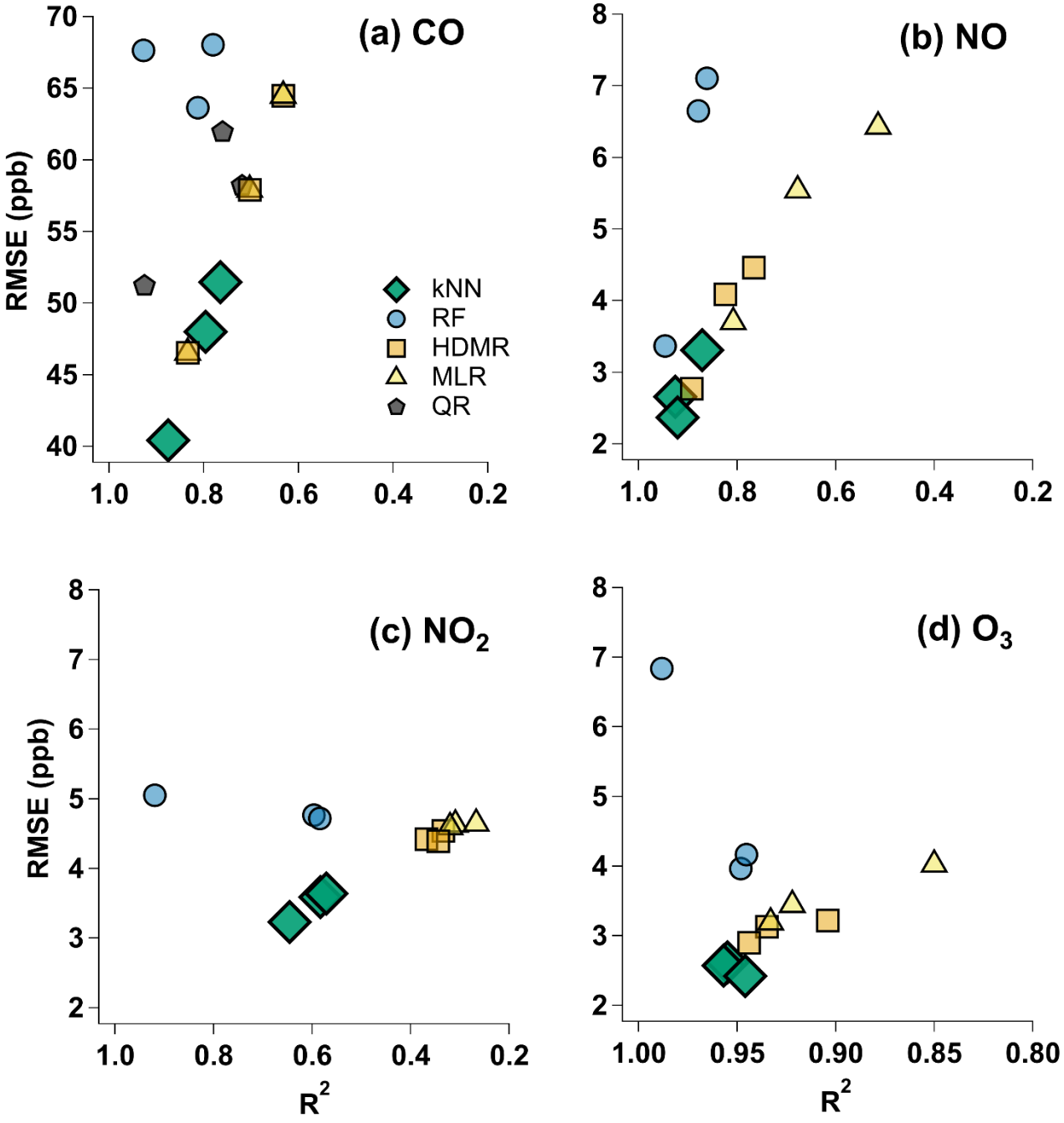


Figure ~~1~~2: Performance comparison of gas sensors (a) CO, (b) NO, (c) NO_2, and (d) O_x as calibrated by the five types of calibration models adopted for this study (kNN hybrid, RF hybrid, HDMR, MLR, QR). The model type is indicated by color and marker shape. An individual data point represents the paired metrics (RMSE and R^2) for one ARISense monitor. Since

there are three ARISense (ARI013, ARI014, ARI015) monitors, there are three markers for each gas sensor-model combination. RMSE is root mean square error. R^2 is the coefficient of determination (-infinity ≤ R^2 ≤ 1). The lower left corner region of each panel indicates the highest performance based on these metrics. ~~Performance comparison of gas sensors grouped by (a) modelling approaches and (b) ARISense monitor. Each data point represents the paired metrics (MAE and R^2) for the four gas sensors (O_3, NO_2, NO, CO) in the three ARISense monitors (ARI013, ARI014, ARI015) as calibrated by five types of calibration models (kNN hybrid, RF hybrid, HDMR, MLR, QR). MAE is mean absolute error. R^2 is the coefficient of determination (infinity ≤ R^2 ≤ 1). The lower left corner region of each panel indicates the highest performance based on these metrics.~~

For the suite of gas sensors in the ARISense monitors, we found the kNN hybrid model ~~models (RF and kNN)~~ to be the best among the modelling approaches used in the pre-deployment collocation testing (Fig. 2). ~~, with the kNN hybrid slightly outperforming the RF hybrid for all four gas sensors (Figure 1a).~~In almost all cases, the kNN hybrid model returned higher R^2 values, slope values closer to 1, 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. 2b-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. 2a) it was only for one of the three ARISense monitors, never all three. Additionally, ~~The other modelling approaches (HDMR, MLR and QR) were similar in overall performance, however~~ the MLR failed to meet target values for some ARISense-gas sensor combinations (Fig. 2a-b). Taken together, these findings suggest the kNN hybrid model is the best choice among these five modelling approaches, given that we sought an approach uniformly applicable to all the gas sensors and all three ARISense.~~The HDMR models were set to allow only first-dimensional interactions, as second-order interactions led to spurious results for data collected outside the bounds of training data (see Sect. 3.2 – on deployment conditions). For the CO sensors, this effectively made the HDMR model equivalent to the MLR model. The statistical metrics achieved by both models were identical, making these points darker in Figure 1. None of the ARISense consistently performed better than the others (Figure 1b); overall performance varied by gas sensor type and modelling approach.~~

3.2 OPC-N2 performance during collocation

Pre-deployment collocation $PM_{2.5}$ measurements in North Carolina (where no reference monitor/data were available) from ARI013, ARI014, and ARI015 suggest the Alphasense OPC-N2 sensors in each monitor responded similarly (R^2 > 0.9) when in the same environment (Fig. S104). 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. ARI015 underestimated the

mass at low concentrations compared to ARI013 and ARI014 (~~negative intercept and~~ non-linear clustering at concentrations < 5 µg m^{-3} in Fig. S10a, c). The OPC-N2 sensors in ARI014 and ARI015 showed the highest similarity (slope = 1 ± 0.05, R^2 = 0.96). ~~Inter-unit variation in the slope and intercept and in low-concentration behaviour suggests the three sensors would require individualized calibrations to account for their differences.~~

~~Figure 5~~Figure 3 shows scatter plots of the ARI023 OPC-N2 and MicroPEM data collected during collocation at the Village 2 site in Malawi. RH-correction partially mitigated the impact of overestimation due to hygroscopic growth but did not remove the artifact entirely (Fig. S12~~6~~). RH-correction improved the precision and error metrics, bringing RMSE~~MAE~~ within the target value (≤ 7 µg m^{-3}) for ~~all averaging intervals~~the 24 h averaged data (Table S7~~1~~). Increased averaging interval had a similar effect, but alone was insufficient to bring RMSE~~MAE~~ within the target range. Linearity was well below the target value (R^2 > 0.7) for all averaging intervals and RH-correction did little improve performance for this metric. For this data set, c~~C~~hanges in bias and linearity appeared driven by averaging interval. For example, the OPC-N2 RH-corrected 1 minute data met the ~~three of the five suggested~~ EPA target ~~values~~ values for slope and intercept, ~~(*m*, *b*, MAE)~~ but the 1 h and 24 h averaged data met neither~~only met one (MAE)~~. Particularly for the 24 h averaged data, t~~T~~he small sample was leveraged by a few points which drove metric values (Fig. 3c).~~.~~ h~~H~~owever, close 1:1 agreement between the instruments was observed for 4 of the 7 data points ~~(Figure 5c)~~. These results highlight the value of longer and more representative collocations. At least two ~~30-day~~30-day collocations would likely be needed, during the hot-dry (Sep to Oct) and ~~during the~~ warm-wet (Nov to Apr) seasons, to characterize this specific site.

Even after RH-correction, the OPC-N2 overestimated mass concentrations compared to the nephelometer when RH was ≥ 70%. Conversely, the OPC-N2 often underestimated mass when RH was ≤ 30%. These effects were most noticeable at higher time resolutions (~~Figure 5~~Fig. 3a-b). The effects of RH were tempered by a longer averaging interval, however for a particularly humid day at this site, the 24 h mass concentration was overestimated by a factor of three (Fig.~~Figure 5~~3c). Notably, the moderate-RH outliers in the 24 h average scatter plot suggest that other factors, in addition to RH, were contributing to error in the OPC-N2 observations.

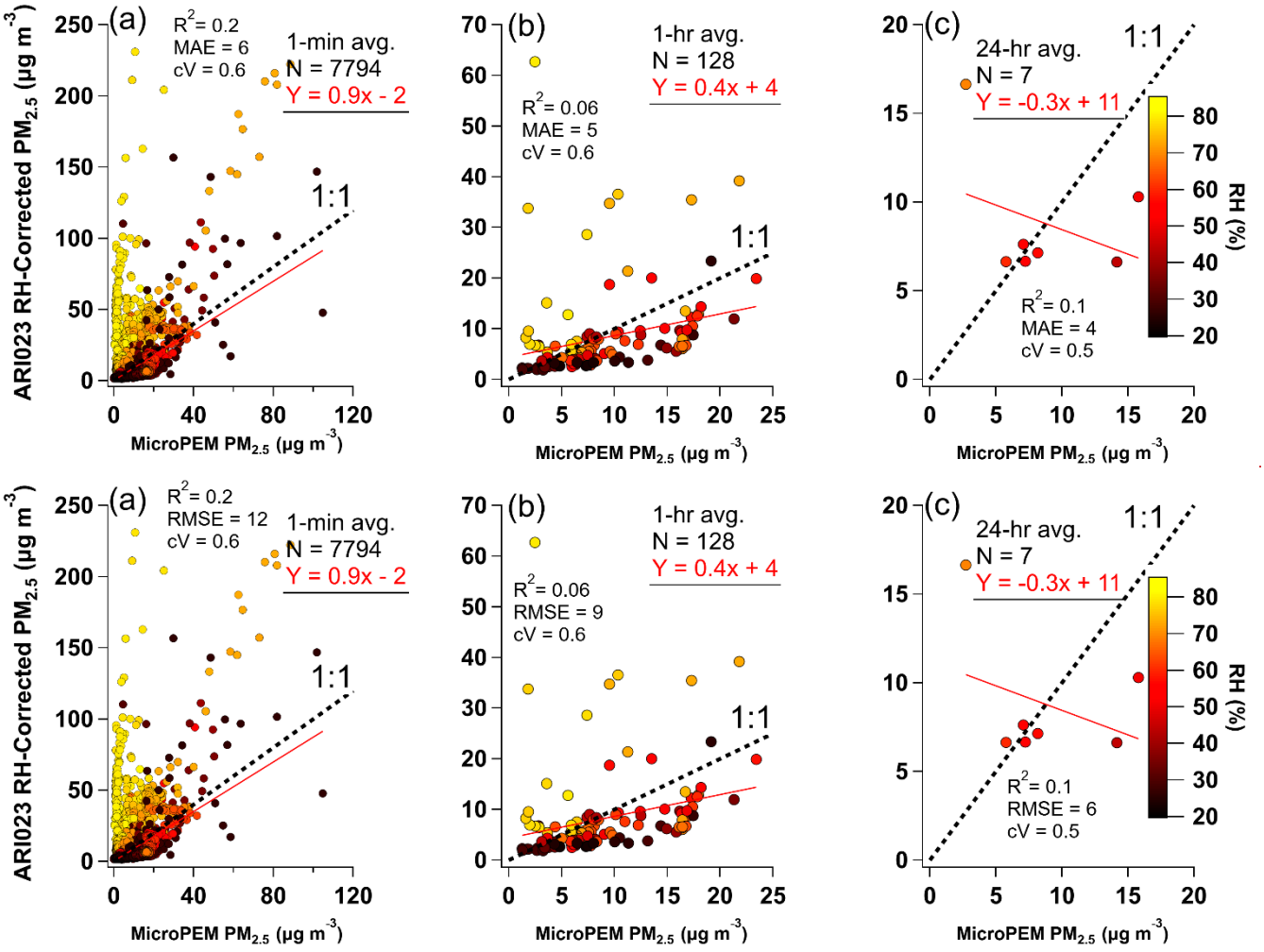


Figure ~~5~~3: Scatter plots of RH-corrected $PM_{2.5}$ mass concentration measurements from the OPC-N2 versus mass-corrected $PM_{2.5}$ measurements from the MicroPEM at 1 min (a), 1 h (b), and 24 h (c) averaging intervals. Data points are colored according to RH (%) conditions. The number of data points (N) and linear fit lines and regression coefficients (*m*, *b*) are given in red as Y = *m*x + *b*. Additional metric values are inset: R^2 is the coefficient of determination, RMSE~~MAE~~ is root mean square~~absolute~~ error assuming the MicroPEM is the reference instrument (units of µg m^{-3}), and cV is coefficient of variation. The black, dashed line is a 1:1 line.

To explore other possible contributors to variable OPC-N2 performance, ~~Figure 6~~Figure 4 shows performance for RH-corrected data stratified by environmental conditions (wind direction, ambient concentration, and RH). ~~Figure 6a-b (W~~wind direction and concentration (Fig. 4a-b~~)~~ were selected to explore the possible ~~characterize the~~ effect of nearby cookstove emissions, while Fig. ~~4~~6c highlights the remaining effect of RH after correction. We hypothesized that ambient concentration and wind direction might impact OPC-N2 performance given that the site was periodically exposed to cookstove emissions from the household kitchen (within 15 m to NW) and from adjacent residences (within 50 m to the S-

SW in Fig. 1d ~~S14~~). Figure ~~4~~6a shows that wind direction was associated with performance variation, although to a lesser degree than RH (Fig. ~~4~~6c). Slightly increased performance was observed for northerly winds. Nearby cookstove use potentially explained the decreased performance associated with southerly winds. Four of the five morning cooking periods observed in the time series data were associated with wind blowing from the SE-S-SW (Fig. S1~~4~~8). ~~Figure 6~~Figure 4b shows that ambient concentration had a modest impact on OPC-N2 performance metrics. Linearity was~~and absolute error were~~ expected to increase with concentration, particularly given that the high-concentration bin (20-105 µg m^{-3}) spanned a larger interval than the other bins~~.~~ Precision within each concentration bin was low. The cV values were well beyond ~~was within or near~~ the recommended target value (cV < 30%). ~~However,~~ T~~t~~he OPC-N2 frequently underestimated the ambient mass concentration compared to the MicroPEM, particularly during higher concentration periods likely dominated by near-field biomass burning (i.e., slope = 0.4 for 20 to 105 µg m^{-3})~~.,~~ This is likely due to the~~its~~ relatively high minimum cut-off diameter of the OPC-N2. During periods of cookstove influence, the size distribution, hygroscopicity, and optical properties of the measured aerosol were likely altered. Assumptions about the source aerosol, used to inform the RH-correction, were found to affect inferred OPC-N2 performance compared to the MicroPEM, though not predictably. For example, higher linearity and lower RMSE~~MAE~~ were observed when the particle composition was assumed to be highly hygroscopic (κ = 1), yet the least bias was observed at the lowest hygroscopicity tested (κ = 0.15). Further, when the aerosol was assumed to be characteristic of wildfire (rather than ammonium nitrate, dust or background in origin), the bias between the OPC-N2 and MicroPEM disappeared (slope = 1.02), yet the error metric was among the highest in ~~among~~ the four aerosol categories and was~~,~~ above the target value (Table S10~~2 S3~~). Summary statistics for each performance assessment metric are given in Tables S8-10 in Sect. 8 of the Supplementary Information.

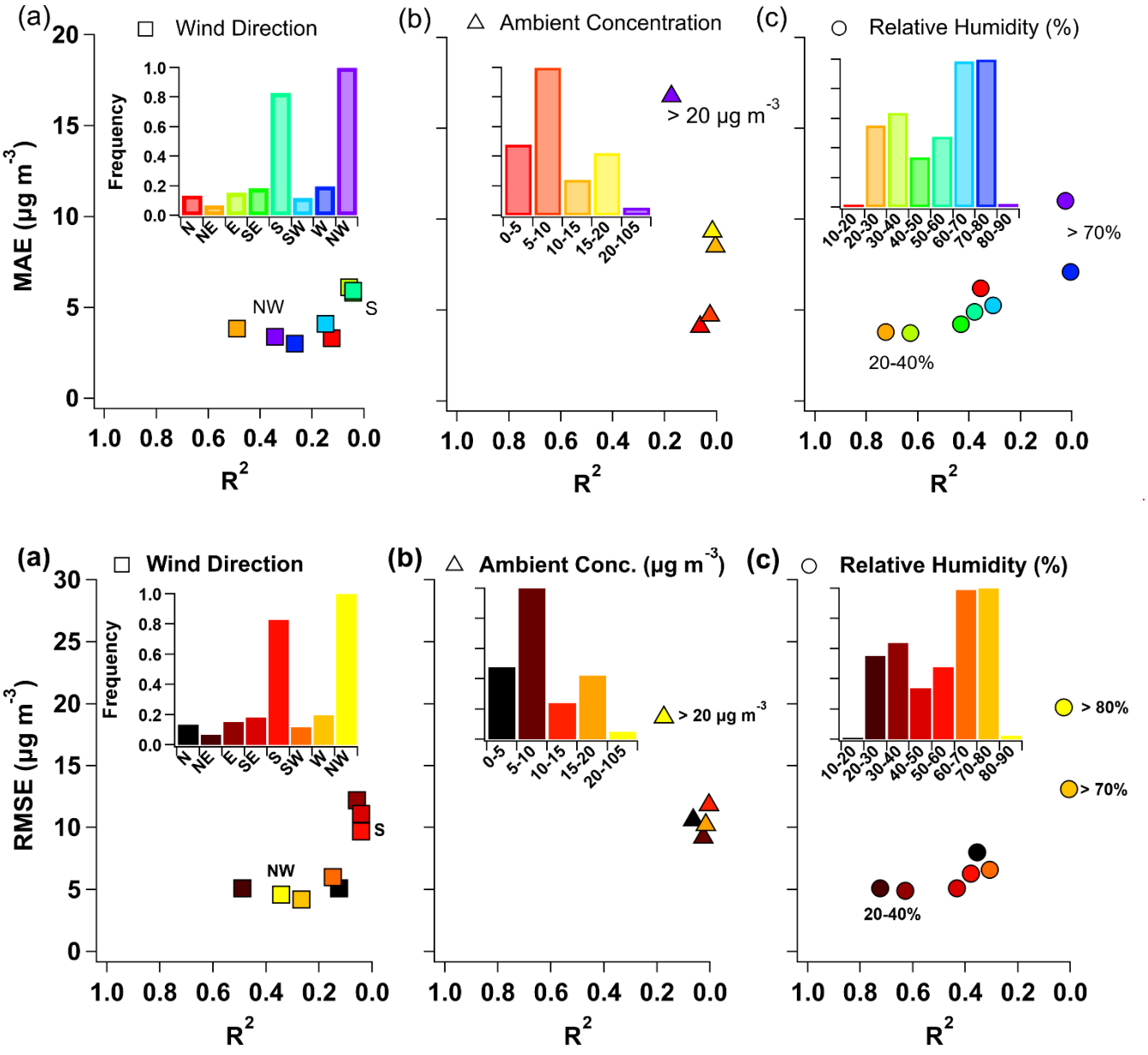


Figure ~~64~~: Performance comparison of the RH-corrected Alphasense OPC-N2 compared to the MicroPEM under different environmental conditions: (a) wind direction, (b) ambient concentration, and (c) relative humidity during collocation at the Village 2 site in Mulanje, Malawi. An individual data point represents the paired metrics (RMSE 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. RMSE is root mean square 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.~~Performance (MAE and R^2) of the Alphasense OPC-N2 compared to the RH-corrected 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. 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 histogram bin colors. 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.~~

In this deployment site, the OPC-N2 performed the best compared to the MicroPEM during dry conditions (20 to 40% RH) and when measuring background aerosol rather than source emissions (Fig. S14~~8~~ - presumed based on time series data). However, this result might be partially due to the coincident effects of high RH (Fig. 7). Figure 4~~6~~c shows OPC-N2 behaviour was largely determined by changes in ambient RH. ~~(Fig. S7).~~ In general, performance decreased with increasing RH, and this effect remained even after RH correction. For RH = 20 to 40%, RH-corrected OPC-N2 performance approached or exceeded the target values for the linearity, error and precision metrics.~~for every metric except for precision~~ (Table S7~~1~~). After RH increased past 70%, the R^2 value approached 0 and the RMSE~~MAE~~ increased beyond the target value. Unfortunately, the inset histogram of Fig. 4~~6~~c shows that an RH range of 60 to 80% was typical for this site during collocation.

We found that the OPC-N2 at this specific site generally underestimated mass concentration compared to the MicroPEM, based on less than unity slope values, and the performance was variable at low ambient concentrations and dependent on RH (Fig. S13~~7~~). However, outside of very humid (RH > 70%) or very dry (RH < 20%) conditions, the RH-corrected OPC-N2 could estimate $PM_{2.5}$ mass concentration within about 10~~6~~ µg m^{-3} of the MicroPEM value for real-time, hourly, and daily monitoring purposes (based on RMSE in Table S7). The findings from this section highlight the importance of quality assurance for low-cost optical particle sensor mass concentration measurements, especially those made in environments with highly variable meteorology and nearby ultrafine aerosol sources. For this site, contextual information on meteorology and emissions sources and their diurnal patterns helped interpret and evaluate the measurements.

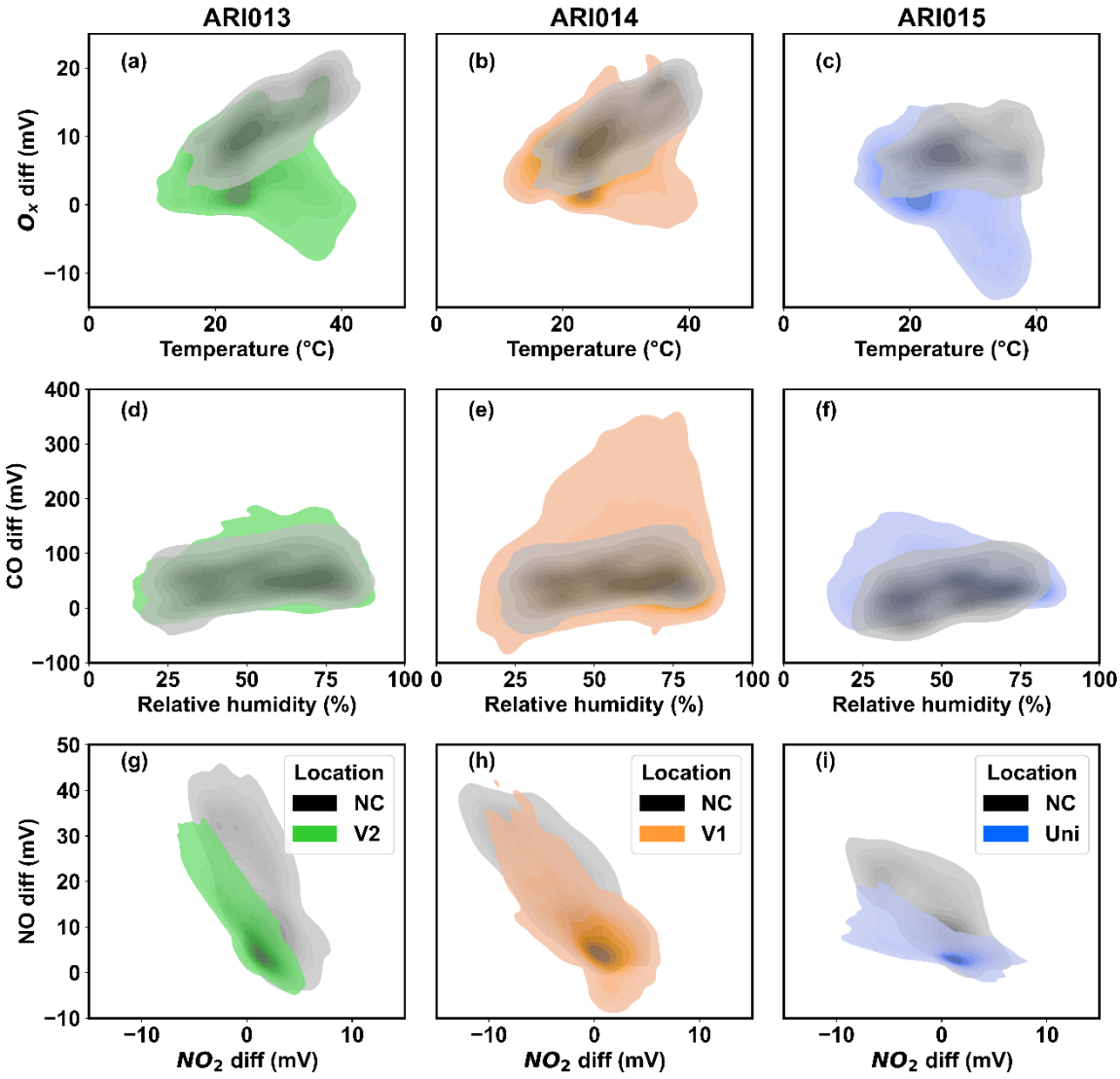
3.3~~2~~ Gas sensor performance during deployment

Given that RH, T, DP, and differential voltage were inputs to the calibration models, the ranges of these values during training and testing should mimic the ranges expected during deployment. Otherwise, the model is required to extrapolate beyond its training bounds, which could lead to non-physical results (e.g., negative concentration values). Further, the ~~testing~~ performance assessment statistics derived from the collocation cannot be expected to hold for conditions far beyond those experienced during the performance characterization. Overall, the collocation and deployment settings exhibited a similar range of environmental conditions (Fig. S15~~9~~-16~~22~~), but T and RH ranges in NC (15 to 40°C and 20 to 80%) were less extreme than in Malawi (10 to 45°C and 10 to 95%). While in Malawi, the ARISense experienced more time at lower temperatures (T < 25°C), lower gaseous concentrations (other than CO), and lower ambient pressure (5 to 15 kPa lower depending on site). Although the ARISense were deployed at a higher elevation in Malawi than during the collocation in

North Carolina (625 m vs 120 m ASL), all models were built using the differential voltages (WE-AE) of each electrochemical gas sensor. Therefore, the pressure-related shifts in the WE and AE baseline were not expected to pose an issue to the calibrated Malawi data (Fig. S16).; The variation in pressure was within the operating range given on the sensor specification sheets (80 to 120 kPa) and was stated not to have long term impacts by the manufacturer (Alphasense FAQs, 2021). Further, others have shown no statistically significant change in electrochemical sensor sensitivity due to changes in pressure (Popoola et al., 2016). Even so, we did not have the laboratory chamber data to investigate this potential issue.

3.3.1 Bivariate histograms

Figure 2Figure 5 shows bivariate distributions of T, RH, and gas sensor differential voltage data collected in NC and Malawi. In addition to capturing interactions between variables, Fig. 5Figure 2 shows that even when in the same environment during the NC collocation, the individual sensors in each ARISense responded differently. Compared to ARI013 and ARI014, the Ox sensor in ARI015 showed weaker temperature dependence (Fig. 52c). Since ARI015 had a shorter collocation period, it could be hypothesized that if ARI015 were present in the collocation environment for the same amount of time as ARI013 and ARI014, its response would look more like the ranges measured by the other sensors. However, this cannot fully explain the variation between individual sensors. For example, there is considerable variation between the ARI013 and ARI014 NO_2 differential voltage ranges (Fig. 52g-h), despite having identical collocation periods. Further, the raw CO sensor data for all three monitors showed much less inter-sensor variation (Fig. 52d-f), even despite the shorter collocation period of ARI015. This inter-sensor variation, which appears largest for the NO_2 sensors, may partially explain the lower performance of this gas sensor group during calibration model performance testing, compared to the other gas sensor types (Figure 1Figure 2).


ARI013
ARI014
ARI015
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
O_x diff (mV)
CO diff (mV)
NO diff (mV)
NO_2 diff (mV)
Temperature (°C)
Relative humidity (%)
Location
NC
V2
V1
Uni

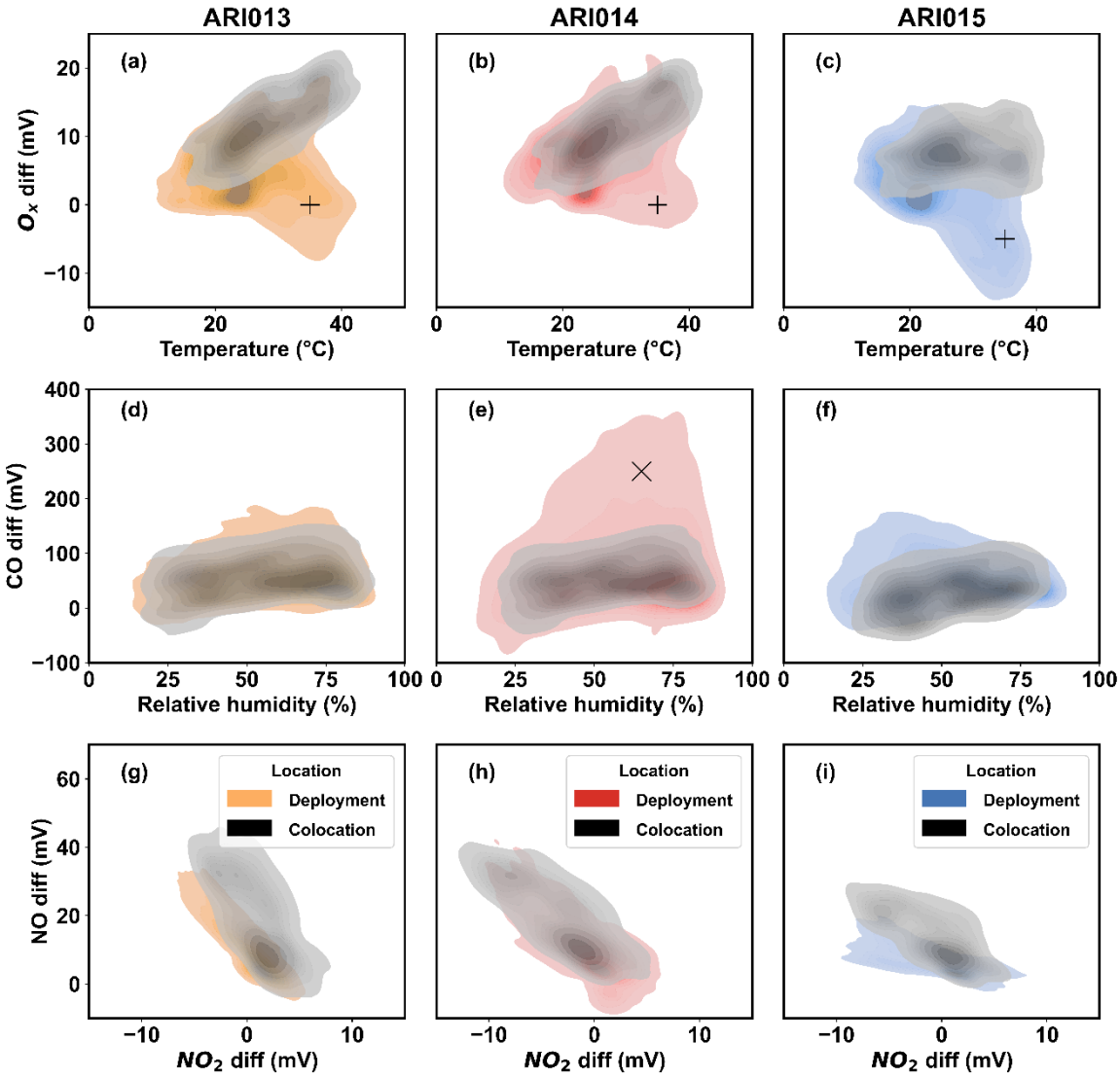


Figure 52: Bivariate distributions of gas sensor calibration model data inputs (RH, T, and O_x, CO, NO, and NO_2 differential voltage) for each ARISense monitor using kernel density estimation. Density is reflected in the color scheme; Darker colors indicate more data points in that region. Training data collected during collocation in North Carolina are shown in grey; data collected during deployment to Malawi are shown in color. ARI013 was deployed to the Village 2 site, ARI014 to the Village 1 site, and ARI015 to the University site. Regions where the deployment distributions overlap with the NC

collocation distributions indicate the regimes for which the calibration models were trained. Regions where the deployment location distributions extend beyond the NC collocation distributions indicate regimes where the calibration models must extrapolate to estimate pollutant concentrations. These regions are indicated by overlaid markers 'x' and '+', discussed in the text. ~~Bivariate distributions of gas sensor calibration model data inputs 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 in North Carolina are shown in grey; data collected during deployment to Malawi are shown in color. Regions where the deployment location distributions overlap with the NC distributions indicate the regimes for which the calibration models were trained. Regions where the deployment location distributions extend beyond the NC distributions indicate regimes where the calibration models extrapolated to estimate pollutant concentrations.~~

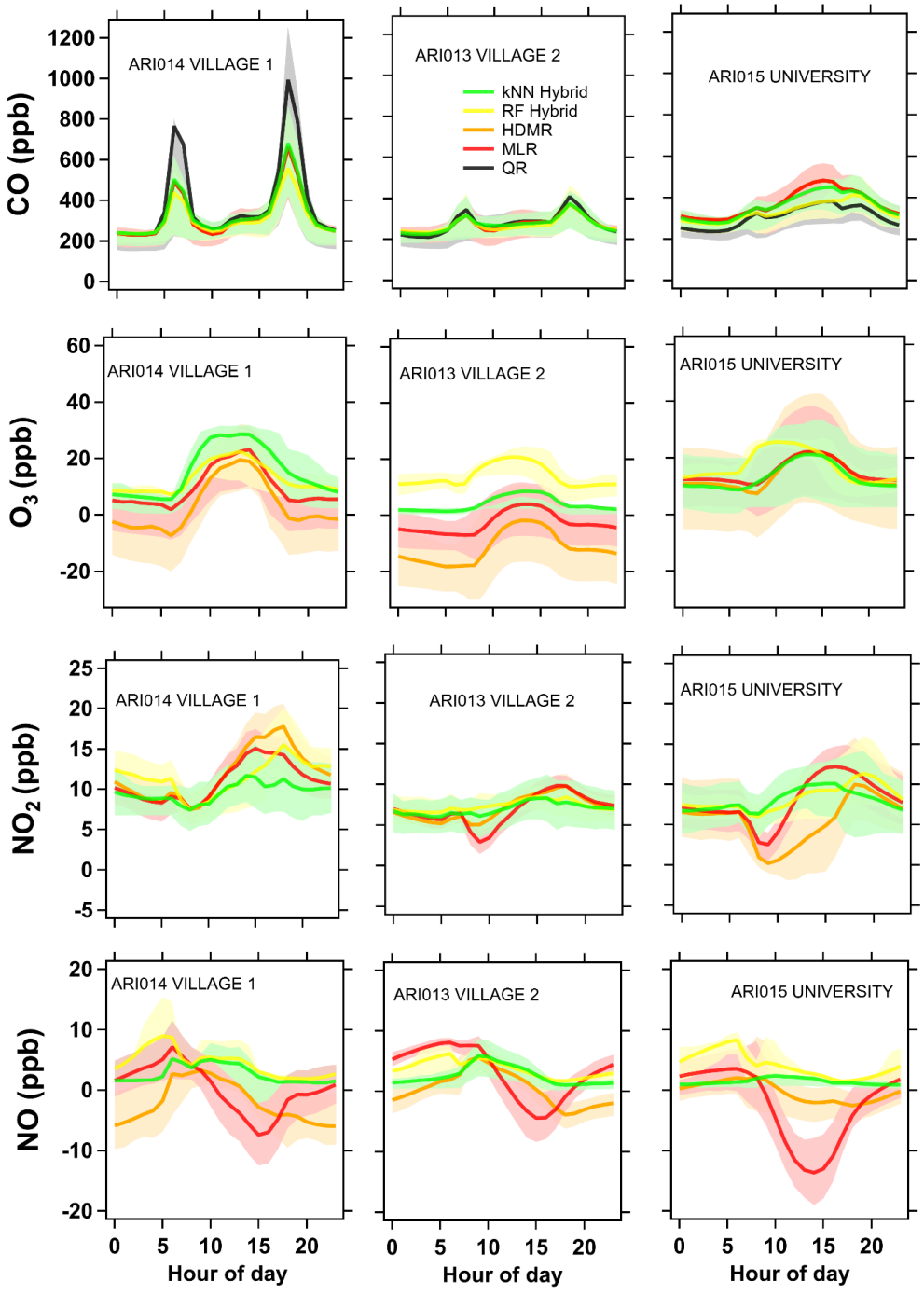
There were notable regimes in Malawi that required the calibration models to extrapolate beyond NC training conditions. NO differential voltage responses in NC and Malawi did not completely overlap (Fig. 5~~2~~g-i), especially in the low-concentration regime (i.e., V near 0 mV) which was more frequent in Malawi. The collocation site in NC was 10 m from an 8-lane freeway (Saha et al., 2018), therefore NO_x concentrations were higher than in rural Malawi where vehicles and industry are rare. However, for ARI014 in Village 1, there was a higher NO_2 response in the deployment environment compared to the collocation environment. This could be partially explained by sensor interference by RH and T, which were more extreme (i.e., beyond the training ranges) in Malawi (Fig. S17~~23~~). Figure 5~~2~~e 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 5e. ~~, but T~~this difference was aligned with observations of nearby sources (Fig. 1c-d). 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~~well controlled~~. ARI014 was deployed in more~~a~~ densely populated V~~v~~illage 1, adjacent to more biomass cookstove activity than ARI013 or ARI015 (Fig. 1c).

The O_x differential voltage ranges were the most dissimilar between the collocation and deployment environments. The most frequent regimes, the heaviest shaded regions in Fig. 5~~2~~a-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 2~~Figure 5a-c shows a high temperature-low ozone regime in Malawi (regions denoted by a '+' marker) that was not present in the NC data. ~~, presumably due to differences in ozone precursor regimes.~~ Further, for all three Malawi sites, the minimum O_x sensor voltages were lower ($-10 < V_{min} < 0$) than minima in the NC collocation.

3.3.2 Diurnal trends

Since the deployment site does not have reference data for quantitative comparison, ~~Given the absence of reference monitors in Malawi,~~ 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, previous observations of ambient diurnal trends in regions with widespread biomass cookstove use (Dionisio et al., 2010; McFarlane et al., 2021; Subramanian et al., 2020), ~~and knowledge of nearby emission sources, regional air quality,~~ and atmospheric chemistry. Non-physical predictions from a given model may indicate that differences between the collocation and deployment environments were too large to reasonably extrapolate and therefore any ~~field~~deployment results calibrated by that model are likely not reliable. Alternatively, coherency among the concentration values and trends estimated by the models may suggest that the ~~field~~deployment results are robust against variation in the modelling approaches. This analysis ~~and~~ can contribute to our confidence in the estimated concentration values and trends, but ultimately cannot address or estimate the quantitative error. ~~.~~ Diurnal trends in Figure 6~~Figure 3~~ suggest the kNN hybrid model was the most transferable for interpreting ~~field~~deployment data for all gas sensors. However, both the kNN and RF hybrid models predicted similar trends and values for most sensors. The MLR and HDMR models also predicted similar trends, but frequently predicted negative values.


ARI014 VILLAGE 1
ARI013 VILLAGE 2
ARI015 UNIVERSITY
kNN Hybrid
RF Hybrid
HDMR
MLR
QR
CO (ppb)
O3 (ppb)
NO2 (ppb)
NO (ppb)
Hour of day

620

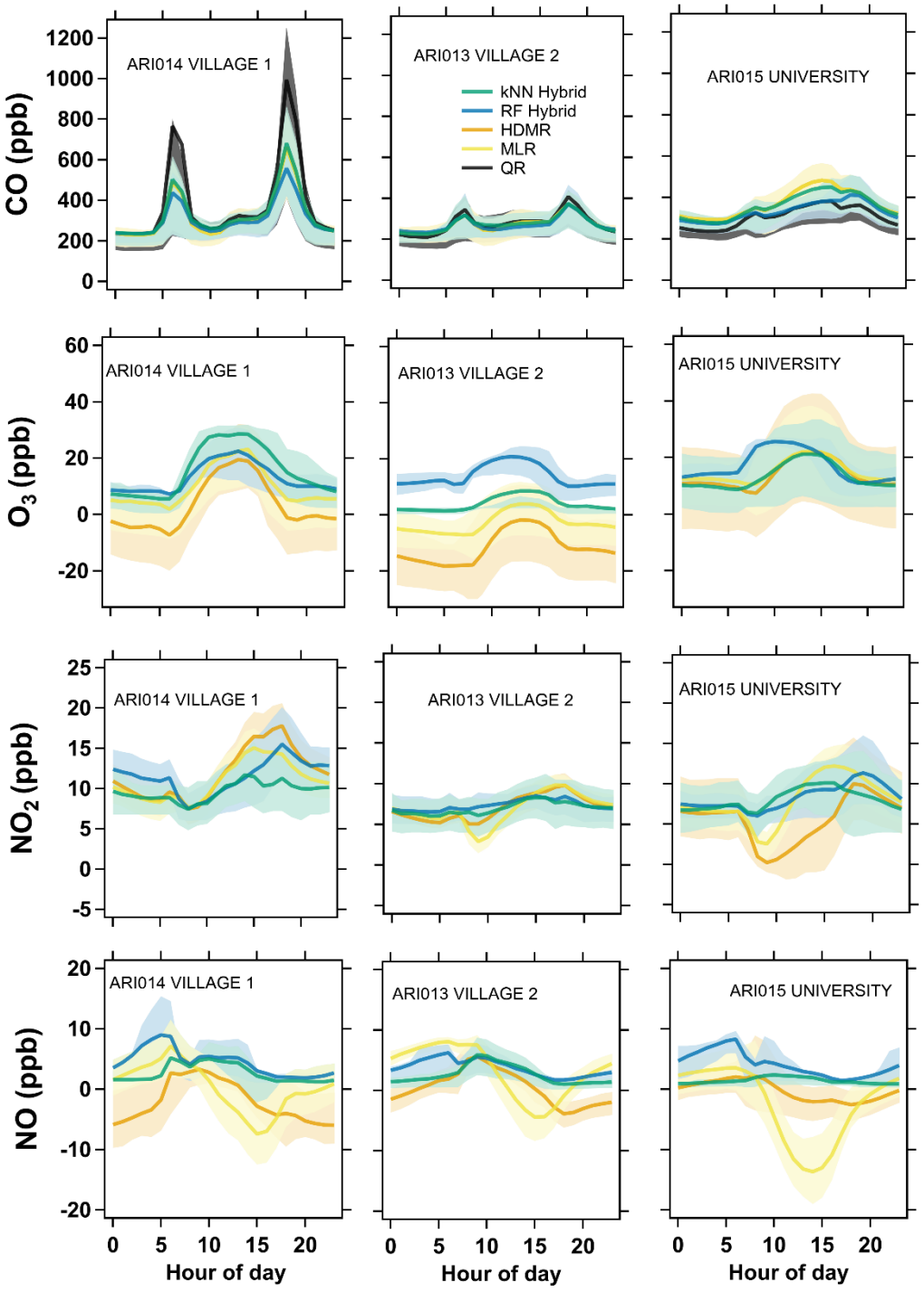


Figure 6~~3~~: Diurnal trends of calibrated gas measurements (rows) at each site (columns) in the three deployment environments. QR model built for and applied to CO data only. The thick line indicates hourly mean, the shaded region indicates interquartile range. Midnight is the zero hour. The hours are in local time.

Calibrated CO data showed the highest coherency across model predictions (Fig. 63). Calibrated CO data were rarely non-physical. All models predicted similar diurnal trends, specific to each site. Knowledge of the nearby emission sources and activity patterns lend support to the calibrated CO data.; For example, the village monitors were adjacent to widespread household biomass cookstove activity, coincident with the concentration peaks seen in the diurnal trend. This diurnal cooking pattern was observed in both CO and OPC-N2 data (Fig. 6 and Figure 1Figure 7, respectively) data at both village sites and was measured in complementary emissions monitoring work (Bittner et al, in prep). Further, ARI014 was in a more densely populated village than ARI013, contributing to higher CO peaks. The QR model overestimated CO peaks compared to other models for the Village 1 data, likely because the model training set did not include high concentration data (Fig. 52e) and the quadratic term was not well constrained. Despite the calibrated CO measurements in Malawi being higher than the concentrations experienced in NC, particularly for ARI014 in Village 1, we expect that the calibrated CO measurements from Malawi are credible. We provide the following reasons for justification: a) the manufacturers report that the sensor response is expected to be linear up to 500 ppm (Alphasense, LTD., 2019), b) we expect that RH/T interference induced on the CO-B4 sensor, ~0.2 mV/ppb (Lewis et al., 2016), has relatively RH/T interference has less influence on overall sensor readings in the higher voltage (i.e., concentration) regime , c) all modelling approaches (other than QR) predicted highly similar diurnal trends and concentration values, and d) there were known CO emission sources, with diurnal usage patterns matching the observed trends, near the monitoring sites. This suggests, for this specific sensor under these conditions, that these modelling approaches (other than QR) could reliably extrapolate beyond the training data limits to provide reasonable measurements in the deployment environment.

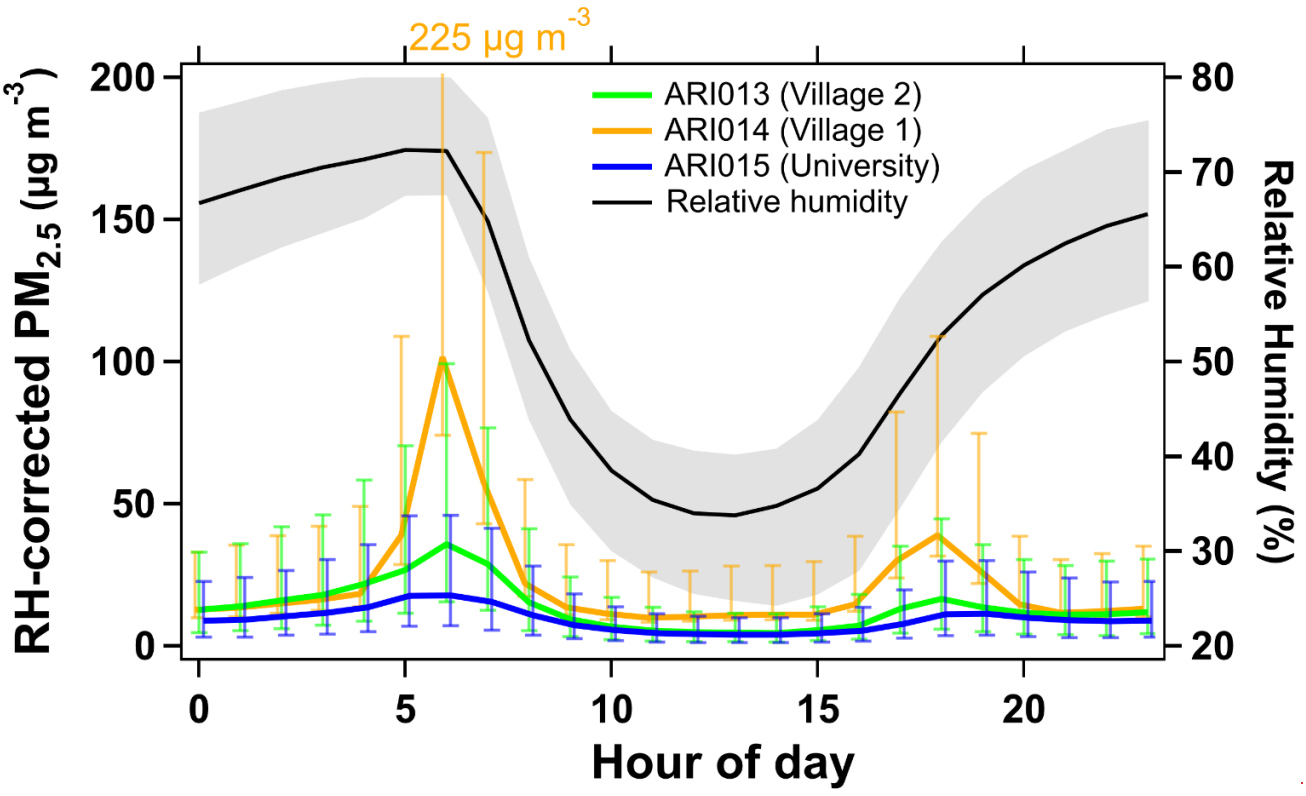
The calibrated NO_x data showed less coherency than the CO data. NO_2 trends were largely similar across the sites and concentrations were rarely negative, but calibrated NO trends varied across models and the lower performing models (HDMR and MLR) often predicted negative values. The better models identified in the NC collocation, kNN and RF hybrid, suggested that mean ambient NO_x levels in Malawi were low (< 15 ppb). We have lower confidence in the calibrated NO_x measurements in Malawi for the following reasons: a) the calibrated observations (5 to 20 ppb) were on the same order of the noise level reported on the sensor specification sheets (15 ppb) and b) the lack of coherency observed between model predictions. Low ambient NO_x levels and a lack of representative data in the NC collocation data likely contributed to the non-physical concentrations predicted by the models in Malawi.

The calibrated O_x3 sensors performed the best during collocation testing compared to the other gas sensors, but in Malawi the calibration models frequently returned non-physical values and showed inconsistent annual diurnal trends between the models and across the sites. For ARI014 and ARI015, the O_3 trends were roughly consistent in shape and magnitude and were aligned with the expected diurnal trend (i.e., peaking at midday). Peaks in the mean concentration were between 10 and 30 ppb, plateauing from 10 AM and 3 PM local timeLT. The RF hybrid model at the ARI015 University site estimated the O_3 peak to occur earlier in the day compared to the other models and sites. This may be the result of a spurious relationship

between O_x voltage and DP in the collocation data set on which the RF Hybrid model was trained, which held at the Village sites but not at the University site. At the Village 2 site (ARI013), there was a change in raw differential voltage response after December 2017 that caused all models to fail for the second half of the deployment. All models either consistently predicted negative values, values < 1 ppb, or failed to reproduce the expected diurnal trend (i.e., peaking around 9am rather than midday). Only O_x ~~ozone~~ data collected before Dec 2017 resulted in reasonable calibrated values and trends (Fig. S18~~24~~). Notably, O_x ~~ozone~~ data collected after Dec 2017 corresponded with the high temperature-low ozone regime (Fig. S19~~25~~) shown in Figure 5~~Figure 2~~a-c. Despite the O_x differential voltage data spanning a similar range in both NC and Malawi, there was little overlap in the ozone~~O_3~~ dimension at comparable concentration, RH, and T conditions. Since ozone is a secondary pollutant driven by complex atmospheric processes and multiple precursors, the ambient conditions that increase or decrease ozone formation in one region may not hold in another environment. Ultimately, although the calibrated O_x ~~ozone~~ sensors performed better than the other gas sensors in NC, the models were tuned for a set of conditions that did not hold in Malawi. This suggests that for these O_x sensors and these modelling approaches, a lack of environmentally similar collocation data compromised our ability to reliably interpret calibrated O_3 measurements in this specific deployment environment.

3.~~4~~5 OPC-N2 performance during deployment

To evaluate the long-term performance of the OPC-N2 during deployment in Malawi, we examined the representativeness of the collocation conditions for the full year of conditions experienced during deployment. Figures S20~~6~~-21~~7~~ show normalized histograms of the T, RH, and $PM_{2.5}$ mass concentration observed during the collocation and the full-year deployment in Malawi, suggesting the two data sets generally spanned a similar range of environmental conditions. However, the collocation occurred during the cool, dry season, and RH minima and maxima (regimes associated with poor performance during collocation – see Section 3.2~~4~~) were more extreme during the 1 year deployment in Malawi.


225 µg m-3
ARI013 (Village 2)
ARI014 (Village 1)
ARI015 (University)
Relative humidity
RH-corrected PM2.5 (µg m-3)
Relative Humidity (%)
Hour of day
0
50
100
150
200
20
30
40
50
60
70
80
0
5
10
15
20

680

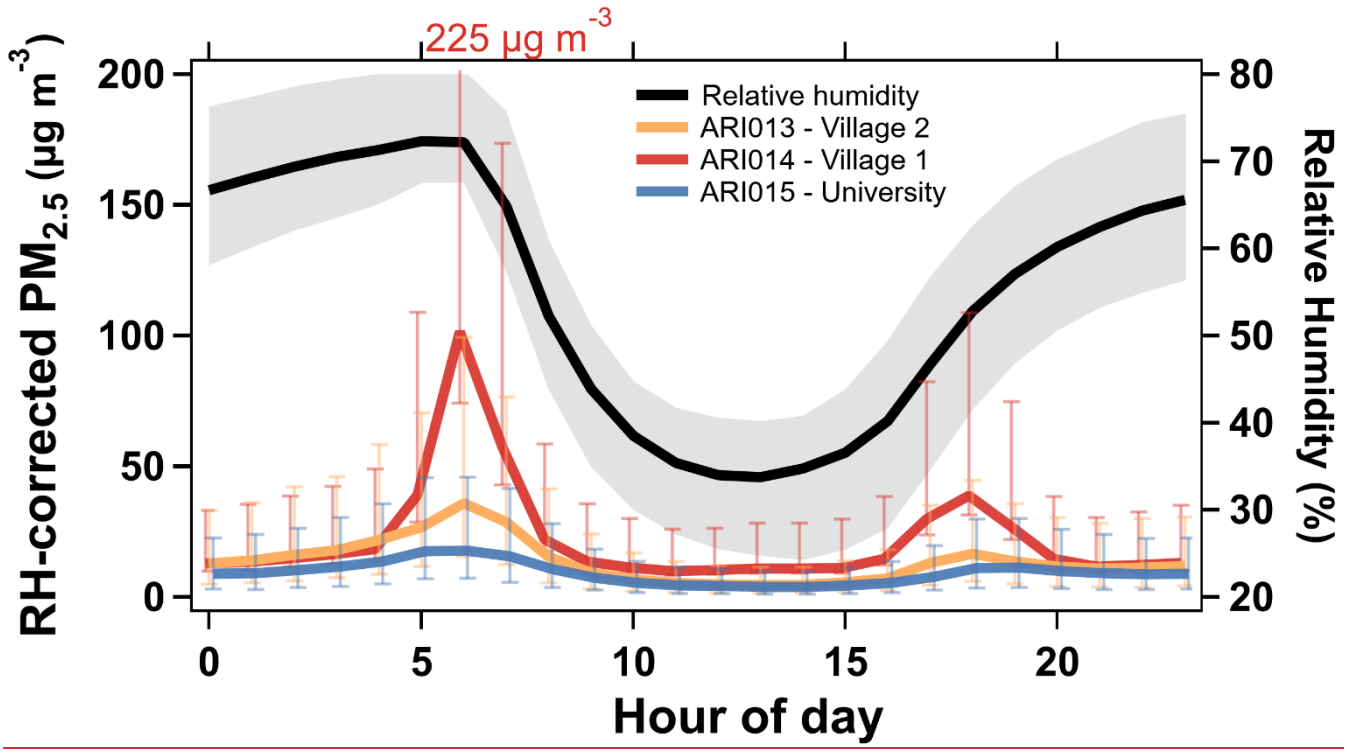


Figure 1~~7~~: Diurnal trends of the (left axis) integrated mean $PM_{2.5}$ mass concentration measured by the OPC-N2 in each ARISense at each deployment site and the annual relative humidity at the Village 2 site~~,~~ (right axis). Error bars represent the calculated 1σ (68%) prediction interval of the hourly mean value. The red~~orange (ARI014)~~ text annotation indicates the upper limit of the Village 1 prediction interval at 6 AM (beyond the range of shown y-axis). Thick lines indicate hourly mean and shaded regions indicate interquartile range.

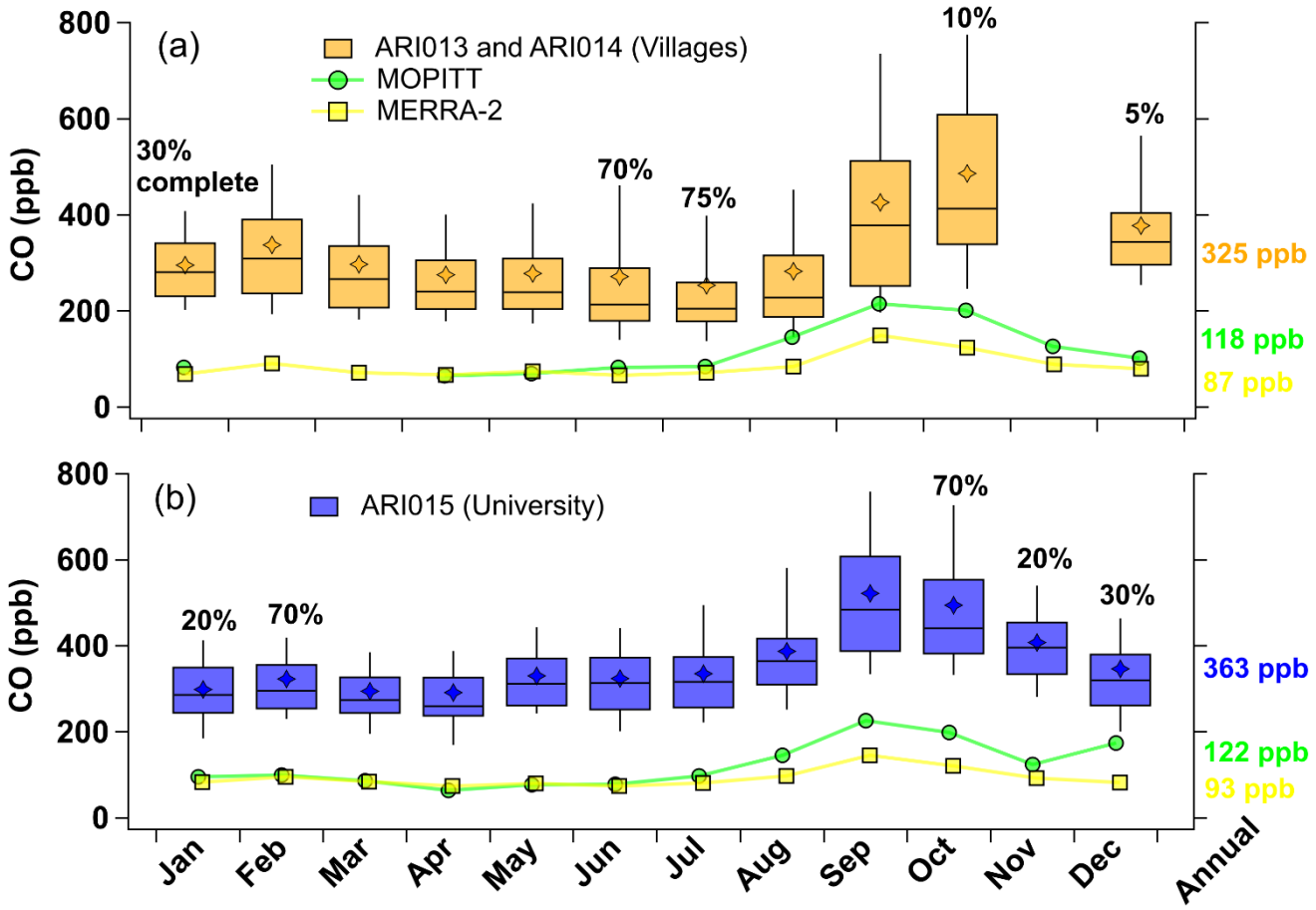
Figure 1~~Figure 7~~ shows the annual diurnal trend of the mean $PM_{2.5}$ mass concentration, with 1-sigma prediction intervals, using hourly-averaged data from each deployment location. Peak $PM_{2.5}$ concentrations were observed around 6 AM local time~~LT~~ at all sites, when morning biomass cookstove activity coincided with high RH (and more atmospherically stable) conditions. ~~Figure 3~~Figure 6 shows that the diurnal trends of ambient CO (also emitted by biomass burning) were similar to the $PM_{2.5}$ diurnal trends at each site. Again, the largest peaks were observed at the more densely populated ARI014 Village 1 site. The prediction intervals were widest between 5 and 7 AM local time~~LT~~, indicating overall low confidence in OPC-N2 measurements during this period. Afternoon and overnight means, coinciding with drier conditions, were similar across all three sites and prediction intervals were narrowest during afternoons. These results suggest background concentrations of $PM_{2.5}$ in rural Malawi were low (5 to 15 µg m^{-3}), but the OPC-N2 could not reliably quantify peak concentrations that were high and variable, dependent on the nearby sources and covariance with ambient meteorology (RH). Despite this, qualitative data from the OPC-N2 sensors was sufficient to identify nearby source activity and indicate periods when ambient

concentrations were likely high enough to be harmful to human health (and at least partially driven by cooking activities associated with much higher exposure concentrations).

3.5 Comparison of ARISense CO to remote sensing and reanalysis data

In the absence of in-situ surface data, we rely on satellites and models to estimate surface air quality for comparison of our results. 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. This analysis was limited to CO, given that the calibrated CO observations were the most reliable of the ARISense gas data and NASA remote sensing data products were more readily available for CO, compared to O_3 or NO_x.

~~Figure 4~~Figure 8 shows the mean monthly CO from the University (ARI015) and Village Mean (average of ARI013 and ARI014) sites compared to that from two area-averaged remote sensing products: CO surface mixing ratio from MOPITT and CO surface concentration from MERRA-2. All three data sets were compared from July 2017 to July 2018, focusing on differences between the peak agricultural burning (Sept to Oct) and non-burning (Dec to Jul) seasons. November and August were excluded from either description (peak burning or non-burning) for the following reasons: (a) a review of fire studies in the region consistently reported Sept and Oct as the dominant months of the burning season (Nieman et al., 2021), (b) Aug and Nov mark the beginning and end of the fire season, respectively, therefore cannot be considered non-burning months, (c) the exclusion of Aug and Nov better captures strong seasonal differences, providing a measurable benchmark to compare the satellite and surface data, and (d) ARISense data for the Village sites was unavailable for Nov 2017 (see Sect. 3.7 - on difficulties in deployment).The MERRA-2 data set was complete for the full year of interest, but MOPITT was missing data for the Village Mean region in February and March 2018. The remote sensing data sets were more similar to one another at the Village Mean site compared to the University site. At both sites, MOPITT reported higher CO concentrations than MERRA-2, especially in the peak burning season.


(a)
ARI013 and ARI014 (Villages)
MOPITT
MERRA-2
CO (ppb)
800
600
400
200
0
30%
complete
70%
75%
10%
5%
325 ppb
118 ppb
87 ppb
(b)
ARI015 (University)
20%
70%
70%
20%
30%
363 ppb
122 ppb
93 ppb
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Annual

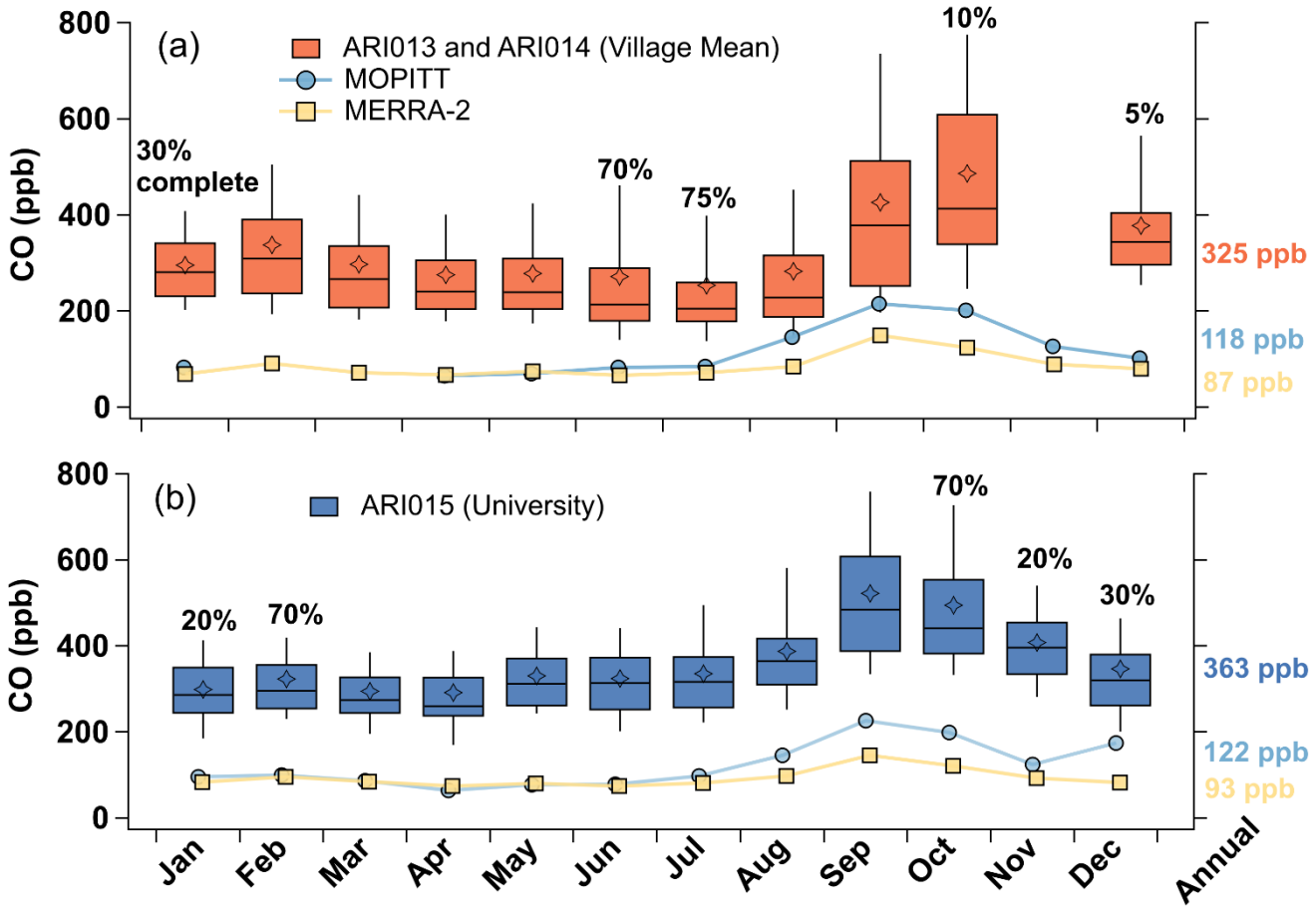


Figure 48: Monthly carbon monoxide (CO) concentration (ppb) reported by the surface ARISense (Tukey box plots) and remote sensing data products (lines and markers indicating mean monthly value) at the (a) Village Means and (b) University sites. Top and bottom of boxes indicate 75th and 25th percentiles, whiskers show 9th and 91st percentiles, midline indicates median, and stars indicate mean. The ARISense surface data were at least 80% complete for each month except where noted with a percentage text label. Data for July 2017 and July 2018 were averaged. Village Means data include datarepresents the average of from ARI014 (Village 1) and ARI013 (Village 2) data. The annual mean from each data source is given on the right axis. MOPITT (Multispectral CO Surface Mixing Ratio Daytime/Descending) is a satellite measurement; MERRA-2 (CO Surface Concentration -ENSEMBLE) is a global reanalysis product.

All three datasets (MOPITT, MERRA-2, and ARISense) indicated that annual mean CO concentrations were slightly higher overall at the University site than at the Village site, although this was less pronounced in MERRA-2. Similarly, all three data sets showed increased ambient concentrations during the peak burning season compared to the non-burning season at both sites. For ARISense, MOPITT, and MERRA-2 observations, respectively, peak season means were larger than non-

burning season means by 160 ppb, 130 ppb, 60 ppb (Village Mean) and 190 ppb, 115 ppb, 50 ppb (University). Although the ARISense indicated larger absolute differences between seasons, the relative increase at both sites was only about 50% of the non-burning season mean, while MOPITT and MERRA-2 reported increases of 125% and 75%, respectively. This could be explained by ARISense proximity to small-scale combustion activity not resolved by satellite imaging. Satellite-based observations approximate ambient background concentrations, which increased during the peak season due to regional agricultural burning. Meanwhile, the ARISense were exposed to ambient background concentrations as well as nearby biomass cookstove emissions, which likely remained consistent throughout the year, showing a lower relative seasonal increase during the peak burning season. Quantitative disagreement between surface and remote CO observations was highest during the burning season, especially at the University site (~~Figure 4~~Fig. 8). Remote sensing data suggested higher CO concentrations at the University compared to the Village Mean during non-burning periods, but during the peak burning season this difference shrank and similar concentrations were observed across both sites. Conversely, differences between ARISense differences grew by about 6% during the peak season. MERRA-2 and MOPITT concentrations were highest in September, consistent with ARISense data at the University site, but not the Village Mean site which peaked in October. However, 90% of the October CO data were missing for the Village site.

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 ±40% between ground-based CO measurements and Aura satellite observations at Cape Point station (Toihir 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. ~~Presently, comparisons between satellite and surface aerosol observations in Africa report moderate to poor agreement (Hersey et al., 2015; Malings et al., 2020(El Nadry et al., 2019)). (Malings et al., 2020)~~

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 ~12,000 square kilometer area (corresponding to 1° 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.

3.6 Comparison to other ambient measurements in SSA

Surface concentrations and diurnal trends of ARISense CO and PM in Malawi were comparable to studies in Kenya, Rwanda, Ethiopia, Uganda and South Africa (Delmas et al., 1999; DeWitt et al., 2019; Laakso et al., 2008; McFarlane et al., 2021; Nthusi, 2017; Scheel et al., 1998; Subramanian et al., 2020; Toihir et al., 2015). However, cComparison of O_3 concentrations suggested the calibrated ARISense observations ~~likely~~ underestimate~~d~~ actual concentrations. ARISense NOx observations were similar to two other studies (Delmas et al., 1999; Laakso et al., 2008), but overall there is ~~L~~little comparable data ~~were~~ available to assess NO_x concentrations in Africa. ~~.~~ The annual median (July 2017 to July 2018) ARISense surface concentrations estimated by the ARISense sensors were 9 to 11 ppb for NO_x, 4 to 15 ppb for O_3 and 240 to 330 ppb for CO.

ARISense CO observations were similar to regional CO concentrations in Central Africa (measured by aircraft), found to be in the range of 250-400 ppb (Delmas et al., 1999). A long-term ambient study at the Rwanda Climate Observatory found a mean CO concentration of 215 ppb from May 2015 to January 2017 (DeWitt et al., 2019), only slightly lower than~~similar to~~ our findings in Malawi. Another LCS study in Kigali, Rwanda observed a range in ambient CO concentrations, from 225 to 500 ppb~~,~~ at their rural and urban sites (Subramanian et al., 2020), spanning the concentration range we observed at our rural and semi-urban sites in Malawi~~.~~

Both studies of Rwanda found mean ambient O_3 concentrations of 30 to 40 ppb (DeWitt et al., 2019; Subramanian et al., 2020). The annual mean ARISense O_3 values were up to a factor of 10 lower, however, we identified quality assurance issues in the calibrated O_3 values, particularly for the second half of the deployment data, therefore the ARISense data are likely to be an underestimate. ~~.~~ For a "relatively clean background site located in dry savannah in South Africa: the annual median (July 2006 to July 2007) trace gas concentrations were equal to 1.4 ppb for NO_x, 36 ppb for O_3 and 105 ppb for CO" (Laakso et al., 2008). Background levels of NO_x and CO at this site were 2 to 5 times lower than the ARISense annual

means, yet background O_3 was in line with the Rwanda studies. This suggests regional ozone concentrations in Central and Southern Africa are on the order of 30-40 ppb.

This South African 'clean' background site had NO_x concentrations up to a factor of 10 lower (1.4 ppb) than ARISense measurements in Malawi (Laakso et al., 2008), but aerial measurements made during intense savanna fire activity in Central Africa found NO_y present in the range of 4-10 ppb (Delmas et al., 1999). Together, these studies suggest that the ARISense NOx concentrations (9-11 ppb) may be reasonable for our non-background, biomass emission influenced sites in Malawi.

Notably, t~~T~~he corresponding PM_1, $PM_{2.5}$ and PM_{10} median concentrations at the clean South Africa background site~~:~~, 9.0, 10.5 and 18.8 µg m^{-3}, respectively (Laakso et al., 2008), were comparable to ARISense observations. The annual median ARISense RH-corrected PM_1, $PM_{2.5}$ and PM_{10} concentrations were 4 to 7, 6 to 10, and 13 to 20 µg m^{-3}, respectively, depending on the site. It is possible that actual concentrations of fine PM were higher at the sites in Malawi, given that concentrations of gasesous emission tracer species (i.e., CO, NO_x) were higher compared to regional background levels found by other studies. However, given the high minimum cut-off diameter of the OPC-N2, this particle sensor would have been unable to detect ultrafine particles emitted from biomass burning. Average ambient $PM_{2.5}$ concentrations (measured with an Alphasense OPC-N2) were found to be 11 to 24 µg m^{-3} at various sites in Kenya, with higher pollution episode concentrations ranging from 35 to 51 µg m^{-3} (Nthusi, 2017). Median ARISense $PM_{2.5}$ concentrations ~~These concentrations~~ were also comparable ~~to measurements from Kenya and~~ to U.S. embassy measurements in Ethiopia and Uganda (DeWitt et al., 2019~~; Nthusi, 2017~~). ~~Average ambient $PM_{2.5}$ concentrations (measured with an Alphasense OPC N2) were found to be 11 to 24 µg m^{-3} at various sites in Kenya, with pollution episode concentrations ranging from 35 to 51 µg m^{-3} (Nthusi, 2017).~~ Taken together, these comparisons suggest PM levels in rural Malawi are comparable to regional measurements made across SSA, but localized impacts from biomass cookstoves can result in higher concentrations of fine PM, which are difficult to accurately quantify with the OPC-N2. In all, although t~~T~~hese comparisons are not a substitute for~~did not allow for~~ quantitative evaluation of the ARISense in Malawi, they provide a benchmark for comparison and suggest that the CO, NOx, and PM ARISense observations are reasonable for this region. At the same time, they cement our conclusion that ARISense O_3 observations are likely erroneous in this environment. ~~but nonetheless improved confidence in the CO and PM data.~~

3.7 Performance of ARISense sensor packages over time

Total data recovery for the 1 year deployment varied by site, season, and sensor, with rates ranging from 30% to 80% (Fig. S22). Average recovery for the 1 year deployment was around 60%, with highest recovery at the University site (~80%) and lowest at Village 1 site (~40%) ~~(Fig. S16)~~. Data across all sites had the highest completeness (>70%) in the cool-dry (Jun-July-Aug 2017 and 2018) and the cool-wet season (Mar-Apr-May 2018). Data losses were mostly explained by power outages, software failures, and sensor equilibration ~~warmup~~ times required after a power outage (Fig. S23). Power outages were common in the warm-wet season (Dec-Jan-Feb) due to insufficient solar intensity resulting from extended periods of

heavy cloud cover. At the ARI014 site, insufficient power led to an unanticipated diurnal cycle wherein the monitor would shut off in the early morning hours and require a few hours of solar power before turning on again. This daily cycle, coupled with the 8-hour long NO sensor re-equilibration time, led to almost 0% NO data recovery in the second half of the deployment for Village 1. In all, nearly 50% of data losses at the ARI014 site were due to system power failure or failure to write data to file. Corrupt USB storage devices, which we were slow to replace due to ongoing civil unrest (The GuardianLilongwe, 2017), resulted in significant data losses in the hot, dry season (Sept-Oct-Nov) at the two Village sites. Individual sensor failure was rare, but ; however, some gas two months of ARI014 O_x data were lost to electrochemical sensor drift and noise artifacts resulting from frequent power outages, and one OPC-N2 (ARI013) failed in the last 3 months of deployment due to an insect nest clogging the OPC-N2 inlet. In all, we recorded 6992 hours of data at the University site (ARI015), 5860 hours for Village 2 (ARI013), and 4720 hours for Village 1 (ARI014). Future deployments should include insect screens over all sensor inlets and improved battery storage and power systems that run at a longer duty cycle in the case of insufficient solar (e.g., power on only once battery is fully charged) to minimize the impact of sensor equilibration times on data recovery.

Since the monitors were deployed to their sites for >1 year, there was some observation overlap in seasonally similar data collected one year apart. To gain insight into sensor stability, we compared the data collected in the first month (July 2017) to the final month (July 2018) of the deployment, given that ambient environmental conditions were similar in July of both years (additional details in Sect. 11 of the Supplementary Information, Sect. 9). It is not possible to know if the range of gas concentrations were significantly different between July 2017 and July 2018.; We explored this analysis on the assumption that inter-annual variability in ambient concentrations was minimal. Bivariate distributions of the raw differential voltage readings from July 2017 and July 2018 showed that the most frequent observations (i.e., heaviest shaded regions) were approximately the same in both years (Fig. S259). Observable differences in the voltage measurements could be partially explained by known environmental differences. For example, the O_x sensor voltages in July 2018 were lower on average than in 2017, but this was consistent with lower temperatures and higher RH in 2018 compared to 2017. However, there was potential evidence of slightly reduced or altered responses in individual sensors, particularly the NO sensors in ARI013 and ARI015 and the CO sensors in ARI013 and ARI014. For these sensors, the 2018 distributions had less spread than the 2017 distributions, suggesting either less variation in ambient concentrations in 2018 or decreased sensitivity in the sensors. Diurnal plots from both years showed that the raw mean voltages and trends were largely consistent (Fig. S2630). However, again the most noticeable differences were in the individual CO and NO sensors identified from the bivariate distributions. For example, the CO peaks measured at mealtimes by ARI013 and ARI014 were about 50 mV lower in 2018 than 2017. These differences could be explained by lower concentrations in 2018 than 2017, changes in the raw sensor response over the one1 year period, or by both. Without reference equipment, we were unable to investigate sensor drift and decay more rigorously. This qualitative preliminary analysis suggests individual sensor responses likely were altered during the one1

year deployment, but there was no clear evidence for systematic deterioration within or across the electrochemical sensor groups used in the ARISense.

In general, the calibrated observations followed the trends identified from the raw sensor voltage readings. Calibrated CO data trends were consistent for both years, with the model responding as expected to the lower voltage readings in 2018 compared to 2017; for ARI013 and ARI014, the calibrated CO peaks at mealtimes were accordingly lower, by about 100 ppb, in 2018 (Fig. S2731). However, although the raw O_x sensor trends in 2018 and 2017 were consistent for all the ARISense (Fig. S26), the kNN hybrid model calibrated O_3 data were highly irregular between the two years (Fig. S2731). For example, the calibrated O_3 data for July 2017 showed the expected diurnal pattern (concentration increasing with solar intensity) with plateaus between 15 and 40 ppb, depending on the site. Yet in July 2018, although the raw O_x diurnal data looked similar to 2017, the calibrated data for ARI013 and ARI015 showed noon-time values between 0 and 5 ppb, and the diurnal trend for ARI013 showed a flat line (i.e., not correlated with solar activity). This finding, that raw O_x sensor voltages were similar year to year while the calibrated O_3 values were not, provides further evidence that the lack of comparable T/RH/ozone~~O_3~~ collocation data contributed to the non-physical O_3 trends observed~~measurements made in~~ during the second half of the deployment at the ARI013 and ARI015 sites.

Before their return to NC, ARI013 and ARI014 were used for high-concentration emissions monitoring experiments ~~in rural Malawi in July and August 2018,~~ after the one~~1~~ year ambient monitoring campaign was completed (Table 2). ~~The details (i.e., number of experiments, duration, approximate CO concentrations) of those experiments are discussed in Supplementary Information Sect. 10. ARI013 and ARI014 were then returned to NC and were collocated with reference instruments at the near-highway NC DEQ site used in the pre-deployment collocation.~~ The reference monitor data from the post-deployment collocation in NC (Aug 2018 to May 2019) were intended to enable investigation of changes in ARI013 and ARI014 raw sensor response and model performance. However, the resulting data instead demonstrated that sensors had been severely degraded during the high-concentration exposures. In the post-collocation data, the raw differential voltage gas sensor responses in ARI013 and ARI014 were well correlated with each other (R^2 = 0.7 to 0.9) (excluding the ARI013 O_x sensor which was clearly degraded~~;~~ Fig. S2832), but less correlated than during the pre-collocation comparison (R^2 = 0.9 to 0.99). To facilitate comparison with the pre-collocation performance metrics shown in Fig. 2 and Tables S4-S6~~1~~, the performance metrics ~~MAE and R^2 values~~ for the post-deployment collocation are given in Table S11 and S12~~7~~. Despite showing inter-sensor consistency, the raw differential voltage sensor measurements (other than CO) made by ARI013 and ARI014 were poorly correlated with reference measurements (Fig. S2933-S304). Inspection of the time series showed that the ARISense NO sensors tracked some spikes in the time-aligned NO reference data, but the NO_2 and O_x sensors did not track reference data trends (Fig. S315-S326). The time series of the differential voltage and temperature data suggest the gas sensors in ARI013 and ARI014 were responding similarly to changes in T and RH, but they were no longer sensitive to changes in the target gas (Fig. S315). This likely explains why the sensors in ARI013 and ARI014 were still well correlated

with each other, but why they were not correlated with reference measurements. The calibrated CO data were the only sensor data still correlated with CO reference measurements, although the calibrated CO data showed aberrant features (Fig. S337-S348). These ambient sensors (except for possibly the CO sensor) ~~were not appropriate for high concentration emissions monitoring and they did not survive extended exposure to near-source concentrations of biomass combustion emissions~~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. 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. ~~,~~ Ultimately, the O_x, NO, NO_2 sensors were permanently altered by the biomass burning emission experiments in Malawi, leading to ~~the~~ poor performance during post-deployment ~~collocation~~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. The partial exception to this was for the kNN hybrid calibrated CO data, which was roughly correlated with the reference data (R^2 = 0.5), suggesting that the CO sensors might retain some function after additional ~~collocation~~collocation and recalibration.

4 Conclusions

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 the calibrated ARISense sensor packages (excluding the NO_2 sensor) would be suitable for supplemental monitoring, based on U.S. EPA target values. 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 testing, 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 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.

~~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, however they suffered individual sensor failures and frequent power losses. Given that 30 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.~~

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 of deployment~~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~~fresh~~ biomass emissions completely degraded the sensors. Key manufacturer specifications indicated that the CO sensor was the most robust.~~;~~ ~~The~~the CO sensor exposure limit was 40 times higher than that of the O_x, NO, and NO_2 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_x, NO, and NO_2 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_2, O_x) and made them unsuitable for future use. ~~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.~~ We caution end users to carefully select an appropriate sensor package given pilot information about the emission sources in their target site.

~~For this deployment site, detailed information on nearby sources, ambient environmental conditions and diurnal emission patterns was needed to help interpret the calibrated measurements from Malawi and assess their validity. Performance assessment in NC suggested that the ARISense sensor package (excluding the NO_2 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. A lack of coherency 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 was unreliable. The collocation data were collected in an urban area near a highway, likely a VOC limited ozone regime, whereas the ozone chemistry in rural Malawi was more likely NO_x limited. This difference in chemical regime could have contributed to the poor performance of the calibration models in the deployment environment. Similarly, poor performance was observed for OPC N2 mass concentration measurements made in a strongly seasonal environment near ultrafine aerosol sources. These findings indicate that the collocation environment should ideally be the same as the deployment environment to simplify calibration and~~

~~improve performance. If in situ calibration is possible, evidence from this study supports the need for at least two collocation periods taking place during different seasons. For example, the high temperature low ozone regime (Fig. 2), which resulted in non-physical concentration predictions (Fig. 3), appeared to be associated with the second half of our deployment period (the wetter months). If in situ calibration is not feasible, laboratory calibration may enable calibration under a variety of environmental conditions and concentrations, however the effectiveness of lab based calibration should be verified. For example, an effective laboratory chamber calibration would require some level of a priori knowledge of typical ambient conditions and local ozone chemistry. 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.~~

~~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. However, the main lesson is clear; 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.~~

A growing body of literature highlights the potential value of LCS technologies ~~for Sub~~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.

Code availability

The basic random forest hybrid and quadratic regression model codes are available in the supplemental information of the original manuscript (doi:10.5281/zenodo.1482011). The k-nearest neighbor, high dimensional model representation, and multi-linear regression model codes are proprietary products of QuantAQ, Inc.; contact David H. Hagan with~~regarding~~ inquiries.

Data availability

The dataset used in this analysis is available as an open-access Dryad depository (doi:10.5061/dryad.cz8w9gj4n). The depository hosts pre-processed ARISense and reference datasets from the pre-deployment and post-deployment collocations, pre-processed RH-corrected OPC-N2 and MicroPEM datasets from the Malawi collocation, and collated ARISense datasets from the 1-year deployment at each of the three monitoring sites in Malawi. Please contact the corresponding author regarding raw data inquiries.

Author contribution

AG was responsible for conceptualization and funding acquisition. AG, EC, DH, and AB developed the methodology. EL, AG, and AB executed the ~~field~~deployment experiments. EC, DH, and AG provided supervision. DH and CM developed software. AB, EC, EL, and AG performed data analytics and visualization. AB wrote the original draft. CM, DH, EL, EC, and AG participated in review and editing.

Competing interests

Eben Cross and David Hagan are the co-founders of QuantAQ, a for-profit company which marketed the ARISense (since discontinued) and is actively developing and marketing sensor-based instrumentation.

Acknowledgements

We would like to acknowledge funding from the National Science Foundation under Coupled-Natural Human Systems Award Number: 1617359. Carl Malings would like to thank Naomi Zimmerman and the Carnegie Mellon University RAMPs Team for their assistance in developing low-cost sensor calibration approaches and acknowledge the EPA funding

source under assistance agreement no. 83628601 and EPA Grant Number R836286, as well as the Heinz Endowment Fund Grants E2375 and E3145. He would also like to acknowledge his support by an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center, administered by USRA through a contract with NASA. This work benefited from State assistance managed by the National Research Agency under the "Programme d'Investissements d'Avenir" under the reference "ANR-18-MPGA-0011" ("Make our planet great again" initiative). Ashley Bittner would like to thank Jeff Bean of Phillips 66 Research Center for his original R code used to generate the prediction intervals presented in Fig. 7, Elliott Hall for his contribution to analysis of the MicroPEM and OPC-N2 collocation data set, Jillian McNaught for her contribution to analysis of the Giovanni data sets, Ky Tanner for conducting the gravimetric filter analysis, Wyatt M. Champion for his assistance in making Fig. 1, Nathan Williams from Carnegie Mellon University for logistical support with ARISense repair, and all members of the Grieshop Atmosphere and Environment Lab for their feedback on figures. For their assistance in coordinating the collocation periods in North Carolina, we would like to thank the North Carolina Department of Environmental Quality and the Environmental Protection Agency and their dedicated employees including Sue Kimbrough (EPA), Richard Snow (EPA), Kay Roberts (NCDEQ), Timothy Skelding (NCDEQ), Joette Steger (NCDEQ), and Vitaly Karpusenko NCDEQ). Finally, we would like to thank all project principal investigators: Dr. Pamela Jagger, Dr. Rob Bailis, Dr. Jason West, and Dr. Adrian Ghilardi,; field assistants: Dominic Raphael and Twana Ghambi, project coordinators and assistants at the Lilongwe University of Agriculture and Natural Resources: Dr. Thabbie Chilongo, Dr. Charles Jumbe, and Misheck Mtaya, and all study participants from the villages of Mikundi and Makaula in Mulanje, Malawi.

References

Amegah, A. K.: Proliferation of low-cost sensors. What prospects for air pollution epidemiologic research in Sub-Saharan Africa?, Environ. Pollut., 241, 1132–1137, https://doi.org/10.1016/j.envpol.2018.06.044, 2018.

Amegah, A. K. and Agyei-Mensah, S.: Urban air pollution in Sub-Saharan Africa: Time for action, Environ. Pollut., 220, 738–743, https://doi.org/10.1016/j.envpol.2016.09.042, 2017.

Alphasense FAQs: https://www.alphasense.com/faqs/, last access: 11 October 2021.

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.

Awokola, B. I., Okello, G., Mortimer, K. J., Jewell, C. P., Erhart, A., and Semple, S.: Measuring Air Quality for Advocacy in Africa (MA3): Feasibility and Practicality of Longitudinal Ambient PM2.5 Measurement Using Low-Cost Sensors, Int. J. Environ. Res. Public Health, 17, https://doi.org/10.3390/ijerph17197243, 2020.

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., and Modzel, P.: Evaluation of Low-Cost Sensors for Ambient PM2.5 Monitoring, J. Sensors, vol. 2018, Article ID 5096540, 16 pages, https://doi.org/10.1155/2018/5096540, 2018.

Bittner, A., Cross, E. S., Hagan, D.H., Malings, C., Lipsky, E., and Grieshop, A.: Data accompanying "Performance Characterization of Lower-cost Air Sensors for Off-grid Deployment in Rural Malawi", Dryad, Dataset, https://doi.org/10.5061/dryad.cz8w9gj4n, 2021.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, J. Roy. Stat. Soc. B Met., 26, 211–252, 1964.

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, Faculty of Science, Medicine and Health - Papers: part A, 1927–1956, https://doi.org/10.5194/amt-10-1927-2017, 2017.

~~Buchwitz, M., Khlystova, I., Bovensmann, H., and Burrows, J. P.: Three years of global carbon monoxide from SCIAMACHY: comparison with MOPITT and first results related to the detection of enhanced CO over cities, Atmos. Chem. Phys., 7, 2399–2411, 2007.~~

Buehler, C., Xiong, F., Zamora, M. L., Skog, K. M., Kohrman-Glaser, J., Colton, S., McNamara, M., Ryan, K., Redlich, C., Bartos, M., Wong, B., Kerkez, B., Koehler, K., and Gentner, D. R.: Stationary and portable multipollutant monitors for high-spatiotemporal-resolution air quality studies including online calibration, Atmos. Meas. Tech., 14, 995–1013, https://doi.org/10.5194/amt-14-995-2021, 2021.

Bulot, F. M. J., Johnston, S. J., Basford, P. J., Easton, N. H. C., Apetroaie-Cristea, M., Foster, G. L., Morris, A. K. R., Cox, S. J., and Loxham, M.: Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment, Sci. Rep-UK, 9, 7497, https://doi.org/10.1038/s41598-019-43716-3, 2019.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., and Bartonova, A.: Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, Environment International, 99, 293–302, https://doi.org/10.1016/j.envint.2016.12.007, 2017.

Chatzidiakou, L., Krause, A., Popoola, O. A. M., Antonio, A. D., Kellaway, M., Han, Y., Squires, F. A., Wang, T., Zhang, H., Wang, Q., Fan, Y., Chen, S., Hu, M., Quint, J. K., Barratt, B., Kelly, F. J., Zhu, T., and Jones, R. L.: Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments, Atmos. Meas. Tech., 12, 4643–4657, https://doi.org/10.5194/amt-12-4643-2019, 2019.

Considine, E. M., Reid, C. E., Ogletree, M. R., and Dye, T.: Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airborne particulate matter sensor network, Environ. Pollut., 268, 115833, https://doi.org/10.1016/j.envpol.2020.115833, 2021.

Crilley, L. R., Shaw, M., Pound, R., Kramer, L. J., Price, R., Young, S., Lewis, A. C., and Pope, F. D.: Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring, Atmos. Meas. Tech., 11, 709–720, https://doi.org/10.5194/amt-11-709-2018, 2018.

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.

Delmas, R. A., Druilhet, A., Cros, B., Durand, P., Delon, C., Lacaux, J. P., Brustet, J. M., Serça, D., Affre, C., Guenther, A., Greenberg, J., Baugh, W., Harley, P., Klinger, L., Ginoux, P., Brasseur, G., Zimmerman, P. R., Grégoire, J. M., Janodet, E., Tournier, A., Perros, P., Marion, Th., Gaudichet, A., Cachier, H., Ruellan, S., Masclet, P., Cautenet, S., Poulet, D., Biona, C. B., Nganga, D., Tathy, J. P., Minga, A., Loemba-Ndembi, J., and Ceccato, P.: Experiment for Regional Sources and Sinks of Oxidants (EXPRESSO): An overview, J. Geophys. Res., 104, 30609–30624, https://doi.org/10.1029/1999JD900291, 1999.

DeWitt, H. L., Gasore, J., Rupakheti, M., Potter, K. E., Prinn, R. G., Ndikubwimana, J. de D., Nkusi, J., and Safari, B.: Seasonal and diurnal variability in O_3, black carbon, and CO measured at the Rwanda Climate Observatory, Atmos. Chem. Phys., 19, 2063–2078, https://doi.org/10.5194/acp-19-2063-2019, 2019.

Di Antonio, A., Popoola, O. A. M., Ouyang, B., Saffell, J., and Jones, R. L.: Developing a Relative Humidity Correction for Low-Cost Sensors Measuring Ambient Particulate Matter, Sensors-Basel, 18, https://doi.org/10.3390/s18092790, 2018.

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.

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.

Duvall, R., Clements, A., Hagler, G., Kamal, A., Vasu Kilar, Goodman, L., Frederick, S., Johnson Barkjohn K., VonWald, I., Greene, D., and Dye, T.: Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors: Use in Ambient, Outdoor, Fixed Site, Non-Regulatory Supplemental and Informational Monitoring Applications, U.S. EPA Office of Research and Development, Washington, DC, 2021a.

Duvall, R., Clements, A., Hagler, G., Kamal, A., Vasu Kilar, Goodman, L., Frederick, S., Johnson Barkjohn K., VonWald, I., Greene, D., and Dye, T.: Performance Testing Protocols, Metrics, and Target Values for Ozone Air Sensors: Use in Ambient, Outdoor, Fixed Site, Non-Regulatory and Informational Monitoring Applications, U.S. EPA Office of Research and Development, Washington, DC, 2021b.

El-Nadry, M., Li, W., El-Askary, H., Awad, M. A., and Mostafa, A. R.: Urban Health Related Air Quality Indicators over the Middle East and North Africa Countries Using Multiple Satellites and AERONET Data, 11, 2096, https://doi.org/10.3390/rs11182096, 2019.

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, J. Geophys. Res-Atmos., 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.

Fullerton, D. G., Semple, S., Kalambo, F., Suseno, A., Malamba, R., Henderson, G., Ayres, J. G., and Gordon, S. B.: Biomass fuel use and indoor air pollution in homes in Malawi, Occup. Environ. Med., 66, 777–783, https://doi.org/10.1136/oem.2008.045013, 2009.

Fullerton, D. G., Suseno, A., Semple, S., Kalambo, F., Malamba, R., White, S., Jack, S., Calverley, P. M., and Gordon, S. B.: Wood smoke exposure, poverty and impaired lung function in Malawian adults, The International Journal of Tuberculosis and Lung Disease, 15, 391–398, 2011.

Giordano, M. R., Malings, C., Pandis, S. N., Presto, A. A., McNeill, V. F., Westervelt, D. M., Beekmann, M., and Subramanian, R.: From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively

calibrating low-cost particulate matter mass sensors, J. Aerosol Sci., 105833, https://doi.org/10.1016/j.jaerosci.2021.105833, 2021.

Gulia, S., Khanna, I., Shukla, K., and Khare, M.: Ambient air pollutant monitoring and analysis protocol for low and middle income countries: An element of comprehensive urban air quality management framework, Atmos. Environ., 222, 117120, https://doi.org/10.1016/j.atmosenv.2019.117120, 2020.

Hagan, D. H. and Kroll, J. H.: Assessing the accuracy of low-cost optical particle sensors using a physics-based approach, Atmos. Meas. Tech., 13, 6343–6355, https://doi.org/10.5194/amt-13-6343-2020, 2020.

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.

Hagan, D. H., Gani, S., Bhandari, S., Patel, K., Habib, G., Apte, J. S., Hildebrandt Ruiz, L., and Kroll, J. H.: Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India, Environ. Sci. Technol. Lett., 6, 467–472, https://doi.org/10.1021/acs.estlett.9b00393, 2019.

Hersey, S. P., Garland, R. M., Crosbie, E., Shingler, T., Sorooshian, A., Piketh, S., and Burger, R.: An overview of regional and local characteristics of aerosols in South Africa using satellite, ground, and modeling data, Atmos. Chem. Phys., 15, 4259–4278, https://doi.org/10.5194/acp-15-4259-2015, 2015.

Jary, H. R., Aston, S., Ho, A., Giorgi, E., Kalata, N., Nyirenda, M., Mallewa, J., Peterson, I., Gordon, S. B., and Mortimer, K.: Household air pollution, chronic respiratory disease and pneumonia in Malawian adults: A case-control study, Wellcome Open Res., 2, https://doi.org/10.12688/wellcomeopenres.12621.1, 2017.

Kelly, K. E., Xing, W. W., Sayahi, T., Mitchell, L., Becnel, T., Gaillardon, P.-E., Meyer, M., and Whitaker, R. T.: Community-Based Measurements Reveal Unseen Differences during Air Pollution Episodes, Environ. Sci. Technol., 55, 120–128, https://doi.org/10.1021/acs.est.0c02341, 2021.

Laakso, L., Laakso, H., Aalto, P. P., Keronen, P., Petäjä, T., Nieminen, T., Pohja, T., Siivola, E., Kulmala, M., Kgabi, N., Molefe, M., Mabaso, D., Phalatse, D., Pienaar, K., and Kerminen, V.-M.: Basic characteristics of atmospheric particles, trace gases and meteorology in a relatively clean Southern African Savannah environment, Atmos. Chem. Phys., 8, 4823–4839, https://doi.org/10.5194/acp-8-4823-2008, 2008.

~~Lewis, A. and Edwards, P.: Validate personal air-pollution sensors, 535, 29–31, https://doi.org/10.1038/535029a, 2016.~~

~~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, 2016a.~~

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~~b~~.

Li, J., Hauryliuk, A., Malings, C., Eilenberg, S. R., Subramanian, R., and Presto, A. A.: Characterizing the Aging of Alphasense NO2 Sensors in Long-Term Field Deployments, ACS Sens., https://doi.org/10.1021/acssensors.1c00729, 2021.

~~Lilongwe, R. in: UN moves staff after mobs kill five in Malawi vampire scare, The Guardian, 9th October, 2017.~~

Liousse, C., Assamoi, E., Criqui, P., Granier, C., and Rosset, R.: Explosive growth in African combustion emissions from 2005 to 2030, Environ. Res. Lett., 9, 035003, https://doi.org/10.1088/1748-9326/9/3/035003, 2014.

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S. P. N., Zimmerman, N., Kara, L. B., Presto, A. A., and R. Subramanian: Supplementary Data for "Development of a General Calibration Model and Long-Term Performance Evaluation of Low-Cost Sensors for Air Pollutant Gas Monitoring" (abridged version) (2.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1482011, 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.

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., and Subramanian, R.: Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation, Aerosol Sci. Tech., 0, 1–15, https://doi.org/10.1080/02786826.2019.1623863, 2019b.

Malings, C., Westervelt, D. M., Hauryliuk, A., Presto, A. A., Grieshop, A., Bittner, A., Beekmann, M., and R. Subramanian: Application of low-cost fine particulate mass monitors to convert satellite aerosol optical depth to surface concentrations in North America and Africa, Atmos. Meas. Tech., 13, 3873–3892, https://doi.org/10.5194/amt-13-3873-2020, 2020.

Mapoma, H. and Xie, X.: State of Air Quality in Malawi, J. Environ. Prot., 4, 1258–1264, https://doi.org/10.4236/jep.2013.411146, 2013.

Marais, E. A. and Wiedinmyer, C.: Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-Africa), Environ. Sci. Technol., 50, 10739–10745, https://doi.org/10.1021/acs.est.6b02602, 2016.

Martin, R. V., Brauer, M., van Donkelaar, A., Shaddick, G., Narain, U., and Dey, S.: No one knows which city has the highest concentration of fine particulate matter, Atmos. Environ.: X, 3, 100040, https://doi.org/10.1016/j.aeaoa.2019.100040, 2019.

MBS: Industrial emissions from mobile and stationary sources-Specifications, Malawi Bureau of Standards, Malawi, 2005.

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 PM2.5 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.

Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J. J., McLeod, M. W., Hodgson, T. F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J. R., and Jones, R. L.: The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks, Atmos. Environ., 70, 186–203, https://doi.org/10.1016/j.atmosenv.2012.11.060, 2013.

Murray, C. J. L., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Raddad, L. J., Abushouk, A. I., Adabi, M., Adekanmbi, V., Adeoye, A. M., Adetokunboh, O. O., Adham, D., Advani, S. M., Agarwal, G., Aghamir, S. M. K., Agrawal, A., Ahmad, T., Ahmadi, K., Ahmadi, M., Ahmadieh, H., Ahmed, M. B., Akalu, T. Y., Akinyemi, R. O., Akinyemiju, T., Akombi, B., Akunna, C. J., Alahdab, F., Al-Aly, Z., Alam, K., Alam, S., Alam, T., Alanezi, F. M., Alanzi, T. M., Alemu, B. wassihun, Alhabib, K. F., Ali, M., Ali, S., Alicandro, G., Alinia, C., Alipour, V., Alizade, H., Aljunid, S. M., Alla, F., Allebeck, P., Almasi-Hashiani, A., Al-Mekhlafi, H. M., Alonso, J., Altirkawi, K. A., Amini-Rarani, M., Amiri, F., Amugsi, D. A., Ancuceanu, R., Anderlini, D., Anderson, J. A.,

Andrei, C. L., Andrei, T., Angus, C., Anjomshoa, M., Ansari, F., Ansari-Moghaddam, A., Antonazzo, I. C., Antonio, C. A. T., Antony, C. M., Antriyandarti, E., Anvari, D., Anwer, R., Appiah, S. C. Y., Arabloo, J., Arab-Zozani, M., Ariani, F., Armoon, B., Ärnlöv, J., Arzani, A., Asadi-Aliabadi, M., Asadi-Pooya, A. A., Ashbaugh, C., Assmus, M., Atafar, Z., Atnafu, D. D., Atout, M. M. W., Ausloos, F., Ausloos, M., Quintanilla, B. P. A., Ayano, G., Ayanore, M. A., Azari, S., Azarian, G., Azene, Z. N., et al.: Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, The Lancet, 396, 1223–1249, https://doi.org/10.1016/S0140-6736(20)30752-2, 2020.

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

Nieman, W. A., van Wilgen, B. W., and Leslie, A. J.: A reconstruction of the recent fire regimes of Majete Wildlife Reserve, Malawi, using remote sensing, Fire Ecol., 17, 4, https://doi.org/10.1186/s42408-020-00090-0, 2021.

Nthusi, V.: Nairobi Air Quality Monitoring Sensor Network Report - April 2017, https://doi.org/10.13140/RG.2.2.10240.64009, 2017.

Petkova, E. P., Jack, D. W., Volavka-Close, N. H., and Kinney, P. L.: Particulate matter pollution in African cities, Air Qual. Atmos. Health, 6, 603–614, https://doi.org/10.1007/s11869-013-0199-6, 2013.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7, 1961–1971, https://doi.org/10.5194/acp-7-1961-2007, 2007.

Pinder, R. W., Klopp, J. M., Kleiman, G., Hagler, G. S. W., Awe, Y., and Terry, S.: Opportunities and challenges for filling the air quality data gap in low- and middle-income countries, Atmos. Environ., 215, 116794, https://doi.org/10.1016/j.atmosenv.2019.06.032, 2019.

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.

Queface, A. J., Piketh, S. J., Eck, T. F., Tsay, S.-C., and Mavume, A. F.: Climatology of aerosol optical properties in Southern Africa, Atmos. Environ., 45, 2910–2921, https://doi.org/10.1016/j.atmosenv.2011.01.056, 2011.

Rahal, F.: Low-cost sensors, an interesting alternative for air quality monitoring in Africa., Clean Air Journal, 30, https://doi.org/10.17159/caj/2020/30/2.9223, 2020.

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Yasar, A., and Rickerby, D.: End-user perspective of low-cost sensors for outdoor air pollution monitoring, Sci. Total Environ., 607–608, 691–705, https://doi.org/10.1016/j.scitotenv.2017.06.266, 2017.

Reid, J. S., Koppmann, R., Eck, T. F., and Eleuterio, D. P.: A review of biomass burning emissions part II: intensive physical properties of biomass burning particles, Atmos. Chem. Phys., 5, 799–825, 2005.

Saha, P. K., Khlystov, A., and Grieshop, A. P.: Downwind evolution of the volatility and mixing state of near-road aerosols near a US interstate highway, Atmos. Chem. Phys., 18, 2139–2154, https://doi.org/10.5194/acp-18-2139-2018, 2018.

Sahu, R., Nagal, A., Dixit, K. K., Unnibhavi, H., Mantravadi, S., Nair, S., Simmhan, Y., Mishra, B., Zele, R., Sutaria, R., Motghare, V. M., Kar, P., and Tripathi, S. N.: Robust statistical calibration and characterization of portable low-cost air

quality monitoring sensors to quantify real-time O_3 and NO_2 concentrations in diverse environments, Atmos. Meas. Tech., 14, 37–52, https://doi.org/10.5194/amt-14-37-2021, 2021.

Scheel, H. E., Brunke, E.-G., Sladkovic, R., and Seiler, W.: In situ CO concentrations at the sites Zugspitze (47°N, 11°E) and Cape Point (34°S, 18°E) in April and October 1994, J. Geophys. Res-Atmos., 103, 19295–19304, https://doi.org/10.1029/96JD04010, 1998.

Sewor, C., Obeng, A. A., and Amegah, A. K.: Commentary: The Ghana Urban Air Quality Project (GHAir): Bridging air pollution data gaps in Ghana, Clean Air Journal, 31, https://doi.org/10.17159/caj/2021/31/1.11172, 2021.

Shikwambana, L. and Tsoeleng, L. T.: Impacts of population growth and land use on air quality. A case study of Tshwane, Rustenburg and Emalahleni, South Africa, S. Afr. Geogr. J.,102, 209–222, https://doi.org/10.1080/03736245.2019.1670234, 2020.

Sousan, S., Koehler, K., Hallett, L., and Peters, T. M.: Evaluation of the Alphasense optical particle counter (OPC-N2) and the Grimm portable aerosol spectrometer (PAS-1.108), Aerosol Sci. Tech., 50, 1352–1365, https://doi.org/10.1080/02786826.2016.1232859, 2016.

Spinelle, L., Gerboles, M., and Aleixandre, M.: Performance Evaluation of Amperometric Sensors for the Monitoring of O3 and NO2 in Ambient Air at ppb Level, Chem. Engineer. Trans., 120, 480–483, https://doi.org/10.1016/j.proeng.2015.08.676, 2015.

Spinelle, L., Gerboles, M., Aleixandre, M., and Bonavitacola, F.: Evaluation of Metal Oxides Sensors for the Monitoring of O_3 in Ambient Air at Ppb Level, Procedia Engineer., 1, 54, 319–324, https://doi.org/10.3303/CET1654054, 2016.

Stevens, T. and Madani, K.: Future climate impacts on maize farming and food security in Malawi, Sci. Rep., 6, https://doi.org/10.1038/srep36241, 2016.

Subramanian, R., Ellis, A., Torres-Delgado, E., Tanzer, R., Malings, C., Rivera, F., Morales, M., Baumgardner, D., Presto, A., and Mayol-Bracero, O. L.: Air Quality in Puerto Rico in the Aftermath of Hurricane Maria: A Case Study on the Use of Lower Cost Air Quality Monitors, ACS Earth Space Chem., https://doi.org/10.1021/acsearthspacechem.8b00079, 2018.

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, Clean Air Journal, 30, https://doi.org/10.17159/caj/2020/30/2.8023, 2020.

Subramanian, R. and Garland, R.: Editorial: The powerful potential of low-cost sensors for air quality research in Africa, Clean Air Journal, 31, https://doi.org/10.17159/caj/2021/31/1.11274, 2021.

The Guardian, Lilongwe, R. in: UN moves staff after mobs kill five in Malawi vampire scare, 9th October, 2017.

Thorson, J., Collier-Oxandale, A., and Hannigan, M.: Using A Low-Cost Sensor Array and Machine Learning Techniques to Detect Complex Pollutant Mixtures and Identify Likely Sources, Sensors, 19, 3723, https://doi.org/10.3390/s19173723, 2019.

Toihir, 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, S. Afr. J. Sci., 111, 1–9, 2015.

Topalović, D. B., Davidović, M. D., Jovanović, M., Bartonova, A., Ristovski, Z., and Jovašević-Stojanović, M.: In search of an optimal in-field calibration method of low-cost gas sensors for ambient air pollutants: Comparison of linear, multilinear and artificial neural network approaches, Atmos. Environ., 213, 640–658, https://doi.org/10.1016/j.atmosenv.2019.06.028, 2019.

Wernecke, B. and Wright, C.: Commentary: Opportunities for the application of low-cost sensors in epidemiological studies to advance evidence of air pollution impacts on human health, Clean Air Journal, 31, https://doi.org/10.17159/caj/2021/31/1.11219, 2021.

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.

Williams, R., Long, R., Beaver, M., and Kaufman, A.: Sensor Evaluation Report, U.S. Environmental Protection Agency, Washington, DC, 2014b.

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, J. Geophys. Res-Atmos., 113, https://doi.org/10.1029/2007JD009229, 2008.

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, Atmos. Chem. Phys., 10, 3479–3494, https://doi.org/10.5194/acp-10-3479-2010, 2010.

Zhang, T., Chillrud, S. N., Pitiranggon, M., Ross, J., Ji, J., and Yan, B.: Development of an approach to correcting MicroPEM baseline drift, Environ. Res., 164, 39–44, https://doi.org/10.1016/j.envres.2018.01.045, 2018.

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.

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., and R. Subramanian: A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, Atmos. Meas. Tech., 11, 291–313, https://doi.org/10.5194/amt-11-291-2018, 2018.