THESIS

AUTONOMOUS LOW-COST OZONE SENSORS: DEVELOPMENT, CALIBRATION, AND APPLICATION TO STUDY EXPOSURE AND SPATIAL GRADIENTS

Submitted by

Dylan M. Giardina

Department of Mechanical Engineering

In partial fulfillment of requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2022

Master's Committee:

Advisor: Shantanu Jathar

Sheryl Magzamen
John Volckens
Samuel Bechara

Copyright by Dylan Michel Giardina 2022

All Rights Reserved

ABSTRACT

AUTONOMOUS LOW-COST OZONE SENSORS: DEVELOPMENT, CALIBRATION, AND APPLICATION TO STUDY EXPOSURE AND SPATIAL GRADIENTS

Ozone (O_3), a criteria pollutant and atmospheric oxidant, is not routinely measured in rural and remote environments and hence exposure to ozone pollution in these regions remains poorly understood. In this work, we built, calibrated, and deployed five low-cost, autonomous ozone sensor systems (called MOOS) in Northern Colorado, a region that is non-compliant for O_3 during the summertime. Each MOOS included the following components: (i) an Aeroqual SM50, a heated metal oxide ozone sensor, mounted inside a custom sensor housing, (ii) a power system that consisted of a 30 W solar panel, 108 Wh lithium-ion battery, and charge controller, (iii) a Particle Boron to acquire, process, and transmit data to the Cloud, and (iv) an environmental sensor to measure temperature, relative humidity, and pressure. In a three-week long collocated study, we found that all MOOS, calibrated using 48 hours of reference data, compared well against reference monitors with a measurement error between 4-6 parts per billion by volume (ppbv). Manufacturer- and laboratory-based calibrations over- and under-estimated ozone levels at higher and lower ozone mixing ratios, respectively. When deployed in Northern Colorado for an additional three weeks to measure O_3 exposure and study O_3 trends across an urban-rural gradient, we found that the MOOS, calibrated using data from the collocated study and calibrated using 48 hours of reference data in the field, demonstrated good sensor performance (RMSE of 3.98 - 8.80 ppbv and MBE of 0.22 - 3.82 ppbv). Compared to the collocated study, the field study resulted in larger measurement errors for all five MOOS (RMSE of 3.66 - 4.00 versus RMSE of 3.98 - 8.80). Furthermore, there was modest variability in the field performance across the different MOOS

(RMSE < 5 ppbv) that could not be explained by environmental differences between the different sites (e.g., proximity of the MOOS to the reference monitor, land use type, temperature). We found that MOOS were able to capture 100% of non-compliant O_3 days during the collocated study and between 25-87% of non-compliant O_3 days during the field study depending on the calibration approach used. Furthermore, both reference monitors and MOOS deployed along the east-west corridor in Northern Colorado were able to capture the negative, west-east O_3 gradients observed in previous aircraft and modeling studies. Overall, our study indicates that the MOOS shows promise as a low-cost O_3 sensor that could be used to supplement routine ambient monitoring and characterize regional ozone pollution.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Shantanu Jathar for being a phenomenal advisor and supporting me through both my graduate and undergraduate career. This work would not be possible without his guidance and support. I would also like to thank my committee members Dr. Sheryl Magzamen, Dr. John Volckens, and Dr. Samuel Bechara for taking the time to provide valuable feedback on this research and giving me the opportunity to present this work.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iv
LIST OF TABLES vi
LIST OF FIGURES ii
CHAPTER 1: INTRODUCTION 1
1.1 Overview of Ozone Chemistry 1
1.2 Ozone Pollution in the Western United States 2
1.3 Motivations for Low-Cost Ozone Sensing 4
1.4 Calibration of Low-Cost Ozone Sensors 5
CHAPTER 2: METHODS 9
2.1 Metal Oxide Ozone Sensor (MOOS) 9
2.2 Reference Monitors 11
2.3.1 Collocated Deployments 13
2.3.2 Field Deployments 14
2.4 MOOS Calibration, Evaluation, and Performance Metrics 15
CHAPTER 3: RESULTS 19
3.1 Collocated Rooftop Study to Evaluate Calibration Methods 19
3.2 Field Deployment and Comparison to Reference Measurements 24
3.3 O_3 Exposure and East-West Gradients 31
CHAPTER 4: CONCLUSIONS, UNCERTAINTIES, AND DIRECTIONS FOR FUTURE RESEARCH 36
REFERENCES 39
APPENDIX 44

LIST OF TABLES

Table 1: Overview of MOOS deployment duration and collocated reference monitors 13

Table S1: Breakdown of component costs for 1 MOOS system. 44
Table S2: Multivariable regression coefficients for MOOS A-E used to develop MCOL..... 54

LIST OF FIGURES

Figure 1: Multi-panel figure that shows the MOOS electrical schematic and 3D cad drawing of the sensor housing 9
Figure 2: Rooftop and field deployment locations across Eastern Colorado 14
Figure 3: MOOS-A O3 mixing ratio estimates using different calibration approaches compared to collocated reference measurements. 19
Figure 4: MOOS-A O_3 mixing ratio estimates corrected using an n-fold cross validation LCOL approach compared to collocated reference measurements 21
Figure 5: MOOS O_3 mixing ratio estimates corrected using LCOL-F, LCOL-Pre, LCOL-24, LCOL-48, and LCOL-96 approaches compared to reference measurements during the field deployment 24
Figure 6: Variation in statistical metrics comparing MOOS O_3 mixing ratio estimates reference measurements 28
Figure 7: MOOS O_3 MDA8 estimates compared to reference MDA8 measurements 31
Figure 8: Difference between reference monitors O_3 measurements and MOOS O_3 estimates compared to the CDPHE reference monitor located in FCW 34

Figure S1: Photo of MOOS deployed in the field 45
Figure S2: Diagram showing active airflow in the sensor housing. 46
Figure S3: Scatter plot of 2B #2060 O_3 analyzer and Thermo O_3 analyzer compared against the other 2B #2061 47
Figure S4: MOOS field deployments 48
Figure S5: Same as Figure 3 but displaying results for MOOS B. 49
Figure S6: Same as Figure 3 but displaying results for MOOS C. 50
Figure S7: Same as Figure 3 but displaying results for MOOS D. 51
Figure S8: Same as Figure 3 but displaying results for MOOS E. 52
Figure S9: Relationship between the LCOL-Full/Reference ratio and ambient temperature and relative humidity during the rooftop deployment. 53
Figure S10: Same as Figure 4 but displaying results for MOOS B. 55
Figure S11: Same as Figure 4 but displaying results for MOOS C. 56
Figure S12: Same as Figure 4 but displaying results for MOOS D. 57
Figure S13: Same as Figure 4 but displaying results for MOOS E. 58
Figure S14: MOOS-A O_3 mixing ratio estimates corrected using n-fold cross validation MCOL 59
Figure S15: Same as Figure S12 but displaying results for MOOS B. 60
Figure S16: Same as Figure S12 but displaying results for MOOS C. 61
Figure S17: Same as Figure S12 but displaying results for MOOS D. 62
Figure S18: Same as Figure S12 but displaying results for MOOS E. 63
Figure S19: CDF plot demonstrating the 90% cutoff used to select LCOL-48. 64
Figure S20: Same as Figure 7 showing the full results for FCW. 65
Figure S21: MOOS O_3 mixing ratio estimates corrected using LCOL-Full and LCOL-Pre approaches compared to reference measurements during the post-field rooftop deployment which spanned roughly 1 week after the field deployment. 66

CHAPTER 1: INTRODUCTION

1.1 Overview of Ozone Chemistry

Ozone (O_3) is a molecule formed by the reaction of free oxygen (O) and molecular oxygen (O_2) and it exists in both the stratosphere and troposphere. While stratospheric O_3 is formed naturally and plays an important role in blocking ultraviolet radiation, tropospheric O_3 is heavily influenced by anthropogenic activity and is a toxic criteria pollutant. O_3 is a highly reactive oxidizing gas and will react with most chemical constituents, including biological tissue. Acute exposure to O_3 at concentrations >80 parts per billion by volume (ppbv) damages the epithelium of the respiratory tract, causing irritation and inflammation of the tissue and has been correlated with epidemiological events such as higher hospital admissions (Lippmann and Schlesinger, 2000; Iriti and Faoro, 2007). Individuals with existing respiratory conditions and those working or recreating outside with higher heart rate have also been shown to have acute and chronic O_3 symptoms exacerbated (Lin *et al.*, 2008). Overall, O_3 is estimated to be responsible for a global mortality burden of 1.23 million deaths, with many lower income areas being disproportionately affected by O_3 pollution (Malley *et al.*, 2017).

The free oxygen (O) needed for O_3 formation is produced via photolysis of nitrogen dioxide (NO_2) into nitric oxide (NO) (collectively $NO + NO_2 = NO_X$). NO_2 is regenerated from NO through the photooxidation of methane (CH_4), carbon monoxide (CO), and volatile organic compounds (VOCs) (Jacob, 1999). O_3 production requires NO_X as a method for free oxygen production, CH_4, CO, and VOCs to enable NO_2 regeneration, and sunlight for NO_2 photolysis. Anthropogenic sources such as power generation, oil and gas (O&G) extraction, traffic, and agriculture emit VOC and NO_X as O_3 precursors. In areas with high anthropogenic emissions of O_3 precursors, high

available sunlight needed for photochemical reactions to take place, and meteorological phenomena causing the stagnation of O_3 or high concentrations of precursors, O_3 mixing ratios can frequently exceed safe levels (>70 ppbv as per National Ambient Air Quality Standards (NAAQS)) and spike to greater than 100 ppbv (Simon *et al.*, 2015). Anthropogenic sources in urban centers are responsible for emitting large quantities of O_3 precursors and have been the primary target of legislation to reduce precursor emissions and therefore O_3 production. While legislation and regulation informed by these monitoring efforts have helped to reduce O_3 pollution in many US urban areas (EPA, 2015), O_3 pollution continues to be a pressing public health problem in areas with ideal conditions for O_3 formation and numerous and varied precursor sources.

1.2 Ozone Pollution in the Western United States

All metropolitan areas in the United States impacted by O_3 pollution have implemented dedicated monitoring efforts to inform air quality advisories and protect populated areas from acute O_3 exposure. A major criterion for designating O_3 monitoring efforts in the US is the nearby metropolitan population. O_3 monitoring efforts in sparsely populated areas are dependent on state and community initiatives and many rural communities impacted by O_3 may not have access to real time O_3 data. Transported O_3 precursors from both domestic sources and Asia are predominantly responsible for baseline rural O_3 in the United States (Cooper *et al.*, 2012), with local O_3 precursors emissions potentially enhancing rural O_3 above baseline levels (Carter and Seinfeld, 2012; Jaffe *et al.*, 2013). O&G extraction is a well-documented source of local O_3 precursors including CO, CH_4, NO_X, and VOCs, and plays an important role in rural O_3 pollution. As an example, O&G activity in the Uintah basin in Utah has caused winter O_3 pollution to frequently exceed NAAQS during winter, despite the area's isolation from metropolitan activity

(Edwards *et al.*, 2014). High wintertime O_3 has been observed in many other basin areas where inversion layers, high snow albedo, and VOC emissions from O&G lead to high O_3 (Field *et al.*, 2015). In the Colorado Front Range, VOC emissions from O&G have been demonstrated to enhance O_3 mixing ratios (L. C. Cheadle *et al.*, 2017), and similar studies have suggested that O&G extraction can increase O_3 roughly 10 ppbv above baseline levels (Rodriguez, Barna and Moore, 2009). Rapid proliferation of O&G extraction efforts across the United States, increased prevalence of summertime wildfires, and transport of O_3 and precursors from urban areas all provide mechanisms to enhance rural O_3 and necessitate the development of dedicated monitoring solutions to provide real time O_3 information to impacted communities. As the primary criteria for designating dedicated O_3 monitoring efforts in the United States is population, a lack of dedicated O_3 monitors in rural areas puts these communities at risk of exposure to high O_3 pollution (*ECFR :: 40 CFR part 58 -- ambient air quality surveillance*).

In addition to informing air quality advisories and regulation, dedicated O_3 monitoring can better quantify the acute and long-term impact of O_3 pollution. Epidemiological studies informed by ground-based, modeled, and satellite O_3 data have been used to assess the impact of O_3 on human health (West *et al.*, 2006; Liu *et al.*, 2018). To illustrate the importance of O_3 monitoring outside of urban areas we will focus on an overview of studies in O&G adjacent regions in the Western United States. The Wyoming Department of Health's implementation of dedicated O_3 monitors in the sparsely populated Sublette County was able to link O&G-influenced wintertime time O_3 events exceeding NAAQS standards to a 3% increase in respiratory hospital admissions over a period of roughly 10 years (Pride *et al.*, 2015). In many areas with enhanced O_3 adjacent to O&G extraction, such as Sublette County, real time monitoring represents the most robust method of assessing local O_3 pollution and public health outcomes. Research efforts attempting to

characterize O_3 at other O&G adjacent regions such as Uintah Basin, Utah, and the Colorado Front Range have demonstrated that modeling and satellite imaging of ground-level O_3 pollution exposure is extremely difficult to perform due the unique meteorology of these areas and diverse precursor sources (Neemann *et al.*, 2015; Sullivan *et al.*, 2016; Flocke *et al.*, 2020). In the Uintah Basin study, modeled variation from snow albedo alone varied O_3 estimates by up to ~20 ppbv. In the Colorado Front Range, strong O_3 precursor point sources such as power plants, NO_X and VOC emissions from traffic, and meteorological transport of emissions from wildfires, O&G, and agriculture make quantifying ground level O_3 via modeling difficult. Likewise, quantifying O_3 using satellites requires O_3 data from in-situ monitors due the complexity of vertical O_3 distributions.

1.3 Motivations for Low-Cost Ozone Sensing

Although the Western United States represents a small subset of O_3 impacted regions, there are several key trends that are demonstrated by this region: 1) anthropogenic activity in rural areas can lead to unsafe O_3 pollution despite isolation from metropolitan areas; 2) developing dedicated monitoring solutions in sparsely populated areas can enable air quality advisories and epidemiological studies (Pride *et al.*, 2015; red, 2019); and 3) modeling and satellite approaches cannot estimate O_3 with sufficient accuracy or spatial resolution to act as tools for air quality advisories. Oftentimes global epidemiological studies have relied on modeled and satellite based O_3 data, which attribute a large portion of the global O_3 mortality burden to underserved lower income regions (Malley *et al.*, 2017). As an example, rapid industrialization in China led to extreme O_3 pollution which remained unmonitored until the development of a national monitoring network in 2013 (Lu *et al.*, 2018), with existing monitoring efforts still falling short compared to

Europe, America, Japan, and Korea as recently as 2018 (Gaudel *et al.*, 2018). Epidemiological studies informed from model and satellite O_3 estimates may not be able to accurately capture O_3 exposure, especially in regions with complex precursor emissions and meteorology. The substantial cost (>USD 10,000), power (>500 W), and operational constraints (operator training, dedicated computer interface, and air-conditioned facility) of reference analyzers makes deploying these instruments in rural and underdeveloped communities both logistically impractical and fiscally prohibitive. In contrast to reference analyzers, low-cost O_3 sensors are characterized by their affordability (< USD 100), small form factor, and low power consumption (<10 W); these qualities allow for low-cost O_3 sensors to be readily integrated into autonomous sensor systems and provide real time pollution information in regions that lack the capital or infrastructure to implement reference monitoring (Piedrahita *et al.*, 2014). In addition to supporting disadvantaged communities, low-cost O_3 sensor integration into research efforts can enable researchers to study pollution at higher spatiotemporal scales, provide in-situ data to validate model and satellite estimates, and better understand the relationship between O_3 and health outcomes through higher spatial density and personal monitoring (Snyder *et al.*, 2013).

1.4 Calibration of Low-Cost Ozone Sensors

The cost, power, and form factor benefits of low-cost O_3 sensing come at the cost of measurement reliability; environmental variables impact low-cost sensor performance and the development of robust low-cost O_3 sensors represents a major hurdle in integrating these sensors into epidemiological and modeling efforts as in-situ monitors. Low-cost O_3 sensors typically rely on either a metal oxide or electrochemical sensing element, with metal oxide semiconductor (MOS) sensors being predominantly used in ambient air monitoring efforts due to their reduced

sensitivity to NO_2 interference (Idrees and Zheng, 2020). In MOS sensors the semiconductor is heated to increase binding affinity with O_3, which results in a sharing of electrons between the MOS surface and reacting gas. This binding generates an electrical signal that is proportional to the O_3 concentration (Zhang *et al.*, 2017). Although the operating principle of these sensors is relatively simple, variation in MOS manufacturing and MOS sensitivity to interferants including temperature, relative humidity, and interfering gases introduces challenges in developing reliable low cost O_3 sensors (Peterson *et al.*, 2017). Innate variation in the MOS can induce measurement variation across "identical" sensors, humidity can alter the conductance of the MOS, ambient temperature fluctuation can influence MOS heating, and non-target gas species can bind to the MOS surface, inhibiting binding with O_3 and leading to fouling of the sensor element over time.

To mitigate the impact of these variables on MOS sensor performance, both low-cost O_3 sensor manufacturers and end-users have worked to develop more reliable sensor hardware and analytical methods to correct for interferants. However, the efficacy and consistency of these efforts to correct low-cost O_3 sensors varies significantly across sensor manufacturers and studies and attempting to standardize low-cost O_3 sensing efforts remains a very difficult endeavor. The vast majority of manufacturers apply proprietary, or "black box," calibrations, meaning that replicating or comparing manufacturer efforts is extremely difficult (Karagulian *et al.*, 2019). Likewise, research efforts to develop calibrations vary by site location, sensor choice, and calibration methodology; however, there are a number of key trends that are consistent across low-cost sensor validation studies. Many low-cost O_3 sensors perform poorly out of the box and require researchers to develop their own calibrations to more accurately estimate O_3 pollution (Isiugo *et al.*, 2018; Masey *et al.*, 2018). These calibration approaches for low-cost O_3 sensors typically involve the use of univariate linear regression, multivariate linear regression, and machine learning

approaches to correct low-cost O_3 sensor mixing ratio estimates against measurements from collocated reference monitors (Karagulian *et al.*, 2019). Multivariate calibration approaches incorporate multiple explanatory variables to develop fixed correction factors for known interferants. Machine learning approaches correct for interferants by iteratively training a calibration algorithm using a portion of a dataset and then applying the calibration algorithm to uncorrected data. Multivariate and machine learning approaches typically demonstrate greater efficacy than univariate approaches due to their ability to correct for interfering factors and in the case of machine learning, sensor drift and sensor stability (Spinelle *et al.*, 2015; Zimmerman *et al.*, 2018; Sayahi *et al.*, 2020). Research has demonstrated that purely laboratory-based approaches to calibrating low-cost O_3 sensors are insufficient, as variation in ambient air conditions, interfering factors, and sensor fouling cannot be simulated in a laboratory environment (Karagulian *et al.*, 2019). Likewise, collocated field calibrations can be ineffective if the final deployment region's climatology is not consistent with the initial test site (Schmitz *et al.*, 2021). As low-cost sensors are intended to be deployed in a range of rural and urban settings with unpredictable ambient conditions, it is critical to understand how these sensors perform in real world environments.

In addition to the low material costs that qualify a "low-cost sensor," the personnel, computational, and logistical costs of implementing these sensors in research and monitoring efforts are important when considering the application of these sensors in air quality research and citizen scientist initiatives. Frequently re-calibrating low-cost sensors in a field setting or incorporating longer duration calibrations periods will increase the reliability of calibrated low-cost sensors. However, these more robust calibration approaches are impractical due to potential infrastructure constraints, personnel costs, and equipment needed to perform multiple field

calibrations, especially in remote field sites. Developing calibration methodologies that can be realistically implemented into large-scale sensor networks and citizen scientist initiatives represents an important objective in low-cost air quality monitoring. As such, the goal of our study is to determine the efficacy of several realistic calibration and engineering approaches for correcting the performance of low-cost O_3 sensors in the field.

In this work, we describe the development of a low-cost O_3 sensor system (MOOS) to evaluate the performance of the Aeroqual SM50 O_3 sensor and assess the efficacy of several practical calibration approaches to estimate O_3 mixing ratios. We deployed five MOOS in a collocated study and at field sites across Northern Colorado in order to evaluate MOOS mixing ratio estimates against collocated reference monitors and study spatial ozone trends. Overall, the MOOS demonstrated efficacy at estimating O_3 pollution and NAAQS compliance, with the sensor network also capturing relevant spatial O_3 trends.

CHAPTER 2: METHODS

In this work, we developed and evaluated an autonomous, low-cost sensor system called the Metal Oxide O_3 Sensor (MOOS) to measure ground-level O_3 concentrations. In the sections below, we discuss the MOOS construction (Section 2.1), deployment (Section 2.3), and calibration approaches (Section 2.4). The reference measurements used for calibration and evaluation are described in Section 2.2.

2.1 Metal Oxide Ozone Sensor (MOOS)

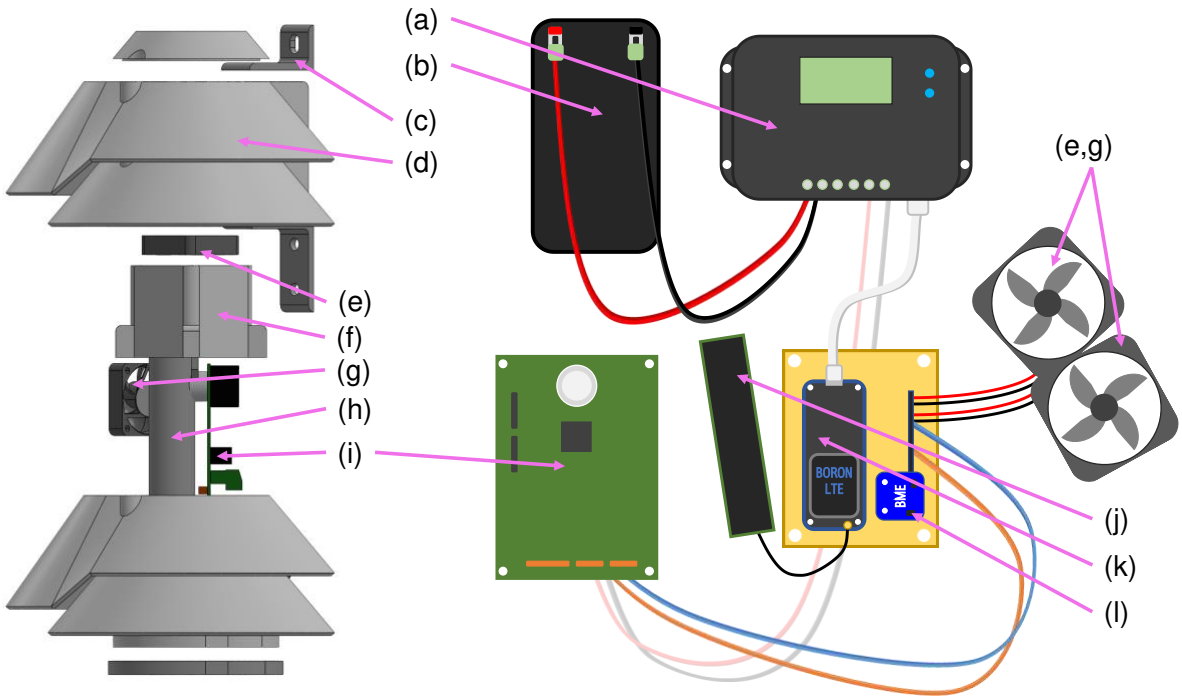

Figure 1: Multi-panel figure that shows the MOOS electrical schematic and 3D cad drawing of the sensor housing. Components include: (a) Renogy Wander Charge Controller, (b) 108 W-hr LiFePo Battery, (c) Sensor housing Mounting Bracket, (d) 3D Printed Sensor housing, (e) Ambient Air Sampling Fan, (f) Fan Mount, (g) Ventilation Fan, (h) PTFE Sampling Tube, (i) Aeroqual SM50, (j) LTE Antenna, (k) Particle Boron, and (l) BME 280.

A diagram of the MOOS component details is shown in Figure 1 and photos of the system are shown in Figure S1. The primary components within the MOOS, their features, and approximate weights and costs are listed in Table S1. The primary sensor used to measure O_3 is the SM50, a MOS sensor manufactured by Aeroqual (New Zealand). This sensor demonstrated the best performance across comparably priced low-cost O_3 sensors evaluated by the South Coast Air Quality Management District and has demonstrated efficacy when implemented in O_3 monitoring efforts (Polidori, Papapostolou and Zhang, 2016; SQAMD). The SM50 was integrated into a sensor housing (described later) but all other components were placed in a sealed polycarbonate box (WQ-57; Polycase, OH), which was mounted on an extruded aluminum frame (25-4115; 80/20, IN) and clamped to an aluminum rod connected to a tripod. For all deployments, the tripod was secured with guywires. A 30 W solar panel (Renogy; RNG International, CA) was used as the primary source of power. The polycarbonate box housed a charge controller (Wanderer; RNG International, CA) to manage voltage and power flows between the power producing (solar panel, battery) and consuming (O_3 sensor, environmental sensor, IoT, and battery) components, a microcontroller with cellular connectivity board (Boron; Particle Inc., CA) to perform data acquisition, real-time processing, and cloud transmission, a battery (BLF-1209WS; Bioenno Tech, LLC, CA) to store and provide power to all components, and an environmental sensor (BME 280; Bosch, Germany) to record temperature, relative humidity, and pressure inside the box. The solar panel and battery were sized using a solar irradiance simulator to comfortably meet the power demands for all months of the year in Colorado. The IoT was programmed to acquire data from the O_3 and environmental sensors at 1 Hz which was averaged across 1-minute intervals before being uploaded to the Particle Cloud (https://www.particle.io/). A MATLAB-based user interface was built on ThingSpeak (https://thingspeak.com/) to request and process data from the Particle

Cloud to monitor minute-averaged O_3 concentrations and other diagnostics produced by the MOOS. MOOS data was post processed from ThingSpeak using the median interpolation function in MATLAB.

To mitigate the impact of temperature and fouling on the SM50 O_3 mixing ratio estimates, an actively cooled sensor housing was developed to maintain Aeroqual SM50 temperatures within 1 °C of ambient and pull ambient air into the sampling tube to minimize sensor exposure to rain and dust events. The body of the sensor housing and all interior mounting hardware were fabricated using 3D printed polyethylene terephthalate glycol (PETG). PETG filament is resistant to damage from ultraviolet light and chemically inert to O_3 (Dutscher, n.d.; Leusink, n.d.), which minimized O_3 wall losses when air was sampled through the sensor housing and prevented degradation of the sensor housing over long field deployments. The sensor housing was mounted to the tripod using aluminum brackets and pipe clamps. The lower half of the housing integrated with the SM50 was actively cooled using a 40 mm computer fan and allowed for ambient air flow through a number of ventilation ports. These ventilation holes were sufficiently covered by the housing overhangs to prevent ingress by rain. A second 40 mm computer fan pulled ambient air through slots in the top of the housing and pushed this through a PTFE tube which interfaced with the SM50 sensor inlet, allowing for the SM50 to sample ambient air without being directly exposed to the elements. A diagram of these gas flows is shown in Figure S2.

2.2 Reference Monitors

O_3 mixing ratio measurements were collected using a combination of Colorado State University (CSU) and Colorado Department of Health and Environment (CDPHE) reference monitors. A Thermo Environmental Instruments Gas Analyzer (Model 49C, Franklin, MA;

Thermo hereafter) and two 2B 202 O_3 Analyzers (Model 202, 2B Technologies, Boulder, CO; 2B hereafter) were used as reference instruments. The limit of detection of the Thermo is 1 ppbv with an accuracy of 1.5% and the 2B monitor limit of detection is 3 ppbv, with an accuracy of 1.5 ppbv or 2%, whichever is higher. Additional real time O_3 data were obtained from CDPHE pollution monitors located at Weld Tower, Christman Field, and CSU through the CDPHE website. All O_3 measurements collected by these instruments were interpolated hourly using the median interpolation function in MATLAB. Laboratory calibrations for the MOOS and CSU monitors were performed using a 2B Ozone Calibration Source (Model 306, 2B Technologies, Boulder, CO, USA). Hourly averaged weather data was obtained from CDPHE monitors with weather data capability and the CSU Weather Station. Based on mixing ratio comparisons between both 2B monitors and the Thermo monitor over the collocated rooftop deployment, there was very high linearity (R^2 = 0.99) between these reference instruments, and no bias between the 2B monitors and -2.14 ppbv of bias between the Thermo and the 2B monitors (Figure S3). The average of both 2B monitors was used as the reference measurement for the collocated rooftop deployment.

2.3.1 Collocated Deployments

Table 1: Overview of MOOS deployment duration and collocated reference monitors.

MOOS	***Site Location***	***Site Index***	***Start Date***	***End Date***	***Collocated Monitor(s)***
A	Powerhouse/CSU	PWR	Jul 21	Aug 11	CSU (2×2B + Thermo)
B	Powerhouse/CSU	PWR	Jul 21	Aug 11	
C	Powerhouse/CSU	PWR	Jul 21	Aug 11	
D	Powerhouse/CSU	PWR	Jul 21	Aug 11	
E	Powerhouse/CSU	PWR	Jul 31	Aug 11	
A	Ft. Morgan	FTM	Aug 12	Sept 6	CSU-2B
B	Weld Tower	TWR	Aug 11	Sept 6	CDPHE
C	Powerhouse	PWR	Aug 11	Aug 23	CSU-Thermo
D	Ft. Collins West	FCW	Aug 11	Sept 6	CDPHE
E	Akron	AKR	Aug 12	Sept 6	CSU-2B

All five MOOS were collocated on the rooftop of the Powerhouse facility at CSU in Fort Collins, CO. MOOS-A through MOOS-D were deployed for a period of roughly 20 days and MOOS-E was deployed for a period of roughly 10 days during the months of July and August (Table 1). All MOOS were deployed within ~15 m of each other and all sampled at a height of ~1.75 m above roof level. Both CSU 2×2B + Thermo monitors were deployed in an air-conditioned office two floors below in order to accommodate the size and power requirements of these reference monitors. The 2B monitors operated continuously outside of a two-day downtime while the Thermo monitor had more frequent downtime due to use in other studies and instrumentation

issues. For the preliminary rooftop deployment, the reference O_3 measurements were represented by the average of the two 2B analyzers.

2.3.2 Field Deployments

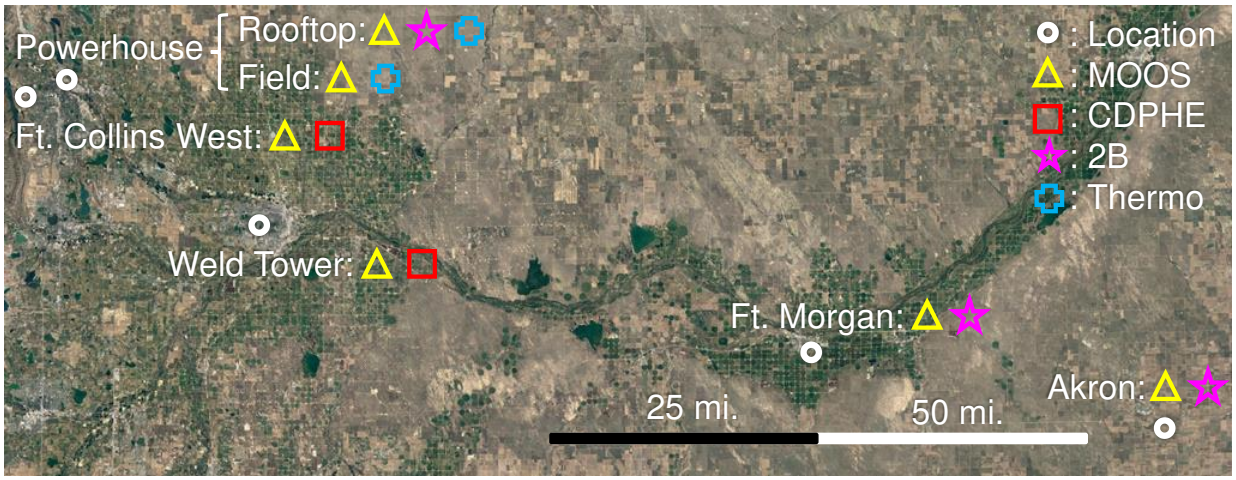

Figure 2: Rooftop and field deployment locations across Eastern Colorado. The Powerhouse was the location for the rooftop study and 1 field location, while all other sites were only locations for the field deployment (Google Earth).

The location of MOOS and reference monitors deployed across Eastern Colorado are shown in Figure 2. The duration for the field deployments is listed in Table 1. The 2B monitors were placed in ventilated enclosures to prevent overheating. MOOS-A was deployed on the Fort Morgan CSU Extension (FTM) office rooftop using a modified tripod at ~4 m above ground level. One of the 2B monitors was placed in an adjacent garage, with the inlet of the sensor being placed ~10 m from the MOOS-A at an elevation of ~3 m above ground level. MOOS-E was deployed at the USDA Central Great Plains Resources Management Research Center in Akron Colorado (AKR) at ~1.75 m above ground level. A 2B monitor was placed inside the building, with the inlet being directly collocated with the MOOS-E at ~1.75 m above ground level. MOOS B was deployed adjacent to the CDPHE Weld Tower monitoring station (TWR) at ~1.75 m above ground level and ~1 m from the inlet of the CDPHE monitor. MOOS-D was deployed ~20 m from the

CDPHE Fort Collins West monitoring station at ~1.75 m above ground level (FCW). MOOS C remained on the Powerhouse rooftop (PWR) along with the Thermo monitor.

All five field locations were exposed to similar ambient conditions, but industrial and motor vehicle activity varied across sites. FCW represented a spacious, dusty open field near a few small roads and the Front Range Foothills. TWR represented a similarly dusty area, however this site was located in an urban area roughly a mile away from a major thoroughfare. The FTM site was located in a dusty area near a railroad. This site had moderate vehicle traffic and was adjacent to several industrial and agricultural buildings. The AKR site was very sheltered, with minimal nearby dust or pollution sources outside of a country road. PWR was the most isolated from ground level across all of the sites, although this site was in an urban setting and near a major thoroughfare, likely experiencing more traffic than FTM and all other sites.

2.4 MOOS Calibration, Evaluation, and Performance Metrics

Four different calibration approaches were used to correct raw MOOS data, including the calibration developed by the manufacturer (MFR), an in-house laboratory calibration developed using a 2B O_3 source (LAB), a linear regression collocated calibration (LCOL), and a multivariate linear regression collocated calibration (MCOL) that accounted for ambient temperature and relative humidity. These represent the predominant methods used to correct low-cost sensors in literature outside of machine learning calibrations (Maag, Zhou and Thiele, 2018).

1. MFR: The MFR calibration represents the black-box laboratory calibration developed by Aeroqual. The SM50 modulates air flow and temperature to create a "zero O_3" event where low airflow and temperature effectively eliminate binding to the MOS surface (Li, 2009) in order to correct for the bulk resistance of the MOS.

2. LAB: A 2B ozone calibrator was interfaced with the SM50 sensor in order to evaluate SM50 performance and develop calibrations based off of known O_3 mixing ratios. The LAB calibration represents a linear regression fit between MOOS and known 2B calibrator output, with O_3 mixing ratios ramping in 20 ppbv increments from 0 to 100 ppbv. This calibration takes the form of $y = B_1x + B_0$ where x represents calibrator output and y represents MOOS estimates, the regress function in MATLAB was used to compute B_1 and B_0 for individual MOOS.
3. The LCOL calibration was developed using a linear regression fit between MOOS O_3 estimates and the reference monitor O_3 measurements in ambient air and takes the form of $y = B_1x + B_0$ where x represents the reference monitor measurements and y represents MOOS estimates. The regress function in MATLAB was used to compute B_1 and B_0. LCOL-Pre represents a calibration developed by taking the coefficients B_1 and B_0 developed during the rooftop deployment and using them to correct field data.
4. The MCOL calibration was developed using a first-degree multivariable polynomial fit between MOOS and reference monitor data including variables for ambient temperature and relative humidity and takes the form of $y = B_0 + B_1x_1 + B_2x_2 + B_3x_3$ where x_1, x_2, and x_3 represent reference measurements, ambient temperature, and relative humidity respectively and y represents MOOS estimates. Coefficients B_0, B_1, B_2, and B_3 were computed using the regress function in MATLAB.

The LCOL and MCOL calibrations were developed using both all time-matched data and an *n*-fold cross-validation approach. Calibrations developed using all data, henceforth referred to as LCOL-*Full* or MCOL-*Full*, were used to develop the best available calibration between MOOS

O_3 estimates and reference measurements. Calibrations developed using a cross-validation approach, henceforth LCOL-*k* or MCOL-*k*, were implemented to represent realistic calibrations strategies, as indefinitely collocating a reference monitor alongside a low-cost sensor is impractical and realistic field calibration efforts will occur over a fraction of the full deployment duration. For LCOL-*k* and MCOL-*k*, *k* represents the duration of the training periods (i.e. LCOL developed using a 48-hour training period is referred to as LCOL-*48*). Individual training periods spanned midnight to midnight local time, with the partial days at the beginning and end of the dataset being parsed out. The duration of training periods varied from 24 hours up to 240 hours, with period duration increasing in increments of 24 hours. Training periods greater than 24 hours in duration were overlapped in order to evaluate more calibrations. Missing data were included when establishing training periods and later removed if missing data comprised more than 25% of the training data; calibration performance did not appear to be impacted when less than 25% of the training data was missing. This was done because removing training periods with any missing data would greatly reduce the number of longer duration training periods. The performance of the LCOL-*k* and MCOL-*k* calibrations developed from training periods was assessed by testing against the rest of the dataset. As O_3 measurements between MOOS and reference monitors are autocorrelated, continuous training periods will introduce bias compared to randomly sampled training periods. This bias allows us to develop a realistic understanding of MOOS performance when a collocated LCOL or MCOL calibration is carried over different training periods. Due to variation in meteorology, sensor performance, and other unknown factors, the performance of calibrations developed using individual periods varied and was graded based on the Root Mean Squared Error (RMSE). In the field deployment, we will be presenting the bottom 10th percentile performing calibration corrected using LOCL-*k* and MCOL-*k* calibrations approaches.

Several statistical metrics were selected to evaluate calibrated MOOS O_3 mixing ratio estimates comparisons to O_3 measurements recorded from reference monitors and CDPHE monitoring stations. These statistical metrics included: 1) Pearson's correlation coefficient (*r*) between individual MOOS and reference; 2) RMSE and Mean Bias Error (MBE) between MOOS and reference; 3) slope and intercept between MOOS and reference. Pearson's *r* was calculated using the corrcoef function in MATLAB (1), slope and intercept were calculated using the polyfit function in MATLAB, and RMSE (2) and MBE (3) were calculated using the equations below, where x_i represents reference O_3 measurements and y_i represents MOOS O_3 estimates.

1) $$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}}$$

2) $$RMSE = \sqrt{\frac{\sum(x_i-y_i)^2}{n}}$$

3) $$MBE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$$

CHAPTER 3: RESULTS

3.1 Collocated Rooftop Study to Evaluate Calibration Methods

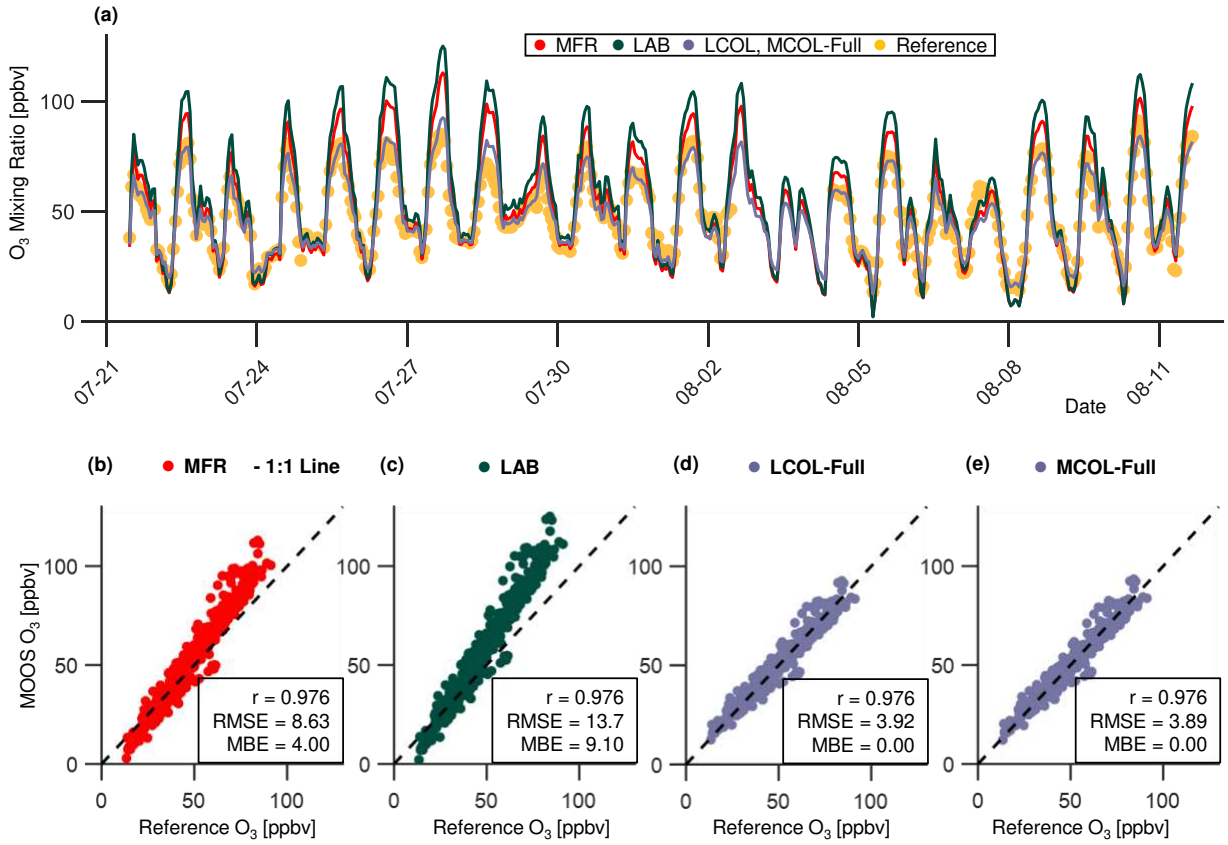

Figure 3: (a) Hourly averaged July and August 2021 MOOS-A O_3 mixing ratio estimates using different calibration approaches compared to collocated reference measurements. Scatter plots for MOOS-A O3 mixing ratios from (b) MFR, (c) LAB, (d) LCOL-Full, and (e) MCOL-Full calibrations compared to reference measurements. Scatter plots for MOOS-A O_3 mixing ratios from (b) MFR, (c) LAB, (d) LCOL-Full, and (e) MCOL-Full calibrations compared to reference measurements. MFR refers to the estimates based on the manufacturer's calibration, LAB refers to the estimates based on laboratory calibrations performed with a Model 306 (2B Technologies), LCOL-Full refers to the estimates based on a linear calibration developed using the entire collocated dataset, and MCOL-Full refers to the estimates based on a multivariable calibration developed using the entire collocated dataset. Time series and scatter plots are shown in Figures S3 to S6 in the SI for the other MOOS.

Results for the O_3 mixing ratio from MOOS-A are compared against reference measurements in Figure 3 for different calibration approaches, with results for all other MOOS being shown in Figure S5 through Figure S8. In Figure 3(a), LCOL-*Full* and MCOL-*Full* results

are presented using the same-colored line since both estimates were nearly identical and therefore overlapped. As noted earlier, the reference measurements are based on the average of the CSU-2Bs deployed during the collocated study. The reference measurements included a small amount of data when only one of the two Model 202s was running (<5% of the data) and excluded a small portion of data when neither of the CSU-2Bs were running (Table 2). All four calibration approaches - MFR, LAB, LCOL-*Full,* and MCOL-*Full* - resulted in strong correlations between MOOS-A and the reference measurements (Pearson's $r = 0.976$; r, henceforth). However, both the MFR and LAB calibrations overestimated O_3 at the higher O_3 mixing ratios observed during the daytime and slightly underestimated the O_3 when the mixing ratios were below 20 ppbv. The MFR and LAB calibrations overestimated the peak O_3 by 22% and 28%, respectively. By design, both LCOL-*Full* and MCOL-*Full* calibrations appeared to correct the bias in high and low O_3 estimates and significantly reduced the RMSE to 3.92 and 3.89 ppbv, respectively, with no change in r. RMSE and MBE for LCOL- and MCOL-full will also provide a benchmark for which to compare LCOL- and MCOL-*k* presented in Figure 4. Although there were some MOOS-to-MOOS differences, the MFR and LAB calibrations consistently overestimated O_3 mixing ratios (Figure S5 through Figure S8) while LCOL-*Full* and MCOL-*Full* performed well across the other sensors. These results suggest that the laboratory-based approach used to calibrate low-cost O_3 sensors in this study induced more measurement error and bias compared to linear collocated calibration approaches. While the SM50 does not directly correct for temperature and relative humidity, the similar performance of LCOL-*Full* and MCOL-*Full* approaches suggest that the SM50 sensors were minimally impacted by these interfering factors. The large temperature and relative humidity error terms (135 and 175 ppbv respectively) and small reference error term (15 ppbv) for the multivariate coefficients used to develop MCOL-*Full* further suggest that MOOS are minimally

impacted by temperature and relative humidity factors (Table S2). Trends in the SM50 performance observed during the rooftop study, which include the manufacturer calibration intra-sensor variance, high linearity with reference measurements, and minimal sensitivity to temperature and relative humidity, were very consistent with existing literature on the SM50 and other similar Aeroqual O_3 sensors (Jiao *et al.*, 2016; Polidori, Papapostolou and Zhang, 2016; Mueller, Meyer and Hueglin, 2017).

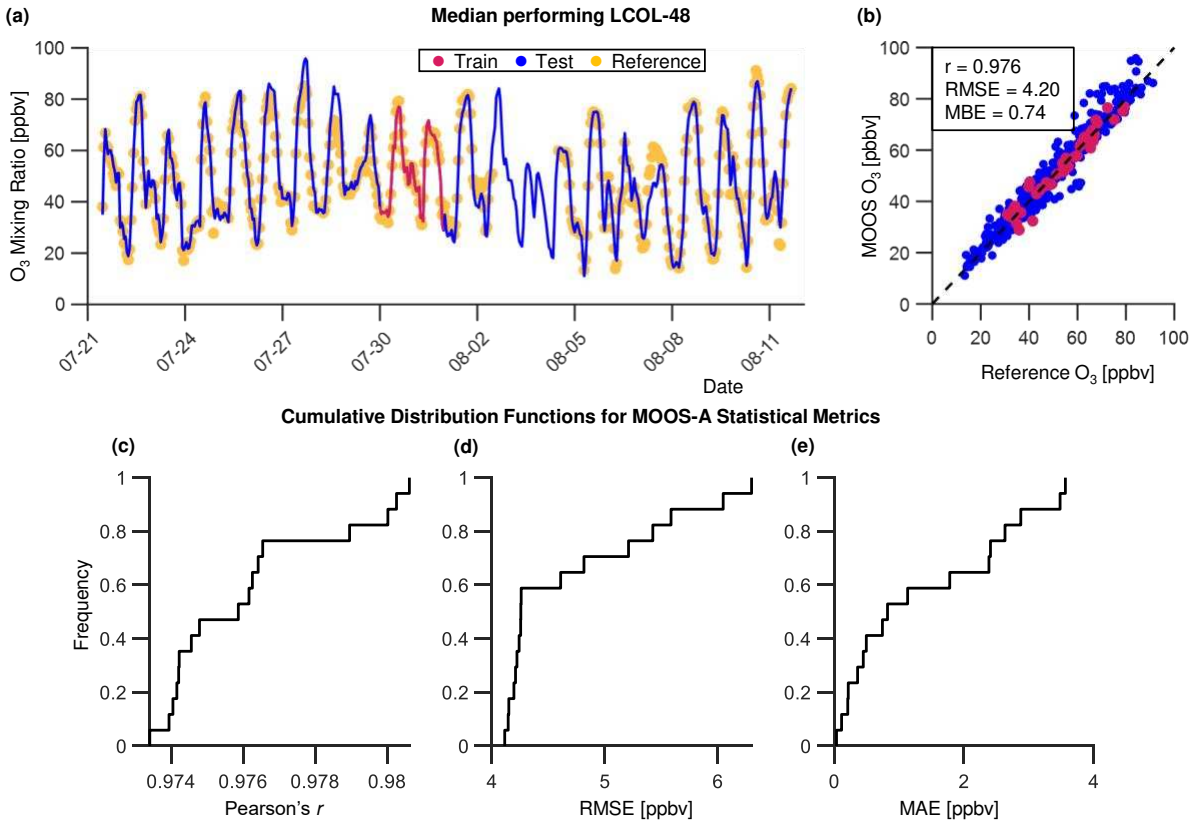

Figure 4: July and August 2021 MOOS-A O_3 mixing ratio estimates corrected using an n-fold cross validation LCOL approach compared to collocated reference measurements as a time series (a) and scatter plot (b). 'Train' refers to the 48-hour interval used as the training period and 'Test' refers to the remainder of the data used to evaluate the calibration. The LCOL-48 approach shown in (a,b) represents the median performing calibration developed from one of the 17 training periods. Statistical metrics r (a), RMSE (b), and MBE (c) from the cross validation of 17 independent LCOL-48 calibrations are shown as cumulative distribution functions. Time series, scatter plots, and CDFs for LCOL-48 are shown in Figures S8 to S11 in the SI for all other MOOS.

Results for the O_3 mixing ratio from MOOS-A corrected using the LCOL calibration approach based on 48 hours of training are compared against reference measurements in Figure 4, with results for all other MOOS being shown in Figure S10 through Figure Figure S13. To remind the reader, the rooftop colocation study included nineteen days of continuous MOOS measurements which resulted in seventeen continuous 48-hour training periods after missing data was dropped. LCOL-*48* calibrations developed from these seventeen 48-hour training periods were then individually tested against the remaining data. For example, the O_3 results in Figure 4(a) show how the MOOS-A was calibrated or 'trained' on reference measurements from August 30-31 and then cross-validated or 'tested' against reference measurements from all other days. LCOL-*48* calibrations developed using these seventeen training periods were characterized by unique *r*, RMSE, and MBE, which describe how well the corrected O_3 estimates compared to reference measurements. These metrics are for "test" data and exclude the training data used to develop LCOL-*48*. Variations in statistical metrics *r*, RMSE, and MBE across different training periods are shown using the cumulative distribution functions in order to compare the performance of all possible LCOL-*48*s (Figure 4c-e), with the individual LCOL-*48* approach demonstrated in Figure 4(a) representing the median performing calibration. *r*, RMSE, and MBE varied between 0.973 and 0.981, 4.11 and 6.30, and -2.87 and 3.56, respectively, across all LCOL-*48*. RMSE and MBE appeared to vary more than *r*, with RMSE and MBE increasing by 0.5 and 1.2 ppbv respectively over the first 60% of training periods and then increasing by 2.1 and 2.6 ppbv over the remaining 40% of periods. Correlation and error between LCOL-*48*-corrected MOOS O_3 estimates and reference measurements did not appear to depend on the relative training period, or the ambient conditions of the training period used to develop LCOL-*48*.

While the MCOL-*Full* approach slightly outperformed LCOL-*Full*, MCOL calibrations developed using an *n*-period validation approach appeared to be slightly inferior to the LCOL calibrations. The median performing MCOL-*48* had a slightly lower correlation and higher error compared to the median LCOL-*48*. Across all MCOL-*48* *r*, RMSE, and MBE varied between 0.907 and 0.977, 4.14 and 9.30, and -2.91 and 4.41 respectively (Figure S14 through Figure S18). MCOL-*48* performed similar to LCOL-*48* across 60% percent of training periods, but performed substantially worse than LCOL-*48* across the remaining 40% of periods. Notably, the least performing MCOL-*48* reduced *r* by 0.066 and increased RMSE by 3 ppbv compared to the least performing LCOL-*48*.

The majority of low-cost O_3 sensors described in the literature incorporate a multivariable regression or machine learning approach to develop calibrations, with these approaches typically outperforming linear calibrations (Karagulian *et al.*, 2019). Multivariable and regression calibrations typically developed over multi-week time periods, which can be impractical when carrying out studies in remote areas without dedicated infrastructure to support reference instruments (Spinelle *et al.*, 2015; Jiao *et al.*, 2016; L. Cheadle *et al.*, 2017; Ferrer-Cid *et al.*, 2019; Miskell *et al.*, 2019). We utilized substantially shorter training intervals to develop field calibrations, which may have been insufficient to train MCOL compared to LCOL. Furthermore, these results suggest that incorporating short term calibrations for interfering factors may deteriorate sensor performance.

3.2 Field Deployment and Comparison to Reference Measurements

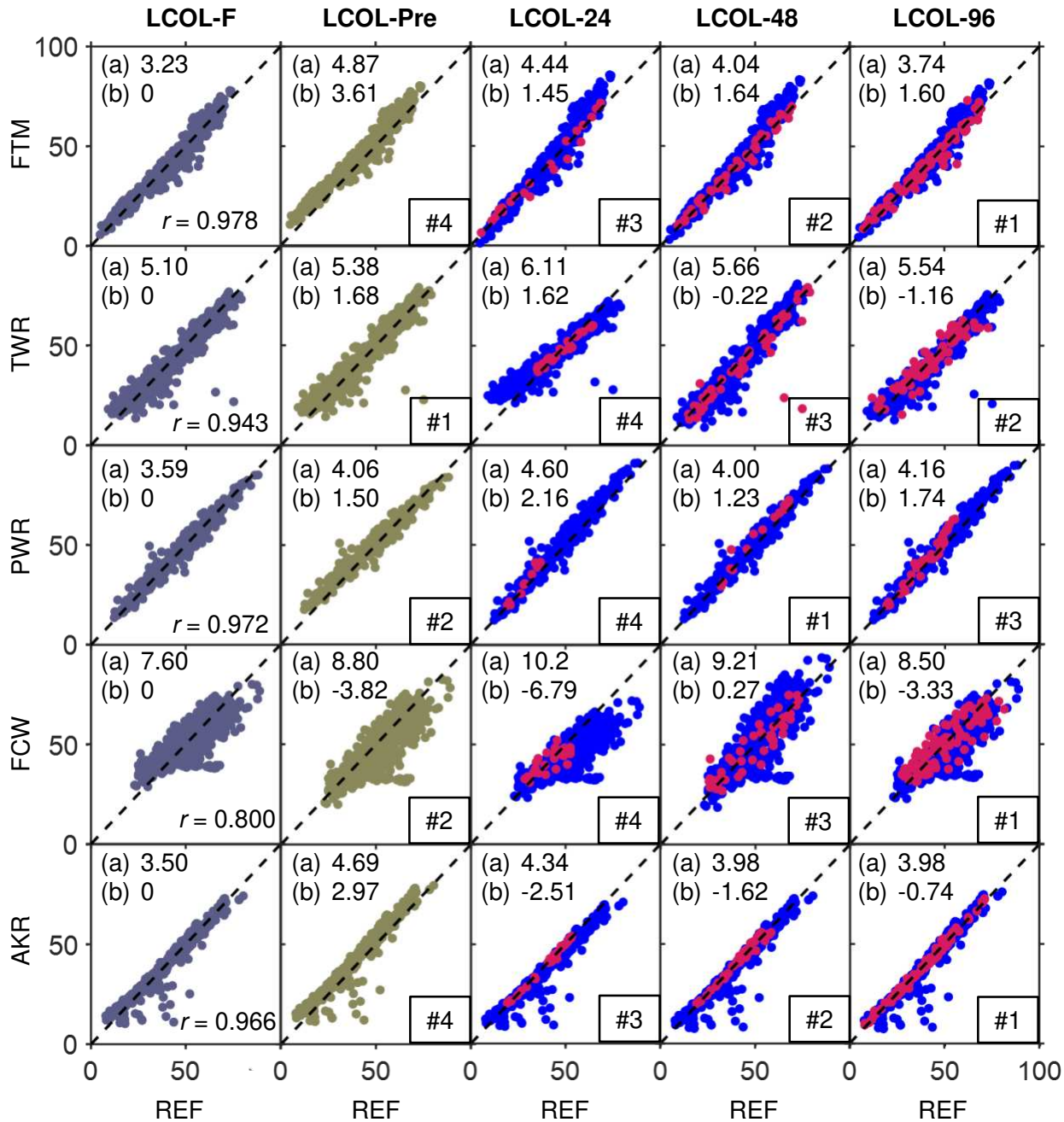

Figure 5: MOOS O_3 mixing ratio estimates corrected using LCOL-F, LCOL-Pre, LCOL-24, LCOL-48, and LCOL-96 approaches compared to reference measurements during the field deployment. LCOL-Full refers to estimates based on a calibration developed using the entire co-located dataset, LCOL-Pre refers to estimates based on a calibration developed from the rooftop deployment, and LCOL-24, LCOL-48, and LCOL-96 refer to estimates based on a n- period cross validation approach that use 24, 48, and 96 hour training periods. LCOL-24/48/96 represent the bottom 10th percentile performing calibration based on RMSE. RMSE (a) and MBE (b) are shown

for each site/calibration combination, with all results being presented in units of ppbv. The number in the bottom right of each plot refers to the rank order for each calibration approach based on RMSE where #1 refers to the best-performing calibration and #4 refers to the worst-performing calibration. r was constant within ±0.005 across all sites and calibration approaches and is shown as a single value in the leftmost plots.

Results of MOOS O_3 mixing ratio estimates from LCOL-*Full*, LCOL-*Pre*, and LCOL-*24/48/96* approaches compared to reference measurements are shown in Figure 5 for all field locations. After 96-hours, increased training duration did not appear to substantially improve correlation, measurement error, or bias error, with more in-depth discussion on calibration length being presented in Figure 6. The duration for the MOOS and reference monitor deployments across locations are shown in Figure 1 and Table 1; after removing data where only one instrument was running, 23, 20, 7, 24, and 23 training periods were developed for the FTM, TWR, PWR, FCW, and AKR sites, respectively. Results for LCOL-*24/48/96* shown in Figure 5 are for the 90th percentile calibration for RMSE for LCOL-*k*, with a visual representation of the RMSE cutoff shown in Figure S19. We chose to present results for the 90th percentile LCOL-*k* calibration in Figure 5 as a lower bound on the MOOS performance. In other words, 10% of the calibrations would be inferior to and 90% of the calibrations would be superior to that shown in Figure 5. LCOL-*Full* is shown as a benchmark for the best possible linear calibration that can be developed at each site.

MOOS deployed at the FTM, TWR, PWR, and AKR sites demonstrated strong correlation with reference measurements across all calibration approaches (r>0.94). At these sites, LCOL-*Pre* and LCOL-*24/48/96* performed nearly as well as LCOL-*Full*, increasing RMSE and MBE by at most 1.64 and 3.61 ppbv. The LCOL-*Pre* and LCOL-*24/48/96* calibration approaches all performed similarly, with RMSE and MBE across these approaches varying by at most 0.73 ppbv and 2.16 ppbv, respectively. The LCOL-*Pre* and LCOL-*24/48/96* calibrations outperformed the

MFR calibration by 0.61 to 4.65 ppbv for RMSE and by up to 4.45 ppbv for MBE. The MOOS deployed at FCW demonstrated poorer overall correlation (r>0.800) and had high RMSE and MBE

The LCOL-*Pre* and LCOL-*24/48/96* calibrations outperformed the MFR calibration by 0.61 to 4.65 ppbv for RMSE and by up to 4.45 ppbv for MBE. The MOOS deployed at FCW demonstrated poorer overall correlation (r>0.800) and had high RMSE and MBE (RMSE<10.2 ppbv, MBE<6.8 ppbv) compared to the MOOS at the other sites. Overall, LCOL-*96* performed marginally better than the shorter duration LCOL-*24/48* calibration approaches and LCOL-*Pre* based on the relative ranking of the calibration approaches. LCOL-*Pre* and LCOL-*48* performed similarly, and LCOL-*24* was the least performing calibration overall. Outside of FCW, all of these calibration approaches improved MOOS O_3 mixing ratio estimates compared to those based on MFR and LAB (not shown).

LCOL-*Pre* was the 1st and 2nd best performing calibration based on RMSE at the TWR and FCW sites respectively, where MOOS were not directly collocated with the reference monitors. LCOL-*Pre* was also the 2nd best performing calibration at PWR, where the MOOS and reference configuration was not altered from the rooftop deployment. LCOL-*Pre* was outperformed by all LCOL-*24/48/96* at FTM and AKR, where MOOS were directly collocated with reference monitors. In general, it appeared that the LCOL-*k* approach developed a more accurate calibration than LCOL-*Pre* when the MOOS and reference monitors were collocated at the same height, while the LCOL-*Pre* approach developed a more accurate calibration when MOOS were at a different elevation and spaced apart from reference monitors. Differences in the MOOS proximity and elevation compared to reference monitors may explain the variable performance of different calibration approaches across the five sites, with larger spatial differences between MOOS and reference monitors potentially introducing variation that could not be captured by our calibration

approach. If we had access to more reference monitors and were able to directly collocate a monitor with all five MOOS, it is possible that the relative efficacy of calibrations developed at TWR and FCW would be similar to those developed at FTM and AKR.

Results shown for LCOL-*24/48/96* are from the 90th percentile calibration, meaning calibrations developed using a median LCOL-*24/48/96* approach will perform better than the results shown in Figure 5. In general, median LCOL-*24/48/96* calibrations reduced RMSE by ~1 ppbv compared to the 90th percentile LCOL-*24/48/96* across all sites. This reduction in error allowed for the median LCOL-24 to outperform LCOL-*Pre* at all sites except TWR (where median LCOL-48 still outperformed LCOL-*Pre*), suggesting that calibrations developed from two days of in-situ calibration will outperform a longer preliminary collocated calibration (~3 weeks) developed in a different region. In real world calibration efforts, the training period selected to develop a calibration will likely depend on the logistics of the study, therefore it is prudent to present calibrations that are representative of all possible outcomes. MOOS exhibited higher correlation and lower measurement error compared to ground-level reference monitors than modeling approaches, which typically reported r <0.9 and RMSE <22 ppbv (Di *et al.*, 2017; Requia *et al.*, 2020; Wang *et al.*, 2020).

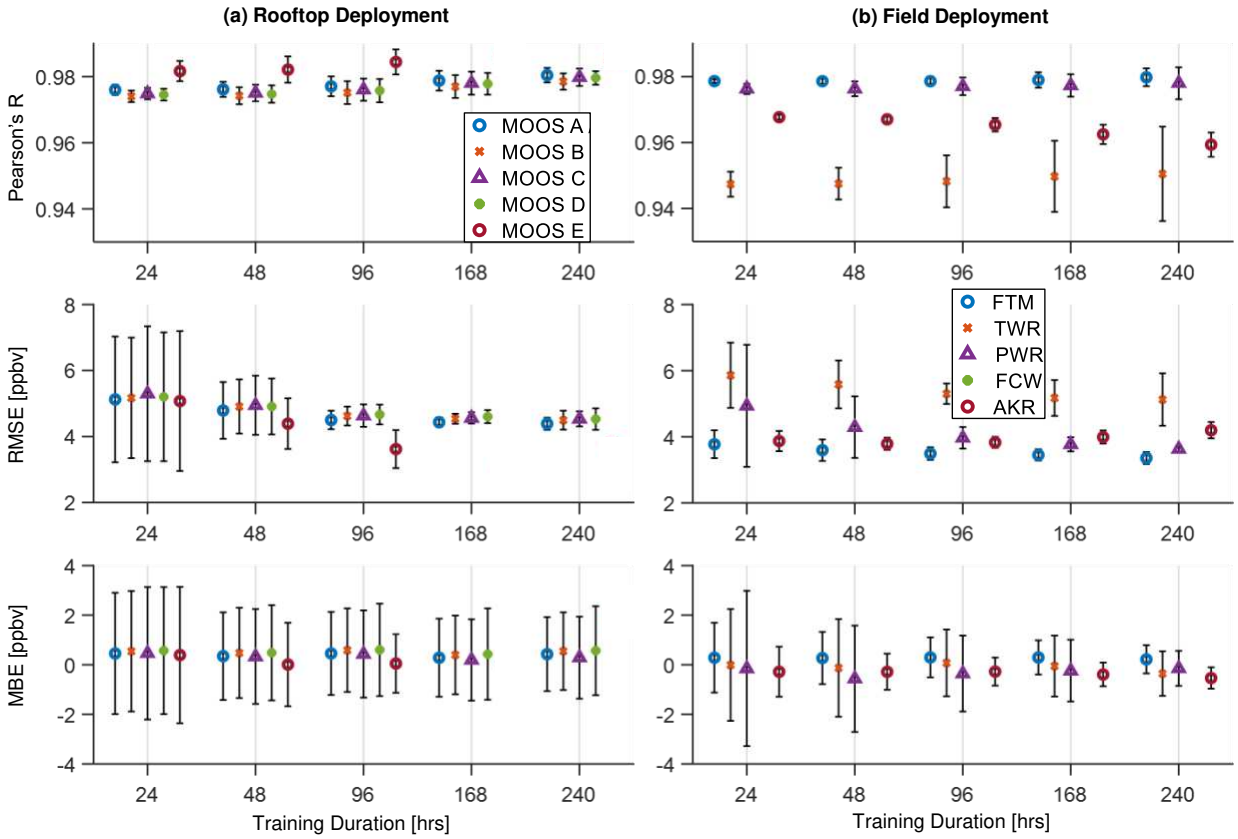

Figure 6: Variation in statistical metrics comparing MOOS O_3 mixing ratio estimates reference measurements across all 5 MOOS shown for the Rooftop Deployment (a) and the Field Deployment (b). Error bars represent 1 standard deviation of the statistical metrics.

Statistical metrics for LCOL-*k* corrected MOOS O_3 estimates varied depending on the duration of the training periods. To remind the reader, these metrics capture the performance of LCOL-*k* applied to "test" data and exclude the data used to "train" the LCOL-*k* calibration. *r*, RMSE, and MBE for all LCOL-*k* calibrations developed over 24- through 240-hour training periods are shown in Figure 6. Results for the 168- and 240-hour training periods are not shown for MOOS-E in the rooftop deployment due to the shorter deployment time (Table 1). Results are not shown for FCW in the field deployment, as this sensor performed substantially worse than all other MOOS and including results for FCW would have made it harder to draw conclusions for the four other MOOS (results for FCW are included in Figure S20). We will first look at trends in

the rooftop deployment. *r* increased slightly with training duration across all MOOS, while RMSE and MBE were less impacted by training period duration. Variability in RMSE decreased from 1.8 ppbv for 24-hour training periods to 0.82 ppbv at 48-hour training periods and was just 0.3 ppbv for 96-hour training periods, while mean RMSE decreased by 0.6 ppbv up to 240-hour training periods. Variation in MBE improved from the 24- to 48-hour training periods by 0.6 ppbv and remained constant after 48-hour training periods, with mean MBE remaining at about 0 regardless of training duration. Overall, intra-sensor performance across training period durations appears to be very consistent, with all 5 MOOS exhibiting minimal intra-sensor variance across all statistical metrics.

In the field deployment, variation in *r* across MOOS was more pronounced and increased with training duration, although this was likely due the fewer training intervals at longer durations. We saw similar trends in both RMSE and MBE compared to the rooftop deployment; variation in RMSE at FRM and AKR decreased from 0.42 ppbv to 0.19 ppbv and mean RMSE decreased by 0.3 ppbv up to 96-hour training periods, and then appeared to either stay the same or slightly increase. Variation in RMSE at TWR and PWR was more impacted by training period duration, decreasing from 0.99 ppbv to 0.31 ppbv and 1.85 ppbv to 0.33 ppbv, respectively up to 96-hour training periods. After 96-hours, variation in TWR then increased up to 240-hour training periods while variation in PWR remained constant. FCW experienced the highest variation in RMSE across all sites of 4.3 ppbv as well as the highest mean RMSE. Variation in MBE decreased up until 240-hour training periods and mean MBE remained constant, although the magnitude of mean MBE varied across sensors. Overall, MOOS correlation (*r*) with reference monitors appeared to be minimally impacted by training period duration. These results suggest a training

period duration between 48 and 96 hours reduces variation in RMSE compared to a 24-hour training period, with >96-hour periods potentially introducing more variation in RMSE.

3.3 O_3 Exposure and East-West Gradients

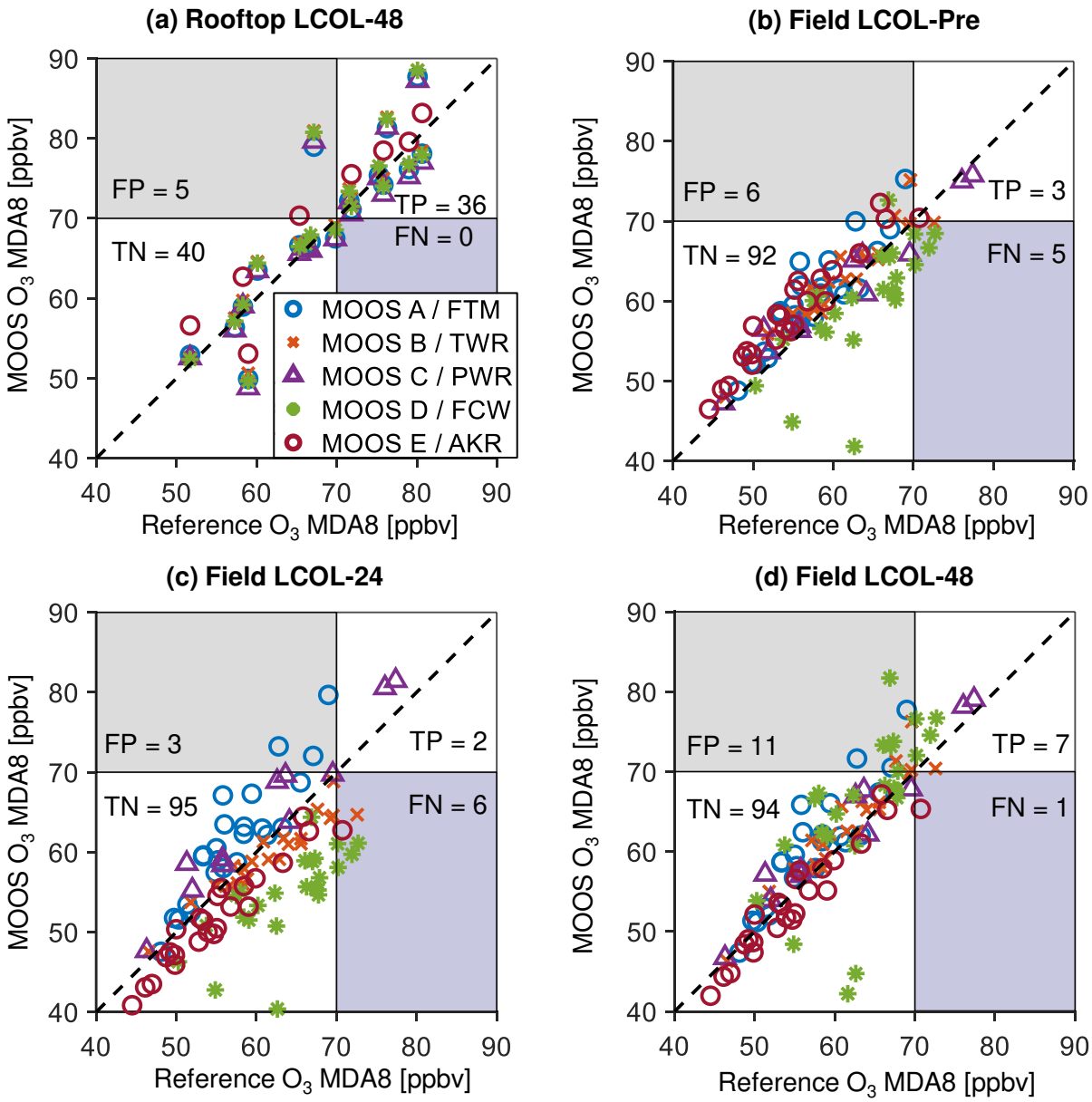

Figure 7: MOOS O_3 MDA8 estimates compared to reference MDA8 measurements across the rooftop and field deployments. Rooftop MDA8 estimates were corrected using the median LCOL-48 approach (a) and field MDA8 estimates were corrected using LCOL-Pre (b), the 90% LCOL-24 approach (c), and the 90% LCOL-48 approach (d). True positives (TP) represent >70 ppbv O_3 events (the NAAQS standard for O_3 exposure) measured by the collocated reference monitor where MOOS estimates also exceed 70 ppbv. True negatives (TN) represent <70 ppbv O_3 events where MOOS estimates also subseeded 70 ppbv. False positives (PN) represent <70 ppbv O_3 events where MOOS estimates exceeded 70 ppbv and false negatives represent >70 ppbv O_3 events where MOOS estimates subceeded 70 ppbv.

In addition to evaluating the MOOS performance based on statistical metrics, we also assessed the ability of MOOS to capture relevant regulatory metrics, specifically the maximum daily 8-hour average O_3 (MDA8). To remind the reader, MDA8 is a metric used to determine compliance with regulation and inform air quality advisories, with MDA8s > 70 ppbv exceeding NAAQS standards and MDA8s < 70 ppbv being compliant with NAAQS standards. MOOS MDA8 estimates corrected using the median LCOL-*48* approach shown in Figure 4(a,b) and Figure S5 through Figure S8(a,b), and the LCOL-*Pre*, LCOL-*24*, and LCOL-*48* approaches shown in Figure 5 are compared to reference MDA8 values in Figure 7. It is important to note that MDA8 estimates corrected using an LCOL-*k* approach includes both testing and training data, with training data making up roughly 5% and 10% of the full dataset for LCOL-*24* and LCOL-*48,* respectively. MDA8 performance was graded based on the prevalence of Type I and Type II errors between the MOOS and reference measurements. Type I error (false positives) represents a scenario where the MOOS MDA8 estimates exceed > 70 ppbv while the true MDA8 is <70 ppbv. Type II error (false negatives) represents a scenario where the MOOS MDA8 estimates are <70 ppbv while the true MDA8 exceeds 70 ppbv. Overall, with a Type II error, the MOOS could lead to unsafe O_3 exposure in communities relying on these monitors for air quality advisories.

MOOS performed well in estimating MDA8 during the rooftop deployment, capturing 100% of >70 ppbv O_3 days measured by the reference monitors and both compliant (<70 ppbv) and non-compliant (>70ppbv) days with an overall error of 6.2%. In the field deployment, MOOS had much higher instances of the Type II error. LCOL-*48* captured 7/8 >70 ppbv O_3 days measured by the reference monitors and classified compliant and non-compliant days with an overall error of 10.7%. LCOL-*Pre* captured only 3/8 of >70 ppbv O_3 days measured by the reference monitor and classified compliant and non-compliant days an overall error of 11.6%, while LCOL-*24* only

captured 2/8 >70 ppbv days measured by the reference monitor and classified compliant and non-compliant days an overall error of 9.3%. Compared to the rooftop study, it is difficult to determine whether the MOOS' inability to capture high O_3 events in the field was due to poor sensor and calibration performance, or a relatively small sample size of high O_3 events which could be captured by the MOOS. During both the rooftop and field deployment, reference monitors recorded eight >70 ppbv events, however in the field only one MOOS could capture each event while on the rooftop five (four when MOOS E was not deployed) MOOS could capture each event. Half of the >70 ppbv days that occurred during the field deployment were recorded at FCW, where the MOOS O_3 demonstrated less measurement reliability than other sites. During the field deployment, only the MOOS deployed at PWR was able to capture 2/2 >70 ppbv days across all calibration approaches. MOOS at both TWR and AKR demonstrated high linearity and low RMSE compared to reference measurements, but were only able to capture 1/2 >70 ppbv events when corrected with LCOL-*Pre*, 1/2 >70 ppbv events when corrected with LCOL-*48*, and 0/2 when corrected with LCOL-*24*.

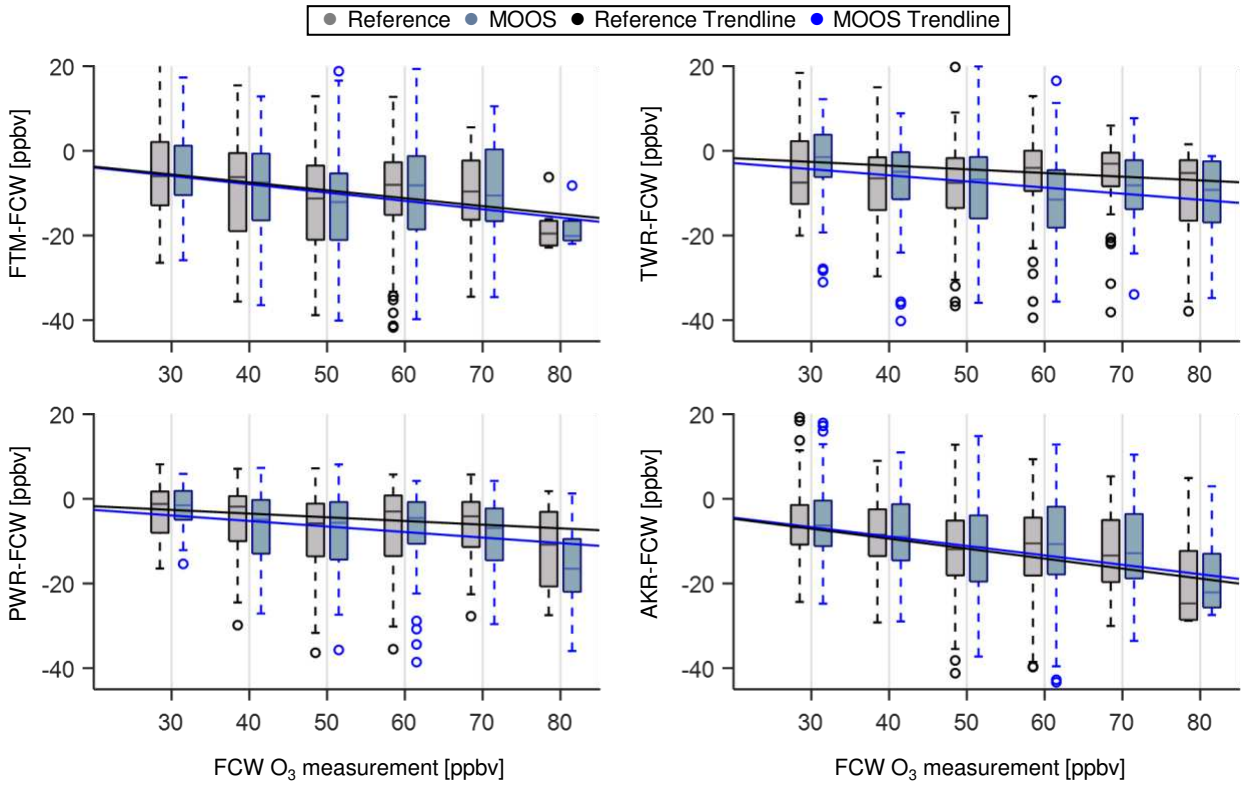

Figure 8: Difference between reference monitors O_3 measurements and MOOS O_3 estimates compared to the CDPHE reference monitor located in FCW. The difference between hourly O_3 events for corresponding reference monitors was taken over the entire field deployment. These deltas were binned based on the O_3 concentration measured at the FCW site, with bin medians at 30 - 80 ppbv. Boxplot boxes and whiskers represent 1 and 2 standard deviations respectively. Trendlines were fitted to the mean of deltas to approximate this difference across all O_3 events.

We also examined the ability of the MOOS to study spatial O_3 trends in Eastern Colorado. The field sites selected in this study represent a roughly 160 km span ranging from the metropolitan Front Range to rural Eastern Colorado, with collocated MOOS and reference monitors deployed at randomly spaced intervals across this gradient (i.e., 1 MOOS per 40 km). In Figure 8, we plot the absolute O_3 between a MOOS site and a datum, where the datum was the westernmost location (i.e., FCW). Results are shown using box plots where the time matched absolute O_3 mixing ratios were aggregated in decadal bins, with bin medians from 30 to 80 ppbv. O_3 events in FCW are from the CDPHE reference monitor, and O_3 events at all other sites were defined by either MOOS or

the collocated reference monitor. The difference between reference O_3 measurements at FCW and the other four reference monitors demonstrated a strong spatial trend. When FCW experienced an 80 ppbv O_3 event, AKR, the easternmost site, experienced 25 ppbv lower O_3, FTM experienced 18 ppbv lower O_3, TWR experienced 11 ppbv lower O_3, and PWR experienced a 12 ppbv lower event, based on the median of the 80 ppbv reference bin. When FCW experienced a 70 ppbv O_3 event, AKR experienced 12 ppbv lower O_3 while TWR experienced only 4.8 ppbv lower O_3. Overall, the differences in absolute O_3 between sites and FCW is statistically significant ($p<0.05$) and suggests there were spatial O_3 differences as we moved East. The O_3 gradient observed by our reference network, albeit for a very short period, has also been seen in the literature. Aircraft O_3 measurements taken at an elevation of ~2300 meters from July 26-28 and August 2-3 were 20 ppbv higher over Fort Collins compared to flight path measurements approximately 100 km away near Akron Colorado (Vu *et al.*, 2016). Modeled surface O_3 also demonstrating a roughly 20 ppbv gradient moving from urban to rural regions (Sullivan *et al.*, 2016). Overall, MOOS were able to capture this spatial gradient similar to the reference monitors. At all bins, FTM and AKR MOOS mean deltas were similar to those for the reference monitors. Variation in 1 and 2 standard deviations were very comparable between MOOS and reference monitors at these sites, with both MOOS and reference monitors also capturing outlier O_3 events. At FCW and TWR, MOOS appeared to predict a sharper mixing ratio gradient than the collocated reference monitors, especially at higher O_3 concentrations.

The ability of MOOS to reliably estimate O_3 mixing ratios demonstrates the potential for these sensors to be implemented as in-situ monitors. Overall, MOOS were able to capture relevant exposure and spatial O_3 trends and compared favorably to modeling approaches.

CHAPTER 4: CONCLUSIONS, UNCERTAINTIES, AND DIRECTIONS FOR FUTURE RESEARCH

In this work, we developed an autonomous, low-cost sensor system (aka MOOS) based on the Aeroqual SM50 O_3 sensor to evaluate the efficacy of several practical calibration approaches and study O_3 trends across the Colorado Front Range. In a three-week collocated study, MOOS demonstrated good correlation with collocated reference monitors and minimal sensitivity to temperature and relative humidity, however manufacturer and laboratory corrected MOOS tended to overestimate O_3 mixing ratios and exhibited substantial intra-sensor variance. These sensors were then deployed for a period of three weeks across Eastern Colorado to further evaluate the sensor performance and study urban and rural O_3 phenomena. Linear calibration approaches based on both the collocated study and 24 to 96 hours of reference data in the field reduced measurement error compared to manufacturer calibrated MOOS and demonstrated the ability to capture spatial O_3 trends. Notably, MOOS were able to capture 100% of noncompliant O_3 events during the collocated deployment and were able to characterize up to 88% of NAAQS compliant and noncompliant O_3 events. MOOS have potential to provide reliable ground level O_3 data in order to better characterize regional O_3 trends, inform air quality advisories, and complement existing monitoring efforts by enhancing spatial monitoring capabilities. However, additional research is needed to quantify the impacts of the sensor housing as well as environmental and ambient air impacts on sensor performance.

Although MOOS demonstrated good performance and minimal sensitivity to temperature and relative humidity, additional work is needed to address MOOS sensitivity to interfering gases, possible impact of the sensor housing on O_3 loss, long term drift, and consistency between MOOS and reference monitors. We were unable to test MOOS sensitivity to NO_X (NO and NO_2) and

deploying the MOOS alongside a NO_X reference monitor could help study this issue in more detail. The sensor housing was developed to keep the Aeroqual SM50 sensor within its operating temperature range as the sensor demonstrated sensitivity to higher temperatures in tests performed in summer 2020. The sensor housing also served to minimize direct exposure of the SM50 sensing element to ambient air, which could cause O_3 wall losses when ambient air contacts the fan and sensor housing. Furthermore, these losses could be exacerbated by the buildup of dust inside the sensor housing over time and should be quantified in future work. This study also did not address long-term sensor drift, useful lifetime of these sensors, and optimal frequency for calibration. The MOOS did not appear to substantially drift over the 2 months of deployment, with <5 ppbv of bias being introduced over this time period (Figure S21), however an increase in offset error will likely occur over time. For shorter deployments during peak O_3 season, it appears that a pre- and post-field deployment calibration approach could correct MOOS estimates, but long-term reliability remains a concern for these sensors and needs to be evaluated further. Finally, there was some inconsistency between the MOOS and reference monitors at some of the field sites. This was most particularly observed at the TWR and FCW sites, which we speculate is from the distance between the reference monitors and the MOOS. Another possible reason for this discrepancy is that FCW is frequently exposed to dust events which could have expedited fouling of the SM50 sensor, a phenomenon which was observed during preliminary tests in 2020. MOOS and the reference monitor height should be set to the minimum EPA O_3 standard for inlet height of 2 meters and close proximity between MOOS and reference monitors should be implemented when possible.

Likewise, it would be important for future work to test and incorporate different calibration approaches in order to further improve the measurement reliability of MOOS systems. LCOL-*Pre* and LCOL-*24/48/96* approaches corrected MOOS estimates within 5 ppbv of reference

measurements, however other calibration approaches could be integrated further improve measurements MOOS. Augmenting these calibration strategies with additional calibrations such as buddy sensors (Miskell *et al.*, 2019), night time proxy calibrations (Bart *et al.*, 2014), and coupling LCOL-*Pre* and LCOL-*k* approaches could further improve sensor performance. "Buddy sensing" involves deploying an additional, calibrated low-cost sensor at field sites in place of a reference monitor, which can be useful when scaling up low-cost sensor initiatives in areas where using a collocated reference monitor for field calibrations would be impractical. During the night, O_3 concentrations tend to normalize, allowing for real-time low-cost sensor bias corrections to nighttime reference measurements over much larger distances, without the need for a direct colocation. Finally, averaging LCOL-*Pre* and LCOL-*k* calibration approaches could reduce some of the bias we saw in field measurements. Incorporating machine learning approaches could potentially improve calibrations developed over poorly performing training periods and develop more sophisticated algorithms for correcting interfering factors (Zimmerman *et al.*, 2018).

Proliferation of O&G activity and increasing urbanization in the Colorado Front Range motivate the development of reliable and low-cost air quality monitoring networks in order to better inform the public on air pollution and capture spatiotemporal trends at high resolution. The autonomous, low-cost sensor developed in this work (MOOS) presents an ideal system to integrate into future air quality monitoring efforts due to the low system cost (<$1000), ease of assembly, autonomous operation, and demonstrated performance in measuring ground-level O_3. These sensors could be incorporated alongside low-cost $PM_{2.5}$, NO_X, and VOC sensors to develop a more holistic understanding of spatial-temporal pollution trends and report real time air quality data to the local population.

REFERENCES

Lippmann, M. & Schlesinger, R. B. Toxicological bases for the setting of health-related air pollution standards. *Annu. Rev. Public Health* 21, 309–333 (2000).

Iriti, M. & Faoro, F. Oxidative stress, the paradigm of ozone toxicity in plants and animals. *Water Air Soil Pollut.* 187, 285–301 (2007).

Lin, S., Liu, X., Le, L. H. & Hwang, S.-A. Chronic exposure to ambient ozone and asthma hospital admissions among children. *Environ. Health Perspect.* 116, 1725–1730 (2008).

Malley, C. S. *et al.* Updated global estimates of respiratory mortality in adults ≥30Years of age attributable to long-term ozone exposure. *Environ. Health Perspect.* 125, (2017).

Jacob, D. J. *Introduction to Atmospheric Chemistry*. (Princeton University Press, 1999).

Simon, H., Reff, A., Wells, B., Xing, J. & Frank, N. Ozone trends across the United States over a period of decreasing NOx and VOC emissions. *Environ. Sci. Technol.* 49, 186–195 (2015).

Epa, U. S. & OAR. Progress cleaning the air and improving people's health. (2015).

Cooper, O. R., Gao, R.-S., Tarasick, D., Leblanc, T. & Sweeney, C. Long-term ozone trends at rural ozone monitoring sites across the United States, 1990-2010. *J. Geophys. Res.* 117, (2012).

Jaffe, D. A. *et al.* Impact of wildfires on ozone exceptional events in the Western u.s. *Environ. Sci. Technol.* 47, 11065–11072 (2013).

Carter, W. P. L. & Seinfeld, J. H. Winter ozone formation and VOC incremental reactivities in the Upper Green River Basin of Wyoming. *Atmos. Environ.* 50, 255–266 (2012).

Edwards, P. M. *et al.* High winter ozone pollution from carbonyl photolysis in an oil and gas basin. *Nature* 514, 351–354 (2014).

Field, R. A., Soltis, J., McCarthy, M. C., Murphy, S. & Montague, D. C. Influence of oil and gas field operations on spatial and temporal distributions of atmospheric non-methane hydrocarbons and their effect on ozone formation in winter. *Atmos. Chem. Phys.* 15, 3527–3542 (2015).

Cheadle, L. C. *et al.* Surface ozone in the Colorado northern Front Range and the influence of oil and gas development during FRAPPE/DISCOVER-AQ in summer 2014. *Elementa (Wash., DC)* 5, 61 (2017).

ECFR :: 40 CFR part 58 -- ambient air quality surveillance. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-58.

West, J. J., Fiore, A. M., Horowitz, L. W. & Mauzerall, D. L. Global health benefits of mitigating ozone pollution with methane emission controls. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3988–3993 (2006).

Liu, H. *et al.* Ground-level ozone pollution and its health impacts in China. *Atmos. Environ.* 173, 223–230 (2018).

Pride, K. R. *et al.* Association of short-term exposure to ground-level ozone and respiratory outpatient clinic visits in a rural location - Sublette County, Wyoming, 2008-2011. *Environ. Res.* 137, 1–7 (2015).

Neemann, E. M., Crosman, E. T., Horel, J. D. & Avey, L. Simulations of a cold-air pool associated with elevated wintertime ozone in the Uintah Basin, Utah. *Atmos. Chem. Phys.* 15, 135–151 (2015).

Flocke, F. *et al.* Air Quality in the Northern Colorado Front Range Metro Area: The Front Range Air Pollution and Photochemistry Éxperiment (FRAPPÉ). *J. Geophys. Res. D: Atmos.* 125, e2019JD031197 (2020).

Sullivan, J. T. *et al.* Quantifying the contribution of thermally driven recirculation to a high-ozone event along the Colorado Front Range using lidar. *J. Geophys. Res.* 121, 10,377–10,390 (2016).

red. Monitoring: Uinta Basin. *Utah Department of Environmental Quality* https://deq.utah.gov/air-quality/monitoring-uinta-basin (2019).

Lu, X. *et al.* Severe Surface Ozone Pollution in China: A Global Perspective. *Environ. Sci. Technol. Lett.* 5, 487–494 (2018).

Gaudel, A. *et al.* Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation. *Elementa (Wash., DC)* 6, 39 (2018).

Piedrahita, R. *et al.* The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques* 7, 3325–3336 (2014).

Snyder, E. G. *et al.* The changing paradigm of air pollution monitoring. *Environ. Sci. Technol.* 47, 11369–11377 (2013).

Idrees, Z. & Zheng, L. Low cost air pollution monitoring systems: A review of protocols and enabling technologies. *Journal of Industrial Information Integration* 17, 100123 (2020).

Zhang, J., Qin, Z., Zeng, D. & Xie, C. Metal-oxide-semiconductor based gas sensors: screening, preparation, and integration. *Phys. Chem. Chem. Phys.* 19, 6313–6329 (2017).

Peterson, P. J. D. *et al.* Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments. *Sensors* 17, (2017).

Karagulian, F. *et al.* Review of the Performance of Low-Cost Sensors for Air Quality Monitoring. *Atmosphere* 10, 506 (2019).

Masey, N. *et al.* Temporal changes in field calibration relationships for Aeroqual S500 O3 and NO2 sensor-based monitors. *Sens. Actuators B Chem.* 273, 1800–1806 (2018).

Isiugo, K., Newman, N., Jandarov, R., Grinshpun, S. A. & Reponen, T. Assessing the accuracy of commercially available gas sensors for the measurement of ambient ozone and nitrogen dioxide. *J. Occup. Environ. Hyg.* 15, 782–791 (2018).

Sayahi, T. *et al.* Long-term calibration models to estimate ozone concentrations with a metal oxide sensor. *Environ. Pollut.* 267, 115363 (2020).

Zimmerman, N. *et al.* A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* 11, 291–313 (2018).

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M. & Bonavitacola, F. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sens. Actuators B Chem.* 215, 249–257 (2015).

Schmitz, S. *et al.* Unravelling a black box: an open-source methodology for the field calibration of small air quality sensors. *Atmos. Meas. Tech.* 14, 7221–7241 (2021).

Polidori, A., Papapostolou, V. & Zhang, H. Laboratory evaluation of low-cost air quality sensors. *South Coast Air Quality Management District: Diamondbar, CA, USA* (2016).

SQAMD Air Quality Sensor Performance Evaluation Center. Air quality sensor summary reports. http://www.aqmd.gov/aq-spec/evaluations/summary-gas.

Leusink, J. Materials ozone resistance chart. https://www.oxidationtech.com/blog/materials-ozone-resistance-chart/.

Dutscher. Labware Chemical Resistance Table.

Maag, B., Zhou, Z. & Thiele, L. A Survey on Sensor Calibration in Air Pollution Monitoring Deployments. *IEEE Internet of Things Journal* 5, 4857–4870 (2018).

Li, J. *35 Years of Chemical Sensors - An Honorary Symposium for Professor Jiri Janata's 70th Birthday Celebration*. (The Electrochemical Society, 2009).

Jiao, W. *et al.* Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmos Meas Tech* 9, 5281–5292 (2016).

Mueller, M., Meyer, J. & Hueglin, C. Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich. *Atmospheric Measurement Techniques* vol. 10 3783–3799 (2017).

Ferrer-Cid, P., Barcelo-Ordinas, J. M., Garcia-Vidal, J., Ripoll, A. & Viana, M. A Comparative Study of Calibration Methods for Low-Cost Ozone Sensors in IoT Platforms. *IEEE Internet of Things Journal* 6, 9563–9571 (2019).

Miskell, G. *et al.* Reliable data from low cost ozone sensors in a hierarchical network. *Atmos. Environ.* 214, 116870 (2019).

Cheadle, L. *et al.* Quantifying Neighborhood-Scale Spatial Variations of Ozone at Open Space and Urban Sites in Boulder, Colorado Using Low-Cost Sensor Technology. *Sensors* 17, (2017).

Requia, W. J. *et al.* An Ensemble Learning Approach for Estimating High Spatiotemporal Resolution of Ground-Level Ozone in the Contiguous United States. *Environ. Sci. Technol.* 54, 11037–11047 (2020).

Di, Q., Rowland, S., Koutrakis, P. & Schwartz, J. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. *J. Air Waste Manag. Assoc.* 67, 39–52 (2017).

Wang, H.-W. *et al.* Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *J. Clean. Prod.* 253, 119841 (2020).

McDuffie, E. E. et al. Influence of oil and gas emissions on summertime ozone in the Colorado Northern Front Range. J. Geophys. Res. 121, 8712–8729 (2016).

Bart, M. et al. High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia. Environ. Sci. Technol. 48, 3970–3977 (2014).

APPENDIX

Item	Description	Quantity	Unit Cost (USD)	Total Cost (USD)
Enclosure				
Polycase WQ-57	Waterproof enclosure for housing power system and sensor	1	54.57	54.57
Polycase WQ-91	Mounting feet for connecting enclosure to mounting frame	1	8.25	8.25
Polycase WQ-57-p01	Backboard for electronics mounting	1	5.01	5.01
HARTING 19000005181	5-9 mm cable gland to allow for sealed cable routing	2	2.97	5.94
Pyle Audio Stand	System stand	1	27.41	27.41
8020 Components	Mounting bracket to allow for mounting of enclosure to pole	1	30.4	30.4
Power System				
Renogy 30 W Panel	Solar Panel	1	59.99	59.99
Bioenno 12V, 9Ah LFP Battery	Battery	1	99.99	99.99
Renogy Wanderer 12V Charge Controller	Solar Charge Controller	1	14.99	14.99
MC4 Solar Panel Cable	MC4 solar panel cable	1	19.49	19.49
Electronics				
SparkFun Level Shifting microSD Breakout	3.3V MicroSD card reader	1	5.5	5.5
Particle Boron	LTE Enabled Microcontroller	1	59.99	59.99
BME 280	Temperature and relative humidity sensor	1	19.95	19.95
32 gb microSD card	Data storage	1	7.47	7.47
Aeroqual SM50	Ozone Sensor	1	325	325
2.54mm MALE & FEMALE HEADER Socket Row Stri	Mounting Electronics to through hole board	1	4.97	4.97
Radiation Shield				
PTEG Fillament Spool	Filament needed to print radiation shield	0.75	22.99	17.24
1" ODE PTFE Tube x 6" length	Sampling tube	0.5	20.47	10.24
40 mm computer fan	Cooling MOOS and pulling in sampling air	2	2.97	5.94
8020 Parts	Mounting brackets for shield	1	15.67	15.67
		Total	764.59	798.0075

Table S1: Breakdown of component costs for 1 MOOS system. Misc. fasteners and wiring costs are not included in the cost breakdown.

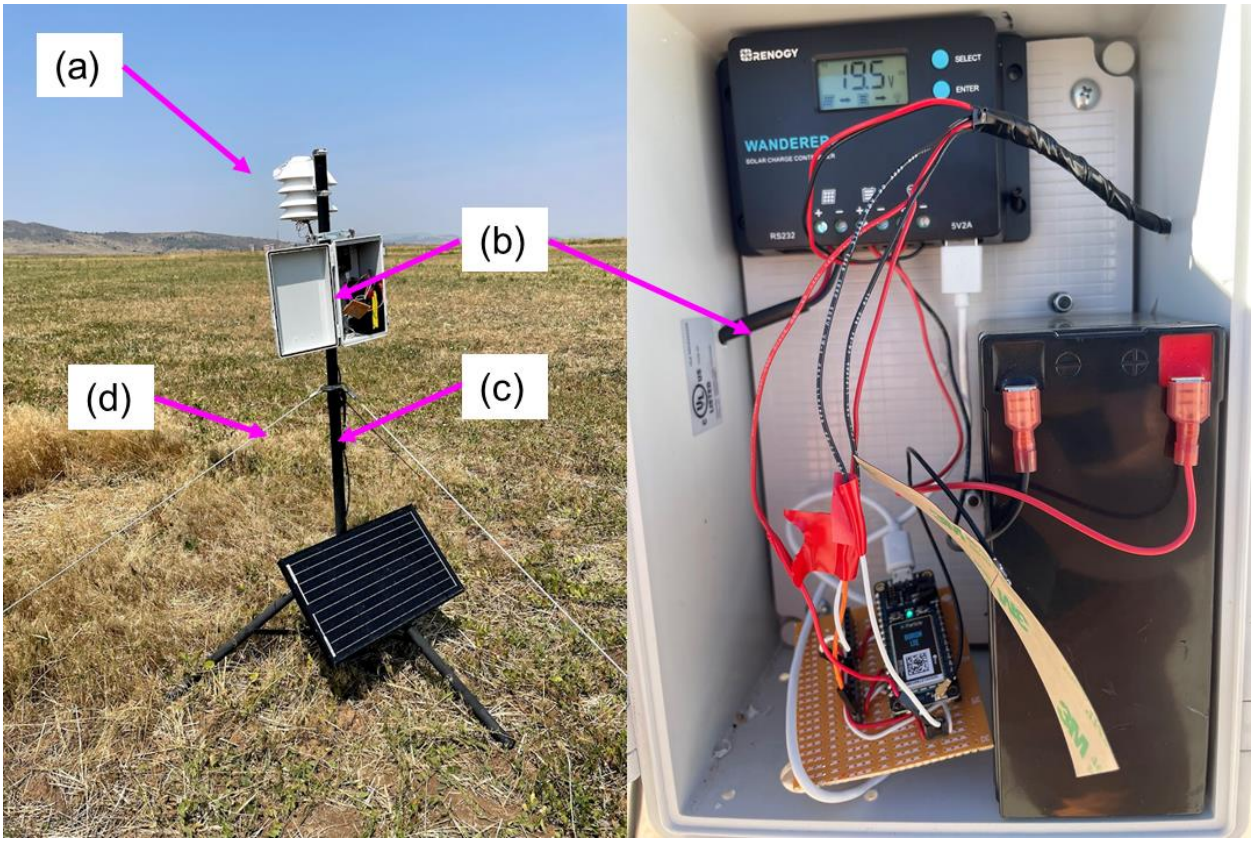

Figure S1: Photo of MOOS deployed in the field including sensor housing (a), interior electrical components (b), stand and solar panel (c), and guy wires (d).

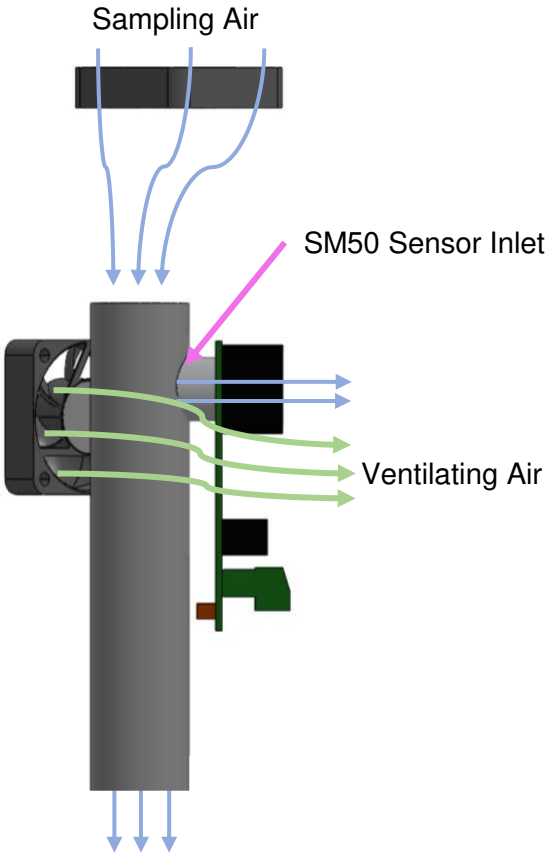

Figure S2: Diagram showing active airflow in the sensor housing. Sampling air is pushed through a fan into the sampling tube where it is then sampled by the SM50. Ventilating air is pushed by a fan over the SM50 to keep the sensor at near ambient air temperature.

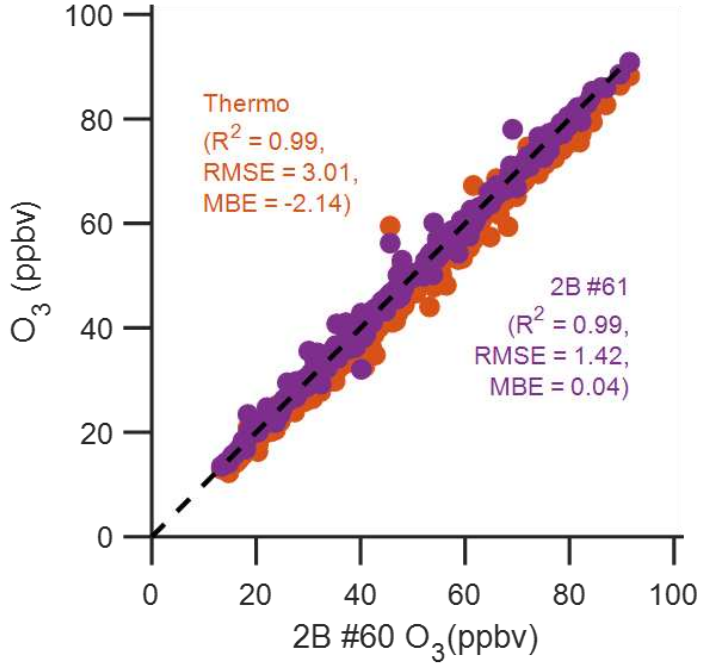

Figure S3: Scatter plot of 2B #2060 O_3 analyzer and Thermo O_3 analyzer compared against the other 2B #2061.

Figure S4: MOOS field deployments at Ft. Morgan (a), Weld Tower (b), Christman Field (c), the Powerhouse rooftop (d), and Akron (e).

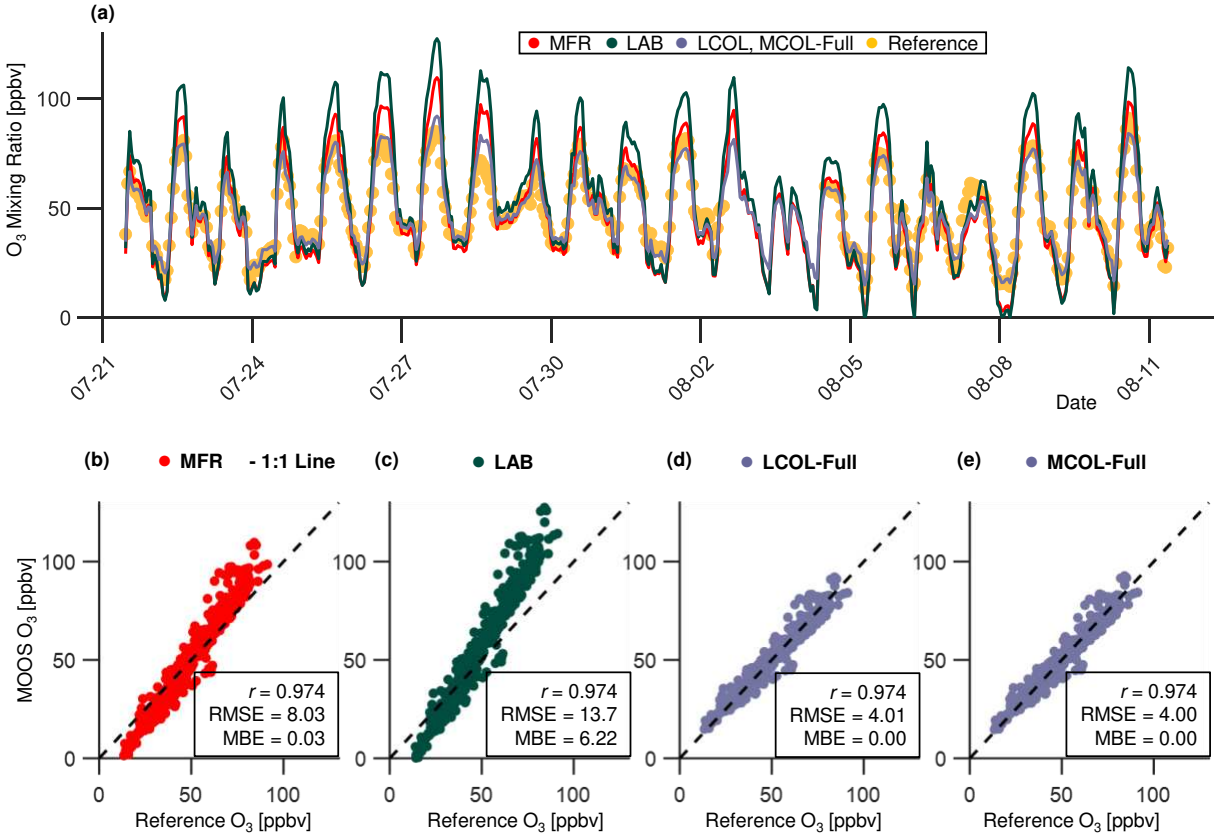

Figure S5: Same as Figure 3 but displaying results for MOOS B.

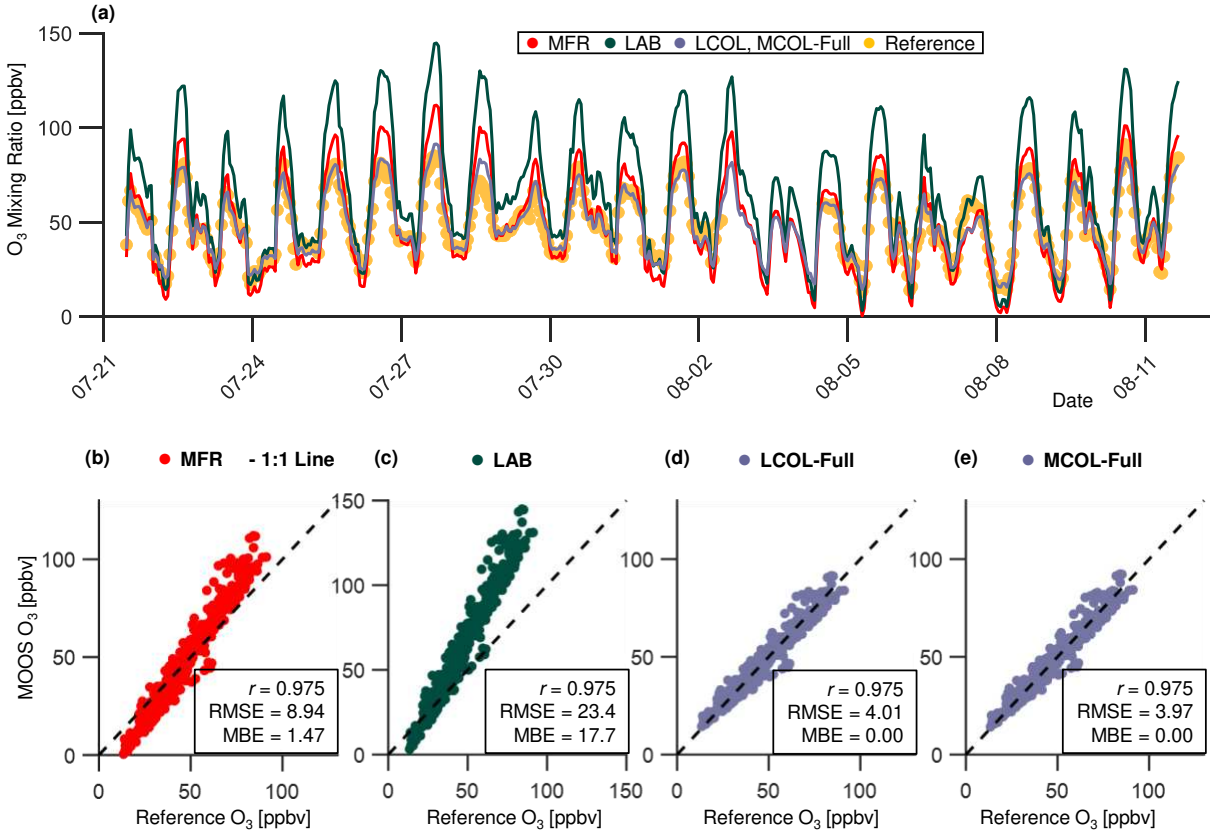

Figure S6: Same as Figure 3 but displaying results for MOOS C.

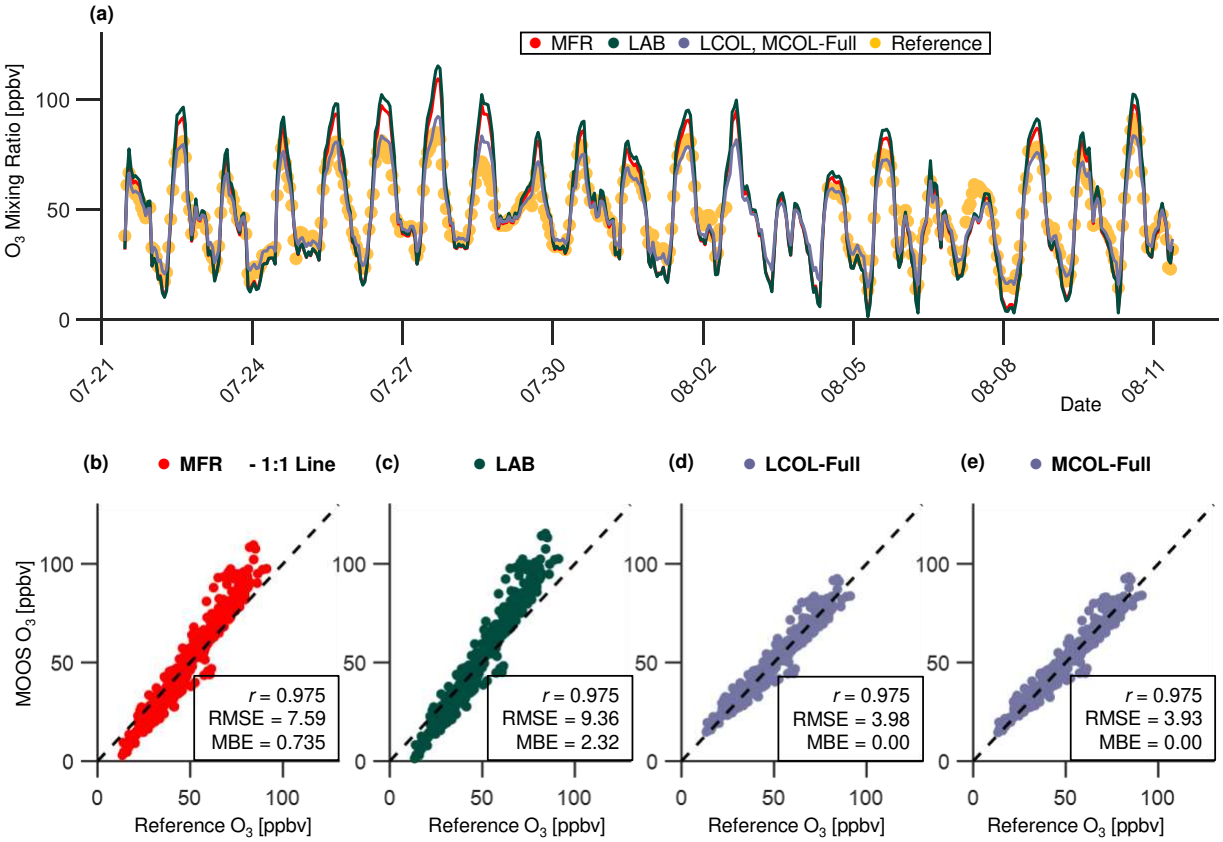

Figure S7: Same as Figure 3 but displaying results for MOOS D.

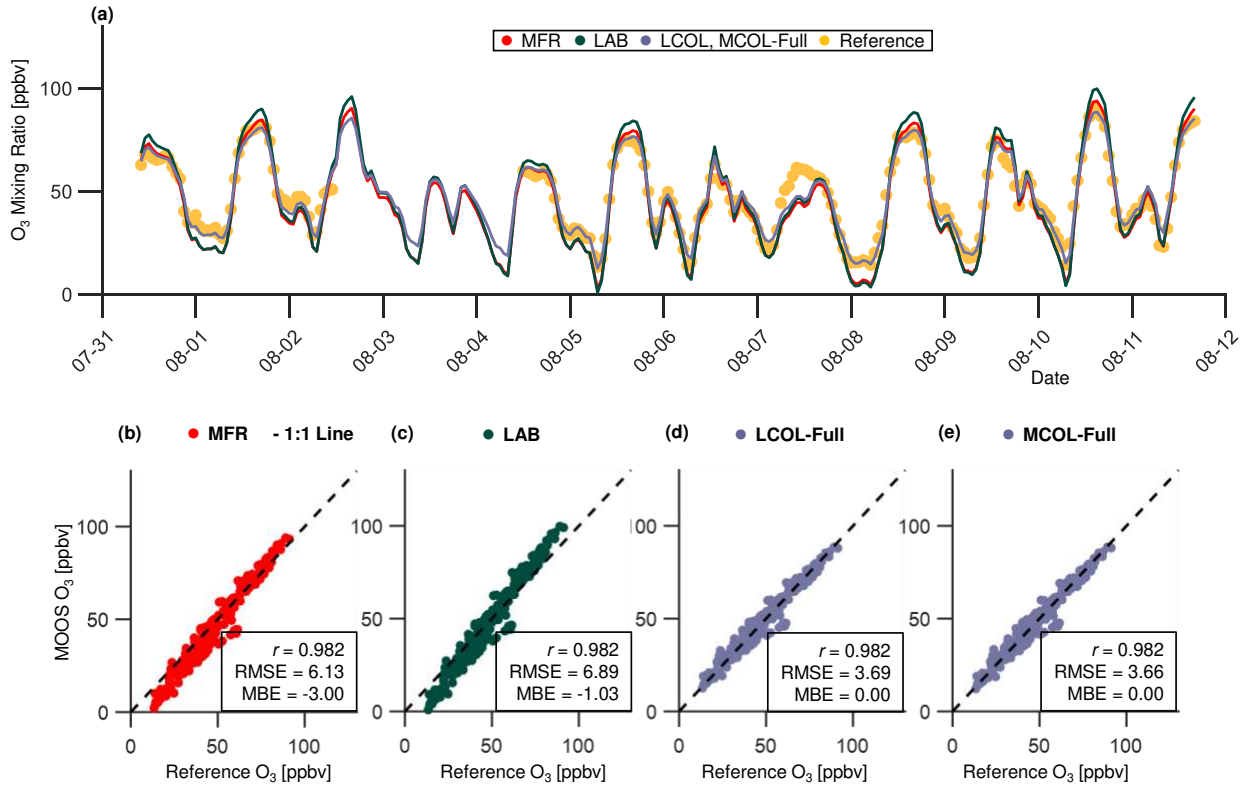

Figure S8: Same as Figure 3 but displaying results for MOOS E.

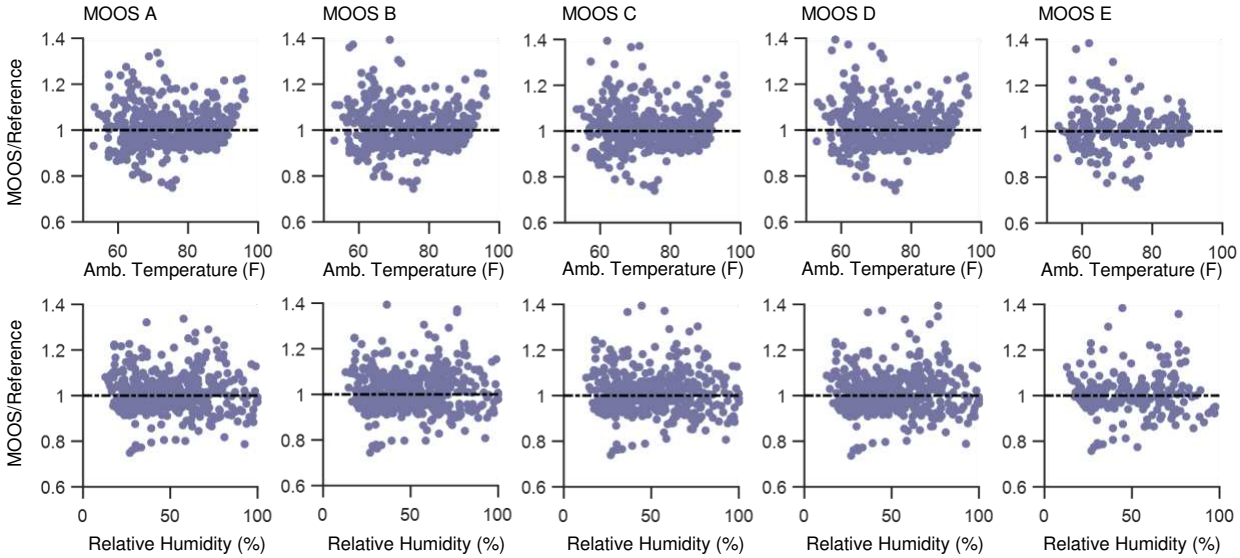

Figure S9: Relationship between the LCOL-Full/Reference ratio and ambient temperature and relative humidity during the rooftop deployment.

Table S2: Multivariable regression coefficients for MOOS A-E used to develop MCOL, model coefficient of determination (r^2) terms for multiple regression and individual predictors, and model error (RMSE) for multiple regression and individual predictors.

MOOS	b_0	b_1	b_2	b_3	r^2_{all}	r^2_1	r^2_2	r^2_3	$RMSE_{all}$	$RMSE_1$	$RMSE_2$	$RMSE_3$
A	14.43	0.75	-0.07	0.00	0.95	0.95	0.58	0.46	15.29	15.45	137.06	174.86
B	17.13	0.73	-0.06	0.00	0.95	0.95	0.57	0.45	16.10	16.16	135.36	174.38
C	21.85	0.72	-0.11	-0.02	0.95	0.95	0.58	0.46	15.91	16.17	137.06	174.86
D	19.16	0.76	-0.10	-0.01	0.95	0.95	0.57	0.45	15.66	15.88	135.36	174.38
E	16.58	0.85	-0.09	-0.01	0.96	0.96	0.71	0.55	13.66	13.66	109.51	169.02

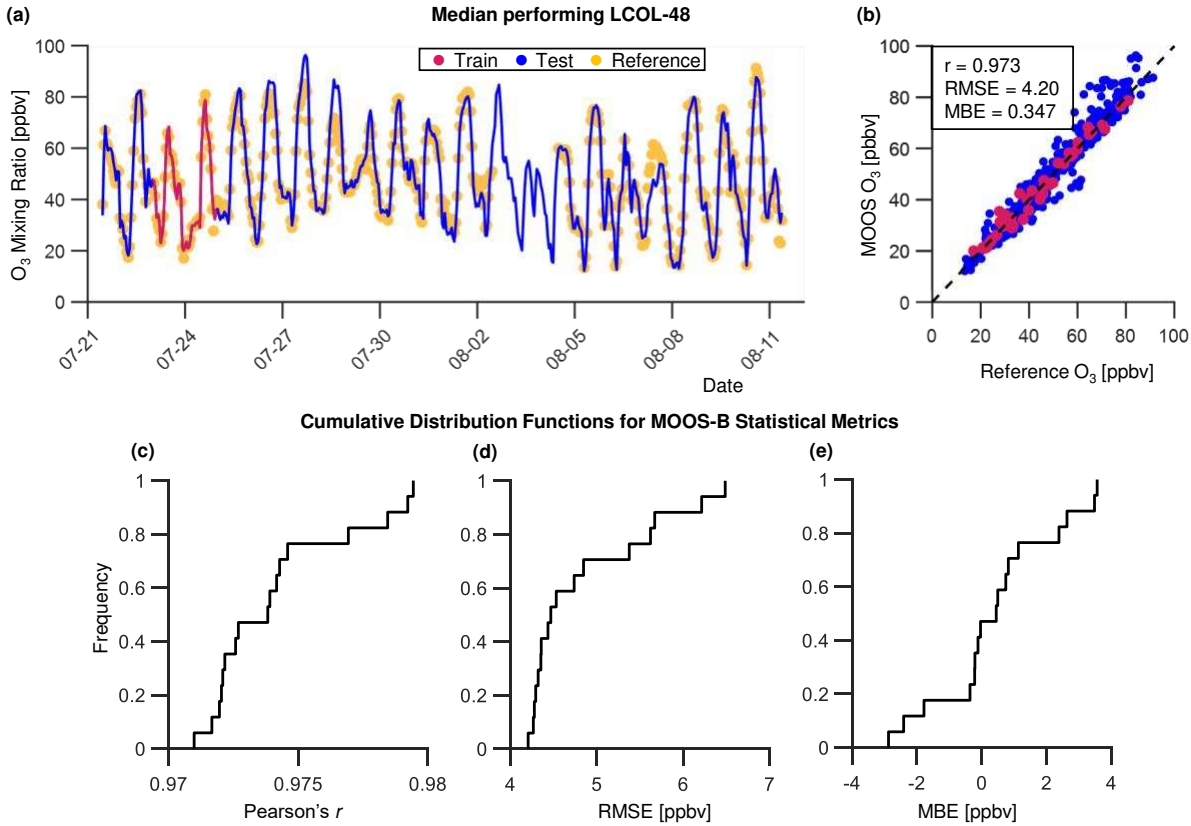

Figure S10: Same as Figure 4 but displaying results for MOOS B.

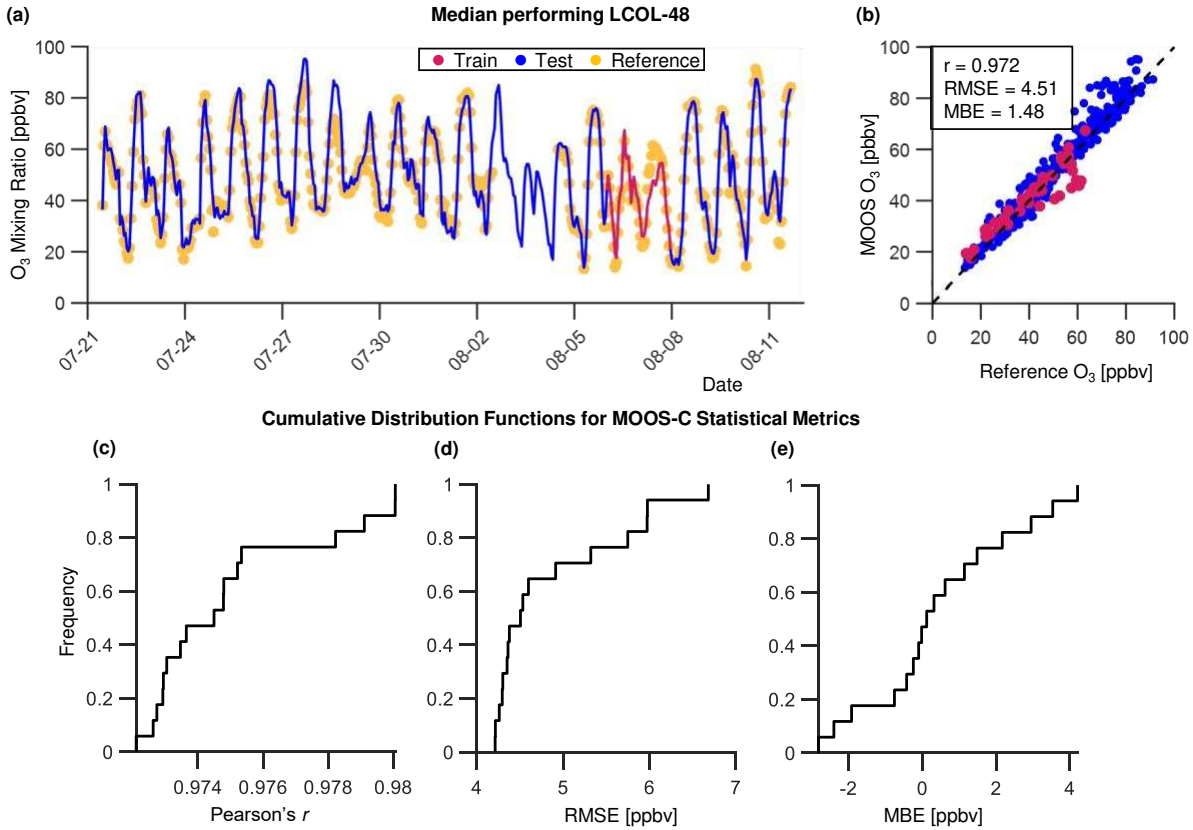

Figure S11: Same as Figure 4 but displaying results for MOOS C.

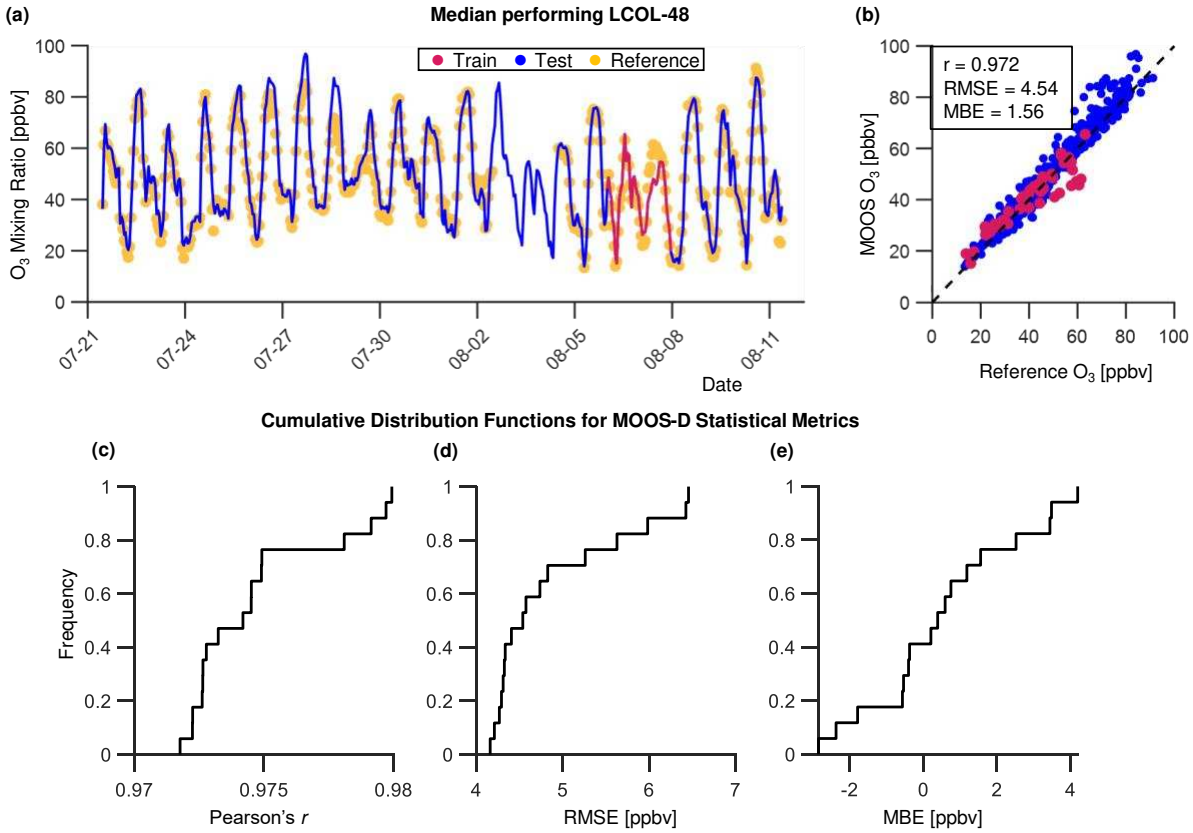

Figure S12: Same as Figure 4 but displaying results for MOOS D.

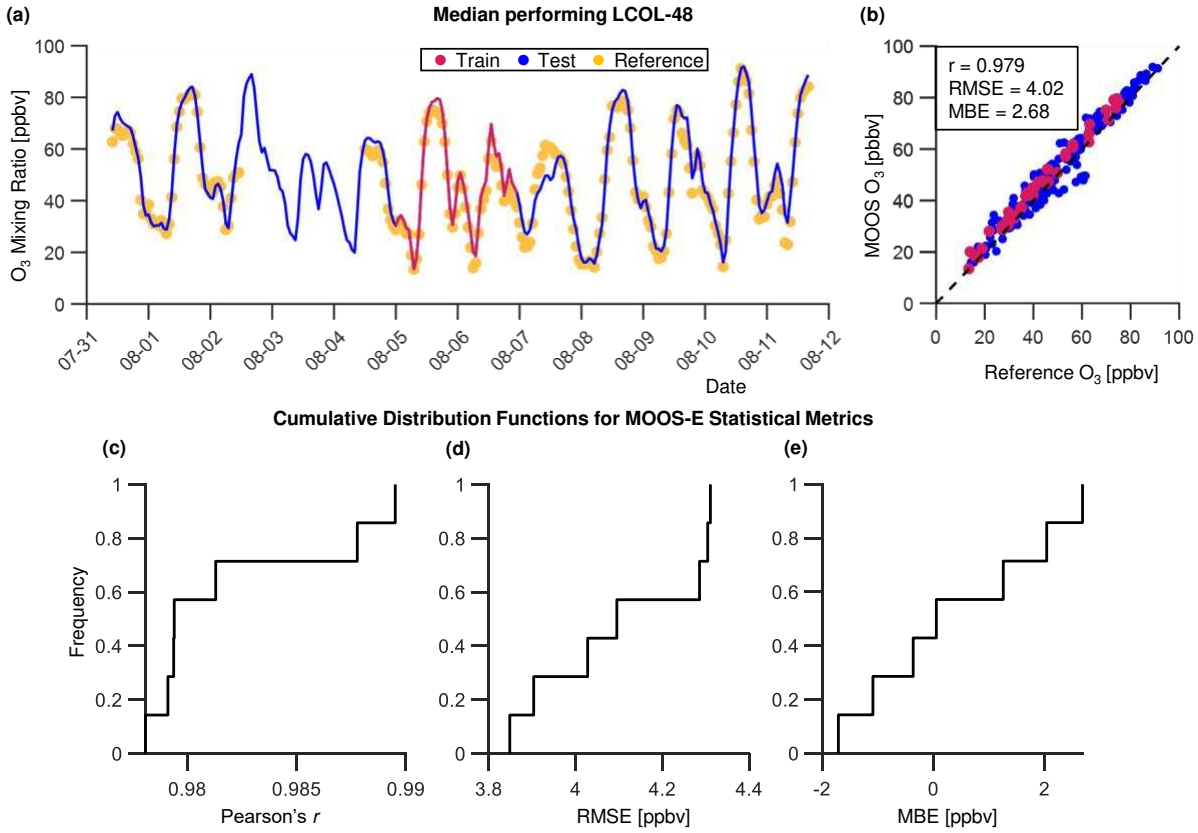

Figure S13: Same as Figure 4 but displaying results for MOOS E.

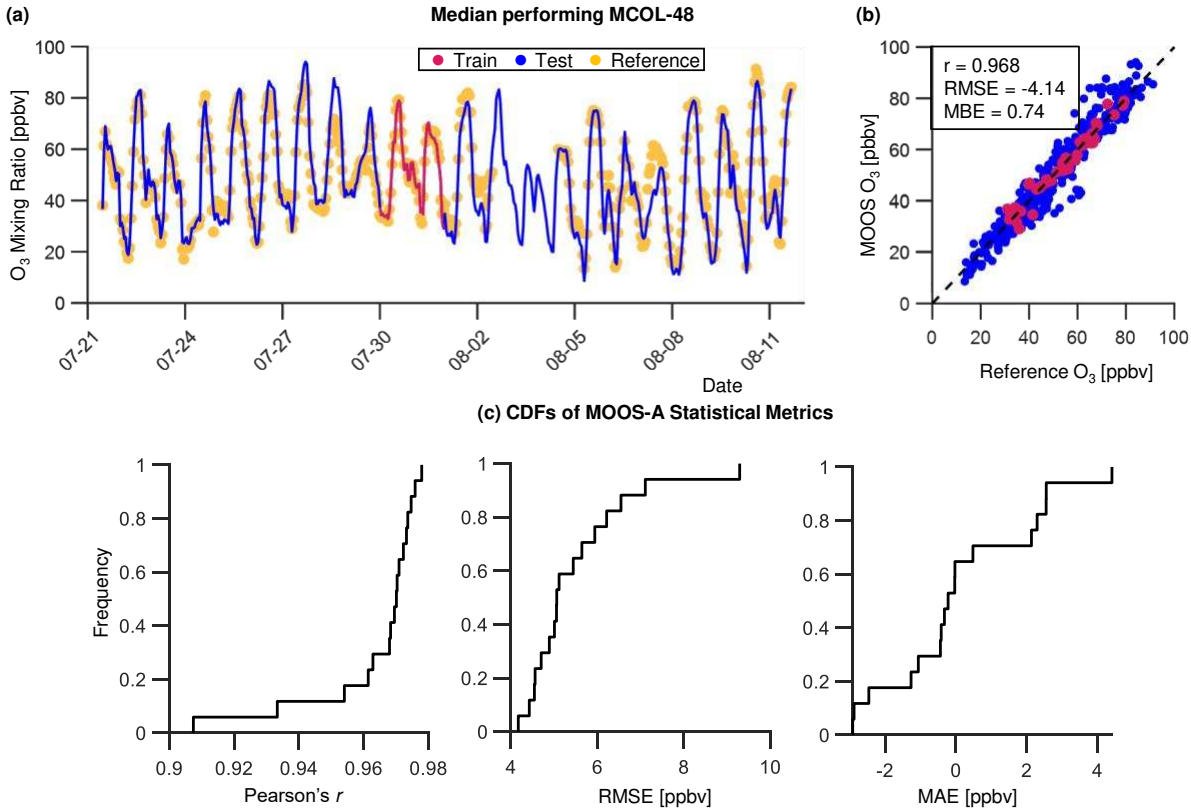

Figure S14: MOOS-A O_3 mixing ratio estimates corrected using an n-fold cross validation MCOL approach compared to collocated reference measurements as a time series (a) and scatter plot (b). 'Train' refers to the 48-hour interval used as the training period and 'Test' refers to the remainder of the data used to evaluate the calibration. The LCOL-48 approach shown in (a,b) represents the median performing calibration developed from one of the 17 training periods. Statistical metrics r (a), RMSE (b), and MBE (c) from the cross validation of 17 independent MCOL-48 calibrations are shown as cumulative distribution functions.

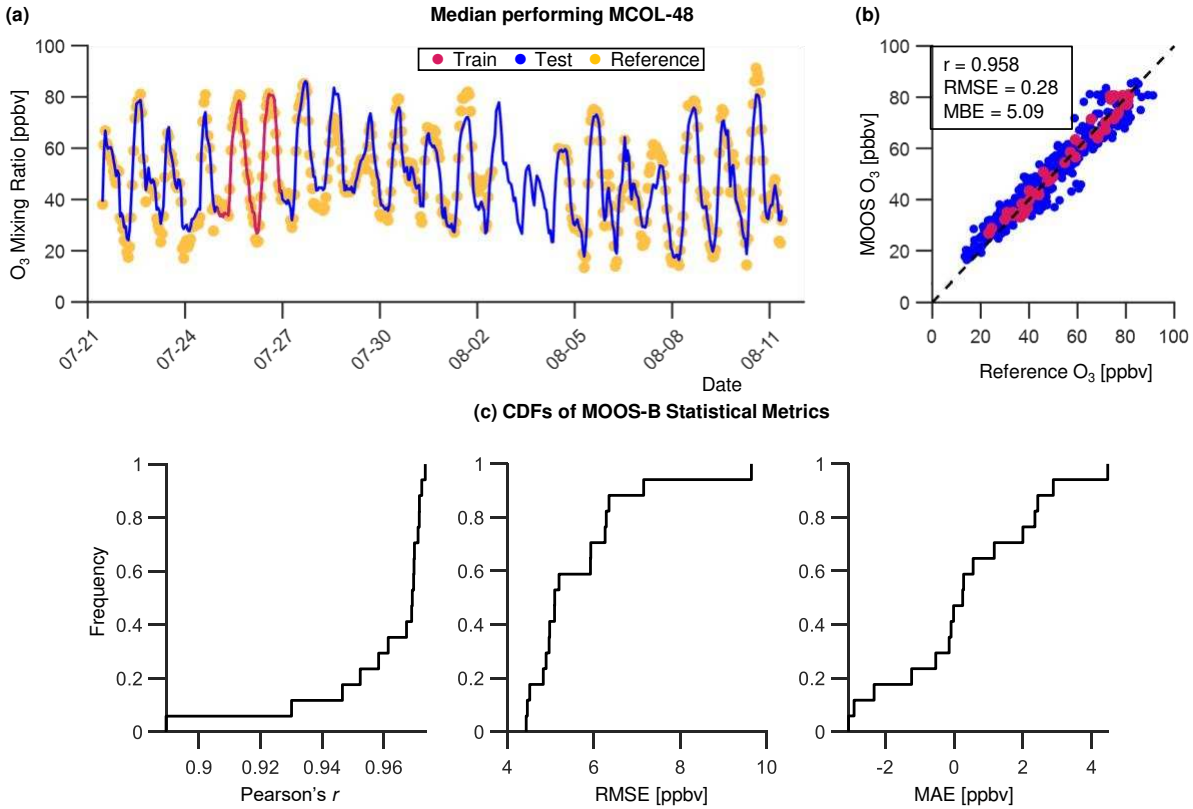

Figure S15: Same as Figure S12 but displaying results for MOOS B.

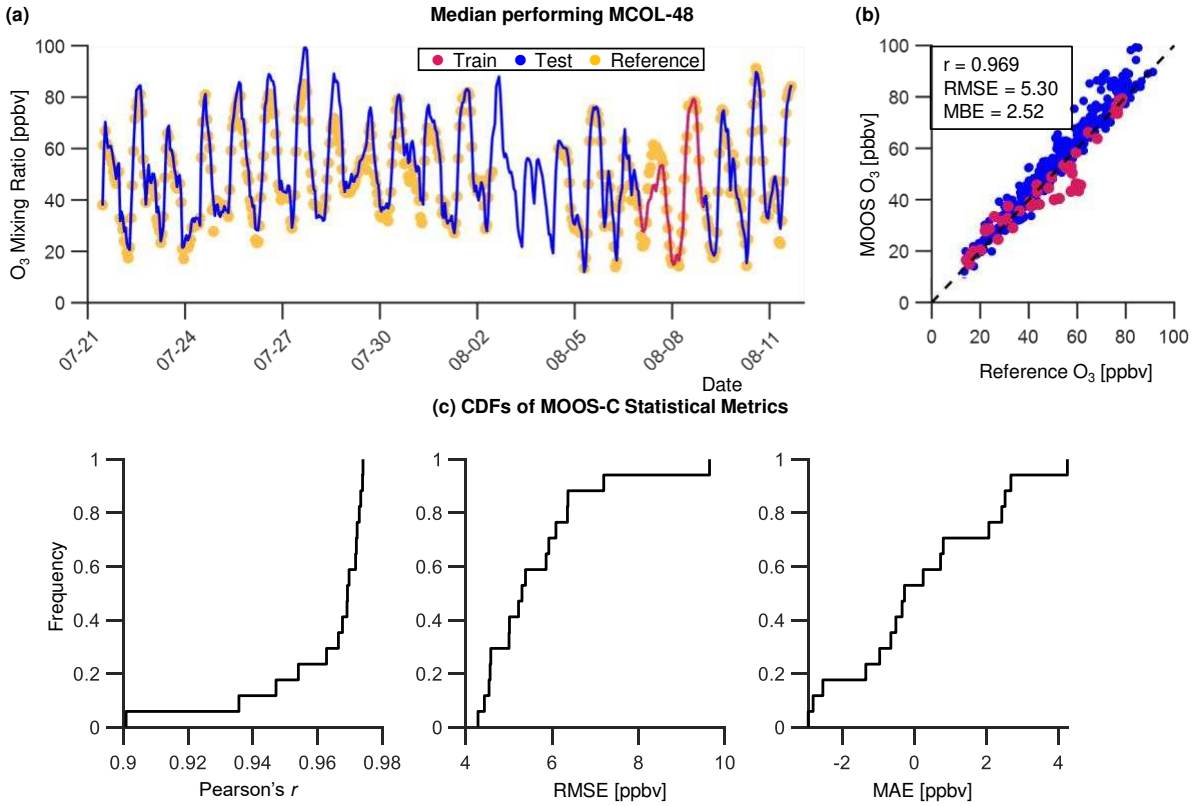

Figure S16: Same as Figure S12 but displaying results for MOOS C.

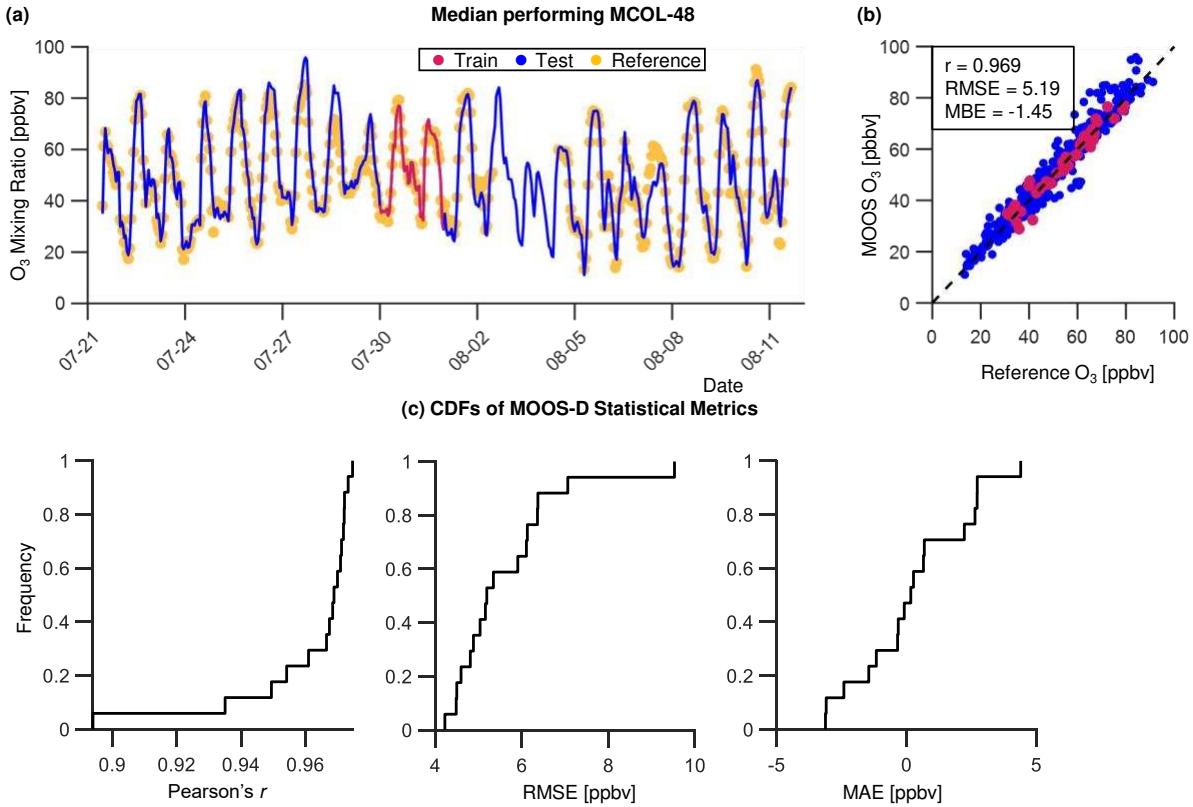

Figure S17: Same as Figure S12 but displaying results for MOOS D.

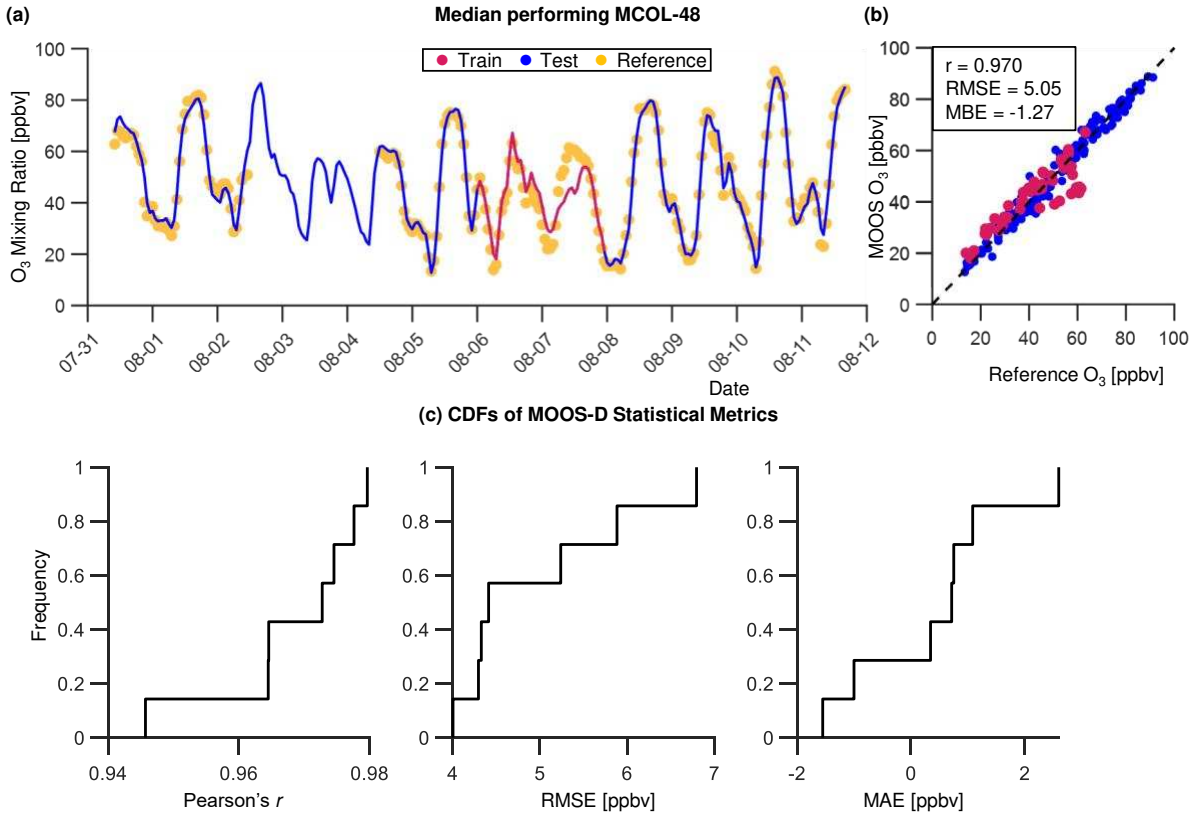

Figure S18: Same as Figure S12 but displaying results for MOOS E.

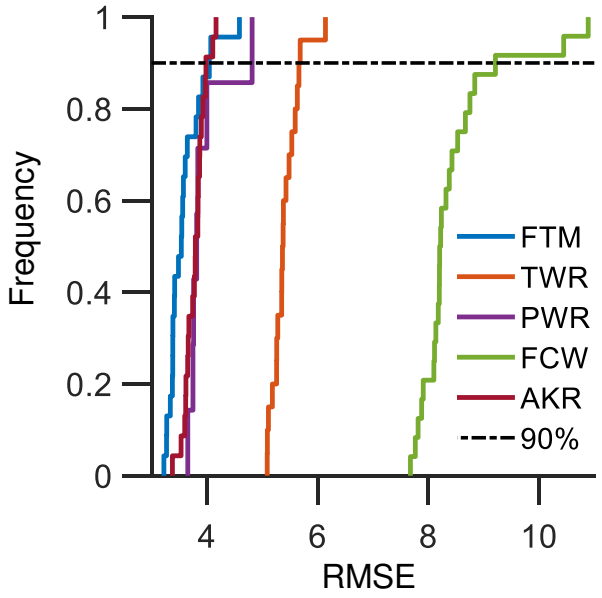

Figure S19: CDF plot demonstrating the 90% cutoff used to select LCOL-48.

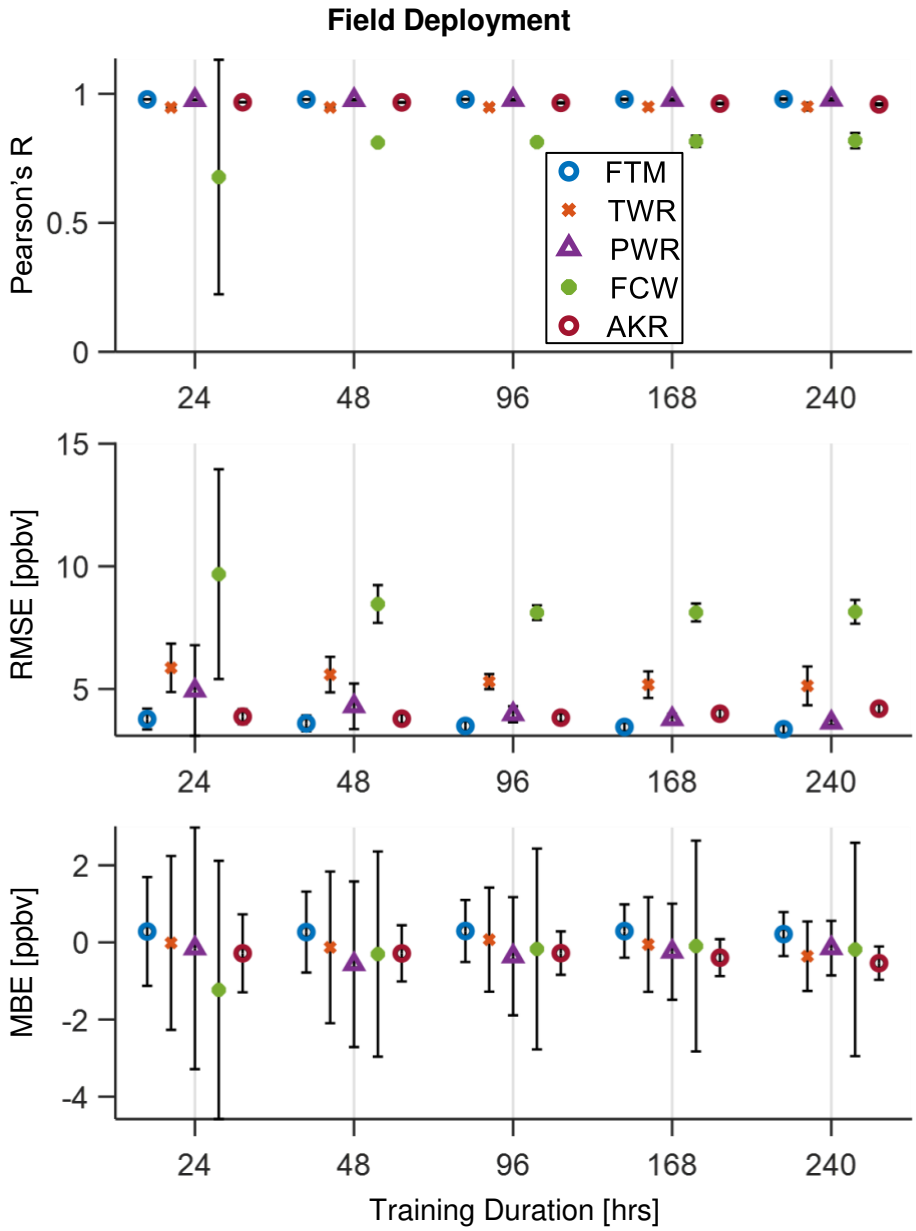

Figure S20: Same as Figure 7 showing the full results for FCW.

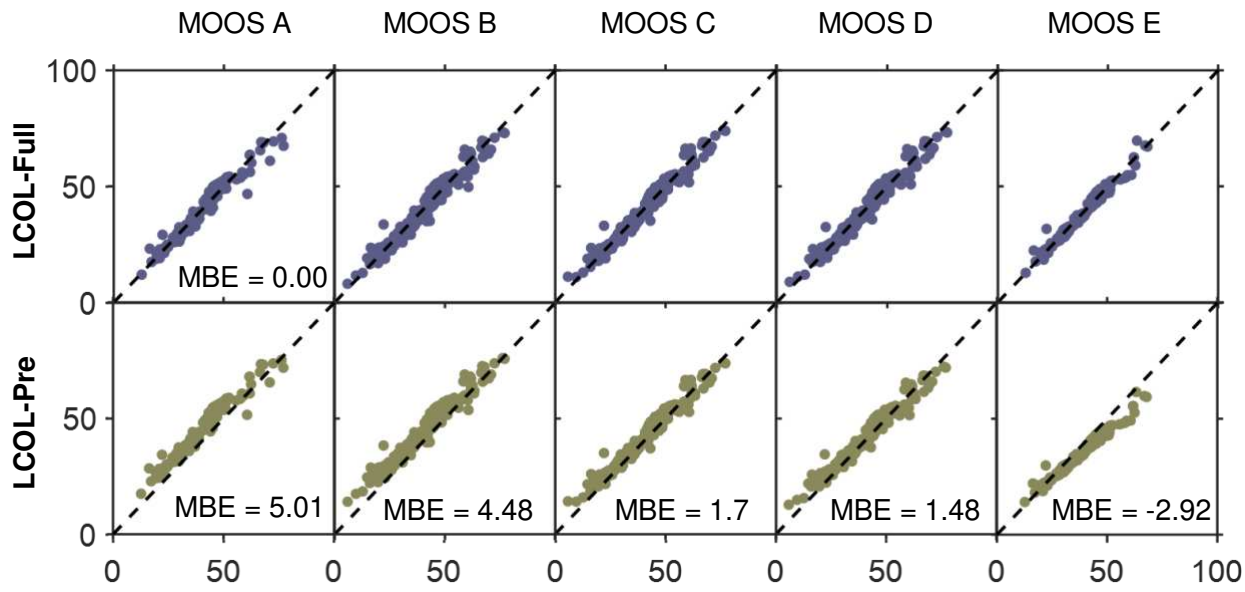

Figure S21: MOOS O_3 mixing ratio estimates corrected using LCOL-Full and LCOL-Pre, approaches compared to reference measurements during the post-field rooftop deployment which spanned roughly 1 week after the field deployment.